BONNER METEOROLOGISCHE ABHANDLUNGEN
Heft 64 (2014) (ISSN 0006-7156)
Herausgeber: Andreas Hense

Michael Weniger

Stochastic Parameterization: A Rigorous Approach to Stochastic Three-Dimensional Primitive Equations

STOCHASTIC PARAMETERIZATION: A RIGOROUS APPROACH TO STOCHASTIC THREE-DIMENSIONAL PRIMITIVE EQUATIONS

DISSERTATION

zur
Erlangung des Doktorgrades (Dr. rer. nat.)

der
Mathematisch-Naturwissenschaftlichen Fakultät

der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

MICHAEL WENIGER

aus

Gifhorn

BONN
JANUAR 2013

Diese Arbeit ist die ungekürzte Fassung einer der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn im Jahr 2013 vorgelegten Dissertation von Michael Weniger aus Gifhorn.

This paper is the unabridged version of a dissertation thesis submitted by Michael Weniger born in Gifhorn to the Faculty of Mathematical and Natural Sciences of the Rheinische Friedrich-Wilhelms-Universität Bonn in 2013.

Anschrift des Verfassers:

Address of the author:

Michael Weniger
Meteorologisches Institut der
Universität Bonn
Auf dem Hügel 20
D-53121 Bonn

---

1. Gutachter: Prof. Dr. Andreas Hense
2. Gutachter: Prof. Dr. Anton Bovier

Tag der Promotion: 29.04.2013

Danksagung

Großer Dank gebührt zuallererst Prof. Dr. Andreas Hense, der meiner vagen Idee als Mathematiker über *"etwas Angewandtes"* zu promovieren aufgeschlossen gegenüberstand und daraus innerhalb von wenigen Gesprächen ein konkretes und spannendes Dissertationthema entwarf. Sein immenses interdisziplinäres Wissen gepaart mit großer wissenschaftlicher Kreativität und der Geduld seine Ideen einem Doktoranden ohne meteorologisches Vorwissen begreiflich zu machen waren maßgeblich für den Erfolg dieser Arbeit. Unerlässlich für den mathematischen Teil war die professionelle und warmherzige Betreuung durch Prof. Dr. Anton Bovier, der mir durch oftmals ebenso spontane wie intensive Diskussionen dabei half etwaige mathematische Hindernisse zu überwinden. PD Dr. Petra Friederichs bin ich sehr dankbar für ihre Fähigkeit und ihr Interesse die Sprachen beider Gebiete zu verstehen. Ihre daraus resultierende konstruktive Betreuung war eine unschätzbar wertvolle Hilfe um auf dem fremdem Gebiet der Meteorologie Fuß zu fassen.

Der gleiche Dank gebührt meinen Kollegen für ihre Geduld mir viele meteorologische Fragen ausführlich zu beantworten. Sie sind dafür verantwortlich, dass ich bei meiner Promotion nicht an erster Stelle an die erfolgreiche Dissertation denke, sondern an die großartigen Jahre auf dem Weg dorthin und die vielen neu gewonnen Freunde. Ein ganz besonderer Dank geht an meine Eltern, die mir stets alle Freiheiten gewährt haben und deren Unterstützung mein Studium und damit diese Arbeit erst ermöglicht haben.

ZUSAMMENFASSUNG

Die Atmosphäre ist ein von starken Nichtlinearitäten geprägtes, unendlichdimensionales dynamisches System, dessen Variablen auf einer Vielzahl verschiedener Raum- und Zeitskalen interagieren. Ein potentielles Problem von Modellen zur numerischen Wettervorhersage und Klimamodellierung, die auf deterministischen Parametrisierungen subskaliger Prozesse beruhen, ist die unzureichende Behandlung der Interaktion zwischen diesen Prozessen und den Modellvariablen. Eine stochastische Beschreibung dieser Parametrisierungen hat das Potential die Qualität der Simulationen zu verbessern und das Verständnis der Skalen-Interaktion atmosphärischer Variablen zu vertiefen.

Die wissenschaftlich Gemeinschaft, die sich mit stochastischen meteorologischen Modellen beschäftigt, kann grob in zwei Gruppen unterteilt werden: die erste Gruppe ist bemüht durch pragmatische Ansätze bestehende, komplexe Modelle zu erweitern. Die zweite Gruppe verfolgt einen mathematisch rigorosen Weg, um stochastische Modelle zu entwickeln. Dies ist jedoch aufgrund der mathematischen Komplexität bisher auf konzeptionelle Modelle beschränkt. Das generelle Ziel der vorliegenden Arbeit ist es, die Kluft zwischen den pragmatischen und mathematisch rigorosen Ansätzen zu verringern. Die Diskussion zweier konzeptioneller Klimamodelle verdeutlicht, dass eine stochastische Formulierung nicht willkürlich gewählt werden darf, sondern aus der Physik des betrachteten Systems abgeleitet werden muss. Ebenso unabdingbar ist eine rigorose numerische Implementierung des resultierenden stochastischen Modells. Diesem Aspekt wird besondere Bedeutung zu Teil, da dynamische subskalige Prozesse oftmals durch zeitabhängige stochastische Prozesse beschrieben werden, die nicht mit deterministischen numerischen Methoden behandeln lassen.

Wir zeigen auf, dass eine stochastische Formulierung der dreidimensionalen primitiven Gleichungen im mathematischen Rahmen abstrakter stochastischer Fluidmodelle behandelt werden kann. Dies ermöglicht die Anwendung kürzlich gewonnener Erkenntnisse bezüglich Existenz und Eindeutigkeit von Lösungen. Wir stellen einen auf dieser theoretischen Grundlage basierenden Galerkin Ansatz zur Diskretisierung der räumlichen und stochastischen Dimensionen vor. Mit Hilfe sogenannter milder Lösungen der stochastischen partiellen Differentialgleichungen leiten wir quantitative Schranken der Diskretisierungsfehler her und zeigen die starke Konvergenz des mittleren quadratischen Fehlers. Unter zusätzlichen Annahmen leiten wir die Konvergenz eines numerischen Verfahrens her, das den Galerkin Ansatz um einer zeitliche Diskretisierung erweitert.

ABSTRACT

The atmosphere is a strongly nonlinear and infinite-dimensional dynamical system acting on a multitude of different time and space scales. A possible problem of numerical weather prediction and climate modeling using deterministic parameterization of subscale and unresolved processes is the incomplete consideration of scale interactions. A stochastic treatment of these parameterizations bears the potential to improve the simulations and to provide a better understanding of the scale interactions of the simulated atmospheric variables.

The scientific community that is dealing with stochastic meteorological models can be divided into two groups: the first one uses pragmatic approaches to improve existing complex models. The second group pursues a mathematical rigorous way to develop stochastic models, which is currently limited to conceptual models. The overall objective of this work is to narrow the gap between pragmatic approaches and the mathematical rigorous methods. Using conceptual climate models, we point out that a stochastic formulation must not be chosen arbitrarily but has to be derived based on the physics of the system at hand. Equally important is a rigorous numerical implementation of the resulting stochastic model. The dynamics of sub grid and unresolved processes are often described by time continuous stochastic processes, which cannot be treated with deterministic numerical schemes.

We show that a stochastic formulation of the three-dimensional primitive equations fits in the mathematical framework of abstract stochastic fluid models. This allows us to utilize recent results regarding existence and uniqueness of solutions of such systems. Based on these theoretical results we propose a Galerkin scheme for the discretization of spatial and stochastic dimensions. Using the framework of mild solutions of stochastic partial differential equations we are able to prove quantitative error bounds and strong mean square convergence. Under additional assumptions we show the convergence of a numerical scheme which combines the Galerkin approximation with a temporal discretization.

Contents

1 Introduction

The atmosphere is a strongly nonlinear and infinite-dimensional dynamical system acting on a multitude of different time and space scales. Numerical models of the atmosphere or the climate system can only treat a finite number of degrees of freedom. Due to scale interactions, the effect of the discretizations is not negligible. Therefore it is necessary to parameterize the effects of the unresolved scales on the resolved ones. In principle the stochastic character of the unresolved scales has to be taken into account. However, due to historic reasons most parameterization of subscale processes are deterministic in nature. They model conditional expectation values of moments (mostly second moments) given the resolved scales [1]. A possible problem of numerical weather prediction (NWP) and climate modeling using deterministic parameterization of subscale and unresolved processes is therefore the incomplete consideration of interactions between the resolved and subgrid scale processes. A stochastic treatment of these parameterizations has the potential to improve the simulations and to provide a better understanding regarding the stochastic characteristics of the simulated atmospheric variables. Furthermore a stochastic model provides a natural framework to describe the state of a chaotic system using probability densities instead of absolute values. This in turn is closely connected to the analysis of stability regimes, tipping points and extreme events, see for instance [2].

The idea of using stochastic climate models has been introduced by Lorenz, who stated: "*I believe that the ultimate climatic models [...] will be stochastic, i.e., random numbers will appear somewhere in the time derivatives*" [3]. However in the same article he made a note of caution: "*If we are truly careful in introducing our random numbers, we can likewise assure ourselves that the probability of producing an ice age, when one ought not to form, is some infinitesimally small number*". In the same year Hasselmann published a path-breaking article [4], which contains the first mathematical formulation of a climate model treating weather effects as random forcing terms. Since then there have been various approaches to incorporate stochastic techniques in existing numerical models for the atmosphere and climate. In many cases this was proposed or done without the appropriate analyses concerning both the numerical implementation and the mathematical and physical properties of the stochastic variables or processes. Most of the numerical issues arise since models involving time dependent stochastic fluctuations require specific numerical schemes. While the numerical flaws are easily avoided in self-contained conceptual models when handled with proper care, it is a nontrivial challenge to implement a stochastic parameterization into more complex, high dimensional models, e.g., a global circulation model (GCM). The second aspect is more subtle but equally important: the occurrence and the character of a stochastic parameterization have to be physically justified. This is essential for the validity and credibility of stochastic climate models.

The scientific community dealing with stochastic meteorological or climate models can be divided into two groups. The first one uses pragmatic approaches to enhance the existing complex models with stochastic terms which are justified afterwards by means of various skill scores. The second group pursues a mathematical rigorous way to develop stochastic models. Due to the complexity of these systems, this approach is currently limited to conceptual models excluding operative GCMs. The overall objective of this work is to narrow the gap between pragmatic approaches and the mathematical rigorous methods with respect to numerical weather prediction and climate models. More specifically we extensively study two conceptual climate models to crystallize the following statements:

- Stochastic terms must not be arbitrarily chosen, but have to be based on the physics of the system.
- Stochastic models require a specific numerical treatment, which fundamentally differs from deterministic schemes.

Furthermore, we demonstrate that the first claim does not require new mathematical tools. In fact, suitable methods to derive stochastic characteristics from data are known since 1927 [5, 6]. In contrast, rigorous numerical schemes for meteorologically relevant stochastic partial differential equations have yet to be derived. Recently, significant progress has been made in the theoretical analysis of stochastic three-dimensional primitive equations (PE), yielding existence and uniqueness of pathwise solutions [7]. This provides a foundation to utilize recently developed numerical schemes based on so called mild solution of stochastic partial differential equations [8]. Before going into further details, we give a brief survey of the current status of research, which serves as both motivation and starting point of this work.

1.1 State of the Art

1.1.1 Pragmatic Approach

The utilization of stochastic terms in NWP is nowadays widely accepted as a tool to improve forecast quality. There is a variety of different techniques currently implemented operationally, for instance:

- Ensemble forecasting: multiple numerical simulations are run with slightly varying initial conditions in order to generate an estimated probability density for the state of the atmosphere at a future time [9].
- Perturbed physics: a variation of the ensemble forecast technique, where parameters of a numerical scheme are described by random variables, i.e. each numerical simulation is based on slightly different model physics [10].
- Stochastic backscattering: numerical integration schemes and parameterizations lead to a systematic unphysical kinetic energy loss. Using autoregressive processes, a fraction of the dissipated energy is injected back into the model [11, 12].

Although these techniques lead to an improved forecasting skill [13], there are some physical and mathematical points of criticism. For instance, the concept of random variables and stochastic processes as well as the application of Itô and Stratonovich calculus is not clearly distinguished in some papers [14]. This is particularly important since stochastic processes require a specific numerical treatment, depending on the kind of calculus used. In a review paper on probabilistic climate predictions at the ECMWF the authors do not mention any stochastic calculus at all [15]. As an example for a physically questionable approach consider the following model, which has been implemented operationally in the ECMWF ensemble weather prediction scheme [15]

$$\frac{\partial}{\partial t}\Psi = A\Psi + \epsilon P\Psi.$$

In this schematic representation Ψ denotes a phase space state variable, e.g., temperature or components of velocity at all grid points, or their projections on specified basis functions. A describes the resolved dynamical terms, P the parameterized influences of the subgrid processes and ϵ is a random variable uniformly distributed in $[0.5, 1.5]$. Motivated by the typical scales for synoptic simulations, the random drawings are constant over a time period of $6h$ and a spatial domain of $10^\circ \times 10^\circ$ in latitude and longitude. Although Buizza et al. [10] show that this scheme has a positive impact on medium-range probability forecast skill scores for precipitation, it is debatable whether this approach is consistent with the physical conditions. Particularly the spatial autocorrelation structure seems questionable since an area of $10^\circ \times 10^\circ$ centered at the equator is about 23 times as large as a $10^\circ \times 10^\circ$ area centered at one of the poles, yet both are treated equally in terms of autocorrelation. A revised version of this scheme is discussed in [13], where ϵ closely follows a Gaussian distribution that is driven by stationary Gaussian auto-regressive processes in the spectral space. However, for perturbations with moderate amplitudes across the entire atmosphere numerical instabilities were discovered. The authors state: "*Further testing*

showed that the cause of the numerical instability are the perturbations in the lowermost part of the atmosphere. The reason is the delicate balance between model dynamics and vertical momentum transport which is established in the lowest model levels on timescales of the order of minutes. As a compromise between numerical stability and high probabilistic skill, the tendency perturbations were reduced towards zero close to the surface" [13, p.4]. These findings point to two possible underlying problems: first, the numerical integration scheme, which has been designed to solve a deterministic system, is not be able to handle the stochastic terms properly. Second, the specific stochastic formulation is not suited to describe the unresolved processes and is therefore not consistent with the model physics. We discuss both issues for the case of a conceptual one-dimensional model in the first half of this work.

1.1.2 Mathematical Rigorous Approach

In the last years awareness for the risks of using stochastic models without careful consideration of physical, mathematical and numerical aspects has grown. Penland and Ewald state: "*Simply replacing the fast term with a Gaussian random deviate with standard deviation equal to that of the variable to be approximated, and then using deterministic numerical integration schemes, is a recipe for disaster*" [16]. The application of mathematical rigorous methods during the development of stochastic models is of great interest and importance not only to investigate the validity of existing models but also to gain further understanding of the model's stochastic behavior. In the following we give a brief overview of approaches and results related to the work at hand.

In the paper "*An applied mathematics perspective on stochastic modeling for climate*" [17], Majda discusses a few systematic strategies for mathematical rigorous stochastic climate modeling. In particular a mode reduction technique (MTV) by Majda, Timofeyev and Vanden-Eijnden [18, 19, 20, 21, 22, 23, 24] used for stochastic modeling of the low-frequency variability of the atmosphere is presented. Starting with a deterministic system that can be orthogonally decomposed into subsystems acting on strongly differing time scales, the fast variables are truncated and represented by nonlinear Itô equations. This technique assumes ergodicity and mixing for the fast modes with Gaussian low-order statistics. Notably, the emerging statistics for the slow variables can very well be non-Gaussian, e.g., Sura applies this technique to explain non-Gaussian sea-surface temperature (SST) variability [25]. Using physically motivated regression fitting strategies [24] the conclusive additive and multiplicative stochastic processes (SPs) are dictated by the systems physics. This allows for a physically intuitive interpretation of the occurring stochastics: the additive noise originates from the linear interaction between the fast modes and the mean state of the slow modes, while the multiplicative noise stems from the advection of the slow modes by the fast modes, a phenomenon known as stochastic drift. This technique is a prime example how SPs can – and should – be based on the model's inherent physics.

On the topic of stochastic treatment of Rossby waves and their propagation on the sphere, Sardeshmukh, Penland and Newman [26, 27] discuss spatially homogeneous noise correlated in time. Monahan, Imkeller and Pandolfo [28] study the orthogonal case of spatially fluctuating noise that is homogeneous in time. Sardeshmukh, Penland and Newman thoroughly characterize suitable multiplicative noise terms for differential equations with a timescale separation ϵ. If the fast variable can be expressed in terms of stationary, centered and bounded stochastic processes, e.g., Ornstein Uhlenbeck processes, the system converges to a Stratonovich SDE for $\epsilon \to 0$ with a specified amplitude for the evolving white noise. The barotropic vorticity equation in spherical coordinates, linearized around the zonal mean flow reads

$$\frac{\partial}{\partial t}\zeta' = -\nabla \cdot (\bar{v}\zeta' + v'\bar{\zeta}) - r\zeta' + S_D.$$

Here ∇ denotes the horizontal gradient operator on the surface of the earth, ζ denotes the vertical component of absolute vorticity, v is the horizontal velocity, r denotes the frictional

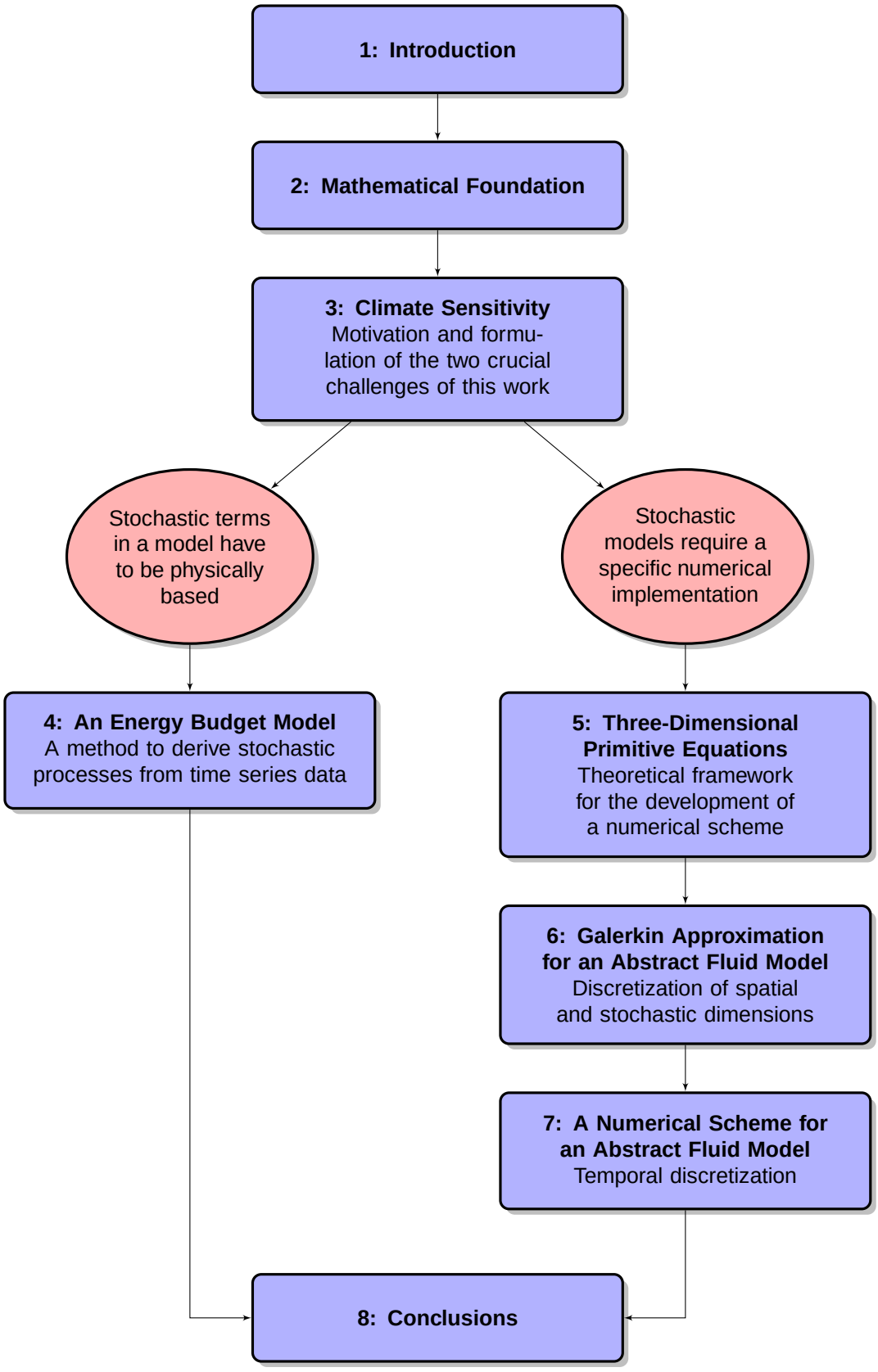
damping rate, and S_D symbolizes a steady Rossby-wave source. Overbars indicate zonal means, and primes denote deviations from the zonal means, i.e. Rossby waves [26, p.6]. Let u denote the zonal velocity component, Ω the Coriolis parameter, and a_e the earth's radius. The system is then modified by fluctuations in the superrotational flow by adding a stochastic contribution to the zonal mean of u:

$$\bar{u} = (u_0 + \eta)\Omega a_e \cos\Theta,$$

where η is, for instance, an Ornstein Uhlenbeck process. Comparing the ensemble mean of several numerical simulations to a deterministic control run shows an additional scale-dependent damping. In contrast, a stochastic disturbance of the frictional damping parameter $r = r_0 + \eta$ leads to a scale independent attenuation of damping. The influence of state dependent fluctuations, i.e. multiplicative noise, on the mean value of a stochastic system is known as *stochastic drift*.

Imkeller, Monahan and Pandolfo [28] tackle the phenomenon of a fluctuating background flow from a different angle by using stationary stochastic processes in the spectral space. They assume a spatially oscillating autocovariance structure that is constant in time. To derive the spectral model, the Fourier transformation was truncated to 50 modes. While sensitivity studies exhibit no change in results for an increased number of modes, a rigorous analysis would require quantitative error bounds for stochastic Galerkin transformations [29]. This is the starting point for the second half of this work, where we discuss this issue for the more general case of the three-dimensional primitive equations (PE) on the sphere.

1.2 Outline


1: Introduction
2: Mathematical Foundation
3: Climate Sensitivity
Motivation and formulation of the two crucial challenges of this work
Stochastic terms in a model have to be physically based
Stochastic models require a specific numerical implementation
4: An Energy Budget Model
A method to derive stochastic processes from time series data
5: Three-Dimensional Primitive Equations
Theoretical framework for the development of a numerical scheme
6: Galerkin Approximation for an Abstract Fluid Model
Discretization of spatial and stochastic dimensions
7: A Numerical Scheme for an Abstract Fluid Model
Temporal discretization
8: Conclusions

After the introduction we construct the mathematical framework necessary for a rigorous discussion of stochastic climate and weather models in Section 2. Starting from definitions of time constant random variables and time dependent stochastic processes we draw our attention to the topic of stochastic integration, with particular consideration of the differences between the two most common formalisms of stochastic integration: Itô and Stratonovich. This allows us to introduce ordinary and partial stochastic differential equations. To conclude the mathematical foundation we examine different concepts of convergence for numerical schemes in the context of stochastic differential equations.

In Section 3 we study stochastically induced instabilities by example of the most basic conceptual climate model, i.e. a one-dimensional linear equation for the global mean temperature, depending only on the climate sensitivity parameter. Several different stochastic formulations for the climate sensitivity are discussed, showing that even the analytical solutions of such a simple model exhibit instabilities in the case of unphysical stochastic terms. Therefore, a stochastic formulation that is inconsistent with the physical setting can result in instabilities regardless of a correct numerical implementation. Conversely, we focus on an instructive one-dimensional model where a physically reasonable stochastic system exhibits an unstable behavior due to an inapt numerical treatment. Although the analyzed scheme is specifically constructed for stochastic differential equations, a seemingly subtle assumption on the growth of its coefficient functions is not met. The so called "*Lipschitz Condition*" states that for every pair of points on the graph of a function, the slope of the line connecting both points is bounded by a global positive constant. The coefficient functions of a system have to obey this condition for the vast majority of numerical schemes. It is, however, not satisfied for many meteorological relevant systems – a fact that is often overlooked when choosing a numerical implementation for a given model. Section 3 should be understood as an extensive motivation emphasizing the two crucial statements

- Stochastic formulations in numerical weather prediction and climate modeling must not be arbitrarily chosen but have to be physically based.
- Stochastic models require a specific numerical treatment which fundamentally differs from deterministic schemes.

7 In Section 4 we focus on the first aspect, i.e. the derivation of a physically based stochastic process. We discuss the construction of specific stochastic terms based on time series data, utilizing well known results for autoregressive processes and spectral time series analysis. As an example, we consider a nonlinear energy budget model with zero spatial dimensions, driven by periodic radiative forcing and subject to varying atmospheric CO_2 levels. These fluctuations motivate a stochastic formulation, which is derived from three different sets of ice core data via a Yule-Walker algorithm for autoregressive processes. The resulting models are analyzed using their sample paths as well as statistical and spectral methods. It turns out that the three different sets of data yield three very distinct model behaviors. This emphasizes that we cannot expect to gain reliable results from models incorporating arbitrarily chosen stochastic formulations.

The second half of this work addresses the need for rigorous numerical schemes for meteorological relevant stochastic differential equations. We concentrate on the three-dimensional primitive equations of the atmosphere (PE), which form the core of virtually every GCM. Since every numerical simulations is limited by a finite grid and time resolution, the PE are subject to various stochastic parameterization approaches in order to account for unresolved subgrid processes or other sources of uncertainty. In Section 5 we derive an abstract operator framework on Hilbert spaces, which allows us to utilize recent results [7] guaranteeing existence and uniqueness of pathwise solutions for the stochastic PE. This abstract approach has the advantage to allow the implementation of a wide class of stochastic terms, including linear, non-linear, additive and multiplicative noise. Based on these theoretical findings we propose a Galerkin scheme for the discretization of spatial and stochastic dimensions in Section 6. The formalism to derive quantitative error bounds is based on the concept of *mild solutions*, which is the foundation for the recent progress

in the numerical treatment of stochastic partial differential equations [8]. However, convergence for the resulting schemes has been shown only for Lipschitz continuous coefficient function, which is not satisfied in the case of the PE. We prove that mean square convergence of a Galerkin approximation for the PE holds true and provide quantitative convergence rates in Theorem 6.10. The key to this result lies in the combinations of the control over the second order derivatives, which is provided by the mild formalism, and anisotrop Sobolev estimates for the nonlinear advection term. In the aforementioned work by Imkeller, Monahan and Pandolfo [28] sensitivity experiments were conducted in order to decide whether a Galerkin approximation using 50 modes yields reliable results. Theorem 6.10 guarantees the convergence of such an approximation as well as quantitative convergence rates and therefore provides a rigorous alternative to the practical approach using sensitivity studies.

In Section 7 we extend the Galerkin approximation by a temporal discretization and derive mean square convergence for the resulting numerical scheme. However, we have to postulate a priori bounds for the numerical solution in order to obtain error bounds for the temporal discretization, which is a well known issue of explicit numerical simulations of non-Lipschitz continuous systems. Under these assumptions we derive quantitative convergence rates regarding discretization of spacial, stochastic and temporal dimensions in Theorem 7.4. We conclude this work with a comprehensive discussion on the realism of the postulated a priori bounds and possible future extensions that hold the promise to yield convergent schemes without the need for additional assumptions.

2 Mathematical Foundation

2.1 Stochastic Processes

In this section we recall the basic vocabulary of probability theory, including random variables and stochastic processes (SPs), which play a central role throughout this work. After each block of definitions we provide some intuitive explanations and instructive examples to illustrate the concept behind the mathematical framework. We start with the classical framework of measure theory by defining

Definition 2.1. σ-Algebra
Let Ω be a set and $\mathcal{P}(\Omega)$ its power-set, i.e., the set off all subsets of Ω. A subset $\mathcal{F} \subset \mathcal{P}(\Omega)$ is called a σ-algebra (over Ω) if

- $\Omega \in \mathcal{F}$ (*$\mathcal{F}$ is not empty*)
- $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ (*$\mathcal{F}$ is closed under complementation*)
- $A_1, A_2, ... \in \mathcal{F} \Rightarrow \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}$ (*$\mathcal{F}$ is closed under countable unions*)

Definition 2.2. Measure
Let Ω be a set and $\mathcal{F}$ a σ-algebra over Ω. A function $\mu : \mathcal{F} \to [-\infty, +\infty]$ is called a measure if

- $\mu(A) \geq 0, \quad \forall A \in \mathcal{F}$ (*nonnegativity*)
- $\mu(\emptyset) = 0$ (*empty set has measure zero*)
- $\mu\left(\bigcup_{i \in \mathcal{I}} A_i\right) = \sum_{i \in \mathcal{I}} \mu(A_i)$, *for all countable index-sets $\mathcal{I}$* (σ-*additivity*)

The pair $(\Omega, \mathcal{F})$ is called a measurable space.

Definition 2.3. Measurable Function
Let $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$ be measurable spaces. A function $f : \Omega \to S$ is $(\mathcal{F}, \mathcal{S})$-measurable if $f^{-1}(A) \in \mathcal{F}$ for every $A \in \mathcal{S}$.

Definition 2.4. Probability Space
Let $(\Omega, \mathcal{F})$ be a measurable space and $P : \mathcal{F} \to [0, 1]$ a measure satisfying $P(\Omega) = 1$. Then P is called a probability measure and the triple $(\Omega, \mathcal{F}, P)$ is called a probability space.

Definition 2.5. P-almost surely
Let $(\Omega, \mathcal{F}, P)$ be a probability space. An event $A \in \mathcal{F}$ occurs P-almost surely (P-a.s.) if $P(A) = 1$.

In the context of probability theory σ-algebras are often identified with the "*available information*"'. Following this line of thought a random variable is measurable if and only if its value is knowable based on the available information. To gain an intuitive understanding of the term "*P-almost sure*"' we consider the exemplary case of a uniform distribution on the interval $[0, 1]$. The probability to hit a single number is zero, since there is an infinite number of possibilities. Nevertheless the event to hit a single number is of course not impossible. As this example suggests, for most finite applications the distinction of "*almost surely*"' and "*surely*"' has no practical consequences. However it is important when dealing with infinite systems, e.g. when studying convergence behavior of random variables or time continuous stochastic processes.

Definition 2.6. Random Variable
Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(S, \mathcal{S})$ be a measurable space. A function $X : \Omega \to S$ is called a S-valued random variable if it is $(\mathcal{F}, \mathcal{S})$-measurable.

Definition 2.7. Stochastic Process
Let $(\Omega, \mathcal{F}, P)$ be a probability space, $(S, \mathcal{S})$ a measurable space and $\mathcal{T}$ a totally ordered set.

- *Define a stochastic process as a family*

  $$X = \{X_t | t \in \mathcal{T}\}$$

  of S-valued random variables X_t on Ω, indexed by $\mathcal{T}$.

- *A family $(\mathcal{F}_t)_{t\in\mathcal{T}}$ of σ-algebras is called a filtration if $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $s, t \in \mathcal{T}$ with $s < t$.*

- *A stochastic process X is adapted to a filtration $(\mathcal{F}_t)_{t\in\mathcal{T}}$ if X_t is $(\mathcal{F}_t, \mathcal{S})$-measurable for all $t \in \mathcal{T}$.*

Definition 2.8. Stationarity
Let $\{X_t\}_{t\geq 0}$ be a stochastic process on a probability space $(\Omega, \mathcal{F}, P)$.

- *X is called weakly stationary if*

  $$\begin{aligned} E[X_{t_1}] &= E[X_{t_1+\tau}] \\ E[X_{t_1} X_{t_2}] &= E[X_{t_1+\tau} X_{t_2+\tau}] \end{aligned}$$

  for all $t_1, t_2, \tau > 0$.

- *For $\tau > 0$, $0 \leq t_1 < \ldots < t_n$ and $x_1, \ldots, x_n \in \mathbb{R}$ let $F_{X_{t+\tau}}(x_1, \ldots, x_n)$ be the cumulative distribution function of the joint distribution of $\{X_{t_1+\tau}, \ldots, X_{t_n+\tau}\}$, i.e.,*

  $$F_{X_{t+\tau}}(x_1, \ldots, x_n) = P(X_{t_1+\tau} \leq x_1, \ldots, X_{t_n+\tau} \leq x_n).$$

  X is called (strongly) stationary if

  $$F_{X_{t+\tau}}(x_1, \ldots, x_n) = F_{X_t}(x_1, \ldots, x_n),$$

  for all $n \in \mathbb{N}$, $\tau > 0$, $0 \leq t_1 < \ldots < t_n$ and $x_{t_1}, \ldots, x_{t_n} \in \mathbb{R}$.

A stochastic process is a collection of random variables, often used to describe the evolution of a random variable or system in time. The distribution of a (strongly) stationary process does not depend on time. However, this does not imply that a "*realization*", or "*sample path*", of such a process is constant in time, i.e.,

$$X \text{ stationary } \not\Rightarrow X_t(\omega) = X_{t+\tau}(\omega), \quad \text{for} \quad \omega \in \Omega.$$

Weakly stationary processes play an important role in signal theory and time series analysis. They only require stationarity of the first two moments of a process. In particular, this leads to a clearly arranged autocovariance structure:

$$\begin{aligned} \text{cov}(X_{t_1}, X_{t_2}) &= E[X_{t_1} X_{t_2}] - E[X_{t_1}] E[X_{t_2}] \\ &= E[X_{t_1-t_2} X_0] - E[X_{t_1-t_2}] E[X_0] \\ &= \text{cov}(X_{t_1-t_2}, X_0). \end{aligned}$$

Therefore the autocovariance does only depend on the "*time lag*", i.e., the difference between two points in time. An important example for a stationary processes is the Ornstein Uhlenbeck process, which is the topic of discussion in Section 3.3.2. Another fundamental stochastic process is the Brownian motion, which is not stationary itself, but has stationary increments, as defined below.

Definition 2.9. Brownian Motion
A time continuous stochastic process $(W_t)_{t\geq 0}$ on a probability space $(\Omega, \mathcal{F}, P)$, adapted to a filtration $(\mathcal{F}_t)_{t\geq 0}$, is called a Wiener process or Brownian Motion if the following holds true $P-$almost surely:

- $W_0 = 0$
- $W_t - W_s \sim \mathcal{N}(0, t-s)$ *is independent of $\mathcal{F}_s$, for all* $0 \leq s < t$
- W *has continuous sample paths.*

Proposition 2.10. Basic Properties of the Brownian Motion
For a Brownian motion W on a probability space $(\Omega, \mathcal{F}, P)$, adapted to a filtration $\mathcal{F}_t$, the following holds true for all $s, t > 0$:

1. $E[W_t] = 0$ *and* $E[W_s W_t] = \min(s, t)$
2. *The process* $s^{-\frac{1}{2}} W_{st}$ *is a Brownian motion with the same distribution as* W_t *for all* $t \geq 0$ *(scaling).*
3. *The process* $W_{s+t} - W_s$ *is a Brownian motion, which is independent of* $(W_u, 0 \leq u \leq s)$ *for all* $t \geq 0$ *(shifting).*
4. *The process* $W_{s-t} - W_s$ *has the same distribution as* W_t *for all* $0 \leq t \leq s$ *(time reversal).*

The basic statistic properties and a few sample-paths of a Brownian motion are visualized in Fig. 1. Before we turn our attention onto the topic of stochastic integration we would like to state Jensen's inequality, which is an essential tool in stochastic analysis and measure theory.

Theorem 2.11. Jensen Inequality
Let X be a real valued random variable on the probability space $(\Omega, \mathcal{F}, P)$ with

$$E[|X|] < \infty.$$

Let f be a convex function. Then

$$f(E[X]) \leq E[f(X)].$$

If f is strictly convex equality holds if and only if X is a constant.

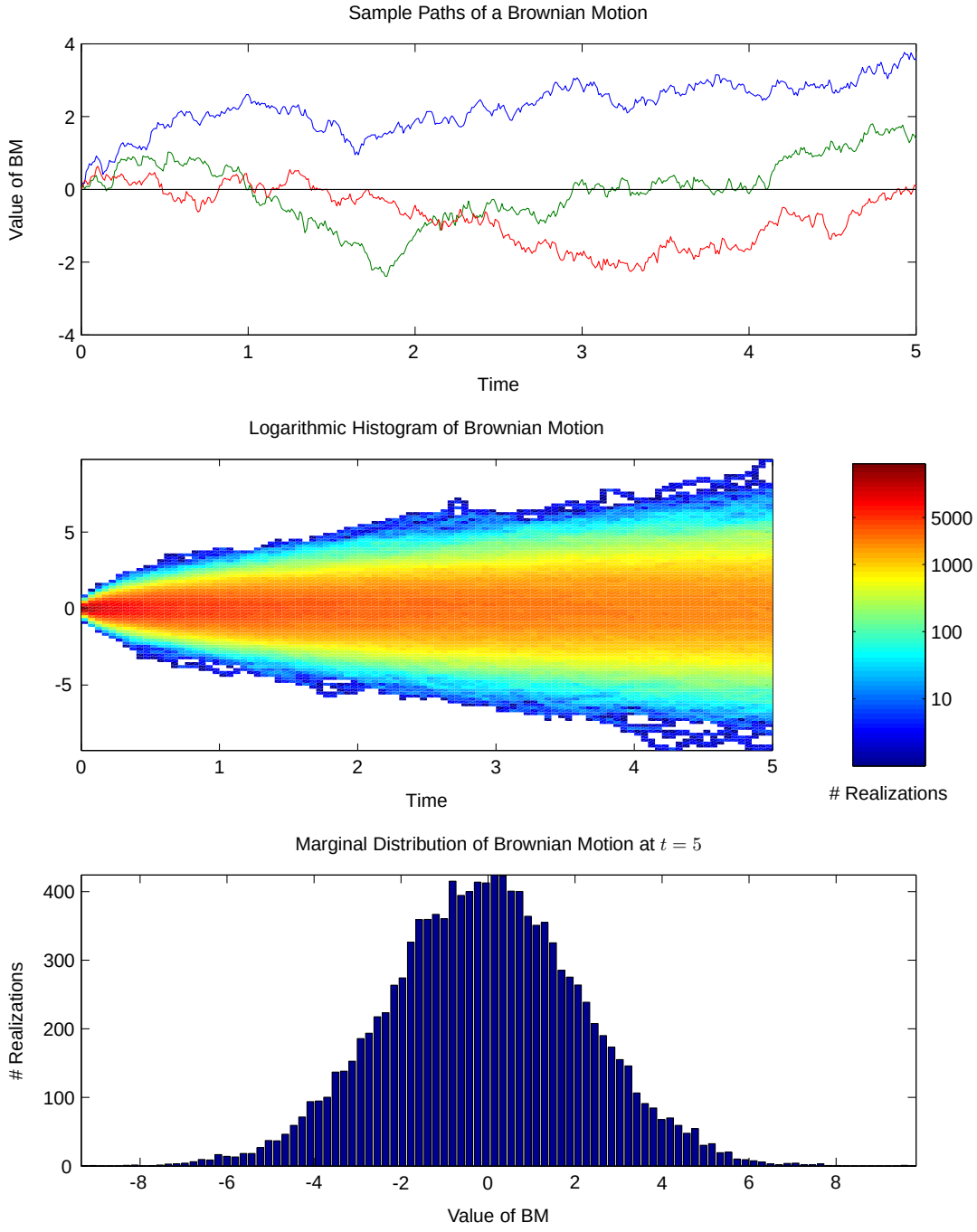


Figure 1: *Visualization of a Brownian Motion*
The first plot shows three realizations of a Brownian motion with their typical rough and noisy fluctuations. The path-density in the second plot, which was derived from 10.000 realizations, indicates Gaussian marginal distributions with zero mean and a variance, which is increasing in time. Finally, the third plot shows that the marginal distribution at $t = 5$ is indeed Gaussian.

2.2 Stochastic Integration

Introducing stochastic processes into differential equations naturally leads to the occurrence of stochastic integrals, e.g.

$$\int f dW_t = ?$$

Unfortunately the assumptions of the classical deterministic approach are not satisfied due to the strongly oscillating paths of Brownian motions. We discuss this issue in more detail to motivate the need for an integral calculus for stochastic processes, which differs from the deterministic calculus.

Definition 2.12. Bounded Variation
A function g on an interval $[a, b] \subset \mathbb{R}$ has bounded variation if its total variation

$$V_a^b(g) = \sup_{\tau \in \mathcal{P}} \sum_{i=1}^{n} |g(t_i) - g(t_{i-1})|,$$

where

$$\mathcal{P} = \{\tau = (t_0, t_1, ..., t_n) \mid \tau \text{ is a partition of } [a, b]\}$$

denotes the set of all possible partitions of the interval $[a, b]$, is finite, i.e.

$$V_a^b(g) < \infty.$$

For a sequence of partitions

$$(\tau_n)_{n \geq 1}, \quad \tau_n = (t_0, t_1, ..., t_n)$$

and intermediate points

$$\{s_i\}_{i=1}^n, \quad s_i \in [t_{i-1}, t_i]$$

we recall the definition of the Stieltje integral

$$\int f dg = \lim_{n \to \infty} \sum_{i=1}^{n} f(t_{i-1})(g(t_i) - g(t_{i-1})),$$

which converges for continuous f and g with bounded variation. A visual description for functions of bounded variation is that their graph does not oscillate too strongly. Unfortunately a well known fact for the Brownian motion states

Proposition 2.13. Regularity of the Brownian Motion
For a Brownian motion W on a probability space $(\Omega, \mathcal{F}, P)$, adapted to a filtration $\mathcal{F}_t$ the following holds true P-a.s.:

- *The paths of W are nowhere differentiable.*
- *The function W has unbounded variation on every interval.*

Therefore we cannot use the deterministic definition without further considerations. It turns out that the values of the convergent sums depend on the choice of the intermediate points $\{s_i\}_{i=1}^n$, see Section 2.3 for an instructive example. Furthermore, the choice of intermediate points in such a pathwise integral definition leads to distinct differential calculi where, for instance, the fundamental chain rule of differentiation does not hold true. In the following we give a descriptive definition of the two most common stochastic integral

types: Itô and Stratonovich. For a rigorous (and necessary more technical definition) we refer to [30] or [31]. Throughout this section, let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathcal{F}_t)_{t\geq 0}$, and let $(S, \mathcal{S})$ be a measurable space. Before defining the different integral types, we introduce two concepts from the mathematical field of time continuous stochastic processes: "*stopping times*"' and "*martingales*"'.

Definition 2.14. Càdlàg Functions
A function f on a set $\mathcal{T}$ is called càdlàg (abbreviation for the French expression "continue à droite, limite à gauche"), i.e., right-continuous with left limits, if

$$\lim_{s\uparrow t} f(s) \quad \textit{exists, and} \quad \lim_{s\downarrow t} f(s) = f(t)$$

for all $t \in \mathcal{T}$.

Definition 2.15. (Conditional) Expectation
Let $X : \Omega \to [-\infty, +\infty]$ be a P-integrable random variable. Then the expectation value of X with respect to P is defined by

$$E[X] = \int_\Omega X(\omega) dP(\omega).$$

Let $\mathcal{A} \subset \mathcal{F}$ be a sub-σ-algebra. A random variable $E[X|\mathcal{A}]$ is called conditional expectation of X given $\mathcal{A}$ if

$$\int_A E[X|\mathcal{A}](\omega) dP(\omega) = \int_A X(\omega) dP(\omega)$$

for all $A \in \mathcal{A}$.

The conditional expectation follows the intuition gained by the concept of conditional probability, e.g. in the context of Bayesian statistics [32]. Let A and B be two events with $P(B) > 0$. Then the conditional probability

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

induces the conditional expectation value for a random variable X

$$E[X|B] = \frac{E[\mathbb{1}_B \cdot X]}{P(B)}.$$

However, in the context of stochastic processes we often need a more general definition. For continuous probability distributions we can observe infinitesimal terms of the form

$$P(A|X = x), \quad P(X = x) = 0,$$

as well as events conditioned on a random variable:

$$P(A|X) \quad \text{or} \quad E[Y|X].$$

Therefore the abstract Definition 2.15 is used, where the Radon-Nikodym Theorem [33] yields the existence of such a conditional expectation value under suitable assumptions (e.g. the existence of the unconditional expectation of X). The conditional probability for an event A, given the information $\mathcal{A}$, can then be derived from Definition 2.15 via

$$P(A|\mathcal{A}) = E[\mathbb{1}_A | \mathcal{A}].$$

Definition 2.16. Stopping Time
Let $\mathcal{T}$ be a totally ordered set, e.g. $[0, \infty)$. A random variable $\tau : \Omega \to \mathcal{T}$ is called a $\mathcal{F}_t$-stopping time if $\{\tau \leq t\} \in \mathcal{F}_t$ for all $t \in \mathcal{T}$, i.e., $\{\tau \leq t\}$ is $\mathcal{F}_t$-measurable.

Recall the interpretation of a σ-algebra as the "*available information*". Then the concept of stopping times formalizes the notion of a "*random time*", which satisfies the following criterion: if a random variable τ is a stopping time one can decide whether or not the event $\{\tau \leq t\}$ has occurred based on the information available at time t. A classical and important example of stopping times are "*hitting times*": consider a Brownian motion W and define τ as the point in time when W leaves the interval $(-a, a)$ for the first time, i.e.

$$\tau = \inf\{t \geq 0 \ : \ |W_t| \geq a\}.$$

Then τ is a stopping time. On the other hand define for a $T > 0$

$$\xi = \inf\{t \in [0, T] \ : \ |W_t| = \sup_{s \in [0,T]} W_s\},$$

i.e., the first time at which the Brownian motion obtains its maximum over a set interval. Then the event $\{\xi \leq t\}$ depends on events during $s \in (t, T]$ and is therefore not knowable based on the information available at time t and consequently no stopping time.

Definition 2.17. Martingale
A $\mathcal{F}_t$-adapted stochastic process $X : T \times \Omega \to S$ is called a martingale if

- $E\left[|X_t|\right] < \infty$ (*L^1-integrable*)
- $X_s = E\left[X_t | \mathcal{F}_s\right]$ (*fair*)

for all $s < t$.

Definition 2.18. Local Martingale
A $\mathcal{F}_t$-adapted stochastic process $X : T \times \Omega \to S$ is called a local martingale if there exists a sequence of $\mathcal{F}_t$-stopping times τ_n such that the following holds true P-a.s.:

- $\tau_{n-1} < \tau_n$, *for all* $n \in \mathbb{N}$ (*increasing*)
- $\lim_{n \to \infty} \tau_n = \infty$ (*divergent*)
- *The stopped process $X(\min(t, \tau_n))$ is a $\mathcal{F}_t$-martingale for all $n \in \mathbb{N}$.*

The etymology of the word "martingale" is not entirely known, but a plausible trail leads to the Provencal expression "*jouga a la martegalo*", which means "to play in an abstract and incomprehensible way" [34]. This is further supported by a French-English dictionary from 1611 [35], where the expression "à la martingale" is translated as "absurdly, foolish". The word "*martingale*" was first used in a mathematical context by Ville in 1939 [36], who stated elsewhere [37] that he borrowed the expression from the vocabulary of gamblers. In fact, the fourth edition of the dictionary of the Académie Francaise, which was published in 1762, states "*to play the martingale is to always bet all that was lost*" [38], a strategy that may indeed seem absurd and foolish. Since then gambling and the mathematical martingale concept has been closely connected. This lead to important results such as the "*optional stopping theorem*" by Doob's [39] as well as interesting interdisciplinary discussions, see e.g. the "*St. Petersburg paradox*" [40]. Furthermore, the context of gambling offers a very intuitive interpretation of a martingale as a "*fair game*". Consider a game where the player starts with a capital M_0 and the win or loss of each round is described by a random variable X_i, for $i = 1, 2, ...$, which are assumed to be pairwise independent. The capital after n rounds is then given by

$$M_n = M_o + \sum_{i=1}^{n} X_i.$$

A game is *fair* if the expected win/loss in each round equals zero, i.e., $E[X_i] = 0$ for all $i = 1, 2, ...$ Suppose we have played n rounds. Then the expected capital after $n+1$ rounds is given by

$$E[M_{n+1}|M_n] = E[M_n + X_{n+1}|M_n] = M_n + E[X_{n+1}] = M_n,$$

which corresponds to Definition 2.17. Regarding the integrability assumption for martingales note that since the function $f(x) = x^p$ is convex for all $x \geq 0$ and $p \geq 1$ Jensen's inequality (Theorem 2.11) implies that any L^p-integrable random variable with $p > 1$ is also L^1-integrable. The "*fairness*" characteristic of martingales is very useful for many calculations involving expectation values of random variables. We note without proof a well known fact regarding martingales of Brownian motions, which is useful for the study of the statistic characteristics of stochastic processes.

Proposition 2.19. Three Martingales w.r.t. a Brownian Motion
Let B be Brownian motion adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$. Then the following processes are $\mathcal{F}_t$-martingales

- $\{B_t |\, t \geq 0\}$ (linear martingale)
- $\left\{(B_t)^2 - t \middle|\, t \geq 0\right\}$ (quadratic martingale)
- $\left\{e^{\lambda B_t - \lambda^2 t/2} \middle|\, t \geq 0\right\}$ (linear martingale)

Square-integrable martingales are the foundation for an abstract framework, which bears their name and rigorously defines stochastic integrals (including well-posedness and existence) for very general stochastic processes. We give a more intuitive, pathwise integral definition, which can be identified with the ones stated within the martingale framework [31]. To this end we need the following generalization of a martingale, which allows for an "*unfair*" but "*well behaved*" contribution.

Definition 2.20. Semimartingale
A real valued, $\mathcal{F}_t$-adapted stochastic process X is called a semimartingale if there exists a decomposition

$$X_t = M_t + A_t,$$

with a local martingale M and a càdlàg, adapted process A with bounded variation. An $\mathbb{R}^n$-valued process is a semimartingale if each of its components is a semimartingale.

Regarding the stochastic integral of a function f with respect to a semimartingale X, this decomposition leads to

$$\int f dX_t = \int f dM_t + \int f dA_t.$$

Then the first contribution is well defined by the aforementioned martingale framework, while the second integral can be defined in a classical pathwise sense due to the bounded variation of A. We can now finally state

Definition 2.21. Itô Integral
Let X, Y be two real-valued stochastic processes on the time interval $[t_0, T]$ and $\{\tau_n\}_{n\geq 1}$ a sequence of partitions

$$t_0 = t_0^n < t_1^n < ... < t_n^n = T$$

such that $\sup_{i\leq n} |t_i^n - t_{i-1}^n| \to 0$ for $n \to \infty$. The stochastic Itô integral is defined by

$$\int_{t_0}^{T} X_t dY_t := \lim_{n\to\infty} \sum_{i=1}^{n} X(t_{i-1}) \left(Y(t_i) - Y(t_{i-1})\right).$$

One can show that the Itô integral is well defined, i.e., independent of the choice of partitions, and converges in probability, if $Y : \Omega \times [t_0, T] \to \mathbb{R}$ is a semimartingale and if $X : \Omega \times [t_0, T] \to \mathbb{R}$ is a left continuous, locally bounded and adapted process.

Definition 2.22. Stratonovich Integral
Under the assumptions of definition 2.21 we define the Stratonovich integral by

$$\int_{t_0}^{T} X_t \circ dY_t := \lim_{n\to\infty} \sum_{i=1}^{n} X\left(\frac{t_i + t_{i-1}}{2}\right)(Y(t_i) - Y(t_{i-1})).$$

Definition 2.23. Stochastic Differential Equation
The stochastic differential equations

$$dX_t = f(t, X_t)dt + g(t, X_t)dY_t$$
$$dX_t = f(t, X_t)dt + g(t, X_t) \circ dY_t$$

should be understood as an abbreviated notation of the corresponding integral equations

$$X_t - X_s = \int_s^t f(u, X_u)du + \int_s^t g(u, X_u)Y_u$$
$$X_t - X_s = \int_s^t f(u, X_u)du + \int_s^t g(u, X_u) \circ dY_u$$

for all $s < t$, with f, g and Y such that the integrals are well defined, i.e., satisfying the assumptions in Definition 2.21.

We study the differences between both integral types in more detail during Section 2.3. It turns out that it is no trivial task to choose the correct integral for a given physical system. However, for a given SDE, i.e., after the integral type was chosen, the following important result allows us to freely choose the calculus in which we want to analyze the system.

Lemma 2.24. Equivalence of Itô and Stratonovich
The Stratonovich SDE

$$dX_t = \underline{a}_t(t, X_t)dt + b_t(t, X_t) \circ dW_t$$

has the same solutions as the Itô SDE

$$dx_t = a_t(t, X_t)dt + b_t(t, X_t)dW_t$$

if and only if

$$\underline{a}_t(t, X_t) = a_t(t, X_t) - \frac{1}{2}b_t(t, X_t)\frac{\partial}{\partial x}b_t(t, X_t).$$

Note that the Stratonovich calculus follows the rules of the deterministic differential calculus whereas the Itô integral exhibits a different behavior, which manifests itself in the famous Itô formula (Theorem 2.28) stated below.

Definition 2.25. Itô Process
Let W be a Brownian motion adapted to the filtration $\mathcal{F}_t$.

1. *A $\mathcal{F}_t$-adapted process B is called a H^2-process on the interval $[0, T]$ if*

$$\int_0^T E\left[B_t^2\right] dt < \infty.$$

2. *A stochastic process X is called Itô process (on the interval $[0, T]$) if it has a decomposition*

$$X_t = X_0 + \int_0^t A_s ds + \int_0^t B_s dW_s, \quad t \in [0, T]$$

with an $\mathcal{F}_t$-adapted and Lebesgue-integrable process A and a H^2-process B.

Theorem 2.26. Itô Isometry
Let W be a Brownian motion and X an Itô-process, both adapted to the filtration $\mathcal{F}_t$. Then

$$E\left[\left(\int_0^t X_s dW_s\right)^2\right] = \int_0^t E\left[X_s\right]^2 ds,$$

for all $t \in [0,T]$.

Theorem 2.27. Martingale Characteristic
Let W be a Brownian motion and X an Itô-process, both adapted to the filtration $\mathcal{F}_t$ with

$$X_t = X_0 + \int_0^t A_s ds + \int_0^t B_s dW_s, \quad t \in [0,T].$$

1. *Then X is a martingale (relative to the filtration $\mathcal{F}_t$) if and only if A_s=0 P-a.s. for almost every $t \in [0,T]$.*
2. *Consequently every Itô integral of a H^2-process B is a martingale. In particular we have*

$$E\left[\int_0^t B_s dW_s\right] = 0,$$

for all $t \in [0,T]$.

Theorem 2.28. One-Dimensional Itô Formula
Let X be an Itô-process with

$$X_t = X_0 + \int_0^t A_s ds + \int_0^t B_s dW_s, \quad t \in [0,T]$$

and $f \equiv f(t,x)$ a function continuously differentiable in t and twice continuously differentiable in x. Define $Y_t = f(t, X_t)$ for all $t \in [0,T]$. Then Y is an Itô-process satisfying

$$Y_t = Y_0 + \int_0^t \frac{\partial}{\partial t} f(s, X_s) + A_s \frac{\partial}{\partial x} f(s, X_s) + \frac{B_s^2}{2} \frac{\partial^2}{\partial x^2} f(s, X_s) ds + \int_0^t B_s \frac{\partial}{\partial x} f(s, X_s) dW_s,$$

for all $t \in [0,T]$.

The Itô formula corresponds to the chain rule of the deterministic (and Stratonovich) differential calculus, where the difference is given by the term

$$\int_0^t \frac{B_s^2}{2} \frac{\partial^2}{\partial x^2} f(s, X_s) ds.$$

This can be understood as an additional drift contribution due to the stochastic integral and is therefore known as "*stochastic drift*". We have encountered a closely related term in Lemma 2.24 in the drift relation between Itô and Stratonovich systems. The Itô isometry and martingale characteristics are the reason why the Itô calculus has proven very successful in both theoretical framework and numerical applications: the Itô isometry yields a natural environment for a "*mean-square-calculus*", offering the derivation of strong convergence results for numerical schemes. Note that there are various generalizations of these results, most notably the multidimensional case (Theorem 4.2.1 in [30]), general L^p-norm estimates, e.g. the Burgholder-Davis Gundi inequality (Theorem 74 in [41]), and semimartingale integrators (Theorem 33 in [41]). The latter together with the martingale characteristic allows for a "*closed*" theoretical framework, which is one of the main reasons for the success of the Itô calculus in the mathematical community.

2.3 Itô, Stratonovich and Beyond

The decision which stochastic integral is appropriate to model a given physical setting has "*attracted considerable attention in the physics community*" during the last 30 years and "*is still as elusive as ever*" [42]. We present a summarizing discussion on this matter, briefly mentioning an alternative approach "*beyond Ito vs. Stratonovich*" [43], which enjoys some success in complex biological models.
As an instructive example consider the SDE

$$dX_t = f(X_t)dt + g(X_t)dW_t,$$

where W is a Brownian motion on a probability space $(\Omega, \mathcal{F}, P)$. Since Itô and Stratonovich only differ for nonconstant g, we treat the most basic case, where g is a function of the noise. This leads to stochastic integrals of the type

$$\int_0^T W_t \, dW_t = \lim_{n\to\infty} \sum_{i=1}^{n} W_{\xi_i} \left(W_{t_i} - W_{t_{i-1}}\right),$$

where τ_n is a partition of $[0, T]$. The integrand W is evaluated at time

$$\xi_i \equiv \xi(\alpha, i) = t_{i-1} + \alpha \left(t_i - t_{i-1}\right) \in [t_{i-1}, t_i], \tag{2.1}$$

for an $\alpha \in [0, 1]$. Itô corresponds to $\alpha = 0$ while we obtain the Stratonovich integral for $\alpha = 1/2$. Using the basic property of Brownian motions (Proposition 2.10)

$$E\left[W_t W_s\right] = \min(s, t),$$

we obtain

$$E\left[\sum_{i=1}^{n} W_{\xi_i} \left(W_{t_i} - W_{t_{i-1}}\right)\right] = \sum_{i=1}^{n} (\xi_i - t_{i-1}) = \sum_{i=1}^{n} \alpha(t_i - t_{i-1}) = \alpha T.$$

Obviously, the value of the integral depends on the choice of α. It is therefore crucial at which point in time we evaluate the integrand when dealing with stochastic integrals. A direct consequence is the zero mean characteristic of the Itô integral, or more precisely: every Itô integral is a local martingale. This is the fundamental reason for the success of the Itô calculus in the mathematical community for two reasons: first, it is very convenient from a technical point of view since every mean value analysis is free of stochastic integrals. Second, it offers the natural and rigorous framework for local martingale integrators. Nevertheless, the Stratonovich integral has the practical advantage, that the rules of the deterministic differential calculus apply. Due to Lemma 2.24 we can transform each Itô system into a Stratonovich system and vice versa. However, when writing down a SDE as a model for a specific physical system, the coefficient functions f and g are usually given by the dynamics of the system. Note that the drift coefficient f changes during the transformation between Itô and Stratonovich, i.e., the choice of α in (2.1) does matter for a *given* f. Therefore, the decision which calculus to choose has to be based on physical (or biological, chemical, etc.) rather than on mathematical considerations.

As indicated above, this dilemma of choice has been subject to debate during the last 30 years. Although there seems to be no universal answer so far, the extensive debate has yielded some guidelines. Define the linear interpolated approximation of the Brownian motion W_t by

$$W_t^n = W_{t_{i-1}} + \frac{W_{t_i} - W_{t_{i-1}}}{t_i - t_{i-1}}, \quad t \in [t_{i-1}, t_i], \quad i \in \{1, 2, ...n\}.$$

Then a theorem of Wong and Zakai (see for instance [44]) states that the sequence $\{X^n\}_{n\geq 1}$ of approximation processes

$$X_T^n = X_0^n + \int_0^T f(X_t^n)dt + \int_0^T g(X_t^n)dW_t^n$$

converges to the solution of the Stratonovich SDE

$$dX_t = f(X_t)dt + g(X_t) \circ dW_t.$$

Miller states in a paper on stochastic processes in oceanography that "*the Stratonovich form should be used if the white noise model is derived as the limit of a sequence of ordinary integrals.*" [45]. Therefore, Stratonovich seems to be the right choice for many continuous physical models, see also [46, 47, 48, 49]. However, most of this work concentrates on stochastic differential equations in finite dimensions since a rigorous stochastic calculus for infinite dimensions was not well developed back then. On the other hand Itô is widely regarded as the right choice for many discrete systems and finance models, see for instance [50] and [51, 52]. Recently there has been a lively discussion following the work of Volpe et al., who published empirical evidence from physical experiments at nanometer scale supporting the case $\alpha = 1$. As a result of these findings a further stochastic integral, the so-called "*A-integration*" [53, 54, 43], is currently a topic of high interest in theoretical biology. An advantage of this type of integration is the "*correspondence between stochastic and deterministic dynamics, for example, fixed points are not changed*" [43]. Furthermore the authors state that "*the new FPE is generally not reachable by the α-type integration in higher dimensions*", which complicates matters from a mathematical point of view, since we cannot expect a convenient transformation analogous with Lemma 2.24. However, a recent review paper [42] on this area, where the authors "*discuss critically some of the most recent contributions*" states "*that some of the new findings are not well based*". Therefore, we restrict the analysis in the present work to Itô and Stratonovich integrals, while the aforementioned $A-$integration points out, that these classical approaches are only two of many possible ways to interpret a stochastic integral. We would like to emphasize that for any practical application the decision which calculus to use has to be made individually. The results are valid for either integration type, although a drift correction may be necessary via Lemma 2.24.

2.4 Convergence of Random Variables

In most meteorological relevant cases we are not able to derive analytical solutions of a SDE, but have to rely on numerical approximations. This begs the question how accurate a numerical scheme solves a given equation and, as a prerequisite, how "*accuracy*" can be measured in the framework of stochastic processes. Regarding the latter we recall commonly used types of convergence for a sequence of random variables. This allows us to give a precise definition of various convergence concepts in the environment of SDE. Let X and $X_1, X_2, ...$ be random variables on a probability space $(\Omega, \mathcal{F}, P)$. One of the natural ways to measure the convergence $X_n \to X$ for $n \to \infty$ is given by the "*convergence on probability*", where the difference $|X_n - X|$, which itself is a random variable, should be arbitrarily close to zero with a probability arbitrarily close to one.

Definition 2.29. Convergence in Probability

$$X_n \xrightarrow{P} X \quad :\Leftrightarrow \quad P(|X_n - X| > \epsilon) \to 0, \quad \textit{for all} \quad \epsilon > 0.$$

The above definition only makes sense if all random variables are defined on the same probability space. To circumvent this assumption we consider the distribution functions $F(x) = P(X \leq x)$ and $F_n(x) = P(X_n \leq x)$, for $n = \{1, 2, ...\}$, which are nondecreasing, right-continuous and bounded to the interval $[0, 1]$.

Definition 2.30. Convergence in Distribution

$$X_n \xrightarrow{d} X \quad :\Leftrightarrow \quad F_n(x) \to F(x),$$

for all x at which $F(x)$ is continuous. This is also known as "convergence in law" and sometimes denoted by $X_n \xrightarrow{\mathcal{L}} X$.

These two concepts are very closely related. In fact, convergence in distribution follows from convergence in probability and both coincide if the sequence of random variables converges to a constant, see Theorem 2.32. The latter case is of interest in the context of numerical schemes since the approximation error is usually estimated by a constant upper bound. Another important concept is the "*convergence in mean*":

Definition 2.31. Convergence in Mean
Let r be a positive constant. Then

$$X_n \xrightarrow{r} X \quad :\Leftrightarrow \quad E\left[|X_n - X|^r\right] \to 0.$$

Note that this is *not* equivalent to $E[X_n] \to E[X]$. Particularly important is the special case $r = 2$, which is known as "*mean square convergence*". First, Theorem 2.26 shows that this concept is the natural one within the Itô calculus, which implies that it is often easy to check. Second, as a direct consequence of the Markov inequality, this type of convergence is stronger then the previous ones, which is summarized in the following result.

Theorem 2.32. Relations Between the Various Convergence Concepts

1. $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$
2. $X_n \xrightarrow{d} c \Leftrightarrow X_n \xrightarrow{P} c$, *for a constant* c
3. $X_n \xrightarrow{r} X \Rightarrow X_n \xrightarrow{P} X$, *for an arbitrary* $r > 0$

2.5 Numerical Treatment of Stochastic Differential Equations

We understand a stochastic differential equation as an abbreviatory notation for a stochastic integral equation, as stated in Definition 2.23. For a rigorous numerical treatment it is indispensable to precisely distinguish between different integral types and their corresponding differential equations. First, we have deterministic ordinary differential equations (ODE) and deterministic partial differential equations (PDE), consisting of Lebesgue-Stieltje integrals of differentiable functions or vector fields. Second, we denote equations where the integrand of a Lebesgue-Stieltje integral contains a stochastic process as "*random ordinary differential equations*" (RODE) or "*random partial differential equations*" (RPDE). Third, "*stochastic ordinary differential equations*" (SODE) and "*stochastic partial differential equations*" (SPDE) may involve stochastic integrals. The difference between RODE and SODE, and RPDE and SPDE respectively, may seem subtle, but is very significant in practical applications. Consider the following integrals corresponding to deterministic, random and stochastic ordinary differential equations:

$$\begin{aligned}
\text{ODE:} \quad & \int_0^t f(s)\, ds \\
\text{RODE:} \quad & \int_0^t g(s, W_s)\, ds \\
\text{SODE:} \quad & \int_0^t g(s, W_s)\, ds + \int_0^t h(s, W_s)\, dW_s,
\end{aligned}$$

for $t > 0$, function f, g, h and a Brownian motion W. Random ordinary and partial differential equations can be solved using deterministic numerical schemes derived for ODE and PDE. However, these schemes converge converge at slower rate for RODE and RPDE since the integrand – a stochastic process – is in general not differentiable but only Hölder continuous, see for instance [55]. Heuristically speaking, the paths of the integrand exhibit strong oscillations leading to less constrained error terms, and hence to a lower convergence order. Finally, SODE and SPDE involve fundamentally different integrals and can

not be treated with deterministic numerical schemes. These results are summarized in Table 1 below.

	ordinary differential equations	partial differential equations
	ODE	PDE
deterministic	$\frac{d}{dt}X_t = f(X_t)$	$\frac{\partial}{\partial t}X_t = \Delta X_t + f(X_t)$
	RODE	RPDE
random	$\frac{d}{dt}X_t = g(X_t, W_t)$	$\frac{\partial}{\partial t}X_t = \Delta X_t + g(X_t, W_t)$
	SODE	SPDE
stochastic	$dX_t = g(X_t)dt + h(X_t)dW_t$	$dX_t = [\Delta X_t + f(X_t)]dt + h(X_t)dW_t$

Table 1: *Different Kinds of Differential Equations*
Examples for different kinds of differential equations, with functions f, g, h, the Laplace operator Δ and a Brownian motion W. While random differential equations can be solved using deterministic numerical schemes, these schemes exhibit a lower convergence order due to the non-differentiable paths of Brownian motions. Stochastic differential equations require schemes specifically derived depending on the type of stochastic integral.

In order to obtain convergence of numerical schemes for a given system we need to have control over the growth of the functions appearing in the differential equation. The most common assumption is Lipschitz continuity defined as follows.

Definition 2.33. Lipschitz and Hölder Continuity
Let X and Y be two Banach spaces endowed with inner products $\langle\cdot,\cdot\rangle_X$ and $\langle\cdot,\cdot\rangle_Y$, respectively. A mapping $f : X \to Y$ is called

- *Lipschitz continuous if there is a $c > 0$ such that*

  $$\|f(x) - f(y)\|_Y \leq c\|x - y\|_X,$$

  for all $x, y \in X$.

- *one-sided Lipschitz continuous if $Y \subseteq X$ and there is a $c > 0$ such that*

  $$\langle x - y,\ f(x) - f(y)\rangle_X \leq c\|x - y\|_X^2,$$

  for all $x, y \in X$.

- *locally Lipschitz continuous if for every $x \in X$ there is a neighborhood U such that f restricted to U is Lipschitz continuous.*

- *Hölder continuous with exponent α if there is a $c > 0$ such that*

  $$\|f(x) - f(y)\|_Y \leq c\|x - y\|_X^\alpha,$$

  for all $x, y \in X$.

for a constant $c > 0$.

A typical example for Lipschitz continuity are differentiable functions with bounded derivatives. Every Lipschitz continuous function $f : X \to X$ is also one-sided Lipschitz continuous due to

$$\langle x - y,\ f(x) - f(y)\rangle_X \leq \|x - y\|_X \|f(x) - f(y)\|_X \leq c\|x - y\|_X^2.$$

Hölder continuity can be regarded as an intermediate step between continuous and differentiable functions. Paths of a Brownian motion are P-a.s. nowhere differentiable but Hölder continuous with exponent $\alpha = 1/2$. This is the underlying reason for the slower convergence of RODE and RPDE using deterministic numerical schemes: for a Brownian motion W, a time-step size $h > 0$ and $t > 0$, we have

$$|W_{t+h} - W_t| \leq c\sqrt{h}.$$

However, a Lipschitz continuous function would yield a discretization error proportional to h. Since stochastic processes are a family of random variables, the convergence concepts for SPs are based on the concepts for random variables discussed above. Let $X_t,\ t \in [0, T]$ be the solution of a SODE or SPDE on the probability space $(\Omega, \mathcal{F}, P)$ and $Y_{t_i}^M$ a numerical approximation with equidistant time steps $t_i = iT/M$, for $i = 1, \ldots, M$ and $M \in \mathbb{N}$. Then three important types of convergence are given as follows (see for instance [44]).

Definition 2.34. Strong Convergence
The numerical solution Y is said to converge with strong order γ if

$$E\left[\left|X_T - Y_M^M\right|\right] \leq cM^{-\gamma},$$

for a constant $c > 0$.

Definition 2.35. Weak Convergence
The numerical solution Y is said to converge with weak order β if

$$\left|E\left[g\left(X_T\right)\right] - E\left[g\left(Y_M^M\right)\right]\right| \leq cM^{-\beta},$$

for a constant $c > 0$ and every polynomial g.

Definition 2.36. Pathwise Convergence
The numerical solution Y is said to converge pathwise with order α if

$$\sup_{i=1,2,\ldots,M} \left|X_{t_i}(\omega) - Y_i^M(\omega)\right| \leq c(\omega)M^{-\alpha},$$

for a constant $c(\omega) > 0$ and for almost all $\omega \in \Omega$.

Intuitively strong convergence implies weak convergence, but which one is more useful for a given setting depends on whether the realizations or only their probability distributions are required to be "*close*". Pathwise convergence is interesting, since most numerical schemes simulate SODE and SPDE on a path-by-path basis. However, the Itô calculus is a mean-square calculus, which makes estimates for pathwise convergence more laborious. Note that the constant appearing in the pathwise estimate (Def. 2.36) does depend on ω and is therefore a random variable. Jentzen and Kloeden state in a recent book that the "*nature of their statistical properties is an interesting question, about which little is known theoretically so far and requires further investigation*" [8]. We study these and other types of numerical convergence concepts by means of an instructive example, which is directly related to the setting in Section 6 and 7. Therefore we postpone this discussion until then.

3 Climate Sensitivity

In this section, we study some fundamental principles, techniques and problems in the context of stochastic climate models. After introducing the necessarily abstract mathematical foundation, we discuss a concrete physical system, which serves as an accessible and instructive showcase model. We study different stochastic formulations including time-constant random variables, time-continuous stochastic processes with different autocorrelation structures and a coupled system, where the effects of two interacting stochastic processes can be observed. Those models are interesting on their own and we will derive results for each of them. Nonetheless, the focus of this section lies in the implications these results have on the broader context of stochastic parameterization.

3.1 Introduction

One of the key issues in climate modeling is the analysis of a system's response to perturbations of control parameters or external forces. A prime example is the study of the climate sensitivity λ, which is defined as the change of (globally and annually averaged) surface temperature T_s in response to a change of radiative forcing F_r, e.g. due to a change in CO_2 concentration:

$$\lambda = \frac{\Delta T_s}{\Delta F_r}. \tag{3.1}$$

This parameter enjoys a huge political interest, since it connects the rather abstract atmospheric concentration of green-house gases to the more tangible mean surface temperature. Although it provides no direct information on regional effects or on consequences on a finer timescale, the directly related "*2 Degree Goal*", i.e., limiting the emission of green-house gases in order to "*limit effectively the increase in global temperatures below 2 degrees Celsius above preindustrial levels*" [56], has become a figurehead in political and public discussion on climate change. It is in fact the only meteorological variable mentioned in the declaration of the G8 summit of Deauville - May 26-27, 2011 [56]. Unfortunately it is by no means obvious how to determine its value, since there is no repeatable experiment where one could change the radiative forcing and measure the response. Therefore the question arises whether or not we can derive λ from characteristics of the earth system without the knowledge of the system's response ΔT_s.[1] Promising tools to a positive answer of this question are the theory of Fluctuation-Dissipation Relation (FDR) and Response Theory (RT). In the framework of statistical mechanics of Hamiltonian systems they yield a quantitative relation between the spontaneous fluctuations and the response of the system to external fields – the Fluctuation-Dissipation Theorem. Based on the FDT, the domain of application has been extended from systems near the thermodynamic equilibrium to chaotic dynamical system, which satisfy certain smoothness assumptions regarding their equilibrium distribution, e.g. mixing or ergodicity. In the following we give a historic overview, based on a review paper by Marconi et al. [57], on this field of research. This topic is closely related to the development of the mathematical field known as stochastic calculus.

3.1.1 Historic Overview

The foundations of the Fluctuation Dissipation Theorem date back to early investigations of atomic motion. At the end of the 19th century, there was no unquestionable evidence of the existence of atoms, which lead to a scientific debate if the atomistic hypothesis should be regarded as an abstract mathematical tool or as an accurate description of nature. One of the participating physicists was Ernst Mach, who stated: "*The atomic theory plays a role in physics similar to that of certain auxiliary concepts in mathematics; it is a mathematical model for facilitating the mental reproduction of facts.*" [58]. Although the expressions

[1] Note that there are some rough estimates in form of historical climate data and ice-core drillings.

for the mean square energy fluctuation were known, J. Williard Gibbs correctly stated that the fluctuations "*would be in general vanishing quantities, since such experience would not be wide enough to embrace the more considerable divergences from the mean values*" [59] and could therefore not be observed. It was in this context that Einstein and Smoluchowski looked into a phenomenon that was for some decades considered as a curiosity. In 1827 the Scottish botanist Robert Brown described a continuous jittery motion of pollen grains suspended in water. Although the Dutch biologist and chemist Jan Ingenhousz described a similar irregular movement of coal dust on the surface of alcohol [60], it was not until 1905/1906 that the significance of this discovery was underlined (independently) by Einstein [61, 62] and Smoluchowski [63]. They found a quantitative relationship between the diffusion coefficient D, a measurable macroscopic quantity, and the Avogadro constant N_A, which is related to the microscopic description:

$$D = \frac{RT}{m\gamma N_A},$$

where R is the universal gas constant, T the temperature, and $m\gamma$ the friction constant. Langevin extended this work using a more mathematical approach in terms of stochastic differential equations [64], which inspired the mathematical field of continuous time stochastic processes. The results of Einsteins theoretical work were supported by experiments of Perrin [65] and gave conclusive evidence for the existence of atoms. More importantly for the application at hand, it can be considered as the first example of a FDR. To this end, we define the mobility of a particle by $\mu = 1/m\gamma$ and note that $R = N_A k_B$, where k_B is the Stefan-Boltzmann constant, leading to

$$D = \mu k_B T.$$

Using a modern notation, we can write the Langevin equation of the Brownian particle in the form of a one-dimensional Itô equation

$$dv = -\frac{1}{\mu} v dt + \sqrt{\frac{2D}{\mu^2}}\, dW_t,$$

where W_t is a Brownian motion (see Definition 2.9). When considering a system in thermodynamic equilibrium, the process defined by this equation is a stationary Ornstein Uhlenbeck process, see Section 3.3.2 below. Then Lemma 3.10 yields

$$\mathsf{cov}\,(v_t, v_s) = \frac{D}{\mu} e^{-\mu^{-1}|t-s|}.$$

Defining the time correlation function C_v of v yields

$$C_v(t) = \mathsf{cov}\,(v_t, v_0) = \frac{D}{\mu} e^{-t/\mu},$$

for $t > 0$. Comparing this with the expression for D, we obtain

$$D = \int_0^\infty \mathsf{cov}\,(v_t, v_0)\,.$$

Now consider a small perturbing force, which is switched on at $t = 0$, i.e.

$$f(t) = F\Theta(t),$$

where $\Theta(t)$ is the Heaviside step function. Denote by Δv the difference between the perturbed ond the original system. Then average response of the system to the perturbing force f is given by

$$E\,[\Delta v] = \mu F,$$

which can be written as

$$\frac{E\,[\Delta v]}{F} = \mu = \frac{D}{k_B T} = \frac{1}{k_B T} \int_0^\infty \mathsf{cov}\,(T_t, T_0)\,.$$

We have thus established a relation between the mobility, which describes the reaction of the system to small perturbations, and the covariance structure of the unperturbed system. If such a relation held true for the climate system, we could calculate the climate sensitivity in (3.1) without the knowledge about the system's response to a change in forcing.

Following the work of Einstein and Smoluchowski, Nyquist published the first formal version of a Fluctuation-Dissipation Theorem in the context of linear electrical networks in 1928 [66]. A few years later, in 1931, Onsager issued his famous paper on reciprocal relations in irreversible processes stating that (the linear approximations of) a macroscopic non-equilibrium perturbation follows the same laws that govern the system's fluctuation in thermodynamic equilibrium [67, 68]. This approach was the foundation for the FDR theorem of Callen and Welton [69] who gave a quantum-theoretical deduction of Nyquist's theorem and showed that it can be applied to a wider class of linear dissipative systems. In the concluding discussion of their article they stated: "*It would appear that a reasonable approach to the development of a theory of linear irreversible processes is through the development of the theory of fluctuations in equilibrium systems.*" This culminated in Kubos linear response theory (LRT) of time dependent correlation functions [70] including the Green-Kubo formula [71, 72].

In the following years, LRT and FDR were used with great success in various fields, surprisingly including applications which did not satisfy the necessary mathematical assumptions. In 1971 Van Kampen pointed out that "*the basic linearity assumption of linear theory is shown to be completely unrealistic and incompatible with basic ideas of statistical mechanics of irreversible processes*" [73]. However it was undisputed that LRT and FDR provided correct expressions for many real-world applications, leading to Van Kampens statement: "*I assert that it [LRT] arrives at these expressions by a mathematical exercise, rather than by describing the actual mechanism which is responsible for the response*". Since Van Kampen analyzed single trajectories (whereas the Green-Kubo formula characterizes the behavior of mean values), it was argued by Kubo that the instability of the trajectories works as a form of mixing in chaotic systems, which in turn stabilizes the distribution functions [74]. This is the underlying idea in the development of a generalized FDR for non-Hamiltonian systems (see for instance Section 3.2 in [57]) and for the Fluctuation Relation (FR) for nonlinear and non-Gaussian systems arbitrarily far from equilibrium. This includes the important Fluctuation Theorem (FT) derived by Evans, Cohen, Morris and Searles in 1993[75, 76, 77]. The FT can be considered as a generalization of the second law of thermodynamics as it quantifies the probability that the entropy of a system away from equilibrium flows in a direction opposite to that defined by the second law of thermodynamics. Note that this is no contradiction since the second law of thermodynamics, in the context of statistical mechanics, is a statistical statement describing the *tendency* for an increase in entropy. Furthermore the FT is consistent with the FDR when equilibrium is approached, and can be considered as a starting point for the development of a theoretical background incorporating complex phenomena such as turbulence or glassy, i.e., non-ergodic, systems and more general concepts of noise, see for instance [78, 79, 80] and the references therein.

3.1.2 Application of FDR on Climate

Although the climate system is a chaotic, highly nonlinear, multi-scale problem with a complex and debated equilibrium structure (see for instance [81]), there has been a profound interest in the application of the FDR in atmospheric and oceanic science, particularly concerning the question of global climate changes. In 1975, Leith [82, 83] emphasized the fact that the climate system does not exhibit a classical Gaussian equilibrium state. However he suggested the use of sensible approximations (which are called "quasi-Gaussian" in [84]) of appropriate variables for which a FDR holds. Following this idea, Bell [85], Carneval [86] and Gritsoun [87, 88, 89], among others, applied quasi-Gaussian FDRs to idealized climate models, while Ruelle [90, 91] and Majda [92, 93, 94], extended the

theoretical framework concerning (hyperbolic) chaotic, dynamic systems far from equilibrium. Furthermore, there are many cases where a FDR has been applied to GCMs and to various data series, see for instance [95, 96, 97]. However, Marconi et al. state that "*in most of these attempts, the FDR has been used in its Gaussian version, which has been acritically considered a reasonable approximation, without investigating its limits of applicability*" [57, p.55]. In order to counteract this carelessness, let us point out that

1. We do not know whether or not the climate system obeys a FDR.
2. We *may* have the theoretical framework to decide whether or not a specific climate model satisfies a quantitative or at least qualitative FDR.

Note that the climate system itself is not ergodic, due to aperiodic external forcing. The strongly debated question whether or not certain climate models satisfy ergodicity or a mixing condition is directly related to the first statement. Regarding the application of a FDR to climate models, we present a brief overview of the discussion in [93]. To this end, consider a conceptual stochastic dynamic system in Itô notation

$$dU_t = F(U_t)dt + \sigma(U_t)dW_t,$$

where $U_t \in \mathbb{R}^N$, $\sigma \in \mathbb{R}^{N\times K}$ and W_t is a K-dimensional Brownian motion, with $N, K \in \mathbb{N}$, $T > 0$ and $t \in [0, T]$. Denote by ρ_{eq} the probability density of the invariant measure. The mean value of an observable $\xi(U)$ of the equilibrial system is given by

$$E\left[\xi(U)\right] = \int \xi(u)\rho_{eq}\, du.$$

For a small $\epsilon > 0$, we introduce the perturbation term $\epsilon \cdot w(U)f(t)$ and consider the perturbed system

$$dU_t^\epsilon = F(U_t^\epsilon)dt + \epsilon w(U)f(t) + \sigma(U_t^\epsilon)dW_t.$$

For sufficiently small ϵ, the FDR and LRT yield that the leading order of the mean response satisfies

$$E\left[\epsilon\xi(U)(t)\right] = \int_0^t R(t-s)\epsilon f(s)\, ds, \tag{3.2}$$

where the linear response operator R is calculated via

$$R(t) = E\left[\xi(U_t)B(U_0)\right], \quad B(U_0) = -\frac{\nabla_U \cdot (w(U)\rho_{eq})}{\rho_{eq}}. \tag{3.3}$$

The linear response operator R only depends on the covariance structure of the unperturbed system. For practical applications in the context of climate change, there are two major obstacles to overcome:

1. The equilibrium probability density ρ_{eq} in (3.3) is usually not known.
2. We are often interested in non-infinitesimal perturbations $\epsilon >> 0$, e.g. a doubling of CO_2 concentration, in contradiction to the assumption in (3.2)

The first issue is addressed by the quasi Gaussian approximation, where the mean and covariance matrix of a Gaussian equilibrium measure is fitted onto the climatology ρ_{eq} and then utilized to calculate (3.3). It is the indiscriminate application of this approximation Marconi criticized [57, p.55]. In the same paper, the second issue is studied in detail showing that a qualitative FDR holds even for for non-infinitesimal perturbations. A quantitative estimate however is usually no longer linear in the perturbation $\epsilon f(t)$, and it would require extensive statistics to resolve the rare events [57, p.50-55]. Despite these difficulties, there exist atmospheric-ocean models, e.g. global quasigeostrophic models, which satisfy a quantitative FDR approximation for low-frequency climate variables with high skill [93]. Qualitative relations provide a tool for a deeper understanding of the interactions between different model variables. For instance, in [98] the dissipation of large

scale variables is analyzed for fluctuating small scale variables in the T21 ECHAM4 atmospheric GCM. The author concludes: "*The stronger the small-scale fluctuations, the stronger are the dissipations of the large-scale variables. This result suggests that the simulation of low-frequency climate variations and the prediction of climate change responses depend on the model representation of smallscale climate components*". This statement further motivates a detailed analysis of unresolved sub grid processes, which is directly connected to stochastic formulations of climate system.

3.1.3 Relaxation Times and Uncertainty

After having evaluated the possibilities and limitations to derive the climate sensitivity by means of LRT and FDR, we take a look at another aspect of the climate sensitivity. Let us consider the linear differential equation

$$\frac{dT}{dt} = -\gamma T + F_0,$$

which can be regarded as the most basic climate model (see Section 4) and has a stable state at $T = F_0/\gamma$. Assuming the system is in equilibrium at $t = 0$ when we turn on an additional constant forcing f, we obtain the solution

$$T_t^f = \frac{F_0}{\gamma} e^{-\gamma t} + \frac{F_0 + f}{\gamma}. \tag{3.4}$$

This leads to a new stable state at $T = (F_0 + f)/\gamma$, i.e., the system's response to the additional force f is given by f/γ, and we obtain

$$\lambda = \frac{\Delta T}{\Delta F} = \frac{1}{\gamma}.$$

The system has a relaxation time proportional to the climate sensitivity λ. This behavior comes as no surprise when dealing with linear differential equations, but, although it is of great practical consequence when studying climate change, it is often, if at all, treated peripherally. The significance in the context of climate change stems from the existence of a multitude of different time scales in the earth system. Due to the huge heat capacity of the oceans numerical models have to be integrated hundreds or even thousands of years for the system to reach a new equilibrium. This fact lead to so called "*transient climate simulations*", where the concentration of green house gases is changed gradually over an integration period of a few hundred years, typically 1850-2100 [99, Ch. 9].

Furthermore there is a great amount of uncertainty connected to the parameter of climate sensitivity. The fourth assessment IPCC report states regarding the expected temperature change: "*It is likely to be in the range 2°C to 4.5°C with a best estimate of about 3°C, and is very unlikely to be less than 1.5°C. Values substantially higher than 4.5°C cannot be excluded, but agreement of models with observations is not as good for those values*" [100, Ch. 2.3]. This significant uncertainty in the context of long relaxation times emphasizes that the characteristics of the dynamical system during the convergence to a new equilibrium are as important as the state of the new equilibrium itself.

In the following section we take a closer look at the system's convergence behavior considering various mathematical rigorous formulations for the presence of uncertainty in the climate sensitivity parameter λ. Aside from the relevant aspects of this subject as stated above, a key purpose of this section is to highlight the imperative of choosing a stochastic framework adapted to describing the underlying physical processes of the model.

3.2 Uncertainties Constant in Time

Let us consider a homogeneous system starting in a non-equilibrial state $T_0 > 0$, defined by

$$\frac{dT}{dt} = -\gamma T, \quad T(0) = T_0, \quad \gamma > 0. \tag{3.5}$$

This setting corresponds to (3.4) with an equilibrium state at $T_\infty = 0$. Throughout this section, the parameter γ can be considered as the inverse climate sensitivity. We are interested in the dynamics of this system when the exact value of γ is not known, but rather a stochastic quantity. In the following, different types of stochastic terms are introduced, which we assume to have equal variance σ^2 and mean value μ, to ensure comparability. To point out the significant differences between solutions of the resulting systems, we calculate the first moments, i.e., the expectation value and the (co-)variance, of the solutions. This is by no means an extensive analysis of the system's stochastic behavior but rather a descriptive way to highlight the importance of choosing a stochastic setting based on the physical conditions.

In the first step we assume that the exact value of γ is not known but constant in time. The system can then be described using random variables. To this end let $(\Omega, \mathcal{F}, P)$ be a probability space and define random variables $\gamma_i : \Omega \to \mathbb{R}$, $i \in \{\mathcal{N}, \mathcal{U}, \Gamma\}$ via their distributions

$$\begin{aligned}
&\gamma_{\mathcal{N}} \sim \mathcal{N}(\mu, \sigma) && \text{(Gaussian distribution)}\\
&\gamma_{\mathcal{U}} \sim \mathcal{U}\big([\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma]\big) && \text{(uniform distribution)}\\
&\gamma_{\Gamma} \sim \Gamma\Big(\frac{\mu^2}{\sigma^2}, \frac{\sigma^2}{\mu}\Big). && \text{(gamma distribution)}
\end{aligned}$$

The probability density functions for these distributions are given by

$$\begin{aligned}
f_{\mathcal{N}}(x; \mu, \sigma) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\\
f_{\mathcal{U}}(x; \mu, \sigma) &= \mathbb{1}_{[\mu-\sqrt{3}\sigma, \mu+\sqrt{3}\sigma]}(x)\\
f_{\Gamma}(x; k, \Theta) &= \frac{1}{\Theta^k} \frac{1}{\Gamma(k)} x^{k-1} e^{-x/\Theta}, \quad \text{with} \quad k = \frac{\mu^2}{\sigma^2},\ \Theta = \frac{\sigma^2}{\mu}.
\end{aligned}$$

Using these expressions, a direct calculation verifies, that $E[\gamma_i] = \mu$ and $\text{var}(\gamma_i) = \sigma^2$ for $i \in \{\mathcal{N}, \mathcal{U}, \Gamma\}$. Denote by T_i the solution of the system

$$dT_i = -\gamma_i T_i dt, \quad T_i(0) = T_0, \quad i \in \{\mathcal{N}, \mathcal{U}, \Gamma\}.$$

The expectation values of each $T_i(t)$ with respect to the distribution of its defining random variable can be obtained analytically. Omitting some intermediate steps, we find

Proposition 3.1. Expectation Values of Random Systems

$$\begin{aligned}
E[T_{\mathcal{N}}(t)] &= T_0 \exp\Big(-\mu t + \frac{\sigma^2}{2} t^2\Big)\\
E[T_{\mathcal{U}}(t)] &= T_0 \exp(-\mu t)\, \frac{\sinh(\sqrt{3}\sigma t)}{\sqrt{3}\sigma t}\\
E[T_{\Gamma}(t)] &= T_0 \left(1 - \frac{-\mu t}{\mu^2/\sigma^2}\right)^{-\mu^2/\sigma^2}
\end{aligned}$$

Regarding the long term behavior we get

$$\lim_{t\to\infty} E[T_{\mathcal{N}}(t)] = \infty$$
$$\lim_{t\to\infty} E[T_{\mathcal{U}}(t)] = \begin{cases}\infty & , \mu < \sqrt{3}\sigma \\ 0 & , \text{ else}\end{cases}$$
$$\lim_{t\to\infty} E[T_{\Gamma}(t)] = 0.$$

We observe unstable systems for Gaussian distributions and uniform distributions where $\mu < \sqrt{3}\sigma$. The cause for this behavior is the positive probability for negative values of γ_i, in contradiction to (3.5). This is a first (very obvious) example for the importance to choose a stochastic formalism consistent with the physical setting.

A more interesting observation can be made when comparing the expectation values of the random solutions to the deterministic system $T(t) = T_0 \exp(-\mu t)$. Since $\sinh(x) > x$ and $(1 + x/y)^y < e^x$ for all $x > 0, y > 0$, we get

$$E\left[T_{\mathcal{N}}(t)\right], E\left[T_{\mathcal{U}}(t)\right], E\left[T_{\Gamma}(t)\right] > T(t), \quad \text{for all} \quad t > 0.$$

In fact a more general result holds true:

Lemma 3.2. Stochastic Drift
Let $(\Omega, \mathcal{F}, P)$ be a probability space and γ an integrable real valued random variable on $(\Omega, \mathcal{F}, P)$ satisfying

$$E[\gamma] = \mu > 0, \qquad \textit{var}(\gamma) = \sigma^2 > 0.$$

Let $T_\gamma(t)$ the solution of a random system and $T(t)$ the solution of a deterministic system defined by

$$dT_\gamma(t) = -\gamma T_\gamma(t)dt, \quad T_\gamma(0) = T_0 > 0$$
$$dT(t) = -\mu T(t)dt, \quad T(0) = T_0.$$

Then

$$E[T_\gamma(t)] > T(t)$$

for all $t > 0$.

Proof. *Lemma 3.2*
Define the function

$$f(x) = e^{-xt}.$$

Since the second derivative satisfies

$$f''(x) = t^2 e^{-xt} > 0 \quad \text{for all} \quad t > 0,$$

f is strictly convex for all $t > 0$. Therefore Jensen's inequality (Theorem 2.11) states

$$E[T_\gamma(t)] = T_0 E\left[e^{-\gamma t}\right] = T_0 E\left[f(-\gamma)\right] > T_0 f\left(E\left[-\gamma\right]\right) = T_0 f(-\mu) = T_0 e^{-\mu t} = T(t),$$

where equality is rules out since γ is not constant due to var $(\gamma) = \sigma^2 > 0$. □

Lemma 3.2 states that it is an intrinsic quality of uncertainty to delay the convergence to equilibrium in system (3.5). The quantitative effect can be seen in Proposition 3.1, where terms depending on the variance σ^2 introduce an additional drift effect in the expectation value. Note that the qualitative result of Lemma 3.2 holds true for arbitrary random variables γ, i.e., it does not depend on the specific characteristics of the stochastic terms.

Remark 3.3. On the Estimate of the Climate Sensitivity Parameter
The effect of stochastic drift has direct implications for the process of estimating the climate sensitivity λ. Let $\{T_t\}_{t\in[0,T]}$ be a temperature time series, on which the model

$$dT = -\bar{\gamma}T dt$$

is fitted in order to estimate a scalar inverse climate sensitivity $\bar{\gamma}$. A common method is given by linear regression, i.e., choose $\bar{\gamma}$ such that ϵ in

$$\log\left(\frac{T_t}{T_0}\right) = -\bar{\gamma}t + \epsilon \qquad (*)$$

is minimized. If we regard λ – and therefore γ – as a random variable, the linearity in (*) implies

$$\bar{\gamma} = E\left[\gamma\right].$$

Comparing the deterministic solution $T_t = T_0 e^{-\bar{\gamma}t}$ with the random variable solution $T_t^\gamma = T_0 e^{-\gamma t}$, Lemma 3.2 yields

$$E\left[T_t^\gamma\right] > T_t.$$

Therefore, neglecting the stochastic character of the climate sensitivity by describing γ purely by its mean value $\bar{\gamma}$, leads to a systematical error.

3.3 Time Dependent Stochastic Processes

3.3.1 White Noise and Exponential Brownian Motion

If the physics of the system give reason to assume that the uncertainty of γ is not constant in time but rather a fluctuating stochastic quantity, we have to use the framework of stochastic processes. Before we use the mathematically rigorous Itô formalism to define stochastic differential equation, we introduce a notation widely-used in physical literature, centered around the *Langevin force*. For a time dependent stochastic parameter γ, we rewrite (3.5) as

$$\frac{d}{dt}T(t) = \big(-\mu + \sigma\Gamma(t)\big)T(t), \quad T(0) = T_0 \qquad (3.6)$$

where $\Gamma(t)$ is a Gauss process satisfying $E[\Gamma(t)] = 0$ and $\mathrm{cov}\big(\Gamma(s)\Gamma(t)\big) = \delta(t-s)$, where δ denotes the Dirac-Delta-function. Since $\Gamma(t)$ is Gaussian, it is uniquely defined by the first two moments of its distribution. While the physical notation has the advantage to fit nicely into the familiar notation of ordinary differential equations, there are some serious shortcomings. First of all, it suggests that the occurring stochastic term can be treated in the same way as any deterministic term, i.e., by means of differential calculus, which is usually not the case. Therefore, we employ the mathematical rigorous notation of stochastic differential equations using Itô- or equivalently Stratonovich-calculus. Following the discussion in Section 2.3, the Langevin equation (3.6) can be written as the Stratonovich stochastic differential equation

$$\begin{aligned} &\mathrm{d}T_t = -\mu T_t dt + \sigma T_t \circ dW_t, \quad T(0) = T_0 \\ \Rightarrow \qquad &T_t = T_0 \exp\big(-\mu t + \sigma W_t\big) \end{aligned} \qquad (3.7)$$

The process defined by (3.7) is called *exponential Brownian Motion*. Using the properties of Brownian motions, we obtain the following statistical quantities.

Proposition 3.4. Statistical Properties of the Exponential Brownian Motion

i) $E[T_t] = T_0 \exp\Big(-(\mu - \frac{\sigma^2}{2})t\Big)$

ii) $\mathit{var}(T_t) = T_0^2\Big[\exp\Big(-2(\mu-\sigma^2)t\Big) - \exp\Big(-2(\mu-\frac{\sigma^2}{2})t\Big)\Big]$

iii) $\mathit{cov}(T_s, T_t) = T_0^2 \exp\Big(-(\mu-\frac{\sigma^2}{2})(s+t)\Big)\Big[\exp\Big(\sigma^2\min(s,t)\Big) - 1\Big]$.

Proof. *Proposition 3.4*
Without loss of generality assume $s \leq t$ and $T_0 > 0$. As a consequence of the exponential martingale (Prop. 2.19), we have

$$E\left[e^{\lambda W_t}\right] = e^{\lambda^2 t/2}.$$

Recall that the increments of a Brownian motion are independent (Def. 2.9) with

$$W_t - W_s = W_{t-s}$$

in a distributive sense (Prop. 2.10). Then for all $0 \leq s \leq t$

$$\begin{aligned} T_0^{-2} E[T_s T_t] &= E\left[e^{-\mu(s+t)+\sigma(W_s+W_t)}\right] \\ &= e^{-\mu(s+t)} E\left[e^{\sigma(2W_s+(W_t-W_s))}\right] \\ &= e^{-\mu(s+t)} E\left[e^{2\sigma W_s}\right] E\left[e^{\sigma(W_t-W_s)}\right] \\ &= e^{-\mu(s+t)} E\left[e^{2\sigma W_s}\right] E\left[e^{\sigma W_{t-s}}\right] \\ &= e^{-\mu(s+t)} e^{2\sigma^2 s} e^{\sigma^2(t-s)/2} \end{aligned}$$

For $s = 0$ we directly receive *i)*. By definition of the covariance

$$\mathrm{cov}(T_s, T_t) = E[T_s T_t] - E[T_s]\,E[T_t]$$

we obtain *iii)*, which yields *ii)* for $s = t$. □

In analogy to the Gaussian random variable case (Prop. 3.1) we observe a weakened damping compared to the deterministic case. But contrary to the time constant model, which shows an unstable long-time behavior for arbitrary $\mu, \sigma > 0$, this model exhibits different stability regimes depending on the choice of μ and σ.

Proposition 3.5. Stability Regimes

$$\begin{aligned} 0 < \sigma^2 \leq \mu \quad &\Longrightarrow \quad \lim_{t\to\infty} E[T_t] = 0,\ \lim_{t\to\infty} \mathit{var}(T_t) \leq T_0^2 \\ \mu < \sigma^2 \leq 2\mu \quad &\Longrightarrow \quad \lim_{t\to\infty} E[T_t] \leq T_0,\ \lim_{t\to\infty} \mathit{var}(T_t) = \infty \\ 2\mu < \sigma^2 < \infty \quad &\Longrightarrow \quad \lim_{t\to\infty} E[T_t] = \infty,\ \lim_{t\to\infty} \mathit{var}(T_t) = \infty \end{aligned}$$

Let us give a heuristic explanation for the first stability regime: for suitable μ, σ, the process $\mu t - \sigma W_t$ not only obtains negative values with merely a small probability but is also likely to leave the negative area soon enough. Therefore the divergent contributions do not govern the system. For $\sigma^2 < \mu$ the drift term $-\mu t$ dominates the fluctuations σW_t and the variance converges to zero for $t \to \infty$, i.e., the system settles in a stable equilibrium. Although we criticized the choice of random variables, which allow negative coefficients γ, as unphysical, this is not necessarily true for the time dependent case. The random variable formulation can be considered as a system with a (properly scaled) time-averaged stochastic process. The first stability regime would then correspond to a physical system

which eventually converges into equilibrium, but where short-lived excursions are possible. An example would be the earth climate starting in the present state. An increase in radiative forcing would in this case lead to a new equilibrium for the surface temperature. Although we expect that the temperature tends to rise until it reaches its steady state, it is likely that we observe a decrease in temperature in the future. This may occur due to natural fluctuation or due to extreme events such as volcanic eruptions.

On the other side of the spectrum for $\sigma^2 > 2\mu$, both expectation value and variance are governed by the more pronounced positive deviations of σW_t leading to a divergent system. Interestingly there exists an interval $\mu < \sigma^2 \leq 2\mu$ where the drift is strong enough to ensure a stable expectation value, but too weak to prevent strong fluctuations at large t. These heuristic arguments are visualized in form of characteristic sample paths (Fig. 2) and histograms (Fig. 3) for each stability regime. This conceptual example points out a significant discrepancy between the expectations on the ability of stochastic meteorological models to quantify the probability of extreme events on the one hand, and their low number of sample paths on the other hand. For instance, consider a meteorological model the state of the second stability regime, and suppose we are interested in the asymptotic characteristics of extreme events. Since $\lim_{t\to\infty} E[T_t] = 0$ we have

$$\forall \delta > 0 \; \exists t_0 > 0 \; \forall t \geq t_0 : \qquad \delta > E[T_t] \,.$$

Let us symbolically regard $\{T_t > 1\}$ as an extreme event. Then $T_t \geq 0$ yields

$$P(T_t > 1) = E[1; T_t > 1] \leq E[T_t; T_t > 1] \leq E[T_t] < \delta.$$

Therefore the probability to witness an extreme event in the longterm behavior is very small. On the other hand, $\lim_{t\to\infty} \mathrm{var}(T_t) = \infty$ implies that these events have to occur with a non-vanishing probability. Now consider a model capable of simulating in the order of 20 sample paths. It is very unlikely that this model captures an extreme event. As a consequence the statistics of the sample paths would suggest that the system follows the first stability regime, i.e., the model might suggest an inapplicable asymptotic behavior. We will encounter a very instructive histogram for this problem in the next section (see Fig. 6). For a detailed and rigorous treatment of these issues we refer to the mathematical field of large deviations [101, 102].

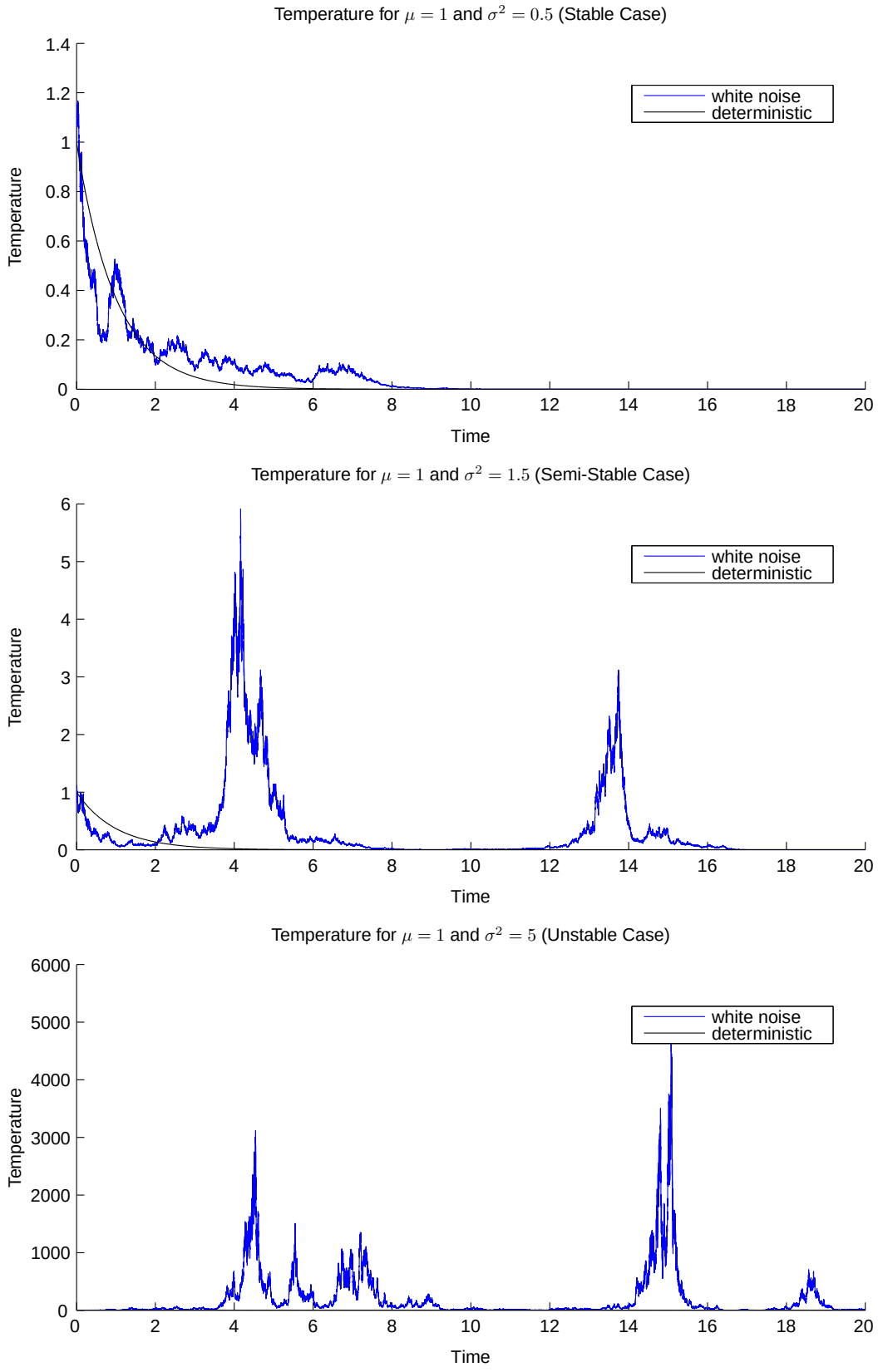


Figure 2: *Characteristic Paths for the White Noise System*
In the stable case, the system converges to $T = 0$ and shows no fluctuations once it reaches the equilibrial state. The second path has significant but not excessive long term fluctuations. The unstable path shows uncontrollable excursions (note the different scale of the vertical axis), which lead to a divergent expectation value and variance.

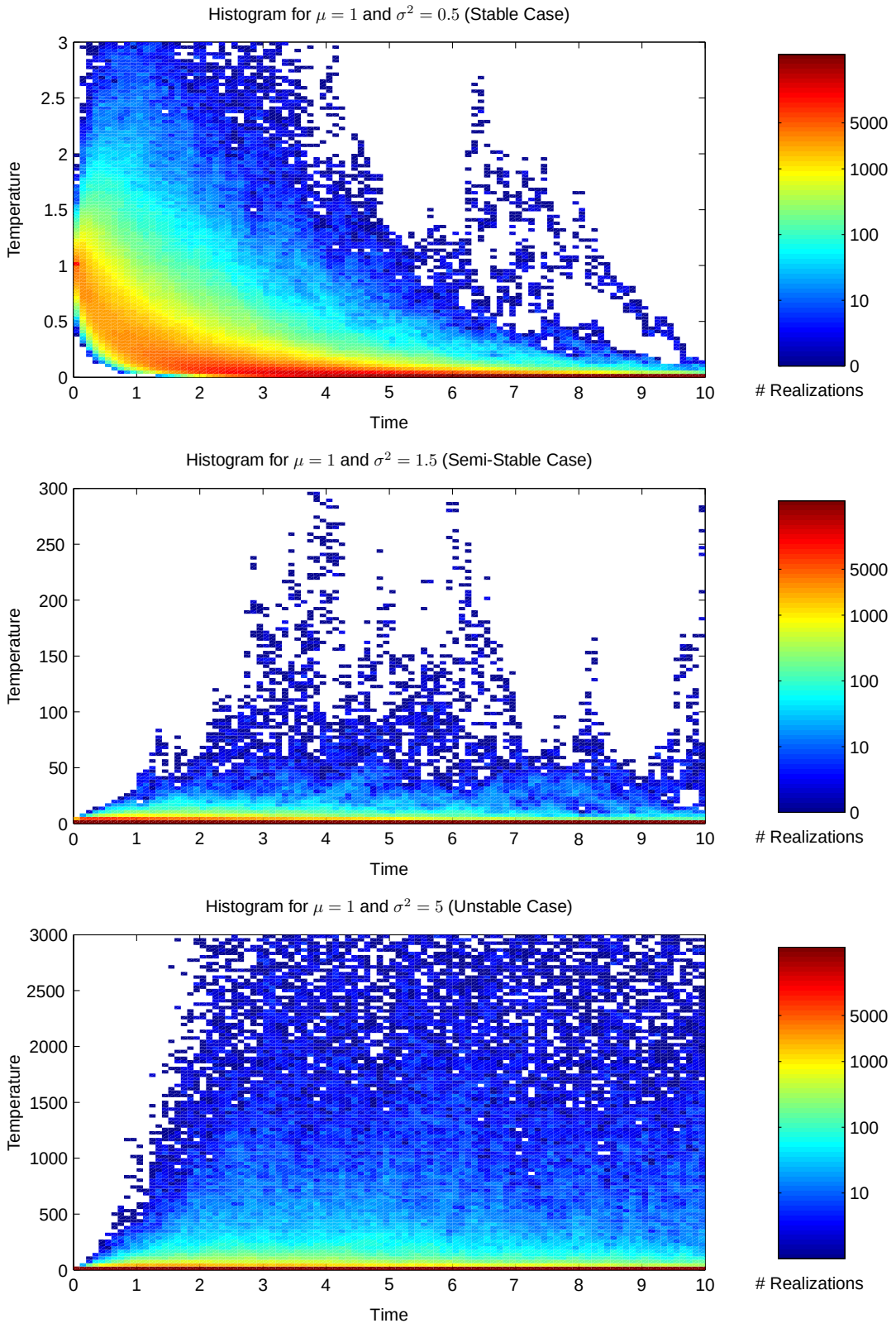


Figure 3: *Histograms for the White Noise System*
The histograms were calculated from 5.000 sample paths. In the stable case only very few, very weak excursions from the stable state $T = 0$ occur. In the second plot there are few but large long term fluctuations. In the unstable case the excursions grow in frequency and amplitude (note the different scale of the vertical axis) leading to a divergent expectation value and variance.

3.3.2 Red Noise and Ornstein Uhlenbeck Processes

We extend the concept of time dependent stochastic processes by allowing nontrivial time correlations. Time correlated Gaussian processes are often denoted as *red* or *colored noise*. This is not a mathematically well defined term but rather a collective term for Gaussian processes, which are not white noise, i.e., delta correlated. The term *colored* is motivated in analogy with the effects of filtering white light, i.e., all frequencies are present with equal amplitude. The term *red noise* takes into account that this type of noise has more energy at lower frequencies (see for instance Fig. 14 in Section 4.3.3). It is closely related to *long-term memory effects*, a notion appearing in various scientific fields. For an elaborate introduction into this subject, we refer to [103]. In the following, we discuss a prime example for red noise – the Ornstein Uhlenbeck process (OUP). Define the red noise model by

$$\begin{aligned} dT_t &= -\epsilon_t T_t dt, \qquad T_0 > 0 \\ d\epsilon_t &= -\Theta\big(\epsilon_t - \mu\big)dt + \sqrt{D}dW_t, \qquad \epsilon_0 \sim \mathcal{N}\left(\mu, \frac{D}{2\Theta}\right). \end{aligned} \tag{3.8}$$

The OUP $\epsilon = (\epsilon_t)_{t\geq 0}$ with mean value μ, drift coefficient $\Theta > 0$ and diffusion coefficient $D > 0$ is defined by an Itô SDE. We choose a normally distributed starting value ϵ_0 to guarantee a stationary Gaussian distribution of ϵ, see Lemma 3.10 below. Before analyzing the statistical characteristics of $T = (T_t)_{t\geq 0}$, we further investigate the properties of the OUP, starting with two representations of ϵ.

Lemma 3.6. Representations of an OUP
The OUP ϵ defined in (3.8) *satisfies*

$$\epsilon_t = \epsilon_0 e^{-\Theta t} + \mu(1 - e^{-\Theta t}) + \sqrt{D}\int_0^t e^{\Theta(s-t)} dW_s \tag{3.9}$$

$$\epsilon_t = \epsilon_0 e^{-\Theta t} + \mu(1 - e^{-\Theta t}) + \sqrt{\frac{D}{2\Theta}} e^{-\Theta t} W(e^{2\Theta t} - 1). \tag{3.10}$$

The second representation turns out to be very convenient for the calculation of statistical quantities of ϵ. In order to prove this lemma we state (without proof) two fundamental theorems from martingale theory as well as a differential version of the Itô formula.

Theorem 3.7. Dubins-Schwarz *(see [104])*
Every continuous local martingale $M = (M_s)_{s\geq 0}$ can be written as a time-changed Brownian motion $(B_{\langle M\rangle_s})_{s\geq 0}$, where $\langle M\rangle = (\langle M\rangle_s)_{s\geq 0}$ is the (continuous) quadratic variation of M.

Theorem 3.8. Quadratic Variation of Semimartingales *(see [31])*
Let X be a semimartingale with quadratic variation $\langle X\rangle$ and ξ a X-integrable process. Then the quadratic variation of $\int_0^t \xi\, dX$ is given by

$$\left\langle \int_0^t \xi\, dX \right\rangle = \int_0^t \xi^2\, d\langle X\rangle .$$

Theorem 3.9. Differential, Time-Dependent Itô Formula *(see [105])*
Let $f(x,t) : \mathbb{R}\times\mathbb{R}^+ \to \mathbb{R}$ be a twice differentiable function and

$$dX_t = a_t dt + b_t dW_t$$

an Itô process. Then $f(X_t, t)$ is an Itô process and satisfies

$$df(X_t,t) = \Big(\frac{\partial}{\partial t}f + a_t\frac{\partial}{\partial x}f + \frac{b_t^2}{2}\frac{\partial^2}{\partial x^2}f\Big)dt + b_t\frac{\partial}{\partial x}f dW_t.$$

Proof. *Lemma 3.6*
Applying Itô's chain rule to the function $f(x,t) = xe^{\Theta t}$ yields

$$\begin{aligned} df(\epsilon_s, s) &= \big(\frac{\partial}{\partial s} f(\epsilon_s, s) - \Theta(\epsilon_s - \mu) f'(\epsilon_s, s) + \frac{D}{2} f''(\epsilon_s, s)\big) ds + \sqrt{D} f'(\epsilon_s, s) dW_s \\ &= \big(\Theta \epsilon_s e^{\Theta s} - \Theta(\epsilon_s - \mu) e^{\Theta s}\big) ds + \sqrt{D} e^{\Theta s} dW_s \\ &= \Theta \mu \epsilon^{\Theta s} ds + \sqrt{D} e^{\Theta s} dW_s. \end{aligned}$$

Integration from 0 to t leads to

$$\begin{aligned} & f(\epsilon_t, t) - f(\epsilon_0, 0) = \Theta\mu \int_0^t \epsilon^{\Theta s} ds + \sqrt{D} \int_0^t e^{\Theta s} dW_s \\ \Rightarrow \quad & \epsilon_t e^{\Theta t} - \epsilon_0 \quad = \Theta\mu \Big(\frac{1}{\Theta} \big(e^{\Theta t} - 1\big) \Big) + \sqrt{D} \int_0^t e^{\Theta s} dW_s \\ \Rightarrow \quad & \epsilon_t \quad = \epsilon_0 e^{-\Theta t} + \mu \big(1 - e^{-\Theta t}\big) + \sqrt{D} \int_0^t e^{\Theta(s-t)} dW_s, \end{aligned}$$

which is (3.9). Regarding (3.10) note that since the Itô integral in (3.9) is a local martingale, Theorem 3.8 applies and we obtain

$$\left\langle \int_0^t e^{\Theta(s-t)} dW_s \right\rangle = \int_0^t e^{2\Theta(s-t)} d\langle W_s \rangle = \int_0^t e^{2\Theta(s-t)} ds = \frac{e^{-2\Theta t}}{2\Theta} \big(e^{2\Theta t} - 1\big).$$

Using Dubins-Schwarz gives us

$$\begin{aligned} \int_0^t e^{\Theta(s-t)} dW_s &= W\left(\left\langle \int_0^t e^{\Theta(s-t)} dW_s \right\rangle \right) \\ &= W\left(\frac{e^{-2\Theta t}}{2\Theta} \big(e^{2\Theta t} - 1\big) \right) \\ &= \frac{e^{-\Theta t}}{\sqrt{2\Theta}} W\left(e^{2\Theta t} - 1\right), \end{aligned}$$

which proves (3.10).

□

Lemma 3.10. Statistical Properties of the OUP

i) For an arbitrarily distributed, real valued random variables ϵ_0 we have

$$\begin{aligned} E[\epsilon_t] &= \mu + e^{-\Theta t}\big(E[\epsilon_0] - \mu\big) \\ \mathit{var}(\epsilon_t) &= \frac{D}{2\Theta} + e^{-2\Theta t}\big(\mathit{var}(\epsilon_0) - \frac{D}{2\Theta}\big) \\ \mathit{cov}(\epsilon_s, \epsilon_t) &= \frac{D}{2\Theta} e^{-\Theta|t-s|} - e^{-\Theta(s+t)}\Big(\mathit{var}(\epsilon_0) - \frac{D}{2\Theta}\Big). \end{aligned}$$

ii) The stationary case, i.e., $\epsilon_0 \sim \mathcal{N}(\mu, \frac{D}{2\Theta})$, obeys

$$\begin{aligned} E[\epsilon_t] &= \mu \\ \mathit{var}(\epsilon_t) &= \frac{D}{2\Theta} \\ \mathit{cov}(\epsilon_s, \epsilon_t) &= \frac{D}{2\Theta} e^{-\Theta|t-s|}. \end{aligned}$$

Proof. *Lemma 3.10*
The stationary case is a trivial consequence of *i)*, which follows from a direct calculation using expression (3.10) and basic properties of Brownian Motions. □

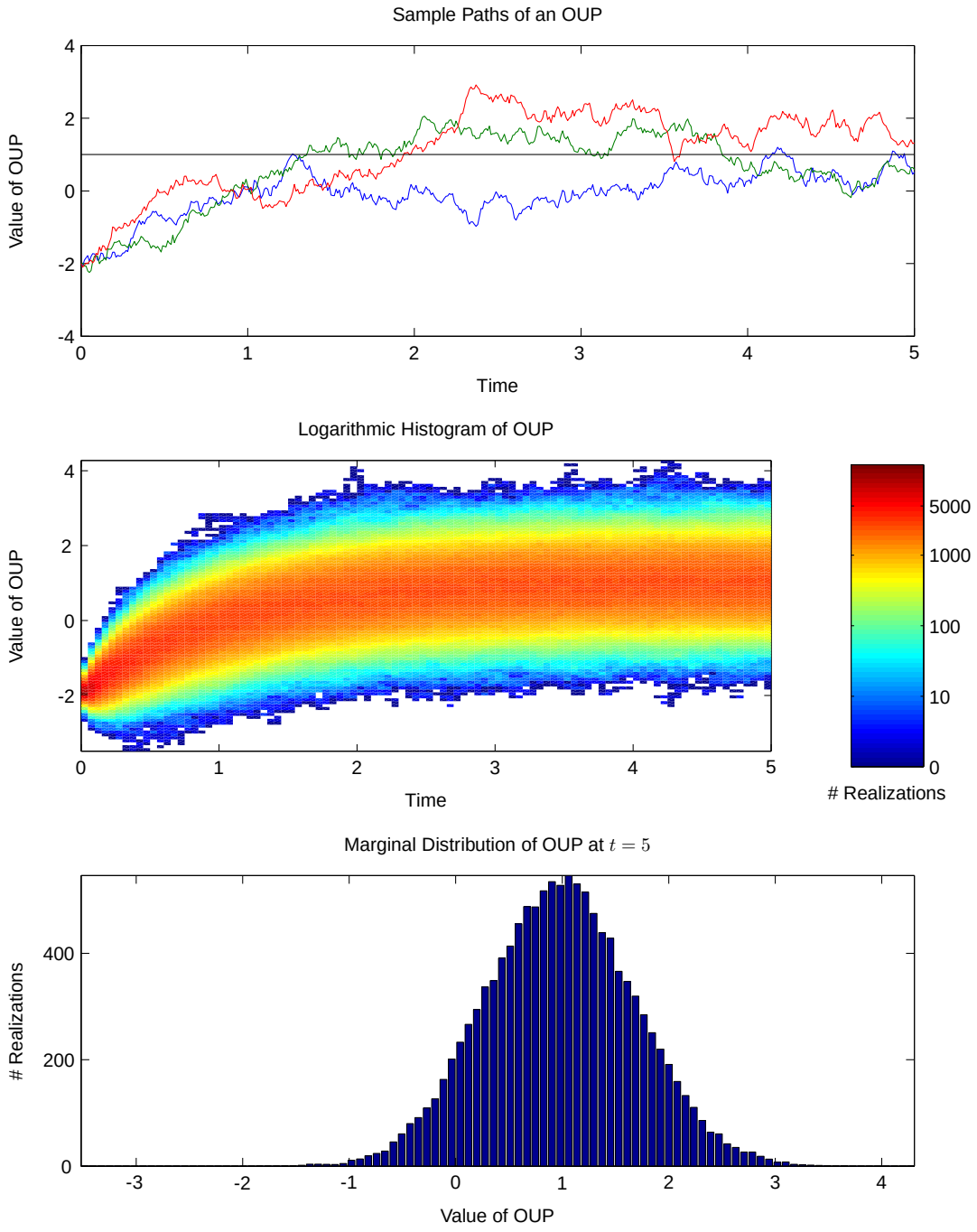


Figure 4: *Visualization of an Ornstein Uhlenbeck Process*
For an OUP with $\mu = 1$, $\Theta = 1$ and $D = 1$ starting in $\xi = -2$ we simulated 10.000 realizations. The upper two plots point out the mean-reverting drift, which steers the process into a stationary distribution. The convergence to this stationary distribution can be observed in the second plot. The third plot shows that this is indeed a Gaussian distribution centered at $\mu = 1$.

Lemma 3.11.
For $\epsilon_0 \sim \mathcal{N}\left(\mu, \frac{D}{2\Theta}\right)$ the stationary OUP is given by

$$\epsilon_t = \mu + \sqrt{\frac{D}{2\Theta}} e^{-\Theta t} W(e^{2\Theta t}) \tag{3.11}$$

Proof. *Lemma 3.11*
It is easily verified, that the first and second moment of (3.10) and (3.11) are identical. The lemma follows since both processes are Gaussian and therefore completely defined by their first two moments. □

In analogy to the cases with random variables and white noise, we assume that $E[\epsilon_t] = \mu$ and var$(\epsilon_t) = \sigma^2$. While this has been a unique characterization for the stochastic variables and processes discussed so far, the time correlated structure of an OUP offers an additional degree of freedom. When we fix the quotient $\frac{D}{2\Theta} = \sigma^2$ by setting $D = 2\sigma^2\Theta$, Lemma 3.11 shows that for every $\Theta > 0$ the process

$$\epsilon_t = \mu + \sigma e^{-\Theta t} W(e^{2\Theta t})$$

satisfies the statistical demands. Since Θ appears in the exponential time argument of the Brownian motion, it can be understood as a measure for the timescale, or inner speed, of the stochastic process. This becomes more concrete when looking at the time lagged covariance (see Lemma 3.10)

$$\text{cov}(\epsilon_t, \epsilon_t + \tau) = \sigma^2 e^{-\Theta\tau}.$$

For large Θ the covariance between two points in time is small even if their distance in time τ is small. For small Θ, however, this expression is significantly greater than zero for long time distances, i.e., the process exhibits a long-term memory. In this heuristic sense, red noise can be understood as a crossover between the random variable case, i.e., infinitely long time correlations corresponding to $\Theta \to 0$, and white noise, i.e., delta correlated noise which corresponds to $\Theta \to \infty$. We will further discuss this interpretation with respect to the statistical properties of T_t. Before we can calculate these properties we need some technical preparations. For greater clarity during the following computations, define

$$\begin{aligned} s \vee t &:= \max(s,t) \\ s \wedge t &:= \min(s,t) \\ \varphi_t &:= \frac{1 - e^{-\Theta t}}{\Theta}, \end{aligned}$$

for $s, t, \Theta > 0$.

Lemma 3.12. Statistical Properties of T
Consider system (3.8) *driven by a stationary OUP ϵ with parameters μ, Θ, D. Then*

$$E[T_t] = T_0 \exp\Big(-\mu t + \frac{D}{2\Theta^2}\big(t - \varphi_t\big)\Big) \tag{3.12}$$

$$\textit{var}(T_t) = T_0^2 \exp\Big(-2\mu t + \frac{D}{\Theta^2}\big(t - \varphi_t\big)\Big)\Big[\exp\Big(\frac{D}{\Theta^2}(t - \varphi_t)\Big) - 1\Big] \tag{3.13}$$

$$\begin{aligned} \textit{cov}(T_s, T_t) = T_0^2 \exp\Big(&-\mu(s+t) + \frac{D}{2\Theta^2}\big(s + t - (\varphi_s + \varphi_t)\big)\Big) \\ &\times \Big[\exp\Big(\frac{D}{2\Theta^2}\big(2(s \wedge t) - (1 + e^{-\Theta|t-s|})\varphi_{s\wedge t}\big)\Big) - 1\Big] \end{aligned} \tag{3.14}$$

Proof. *Lemma 3.12*
Due to its technical nature this proof is carried out in Appx. A. □

Let us take a closer look at the results of Lemma 3.12, particularly with consideration of the long-term memory effect mentioned above. Recall that we can fixate a stationary OUP with mean value μ and variance σ^2 and still retain a degree of freedom in form of the parameter Θ. This parameter is directly related to the relaxation time of the time lagged covariances

$$\mathsf{cov}\,(\epsilon_t, \epsilon_t + \tau) = \sigma^2 e^{-\Theta\tau}.$$

Without loss of generality assume $T_0 = 1$. Then the expectation values of the systems considered so far read:

$$E\left[T_t^R\right] = \exp\Big(-\mu t + \frac{\sigma^2}{\Theta}\big(t - \varphi_t\big)\Big) \qquad \text{(red noise driven)}$$

$$E\left[T_t^W\right] = \exp\Big(-\mu t + \frac{\sigma^2}{2} t\Big) \qquad \text{(white noise driven)}$$

$$E\left[T_t^{\mathcal{N}}\right] = \exp\Big(-\mu t + \frac{\sigma^2}{2} t^2\Big). \qquad \text{(Gaussian random variable)}$$

Since we aim to analyze stability regimes analogous with the white noise case, we take a look at the asymptotic behavior of φ. Note that φ is monotonically increasing in t. For a fixed $\Theta > 0$ we have

$$\varphi_\Theta(0) = 0$$
$$\varphi_\Theta(t) = \frac{1 - e^{-\Theta t}}{\Theta} \longrightarrow \frac{1}{\Theta} \quad \text{for} \quad t \to \infty.$$

Θ is inversely proportional to the amplitude of the stochastic drift, and it induces a time lag, which is (for large t) equal to $1/\Theta$ as well. For fixed variance, the relaxation time of the covariance structure - or heuristically: the memory - of the noise process ϵ therefore determines the stochastic drift. Following this interpretation, consider the scaling limes, where $\Theta \to 0$ and $D = 2\sigma^2\Theta$, such that $D/2\Theta = \sigma^2$ remains constant. Using representation (3.10) of ϵ immediately gives us $\epsilon_t \equiv \epsilon_0$ for all $t > 0$. Since we are in the stationary case, ϵ_0 has to be normally distributed with mean μ and variance σ^2. Indeed, using this scaling we obtain

$$\frac{1}{\Theta}\left(t - \varphi_t(\Theta)\right) \longrightarrow \frac{t^2}{2},$$

for all $t > 0$. This leads to consistent expectation values for the scaling limes of the red noise system and the system defined with a Gaussian random variable. In this heuristic sense we can interpret red noise as a conjunction of the rapidly varying white noise (no memory) and the time constant random variable case (infinite memory). We refer to Fig. 5 for a graphical representation of these interpretations. It comes as no surprise to find familiar stability regimes.

Proposition 3.13. Stability Regimes

$$0 < \frac{D}{\Theta^2} \le \mu \quad \Longrightarrow \quad \lim_{t\to\infty} E\left[T_t\right] = 0,\ \lim_{t\to\infty} \mathsf{var}\left(T_t\right) < \infty$$

$$\mu < \frac{D}{\Theta^2} \le 2\mu \quad \Longrightarrow \quad \lim_{t\to\infty} E\left[T_t\right] \le T_0,\ \lim_{t\to\infty} \mathsf{var}\left(T_t\right) = \infty$$

$$2\mu < \frac{D}{\Theta^2} < \infty \quad \Longrightarrow \quad \lim_{t\to\infty} E\left[T_t\right] = \infty,\ \lim_{t\to\infty} \mathsf{var}\left(T_t\right) = \infty$$

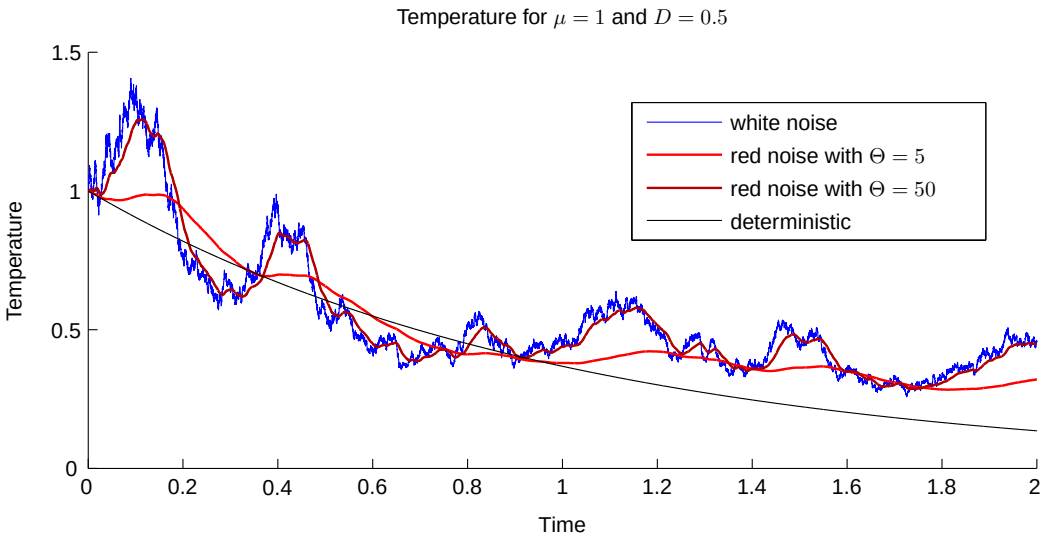


Figure 5: *Sample Paths for the Red Noise System*
For the simulation of these realizations we used the same set of random numbers for each path and fixed the variance of the underlying multiplicative noise process, i.e., we chose the parameter D depending on the value of Θ. Both the smoothing characteristic as well as the time lag for small Θ can be easily recognized.

Naturally these regimes are not solely determined by the relation between mean value and variance, but include the memory effect through $1/\Theta$. Since the heuristic interpretation follows the one given for the white noise system, we omit further discussions and refer directly to the histogram plots (Fig. 6). The histogram for the semi-stable regime is an excellent example for the issue of large deviations for models which are only capable of simulating a small number of sample paths. As mentioned in the concluding discussion of the white noise model, there is a high probability for such a model to miss the large excursions, which would wrongly imply a stable asymptotic behavior.

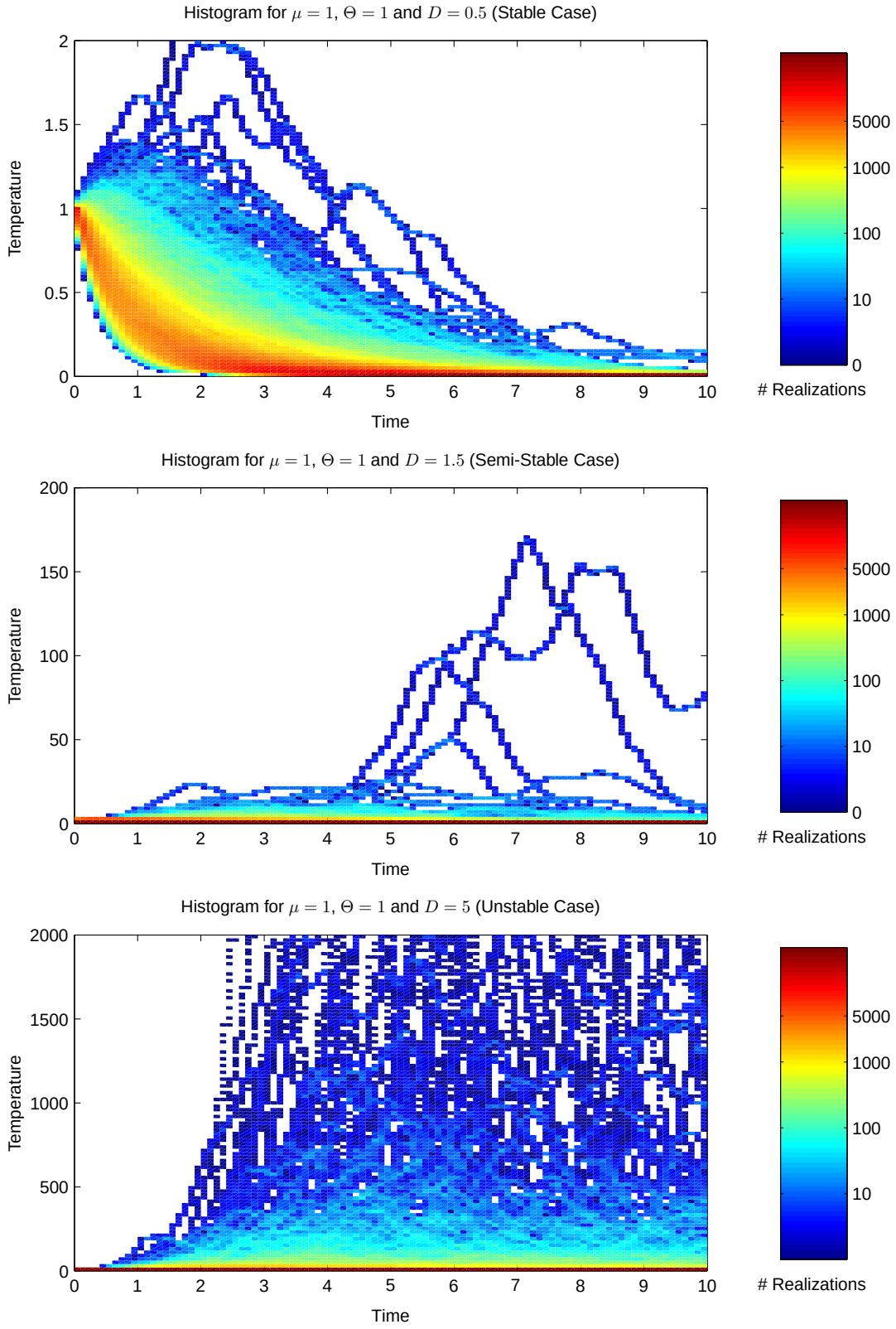


Figure 6: *Histograms for the Red Noise System*
The histograms were calculated from 5.000 sample paths. Completely analogous with the white noise model, in the stable case only very few, very weak excursions from the stable state $T = 0$ occur. In the second plot there are few but large long term fluctuations. In the unstable case the excursions grow in frequency and amplitude (note the different scale of the vertical axis) leading to a divergent expectation value and variance.

3.4 Coupled Noise

Thus far we have discussed the essential cases of systems driven by random variables and white noise, as well as a more elaborate system driven by time-integrated red noise. For these models we were able to find exact analytical solutions and statistical properties. However most meteorologic applications are too complex to be treated analytically, forcing us to derive suitable numerical methods. In this section we discuss a system based on the previous considerations, which is complex enough to justify a numerical treatment and to show some interesting effects. Furthermore it is a prime example for the failure of explicit numerical schemes for non-Lipschitz continuous systems (see Definition 2.33), which is discussed in Section 3.5.1.

The stability regimes of the models driven by white or red noise show a clear discrepancy between the statistical results on the one hand and our physical notion on the other hand: we expect that even a strongly fluctuating system has a finite mean temperature. This suggests that the mathematical formulation of these models is not well suited to describe the underlying physical processes. The root of these unphysical results is the strictly positive probability for the occurrence of a negative climate sensitivity, i.e., reducing the radiative forcing would increase the temperature. However, we would like to include the possibility of short-term fluctuations with arbitrary sign both during the convergence to equilibrium and in the equilibrial state. To this end we identify the multiplicative process γ with a nonnegative, non-delta correlated Gaussian process: the square of a stationary Ornstein Uhlenbeck process. The short-term fluctuations f are described by an additive white noise process, where we implicitly assume that they occur on a time scale much smaller than the time scale of γ. In order to determine a suitable coefficient, suppose that we have knowledge of the amplitude of these fast paced, zero mean stochastics in equilibrial state, for instance, in form of temperature time series of ice core drillings

$$E\left[\Delta T^2\right] = \sigma^2.$$

Due to the time scale separation, it is an acceptable approximation to consider $\gamma = \bar{\gamma}$ as constant during a small time interval. This leads to the abstract equation

$$dT_t = -\bar{\gamma} T_t dt + f,$$

for all t in a small interval during equilibrium. The short term stochastics f now result in temperature fluctuations

$$\Delta T = \frac{f}{\bar{\gamma}}.$$

Since we do not know the value of $\bar{\gamma}$, f has to be a function of γ in the fully time dependent model. This motivates the term "coupled noise system", which we define, using Stratonovich calculus, by

$$\begin{aligned} dT_t &= -\gamma T_t dt + \underbrace{\sigma \gamma_t \circ dW_t^1}_{\hat{=} f} \\ d\gamma_t &= -2\Theta\gamma_t dt + 2\sqrt{D\gamma_t} \circ dW_t^2, \quad \gamma(0) = \epsilon_0^2, \quad \epsilon_0 \sim \mathcal{N}\left(0, \frac{D}{2\Theta}\right), \end{aligned} \tag{3.15}$$

where $\gamma = (\gamma_t)_{t \geq 0}$ is the square of the OUP ϵ defined in (3.8) with $\mu = 0$, and W^1, W^2 are two independent Brownian Motions on the probability space $(\Omega, \mathcal{F}, P)$. We derived the Stratonovich equation for γ from the Itô equation for ϵ by using Lemma 2.24, which implies that an OUP has the same drift in its Itô and Stratonovich representation due its constant diffusion coefficient $\sqrt{D}$. Note that we made a technical simplification by assuming $\mu = 0$ in the definition of γ, which is synonymous with the fact that the positive mean value of γ is based solely on the squared fluctuations of the centered OUP ϵ (see Lemma 3.16). This choice was made for greater clarity especially during the discussion of a numerical treatment of this model (see section 3.5). All results of this model can easily be modified for the cases where $\mu > 0$.

Remark 3.14. The abstractness of the model
In this very abstract setting we use major simplifications, which do not necessarily correspond to the physical conditions. First we assume that the timescales of the state variable and the additive noise process can be separated. The permissibility of this approximation depends on the the (hypothetical) data describing the equilibrial temperature fluctuations. In this context another question arises: which physical processes are represented by an additive noise? Most processes within the earth system depend on the temperature, which suggests a multiplicative noise structure to describe their fluctuations. Strictly additive forces would therefore be external, e.g. (fluctuations of) the Milankovich cycles [106] describing variations in solar forcing. Another conceivable possibility is to treat the additivity itself as an approximation. Consider for instance a process with very weak dependence on the temperature, i.e., its response to small temperature changes is negligible. Following this line of thought, we perceive both the time scale separation and the additivity of the white noise process as mathematical idealizations for a strongly abstracted model.

This abstracted model is well suited to showcase the two main aspects of this section:

- The development of a basic method to derive a stochastic process from given data.
- The derivation of basic numerical methods including a typical issue of SODE.

More precisely, we derive the parameters Θ and D of the OUP-square γ for (hypothetically) given time series in the following Section 3.4.1. We simulate sample paths, i.e., solutions for a single realization of the stochastic quantities, using explicit and implicit Milstein schemes. In this context, we encounter in Section 3.5 a typical and descriptive issue of explicit numerical schemes for systems which do not satisfy Lipschitz conditions.

3.4.1 The OUP-Square Process

Suppose a set of data $\hat{\gamma}$ for the parameter γ is given, indicating stationarity and yielding sample expectation value $\hat{E}$ and variance $\hat{V}$ as well as an estimate for the decorrelation time $\hat{\tau}$, i.e., the time at which $\mathrm{cov}\,(\hat{\gamma}_t, \hat{\gamma}_{t+\hat{\tau}}) = \frac{1}{2}\mathrm{var}\,(\hat{\gamma}_t)$. In order to simulate sample paths of the coupled noise model, we have to derive the parameters Θ and D for the OUP-square process. To this end, based on a given OUP ϵ, the statistical properties of $\gamma = \epsilon^2$ are calculated. Since the OUP ϵ is Gaussian, it is uniquely defined by its first two moments. Therefore we may identify the statistical values of the abstract OUP-square γ with the ones obtained from data $\hat{\gamma}$ to derive the necessary parameters for our model.

Lemma 3.15. Statistical Properties of an OUP-Square Process
Let ϵ be a stationary OUP with parameters μ, Θ and D. Then the square process $(\gamma_t)_{t\geq 0} = (\epsilon_t^2)_{t\geq 0}$ satisfies

i) γ_t *is a stationary process*

ii) $E\,[\gamma_t] = \mu^2 + \frac{D}{2\Theta}$

iii) $\mathrm{var}\,(\gamma_t) = \frac{2D}{\Theta}\mu^2 + \frac{D^2}{2\Theta^2}$

iv) $\mathrm{cov}\,(\gamma_s, \gamma_t) = \frac{2D}{\Theta}\mu^2 e^{-\Theta|t-s|} + \frac{D^2}{2\Theta^2}e^{-2\Theta|t-s|}$

v) $\tau = -\frac{1}{\Theta}\ln\left(\sqrt{\frac{4\Theta^2}{D^2}\mu^4 + \frac{2\Theta}{D}\mu^2 + \frac{1}{2}} - \frac{2\Theta}{D}\mu^2\right)$.

Proof. *Lemma 3.15*
Due to its technical nature this proof is carried out in Appx. A. □

To gain an better idea of the basic characteristics of this process, we refer to Fig. 7 and the remarks in the caption.

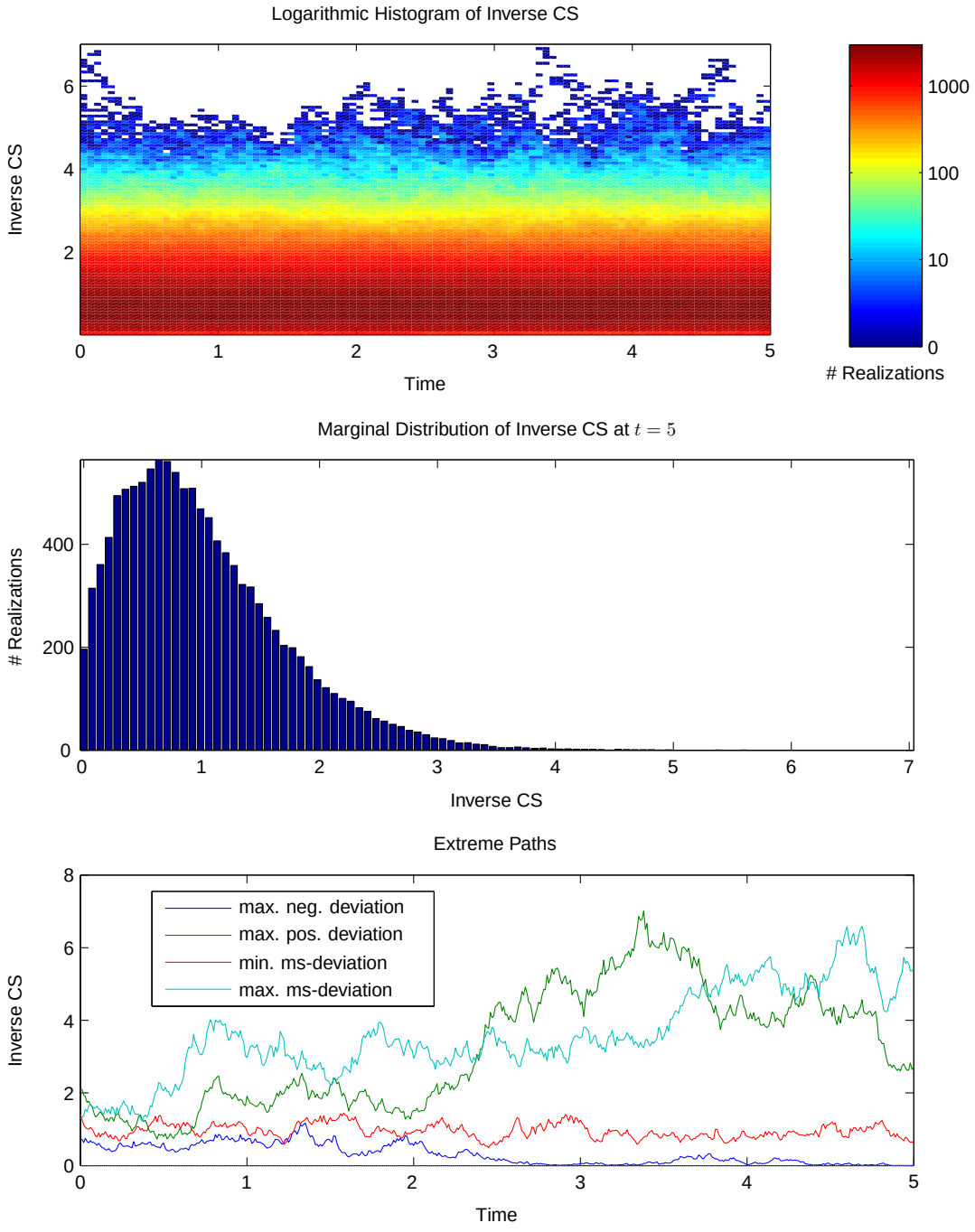


Figure 7: *Visualization of an OUP-Square Process*
The inverse climate sensitivity is represented by an OUP-square process with expectation value $E = 1$, variance $var = 0.5$ and decorrelation time $\tau = 1$. The upper two plots outline the stationary distribution based on 10.000 simulated realizations, while the third plot shows paths with exceptional deviation from the mean value. The paths with maximal negative and positive deviation are the ones with the largest pointwise difference between the realization and the mean value. They differ from the paths with maximal and minimal mean-square deviation.

We are now in a position to derive the OUP parameters μ, Θ and D from the aforementioned values $\hat{E}, \hat{V}$ and $\hat{\tau}$ obtained from data.

Lemma 3.16. Derivation of OUP-Square Parameters from Data
Let $\hat{E}, \hat{V}, \hat{\tau} > 0$ such that $2\hat{E}^2 \geq \hat{V}$. Then the parameters μ, Θ and D of an OUP square process $(\gamma_t)_{t\geq 0}$ are given by

i) $\mu = \sqrt{\hat{E} - \frac{c}{2}}$

ii) $\Theta = -\frac{1}{\tau} \ln\left(\sqrt{\frac{4\mu^4}{c^2} + \frac{2\mu^2}{c} + \frac{1}{2}} - \frac{2\mu^2}{c}\right)$

iii) $D = c\Theta$,

where $c := 2\hat{E} - \sqrt{4\hat{E}^2 - 2\hat{V}}$.

Proof. *Lemma 3.16*

i) Follows directly from *iii)* and Lemma 3.15 *ii)*.

ii) Using *iii)* and Lemma 3.15 *v)* one gets

$$\begin{aligned}\Theta &= -\frac{1}{\tau} \ln\left(\sqrt{\frac{4\Theta^2}{D^2}\mu^4 + \frac{2\Theta}{D}\mu^2 + \frac{1}{2}} - \frac{2\Theta}{D}\mu^2\right) \\ &= -\frac{1}{\tau} \ln\left(\sqrt{\frac{4}{c^2}\mu^4 + \frac{2}{c}\mu^2 + \frac{1}{2}} - \frac{2}{c}\mu^2\right)\end{aligned}$$

iii) Lemma 3.15 *ii) and iii)* yield

$$\begin{aligned}c &= 2E - \sqrt{4E^2 - 2V} \\ &= 2\mu^2 + \frac{D}{\Theta} - \sqrt{\left(2\mu^2 + \frac{D}{\Theta}\right)^2 - \left(4\mu^2\frac{D}{\Theta} + \frac{D^2}{\Theta^2}\right)} \\ &= 2\mu^2 + \frac{D}{\Theta} - \sqrt{4\mu^4 + 4\mu^2\frac{D}{\Theta} + \frac{D^2}{\Theta^2} - 4\mu^2\frac{D}{\Theta} - \frac{D^2}{\Theta^2}} \\ &= \frac{D}{\Theta}\end{aligned}$$

□

Since this result allows us to construct the stochastic process from a given set of data, we are now in a position to numerically treat the coupled noise system (3.15).

3.5 A Numerical Example

Before we begin with a numerical treatment of the coupled noise system, we recall Definitions 2.21 and 2.22 of stochastic integrals

$$\int_0^t G_s dW_s = \lim_{|\tau_n|\to 0} \sum_{i=1}^{n} G(t_{i-1} + \alpha(t_i - t_{i-1}))\left[W(t_i) - W(t_{i-1})\right],$$

for a Brownian motion W. As we have seen in Section 2.3 and contrary to the case of Lebesgue-Stieltjes integrals, the choice of α changes the value of the stochastic integral. This has immediate consequences for the numerical treatment of SDE. Not only do we need different numerical schemes for different values of α, we also have to expect the failure of deterministic numerical schemes. The most basic scheme for SODE is the Euler-Maruyama scheme, followed by the Milstein scheme and more elaborate methods such as stochastic Runge-Kutta schemes. For an extensive treatment regarding SODE, we refer to [44]. In the following, we discuss the two-dimensional explicit Milstein scheme [44, p. 346f]. We point out a critical failure of this numerical scheme and introduce the implicit Milstein scheme, which prevents this error. For a system of SODE

$$d\xi_i = h_i(\boldsymbol{\xi}, t)dt + \sum_{k=1}^{m} g_{ik}(\boldsymbol{\xi}, t) \circ dW_t^k, \qquad \boldsymbol{\xi} = (\xi_k)_{k=1}^m, \quad i \in \{1, ..., d\}, \quad m, d \in \mathbb{N}$$

define for $k \in \{1, ..., m\}$, $n \in \{1, ..., N\}$, $N \in \mathbb{N}$ and $T > 0$

$$\begin{aligned}
t_n &= \frac{n}{N}T \\
\Delta_n &= t_{n+1} - t_n \\
\Delta W_n^k &= W^k(t_{n+1}) - W^k(t_n) \\
\boldsymbol{\xi}_n &= (\xi_n^k)_{k=1}^m,\ \xi_n^k = \xi_k(t_n) \\
h_n^k &= h_k(\boldsymbol{\xi}_n, t_n) \\
g_n^{ij} &= g_{ij}(\boldsymbol{\xi}_n, t_n) \\
J_n^{j_1,j_2} &= \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_2} \circ\, dW_{s_1}^{j_1} \circ dW_{s_2}^{j_2}
\end{aligned}$$

Then the explicit Milstein scheme is given by

$$\xi_{n+1}^k = \xi_n^k + h_n^k \Delta_n + \sum_{j=1}^{m} g_n^{ij} \Delta W_n^j + \sum_{j_1,j_2=1}^{m} \sum_{i=1}^{d} g_n^{ij_1} \frac{\partial}{\partial x_i} g_n^{kj_2} J_n^{j_1,j_2}. \tag{3.16}$$

One can well imagine that especially the multiple integration terms J are a potential source of problems. Applying this scheme to the coupled noise system (3.15), we obtain $m = d = 2$ and

$$\begin{aligned}
h_n^1 &= -\xi_n^2 \xi_n^1 \\
h_n^2 &= -2\Theta\xi_n^2 \\
G = (g_n^{ij})_{ij=1}^2 &= \begin{pmatrix} \xi_n^2 \sigma & 0 \\ 0 & 2\sqrt{D\xi_n^2} \end{pmatrix},
\end{aligned}$$

which leaves only the terms containing $J_n^{2,1}$ and $J_n^{2,2} = \frac{1}{2}(\Delta W_n^2)^2$ in the last sum and yields the discretized model

$$\begin{aligned}
\xi_{n+1}^1 &= \xi_n^1 - \xi_n^2 \xi_n^1 \Delta_n + \sigma\xi_n^2 \Delta W_n^1 + 2\sqrt{D\xi_n^2}\sigma J_n^{2,1} \\
\xi_{n+1}^2 &= \xi_n^2 - 2\Theta\xi_n^2 \Delta_n + 2\sqrt{D\xi_n^2}\Delta W_n^2 + D(\Delta W_n^2)^2.
\end{aligned} \tag{3.17}$$

A detailed derivation for a suitable approximation of the multiple Stratonovich integral $J_n^{j_1,j_2}$ via Fourier expansions is given in [44, Ch. 5.8]. This leads to

$$\begin{aligned} J_{n,p}^{i,j} &= \frac{1}{2}\Delta W_n^i \Delta W_n^j + \sqrt{\rho_p \Delta_n}\left(\mu_n^{i,p}\Delta W_n^j - \mu_n^{j,p}\Delta W_n^i\right) \\ &\quad + \frac{\Delta_n}{2\pi}\sum_{r=1}^{p}\frac{1}{r}\left[\zeta_n^{i,r}\left(\sqrt{\frac{2}{\Delta_n}}\Delta W_n^j + \eta_n^{j,r}\right) - \zeta_n^{j,r}\left(\sqrt{\frac{2}{\Delta_n}}\Delta W_n^i + \eta_n^{i,r}\right)\right] \\ \rho_p &= \frac{1}{12} - \frac{1}{2\pi^2}\sum_{r=1}^{p}\frac{1}{r^2} \end{aligned}$$

with $\mathcal{N}(0,1)$ distributed random variables $\mu_n^{i,p}, \eta_n^{i,r}, \zeta_n^{i,r}$. Since the Fourier expansion is truncated at $p \in \mathbb{N}$, this parameter defines the quality of the approximation.

Given initial values $\xi_0^1 = T_0 > 0$ and $\xi_0^2 = \epsilon_0^2$, $\epsilon_0 \sim \mathcal{N}(0, 2D/\Theta)$, we are now in a position to simulate sample paths for the coupled noise system (3.15). However, it turns out that the numerical scheme constructed above fails. The reason for this failure is the infraction of a Lipschitz continuity requirement for drift- and diffusion coefficients, i.e., $\sqrt{D\xi_2}$ is not Lipschitz continuous in ξ_2 at zero. The reasons why we still present this inappropriate scheme are twofold:

- The implicit Milstein scheme discussed below does not fail and is easy to derive from the explicit Milstein scheme.
- In the majority of nonlinear meteorological applications, the assumption of Lipschitz continuity is not satisfied, e.g. for the primitive and Navier-Stokes equations. However this issue is seldom raised, even if explicit schemes for SODE are used. The coupled noise model is therefore a good showcase why the Lipschitz condition is not just a mathematical sophistry but essential in order to guarantee convergence.

3.5.1 The Failure of the Explicit Milstein Scheme

Due to the simplicity of this showcase model, we are able to pinpoint the term which causes the explicit Milstein scheme to fail. We can even give a failure probability for each time step. To this end, recall that we defined the process $\gamma \equiv \xi^2$ as the square of an OUP, which is naturally a nonnegative process. However, one gets from (3.17)

$$\begin{aligned} \xi_{n+1}^2 &= \xi_n^2 - 2\Theta\xi_n^2\Delta_n + 2\sqrt{D\xi_n^2}\Delta W_n^2 + D(\Delta W_n^2)^2 \\ &= \xi_n^2 - 2\Theta\xi_n^2\Delta_n + 2\sqrt{D\xi_n^2\Delta_n}\eta + D\Delta_n\eta^2, \qquad \eta \sim \mathcal{N}(0,1) \\ &= \left(\sqrt{D\Delta_n}\eta + \sqrt{\xi_n^2}\right)^2 - 2\Theta\Delta_n\xi_n^2. \end{aligned}$$

This leads to

$$\begin{aligned} & \xi_{n+1}^2 < 0 \\ \Leftrightarrow \quad & \left|\sqrt{D\Delta_n}\eta + \sqrt{\xi_n^2}\right| < \sqrt{2\Theta\Delta_n\xi_n^2} \\ \Leftrightarrow \quad & \left|\eta + \sqrt{\frac{\xi_n^2}{D\Delta_n}}\right| < \sqrt{\frac{2\Theta\xi_n^2}{D}}. \end{aligned}$$

Since η is a $\mathcal{N}(0,1)$ distributed random variable, the probability $P(\xi_{n+1}^2 < 0 \mid \xi_n^2)$ can be expressed as

$$P(\xi_{n+1}^2 < 0 \mid \xi_n^2) = P\left(|\tilde{\eta}| < \sqrt{\frac{2\Theta\xi_n^2}{D}}\right), \qquad \text{where} \quad \tilde{\eta} \sim \mathcal{N}\left(\sqrt{\frac{\xi_n^2}{D\Delta_n}}, 1\right), \quad (3.18)$$

which does not vanish for $\xi_n^2 > 0$.

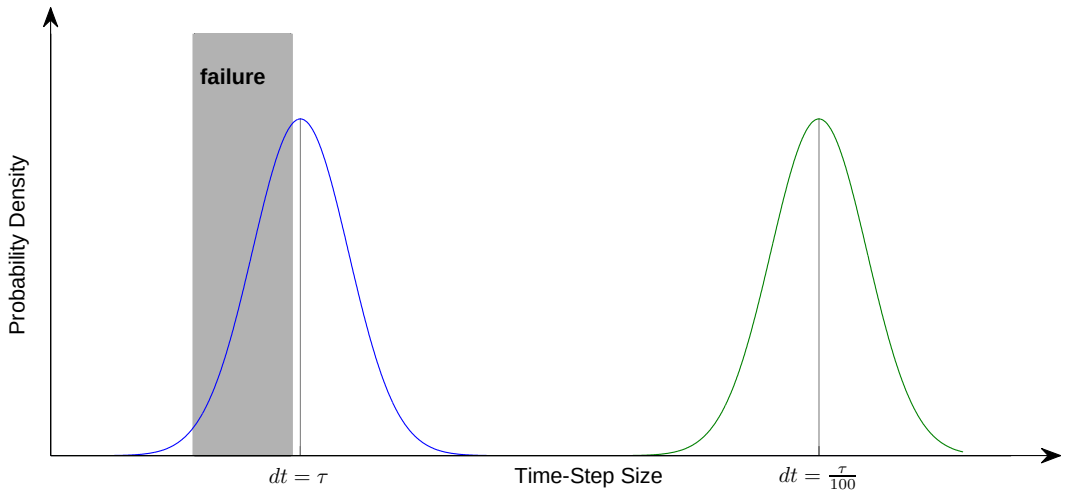


Figure 8: *Failure Probability for the Explicit Milstein Scheme*
The plot shows the failure probability of the $(n+1)st$ time step for distinct time-step sizes, where the value of all other variables, including the stochastic processes up to time-step n, are assumed to be known. Smaller time-step sizes lead to a rapidly decreasing failure probability. Note that this interpretation only applies to the conditional probabilities for fixed ξ_n^2.

Therefore, the discretization (3.17) has the potential to yield negative values for ξ^2. This is particularly problematic because one has to use the square root $\sqrt{\xi^2}$, resulting in the failure of the explicit Milstein scheme. The distribution of $\tilde{\eta}$ in (3.18) depends on the time-step size Δ_n, which might lead to the idea that it becomes small enough to ignore in practical application for sufficiently small time-step size. We discuss in the following that this is not the case. Although the schematic diagram for the failure probability (Fig. 8) seems to support this idea, we have to consider that this diagram assumes a fixed ξ_n, which is a random variable in the numerical scheme. Furthermore, for a fixed time interval a finer time-step size implies that the scheme has more "*chances*" to fail, which potentially cancels the effect of a smaller failure probability. Numerical experiments for 10.000 realizations, which are stopped at the first time the numerical scheme failed, show that the aforementioned effects indeed cancel each other. This leads to a median of the "*first-error time*", which is independent of the time-step size (Fig. 9).

3.5.2 The Implicit Milstein Scheme

The issue (3.18) for non-Lipschitz continuous systems does not occur if the following implicit Milstein Scheme [44, p. 400] is used:

$$\begin{aligned}\xi_{n+1}^k = \xi_n^k &+ \left(\alpha^k h_{n+1}^k + (1-\alpha^k) h_n^k\right) \Delta_n \\ &+ \sum_{j=1}^{m} g_n^{ij} \Delta W_n^j + \sum_{j_1,j_2=1}^{m} \sum_{i=1}^{d} g_n^{ij_1} \frac{\partial}{\partial x_i} g_n^{kj_2} J_n^{j_1,j_2},\end{aligned}$$

where the parameters $\alpha^k \in [0,1]$ indicate the degree of implicitness. Note that, compared to the explicit Milstein scheme (3.16), only the deterministic drift terms of the discretization are modified. Using this scheme yields the discretized model

$$\begin{aligned}\left(1+\alpha^1 \Delta_n \xi_{n+1}^2\right)\xi_{n+1}^1 &= \left(1-(1-\alpha^1)\Delta_n \xi_n^2\right)\xi_n^1 + \sigma \xi_n^2 \Delta W_n^1 + 2\sqrt{D\xi_n^2}\sigma J_n^{2,1} \\ \left(1+2\alpha^2 \Theta \Delta_n\right)\xi_{n+1}^2 &= \left(1-2(1-\alpha^2)\Theta\Delta_n\right)\xi_n^2 + 2\sqrt{D\xi_n^2}\Delta W_n^2 + D(\Delta W_n^2)^2.\end{aligned} \tag{3.19}$$

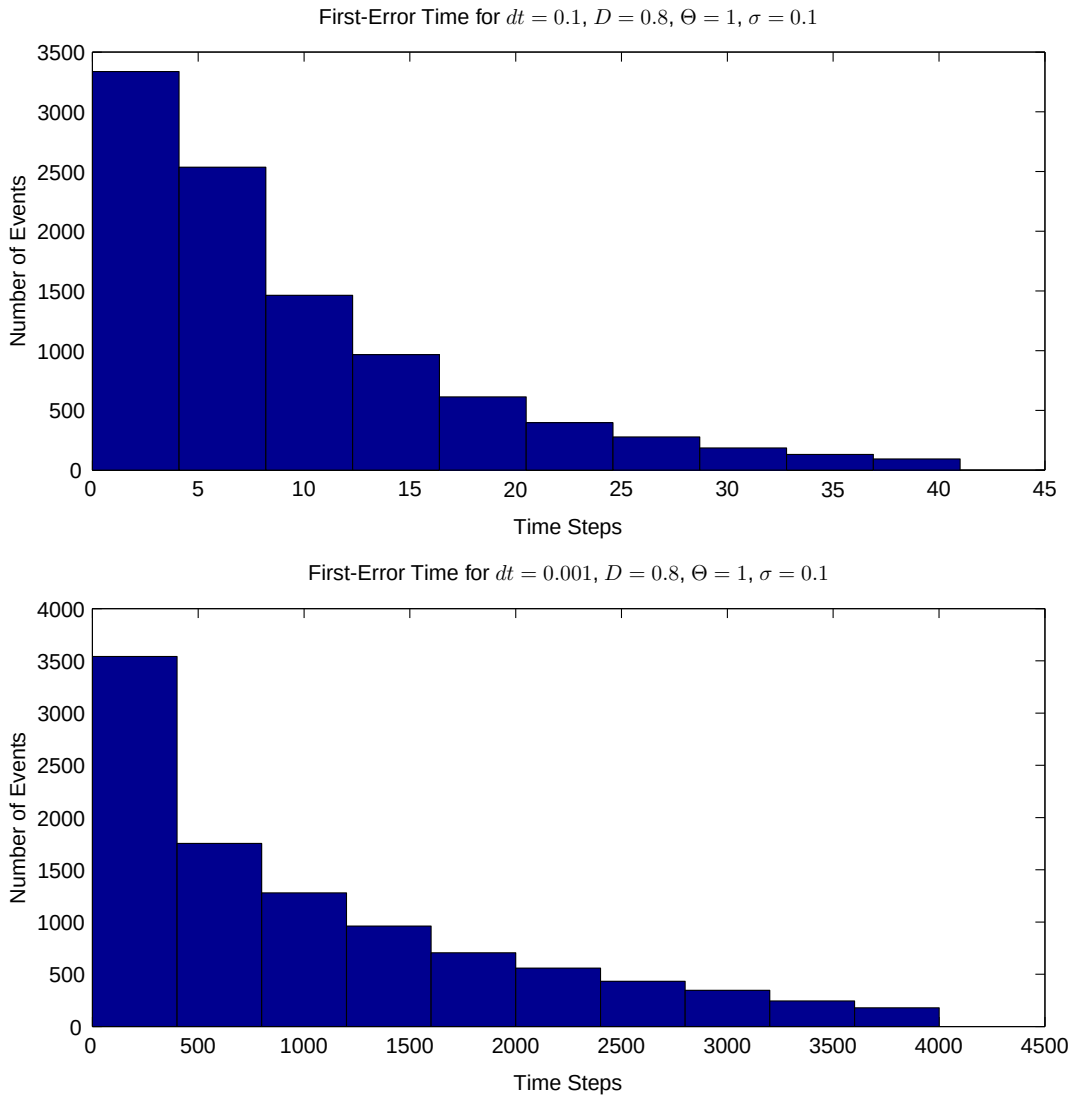


Figure 9: *First-Error Time for the Explicit Milstein Scheme*
The explicit Milstein scheme 3.17 was used to simulate 10.000 realizations of the coupled noise model. The plots show the empirical distribution of the first-error time step, i.e., the time step at which a realization of ξ^2 obtains a negative value for the first time. Note that the time-step size size of the lower plot is smaller by a factor of 100. The first-error time coincides with a median of 0.7 for the first, and 0.72 for the second case. The failure of the explicit Milstein scheme does not depend strongly on the time-step size.

In analogy to the previous section the second equation yields

$$\begin{aligned}
& \xi_{n+1}^2 < 0 \\
\Leftrightarrow \quad & \big(1 - 2(1-\alpha^2)\Theta\Delta_n\big)\xi_n^2 + 2\sqrt{D\xi_n^2}\Delta W_n^2 + D(\Delta W_n^2)^2 < 0 \\
\Leftrightarrow \quad & \left(\sqrt{D}\eta + \sqrt{\xi_n^2}\right)^2 - \xi_n^2 + \big(1 - 2(1-\alpha^2)\Theta\Delta_n\big)\xi_n^2 < 0, \qquad \eta \sim \mathcal{N}(0,1) \\
\Leftrightarrow \quad & \left|\eta + \sqrt{\frac{\xi_n^2}{D}}\right| < \sqrt{\frac{2(1-\alpha^2)\Theta\Delta_n\xi_n^2}{D}}.
\end{aligned}$$

Therefore, the probability $P(\xi_{n+1}^2 < 0 \mid \xi_n^2)$ vanishes if and only if $\alpha^2 = 1$, i.e., if one uses a Milstein scheme fully implicit in the second component. Due to the one-sided coupling of the two SDE (3.19), i.e., ξ^1 does not appear in the equation defining ξ^2, the implicit scheme can be implemented instead of the explicit one without additional computation costs. We would like to emphasize that this analysis only shows that one particular issue (3.18) is solved by an application of the implicit scheme. It does in no way guarantee the actual convergence of the Milstein scheme. Note however, that we can rewrite the coupled noise model to obtain the form

$$\begin{aligned}
dT_t &= -\epsilon_t^2 T_t dt + \sigma\epsilon_t^2 \circ dW_t^1 \\
d\epsilon_t &= -\Theta\epsilon_t dt + \sqrt{D} \circ dW_t^2, \quad \epsilon_0 \sim \mathcal{N}\left(0, \frac{D}{2\Theta}\right).
\end{aligned}$$

The coefficients $-\Theta\epsilon_t$ and $\sqrt{D}$ in the defining equation of the OUP ϵ are Lipschitz continuous. We can thus use an explicit or implicit Milstein scheme to simulate sample paths of ϵ. Furthermore, the coefficients $-\epsilon_t^2 T_t$ and $\sigma\epsilon_t^2$ of the state variable T are Lipschitz continuous in T, which again allows the application of the aforementioned numerical schemes [44]. We are in an unusually comfortable position, since the equations are only coupled in one direction, permitting the Lipschitz conditions to be considered separately. For more complex models, it is a highly nontrivial task to derive convergent numerical schemes, which is discussed in detail in Sections 6 and 7.

Using the implicit Milstein scheme (3.19), we can simulate sample paths of the coupled noise model. Fig. 10 shows a single realization of the temperature process and the two driving noise processes, i.e., the multiplicative inverse climate sensitivity and the additive Brownian motion. Since the multiplicative process is nonnegative, it can only cause a convergence to the equilibrial state $T = 0$. The additive noise, however, is not restricted in this way and can lead to positive and negative temperature fluctuations. The coupling of the additive and multiplicative process, i.e., the strength of the additive process depends on the value of the multiplicative one, exhibits two interesting results. First, we can observe time windows, in which the system is virtually force free, resulting in "*plateaus*" of constant temperature, see for instance the interval $t \in [2.1, 2.5]$ in Fig. 10. This is the case if the inverse climate sensitivity is very small. Second, if the multiplicative process is very large, the system is pushed in the direction of its equilibrium. Simultaneously, the strength of the additive process is proportionally large, resulting in strong fluctuations in either direction. For a constant inverse climate sensitivity, this behavior corresponds exactly to the one we observe for an OUP: a mean-reversing drift combined with an additive white noise. It is therefore no surprise to find a similar stationary distribution for the temperature process, see Fig. 10. Although the marginal distribution is Gaussian, the aforementioned plateaus cannot be observed for an OUP. This emphasizes the importance to consider both pathwise and statistical analysis of a stochastic dynamic system in order to understand its characteristics.

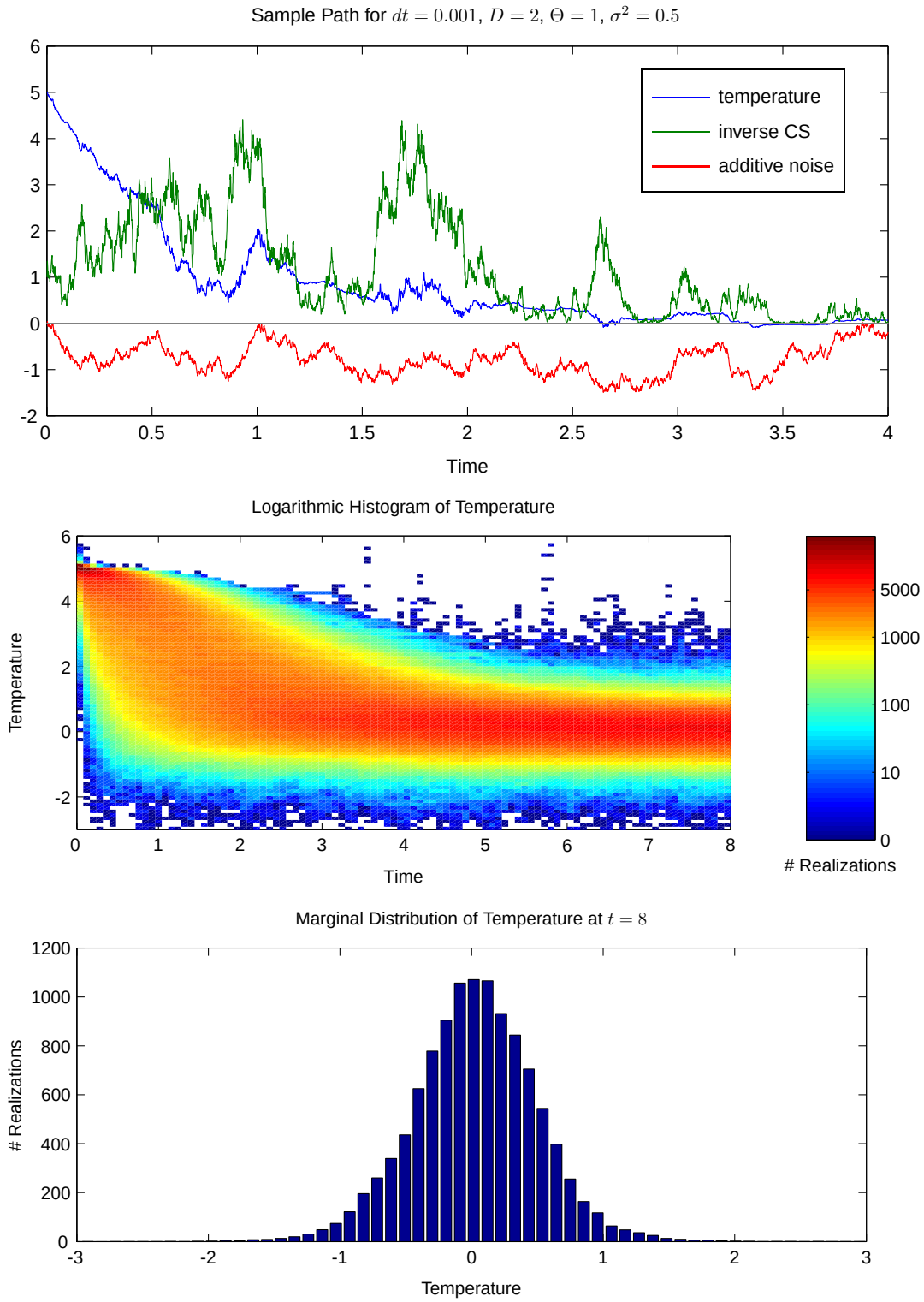


Figure 10: *Histogram and Marginal Distribution for the Coupled Noise Model*
Using the implicit Milstein scheme (3.19) with parameters $\mu = 0$, $\Theta = 1$, $D = 2$, $\sigma^2 = 0.5$ and time-step size $dt = 0.001$, we simulated 10.000 realizations of the coupled noise model. The first plot shows a sample path, including temperature and both driving stochastic processes. The lower two plots show the convergence of the temperature process into an equilibrial state at $T = 0$. Furthermore, the resulting marginal distribution in equilibrium at $t = 8$ is Gaussian.

3.6 Conclusion

We studied various formulations for uncertainties of the climate sensitivity parameter during convergence to equilibrium. Climate sensitivity and the resulting equilibrial states are two of the most discussed topics in the debate of climate change and the convergence process towards new equilibria is of great importance in this context. There is a close connection to the analysis of transient climate projections, which incorporate the fact that, due to the large heat capacity of the oceans, the convergence process takes hundreds or even thousands of years [99, Ch. 9]. The convergence process itself is therefore crucial for discussions centered around political decision making concerning the next decades. One of the reasons for the popularity of the climate sensitivity parameter is its easy accessibility due to a very basic mathematical structure, i.e., a one-dimensional linear differential equation. In the context of this work, the simple structure allows us to point out some fundamentally important phenomena arising during the treatment of stochastic meteorological systems.

The analytical results in Sections 3.2 and 3.3 show the occurrence of a stochastic drift in systems where uncertainties are present in a multiplicative way. Furthermore, Lemma 3.2 demonstrates that the mere presence of uncertainties leads to a slower convergence into equilibrium (in a mean value sense), regardless of the particular characteristics of the stochastic terms. For non-delta correlated noise, i.e., for systems where the fluctuations exhibit a time dependent correlation structure, we observe a time lag in the stochastic drift term as well as memory effects. The occurrence of different stability regimes for the different stochastic models emphasizes the importance to choose stochastic terms (including the type of stochastic integral) very carefully, see Propositions 3.1, 3.5 and 3.13. This decision-making process already begins with the basic mathematical formulation of a model. Consider the following models:

$$dT = -\frac{1}{\lambda} T dt \quad \text{and} \quad dT = -\gamma T dt,$$

using the climate sensitivity λ and its inverse γ, respectively. Although the climate sensitivity seems to be the physically more meaningful parameter, the first representation automatically excludes many stochastic formulations, e.g., ones with a positive probability for $\lambda = 0$. While this seems to be a trivial observation in this simple case, it may not be so obvious in more complex models. Conclusively, a stochastic formulation must not be arbitrarily chosen but has to be based on the physics of the system at hand.

Regarding the numerical simulation of stochastic models, we have addressed the discrepancy between the wish to capture extreme events and the usually low number of simulated sample paths. Furthermore, we have seen that the numerical implementation of the rather simple coupled noise model has to be done carefully, see Section 3.4. Even the explicit Milstein scheme – a numerical scheme specifically designed for SODE – produces instabilities, because a seemingly subtle condition is not satisfied: the Lipschitz continuity of all coefficient functions. The fact that this is a very simple model compared to meteorological relevant system, e.g., Navier-Stokes equations or Primitive equations, emphasizes the necessity to use a numerical implementation based on the type of the occurring stochastic integrals, which is capable to deal with the – often nonlinear – coefficients. This is particularly problematic in systems involving stochastic partial differential equations (see Sections 6 and 7). It should go without saying that this excludes any and all integration schemes designed for deterministic differential equations.

In the following Section 4 we propose a concrete method to fit a stochastic process on a given set of ice core data. This is illustrated by means of an energy budget model, which can be considered as a natural extension of the linear system studied above. In Sections 6 and 7 we address the mathematically demanding issue of a rigorous numerical scheme for the three-dimensional primitive equations of the atmosphere. These equations are of high meteorological relevance, since they constitute the dynamical core of virtually every GCM.

4 An Energy Budget Model

In the previous section we have discussed the importance of carefully choosing stochastic terms representing physically meaningful processes. In the following we present a method to do this on the basis of a concrete example using carbon dioxide data from ice core drillings to derive a stochastic formulation of an energy budget model by Budyko and Sellers [107, 108], describing the evolution of the global mean temperature. The one-dimensional differential equation

$$c\,\frac{dT_t}{dt} = S_t(1-\alpha(T)) - k_B(T_t - \Delta T)^4,$$

which is studied in Section 4.2, can be considered the natural nonlinear extension of the linear climate sensitivity system. Instigated by Hasselmann's paper [4], it was Fraederich, who introduced stochastic forcing in form of an additive white noise term into this model [109]. Since then it has been an active field of research, highlighted in the review papers [110, 111] and closely connected to the phenomenon of stochastic resonance [112].

After defining the deterministic framework in Section 4.2 we point out physical issues of the model derived from a white noise ansatz in Section 4.3.1. A different approach using physically based stochastics is proposed with a focus on spectral ice core data analysis in Section 4.3.2. It turns out that the six different sets of ice core data lead to three distinct parameter classes for the stochastic terms. We consider one exemplary set of data for each parameter class. The simple mathematical structure of the model allows very cost effective numerical simulations of the nonlinear stochastic terms in Section 4.3.4, paving the way for comprehensive statistical evaluations in Section 4.4. Particular emphasis is placed on the pathwise analysis of the model since basic statistical properties are unable to capture some of the system's crucial characteristics, i.e., bifurcation structure, timescales of crossings between stable states and correlation between insolation and temperature. We develop a basic understanding of the model's behavior by studying single realizations and marginal distributions of the temperature at local insolation extrema. A coherence analysis in Section 4.4.3 yields solid evidence on the differences between the three models, i.e., depending on the choice of ice core data we obtain fundamentally differing temperature processes. Since there is no reason why a particular set of data should be preferred, we come to the conclusion that the model is inapt to be fitted on the type of data available, which is discussed in detail in Section 4.5. This emphasizes once more the importance of choosing the stochastic terms very careful even in such a basic framework. We would like to underline that the main focus of this section does not lie in the concrete results for the stochastic EBM, but rather in the methods used to derive them:

1. Identification of a variable suited to be described stochastically
2. Spectral analysis of data to determine the distribution of the stochastic process (SP)
3. A rigorous numerical implementation
4. Statistical and pathwise analysis of the stochastic dynamic model

This methodical framework can and should be used for stochastic formulations of more complex models. However, we discuss in Sections 6 and 7, that the development of rigorous numerical schemes is a challenging task in the context of nonlinear stochastic partial differential equations.

4.1 A Method to Derive Physically Based Stochastic Models

In most cases the starting point for the development of a stochastic model is a deterministic one, which captures the governing dynamics. Time dependent SPs bear the potential to describe the dynamics of fluctuations and uncertainties of model variables, for instance due to unresolved sub grid processes, with an emphasis on the word "*dynamics*". While

stochastic parameterizations can be used to represent uncertainties of model parameters, initial values and other time independent variables, stochastic processes provide a tool to lay a hand on time and state dependent interactions between sub grid fluctuations and model variables. It is indispensable to carefully choose a SP based on the physics of a given model. In a more recent article Penland and Ewald state, that "*simply replacing the fast term with a Gaussian random deviate with standard deviation equal to that of the variable to be approximated, and then using deterministic numerical integration schemes, is a recipe for disaster*"[16]. While the use of a correct stochastic numerical approach is equally important we focus on the problem of finding a physically based SPs in this section. As the first step we have to identify unresolved physical processes which may have an effect on the model variables. To this end we take a look at each occurring physical variable whose dynamics are not captured by the model, especially the ones which are approximated by constant values. For two kinds of physical quantities a stochastic representation is uncalled-for: "*true*" constants, e.g. the Stefan-Boltzmann-constant k_B, and nonconstant variables that can accurately be described in a deterministic way, e.g. incoming solar radiation depending on Milankovich cycles. The interesting variables are inexact approximations of processes that cannot be described in a deterministic way within the models framework. Stochastic representations are also limited by the models space and time scales, providing another criterion in the selection process. When a eligible variable, or a set of eligible variables, is found we need input on the distribution of the stochastic process. This may be information based on finer resolved models, theoretical insights or data. Deducing the complete distribution of the SP can be a nontrivial task and should be done carefully, since even small variations may lead to significant changes in the dynamic structure of the system in the context of chaotic systems.

4.2 Deterministic Framework

The deterministic energy budget model (EBM) is given by an ordinary differential equation (ODE) for the global mean temperature T, characterizing the radiation balance between net short wave radiation and outgoing long wave black body radiation

$$c\, dT_t = \big[(1-\alpha(T))S_t - k_B(T_t - \Delta T)^4\big]dt. \tag{4.1}$$

Here $c = 3 * 10^8 J/(m^2K)$ denotes the global heat capacity, S_t the (time dependent) insolation, $\alpha(T)$ the albedo, $k_B = 5.67 * 10^{-8} W/(m^2K^4)$ the Stefan-Boltzmann-constant and $\Delta T = 32.6K$ the temperature offset between surface and the top of the atmosphere. Since we are dealing with a model with spatial dimension zero, obliquity and precession are not accounted for and the solar forcing is solely dependent on the eccentricity of the earth orbit (right plot of Fig.11). Note that the absolute value of S_t only varies within the small interval between $341.5 W/m^2$ and $342.0 W/m^2$. We choose an albedo parameterization so that the model to reproduces stable states at typical ice age and warm age temperatures obtained from ice core data (for instance [113]). More specifically we set:

- $\alpha(T_1) = 0.3901$ for $T_1 = 278.9K$ (*ice age*)
- $\alpha(T_2) = 0.2950$ for $T_2 = 288.0K$ (*present state*)
- $\alpha(T_3) = 0.2676$ for $T_3 = 290.3K$ (*warm age*)

with a linear interpolation in between, as shown in the left plot of Fig.11. Regarding the systems dynamic structure, our primary interest lies in the study of equilibria, i.e., the roots of the right hand side of equation (4.1):

$$S_t(1-\alpha(T)) - k_B(T_t - \Delta T)^4 = 0.$$

As a result of the specific albedo parameterization we have three stable and two instable equilibria. The outer two stable equilibria are located at $T_{ice} = 278.9K$ and $T_{warm} = 290.3K$, as shown in Fig. 12. The location of these two equilibria is insusceptible to variations in insolation. However, the location of the third stable equilibrium strongly depends on the solar forcing and varies within the interval $[283.5K, 288.0K]$. Note that a more strongly fluctuating insolation would create areas where one or more of these equilibria would not exist, which would cause the system to "*jump*" to the nearest existing stable state. However, in this deterministic setting the variations in solar forcing are weak. Therefore the system eventually settles in one of the stable states without the possibility to cross between these states.

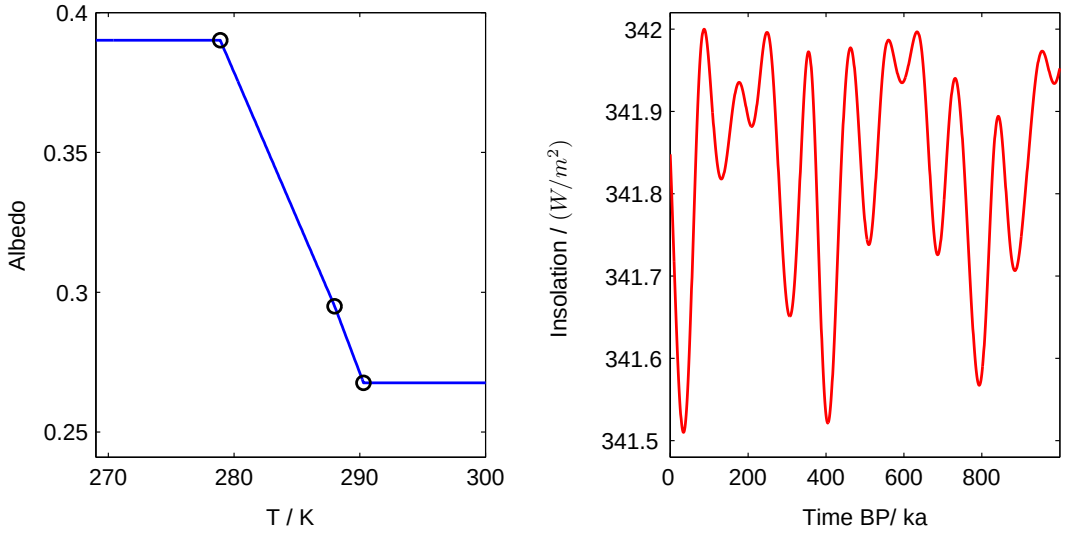


Figure 11: *Albedo Parameterization and Insolation Cycle*
Albedo parameterization (left) and insolation cycle (right) used in the deterministic framework. The albedo was derived from stable ice- and warm-age states as well as the intermediate current state, indicated by circles in the plot on the left hand side.

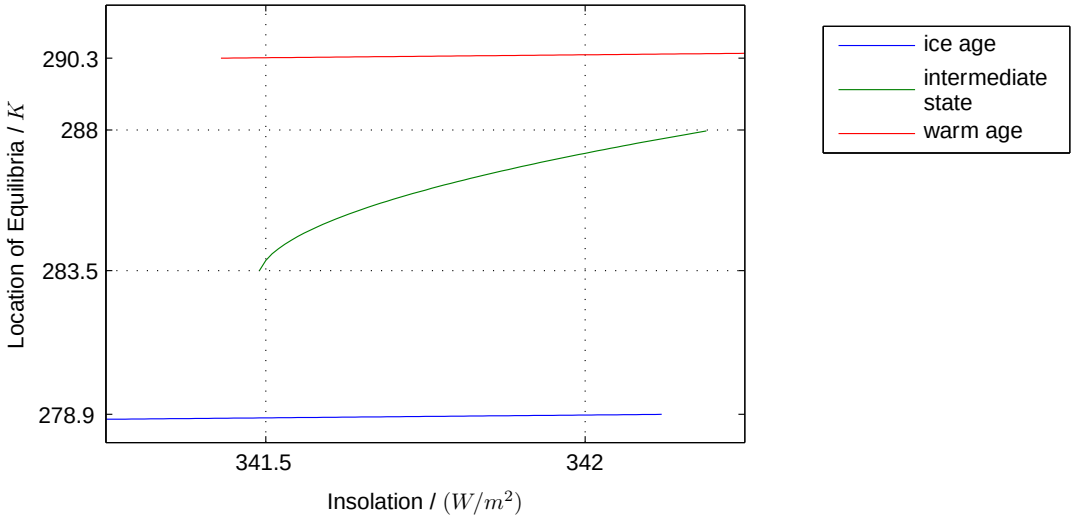


Figure 12: *Equilibrial Structure of the EBM*
The three stable equilibria of the EBM system are plotted against the insolation S_t with values in the interval $[341.5, 342.0]W/m^2$, marked by vertical grid lines. Obviously the outer equilibria are insusceptible to the fluctuations in insolation, whereas the location of the intermediate steady state strongly depends on the value of the solar forcing.

4.3 A Stochastic EBM

4.3.1 Motivation

Contrary to the aforementioned steady state behavior, time series from ice core data reveal that the temperature is not constant but rather a strongly fluctuating process, varying between ice- and warm-ages. This motivates a stochastic formulation, since the deterministic framework is inapt to capture this dynamic. One common approach is the addition of stochastic forcing in form of an additive white noise term [109]. While being a reasonable first step, this ansatz is disputable concerning

- *additivity*
  There is no temperature independent variable in the model, begging the question what kind of physical process would be described by a temperature independent noise term.
- *whiteness*
  Since the physical processes described by this model act on large space- and timescales it is hard to imagine why a non-correlated noise should adequately capture the occurring fluctuations.

Therefore we follow a different approach: rather than using a purely mathematical tool to modify the model in order to produce the desired effect, we take a look at the physical variables, which are already present in the deterministic EBM, with the intention to describe one of them as a SP. For a successful stochastic formulation a variable has to satisfy the following conditions: foremost there has to be a significant fluctuation of the global mean since the model does not resolve spatial dimensions. These fluctuations have to act on a timescale that can be covered by the model, i.e., not smaller than one year. Finally we need evidence on the distribution of the SP. This can be achieved by means of a more sophisticated model, which resolves the variable in question, or through data analysis. Due to the simplicity of the deterministic model our choices are very limited and we identify ΔT as a suitable candidate, where its variations correspond, for instance, to varying concentrations of greenhouse gases.

4.3.2 Time Series Data

In order to derive the characteristics of the fluctuations we use the following sets of ice core data: Vostok [113], DomeFuji [114], Byrd [115], EPICADome [116], TaylorDome[117], TaylorDome$_2$ [118], which were chosen due to their good resolution and large time span. Graphs of Vostok, Byrd and Taylor data (Fig.13) exemplary show that the scale-demands are met and provide direct input on the distribution of the SP. Since there is a strong correlation between CO_2 concentration and temperature, we face the danger of predefining the models equilibrium structure if the data contains crossings between warm- and ice-ages. This issue arises primarily from the simplicity of the model, which results in very few degrees of freedom. Hence the SP should describe the "*natural*" small fluctuations of ΔT excluding feedback effects with the temperature, i.e., a distribution independent of the state variable T. Therefore we study detrended, dimensionless time series in time intervals containing as few crossings between ice- and warm-ages as possible. The first step is the conversion of CO_2 data to temperature anomalies, which directly involves the climate sensitivity parameter. Here we assume that a doubling in CO_2 concentration yields a constant and deterministic ΔT increase of $2K$ and note that this value is obviously subject to large uncertainty (see Section 3).

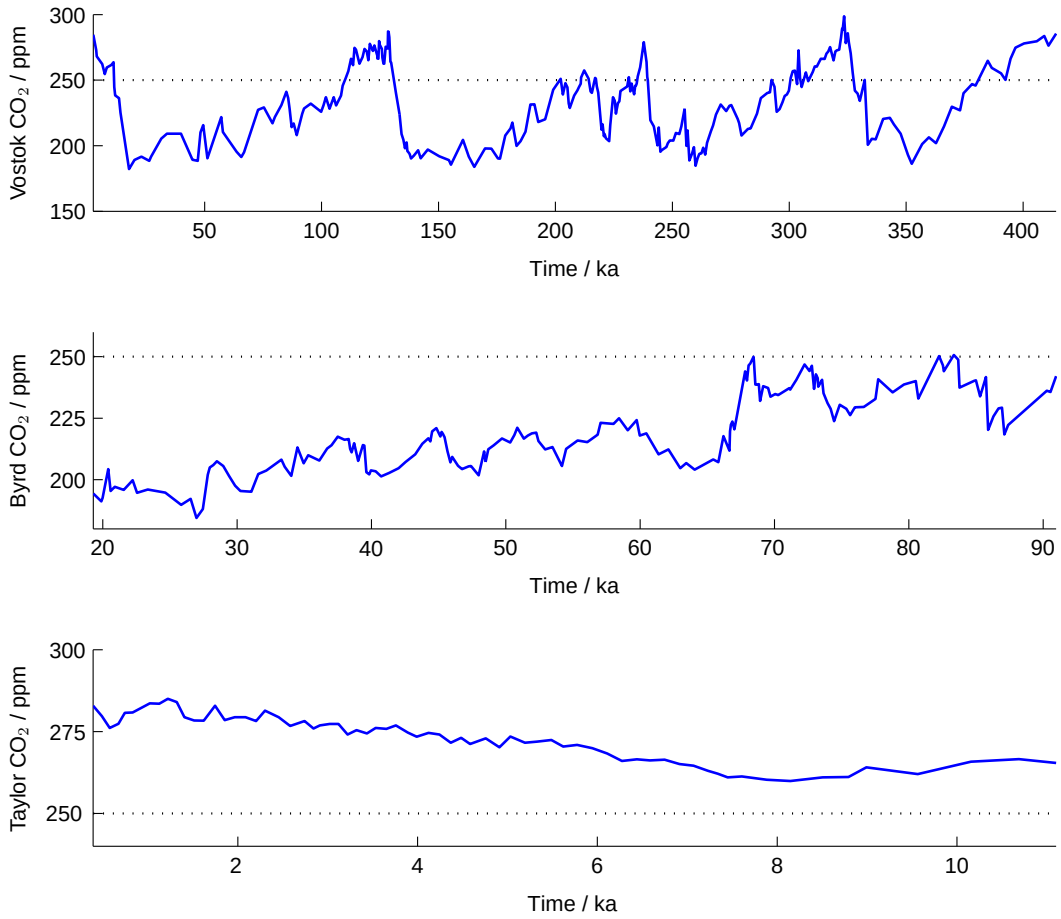


Figure 13: *Ice Core Time Series*
The plots show the CO_2 concentration from three exemplary ice core time series: Vostok, Byrd and TaylorDome. Note the different horizontal and vertical scales. Obviously both amplitude and frequency of the fluctuations are not homogeneous among the different sets of data.

We obtain

$$\Delta T' = \frac{2}{\log 2} \log\Big(1 + \frac{\text{detrend}(CO_2)}{\text{mean}(CO_2)}\Big),$$

where $\Delta T'$ is the desired data of temperature fluctuations, *detrend*(CO_2) the lineary detrended data from ice core drillings and *mean*(CO_2) the sample mean value of the CO_2 time series. In order to characterize the SP we use a spectral analysis of the $\Delta T'$ time series. It turns out that the distribution closely follows an OUP

$$d\Delta T_t = -\Theta\big(\Delta T_t - \mu\big)dt + \sqrt{D}\, dW_t,$$

with parameters Θ, μ and D depending on data. Therefore, the distribution of the SP is (close to) a stationary Gaussian distribution (see Fig. 14), which we discuss in more detail in Section 4.3.3. Note that Ito and Stratonovich calculus are identical in this case, due to the constant diffusion coefficient $\sqrt{D}$, see Lemma 2.24. For a stationary process the expectation value μ should naturally be equal to the value of ΔT used in the deterministic model. The remaining parameters, which are responsible for the fluctuations, are fitted to the spectral data in order to confirm the validity of the red noise assumption. Note however that the confidence interval for the parameter values of each time series is quite large due to low data size. In addition the parameters calculated from different time series differ strongly and it is not obvious which parameters we should trust. We denote the sample decorrelation time by τ and the sample variance of each time series by *var*.

Data	$\tau[ka]$	var $\cdot 10^{-3}$	timespan $[ka]$	# data points
Vostok	16.7	10.8	410	283
DomeFuji	14.0	10.6	340	290
Byrd	2.2	11.2	70	171
TaylorDome$_2$	2.3	7.2	42	73
EPICADome	2.0	11.4	12	72
TaylorDome	0.6	1.0	11	69

Table 2: *Statistical Properties of Ice Core Data*
Statistical properties of dimensionless, detrended ice core data, which can be divided into three distinct classes depending on decorrelation time and variance. One representative of each class was chosen for further analysis: Vostok, Byrd and TaylorDome.

Following Table 2, we can divide these time series into the three classes [DomeFuji and Vostok], [Byrd, EPICADome and TaylorDome$_2$] and [TaylorDome]. There are pro and contra arguments for each class: the first one consists of very large data sets covering long time periods. While the high amount of data points is certainly a positive aspect it implies the problem of embedded crossings between ice- and warm-ages. The second class contains the majority of time series where each member yields similar parameters. The third class consists of only one scarce set of data with only 69 data points. Nevertheless it leads to the weakest noise signal, so one would be "on the safe side" using these parameters. Since there is no good argument to exclude either class we analyzed the system for one time series of each category: Vostok,Byrd and TaylorDome.

4.3.3 Deriving the Parameters of the OUP

We have already discussed one possible method to fit an OUP onto given data by directly calculating its parameters from the statistical characteristics of the time series (see Lemma 3.16). However this approach needs the rather utopian assumption of a stationary Gaussian distributed set of data. In practical applications we might be lucky enough to have access to time series, which are "*close to*" a Gaussian distribution. Therefore we need a method, which works equally well in the ideal case of Gaussian distributions, but also yields solid results for other types of data. We observe that an OUP can be considered as the continuous time analog to a discrete-time first order autoregressive process.

Definition 4.1. AR(p) Process
The autoregressive process of order $p \in \mathbb{N}$ is denoted by AR(p) and defined by

$$X_n = c + \sum_{k=1}^{p} \phi_k X_{n-k} + \sigma\varepsilon_n, \quad n \geq 0$$

for constants $c, \phi_1, ..., \phi_p, \sigma \in \mathbb{R}$ and independent, identically distributed (iid) random variables ε_n with $E[\varepsilon] = 0$ and var $(\varepsilon) = 1$.

Remark 4.2. OUP as an AR(1) Process
The parameter p in Def. 4.1 determines the memory of the process. For $p = 1$ the next step of the process only depends on its current value, but not on the past, which is known as "Markov property". A famous result by Doob states that every OUP is a Markov process. Vice versa, every stationary Gaussian process $(Y_t)_{t\geq 0}$, which satisfies the Markov property and is continuous in probability, i.e.

$$P(|Y_t - Y_s| > \epsilon) \to 0 \quad \textit{as} \quad s \to t,$$

for all $\epsilon, t > 0$, is an OUP [119]. Rewriting Def. 4.1 for $p = 1$ yields

$$\begin{aligned} & X_n = c + \phi_1 X_{n-1} + \sigma\varepsilon_n \\ \Leftrightarrow\ & X_n - X_{n-1} = -(1-\phi_1)\left(X_{n-1} - \frac{c}{1-\phi_1}\right) + \sigma\varepsilon_n. \end{aligned}$$

For $\Theta := 1 - \phi_1$ and $\mu := c/(1-\phi_1)$ we obtain

$$\Delta X_n = -\Theta\left(X_{n-1} - \mu\right) + \sigma\varepsilon_n,$$

which is the discrete analog to the defining SDE of the OUP

$$dY_t = -\Theta\left(Y_t - \mu\right)dt + \sigma dW_t.$$

Since AR(p) models play an important role in time series analysis, we can take advantage of a rich theoretical and numerical knowledge from meteorology [120], physics [121] and economics [122]. In order to derive the parameters of the OUP from ice-core data we use a method named after Yule [5] and Walker [6], which has lost none of its appeal since its publication 80 years ago. Without loss of generality we assume $c = 0$ for greater clarity during the following discussion. Then an AR(p) process X satisfies

$$\sum_{k=0}^{p} \phi_k X_{n-k} = \sigma\varepsilon_n, \quad n \geq 0$$

with $\phi_0 := 1$. This leads to

$$E\left[X_n \sigma\varepsilon_n\right] = E\left[X_n \sum_{k=0}^{p} \phi_k X_{n-k}\right] = \sum_{k=0}^{p} \phi_k c_k,$$

where $c_k = E\left[X_n X_{n-k}\right]$ is the autocovariance of X with time lag k. On the other hand we have

$$\begin{aligned} E\left[X_n \sigma\varepsilon_n\right] &= E\left[\left(\sigma\varepsilon_n - \sum_{k=1}^{p} \phi_k X_{n-k}\right)\sigma\varepsilon_n\right] \\ &= \sigma^2 E\left[\varepsilon_n^2\right] - \sigma \sum_{k=1}^{p} \phi_k E\left[X_{n-k}\varepsilon_n\right] \\ &= \sigma^2 E\left[\varepsilon_n^2\right] - \sigma \sum_{k=1}^{p} \phi_k E\left[X_{n-k}\right] E\left[\varepsilon_n\right] = \sigma^2, \end{aligned}$$

where we used the independence of ε_n and X_{n-k} for all $k \geq 1$. Following this line of thought yields for $l \geq 1$

$$0 = E\left[X_n \sigma\varepsilon_{n+l}\right] = E\left[X_n \sum_{k=0}^{p} \phi_k X_{n+l-k}\right] = \sum_{k=0}^{p} \phi_k c_{l-k}.$$

Writing these identities for $l = 1, \ldots, p$ in matrix form, we obtain the "*Yule-Walker equations*"

$$\begin{pmatrix} c_0 & c_1 & \ldots & c_p \\ c_1 & c_0 & \ldots & c_{p-1} \\ \vdots & \vdots & \ldots & \vdots \\ c_p & c_{p-1} & \ldots & c_0 \end{pmatrix} \begin{pmatrix} 1 \\ \phi_1 \\ \vdots \\ \phi_p \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{4.2}$$

Let Σ_p denote the $(p+1)\times(p+1)$ covariance matrix in (4.2). Note that Σ_p is a nonnegative definite Toeplitz matrix, since

$$\begin{aligned} v^\top \Sigma_p\, v &= \sum_{j=0}^{p}\sum_{k=0}^{p} v_j c_{j-k} v_k \\ &= \sum_{j=0}^{p}\sum_{k=0}^{p} v_j v_k \mathsf{cov}\,(X_{n-j}, X_{n-k}) \\ &= \mathsf{var}\left(\sum_{k=0}^{p} v_k X_{n-k}\right) \geq 0. \end{aligned}$$

As a consequence we have access to numerical schemes solving (4.2) in very cost effective way. Note that the last calculation does not require that X is an AR(p) processes. This motivates the question how well the Yule-Walker algorithm performs for arbitrary (weakly) stationary processes. To this end let $(X_n)_{n\geq 0}$ be a weakly stationary zero mean process with autocovariance sequence $\{c_k\}$. Identifying $X_{n-p},\ldots,X_{n-1}$ with time series data, we want to find the best linear predictor of the process X. Defining the predictor

$$\hat{X}_n = -\sum_{k=1}^{p} v_k X_{n-k}$$

and $v_0 = 1$, we obtain the mean square prediction error

$$\begin{aligned} E\left[\left(X_n - \hat{X}_n\right)^2\right] &= E\left[\left(\sum_{k=0}^{p} v_k X_{n-k}\right)^2\right] \\ &= E\left[\sum_{j=0}^{p}\sum_{k=0}^{p} v_j X_{n-j} X_{n-k} v_k\right] \\ &= E\left[\sum_{j=0}^{p}\sum_{k=0}^{p} v_j c_{j-k} v_k\right] = v^\top \Sigma_p v. \end{aligned}$$

In the following we show that the expression $v^\top\Sigma_p v$ is minimized for $v=\phi$, where ϕ is the solution of the Yule-Walker equation (4.2), and that the minimal value of the mean square prediction error equals σ^2. For an arbitrary v with $v_0 = 1$ we have

$$\begin{aligned} v^\top\Sigma_p v &= (\phi + (v-\phi))^\top \Sigma_p (\phi + (v-\phi)) \\ &= \phi^\top\Sigma_p\phi + 2(v-\phi)^\top\Sigma_p\phi + (v-\phi)^\top\Sigma_p(v-\phi). \end{aligned}$$

Due to (4.2) we get $\Sigma_p\phi = (\sigma^2, 0, \ldots, 0)^\top$ and hence the first term equals σ^2. For the second term we obtain

$$2(v-\phi)^\top\Sigma_p\phi = 2(v-\phi)^\top(\sigma^2,0,\ldots,0)^\top = 0$$

since the first entry $v_0 - \phi_0 = 1 - 1 = 0$ vanishes. This yields

$$v^\top\Sigma_p v = \sigma^2 + (v-\phi)^\top\Sigma_p(v-\phi).$$

Due to the definiteness of Σ_p the last term is nonnegative and equals zero for $v=\phi$. Obviously this leads to a minimal value of the mean square error equal to σ^2. This rather elementary fact is one of the main reasons for the success of the Yule-Walker method:

Remark 4.3. Effectiveness of the YW-Algorithm
By solving (4.2)*, which can be done in very cost efficient ways, we obtain the coefficients of the best linear predictor of X_n based on the past $X_{n-p}, \dots, X_{n-1}$ with a corresponding minimal mean square error of σ^2, even if X itself is no AR(p) process.*

However, we would like to mention that there are some cases where the YW method fails. The issue arises since – in practical applications – the parameter p is not given but has to be estimated, for instance, using Akaike's criterion [123]. For nearly periodic signals the combination of Akaike and Yule-Walker can lead to incorrect parameter estimates [124]. De Hoon et al. [124] suggest the use of the closely related Burg algorithm [125, 126] to avoid this issue. Fortunately MatLab has build-in functions for both methods, yielding nearly indistinguishable parameters for the present sets of data. Using the results of Rem. 4.2 and the MatLab function "*aryule*", i.e., equation (4.2), we obtain the following parameters for the fitted OUP (Table 3), which are visualized in Fig. 14.

Data	$\Theta \cdot 10^{-6}$	$D \cdot 10^{-6}$	# data points
Vostok	41.5	8.99	283
Byrd	312.7	6.98	171
TaylorDome	1079.4	2.23	69

Table 3: *Parameters for an OUP fitted on Ice Core Data*
Parameters of the fitted OUP and number of available data points for each time series data. Obviously the three corresponding stochastic processes differ strongly.

4.3.4 Numerical Aspects

The stochastic EBM is a two-dimensional system containing one RODE for T_t and one SODE for ΔT_t

$$\begin{aligned} c\,\mathrm{d}T_t &= \big[S_t(1-\alpha(T_t)) - k_B(T_t - \Delta T_t)^4\big]\mathrm{d}t \\ d\Delta T_t &= -\Theta\big(\Delta T_t - \mu\big)\mathrm{d}t + \sqrt{D}\,\mathrm{d}W_t. \end{aligned}$$

We already solved the equation defining the OUP analytically in Lemma 3.6, yielding

$$\Delta T_t = \Delta T_0 e^{-\Theta t} + \mu(1 - e^{-\Theta t}) + \sqrt{\frac{D}{2\Theta}} e^{-\Theta t} W(e^{2\Theta t} - 1).$$

This allows us to simulate the process ΔT_t in a direct and computational cost effective way by generating Gaussian random variables with variances corresponding to the exponential time scale $e^{2\Theta t}$. The remaining RODE can be solved by common deterministic numerical schemes, although they converge at a slower rate, since the process ΔT_t is only Hoelder continuous and not differentiable [55]. We used a fourth order Runge-Kutta scheme with a time step of one year.

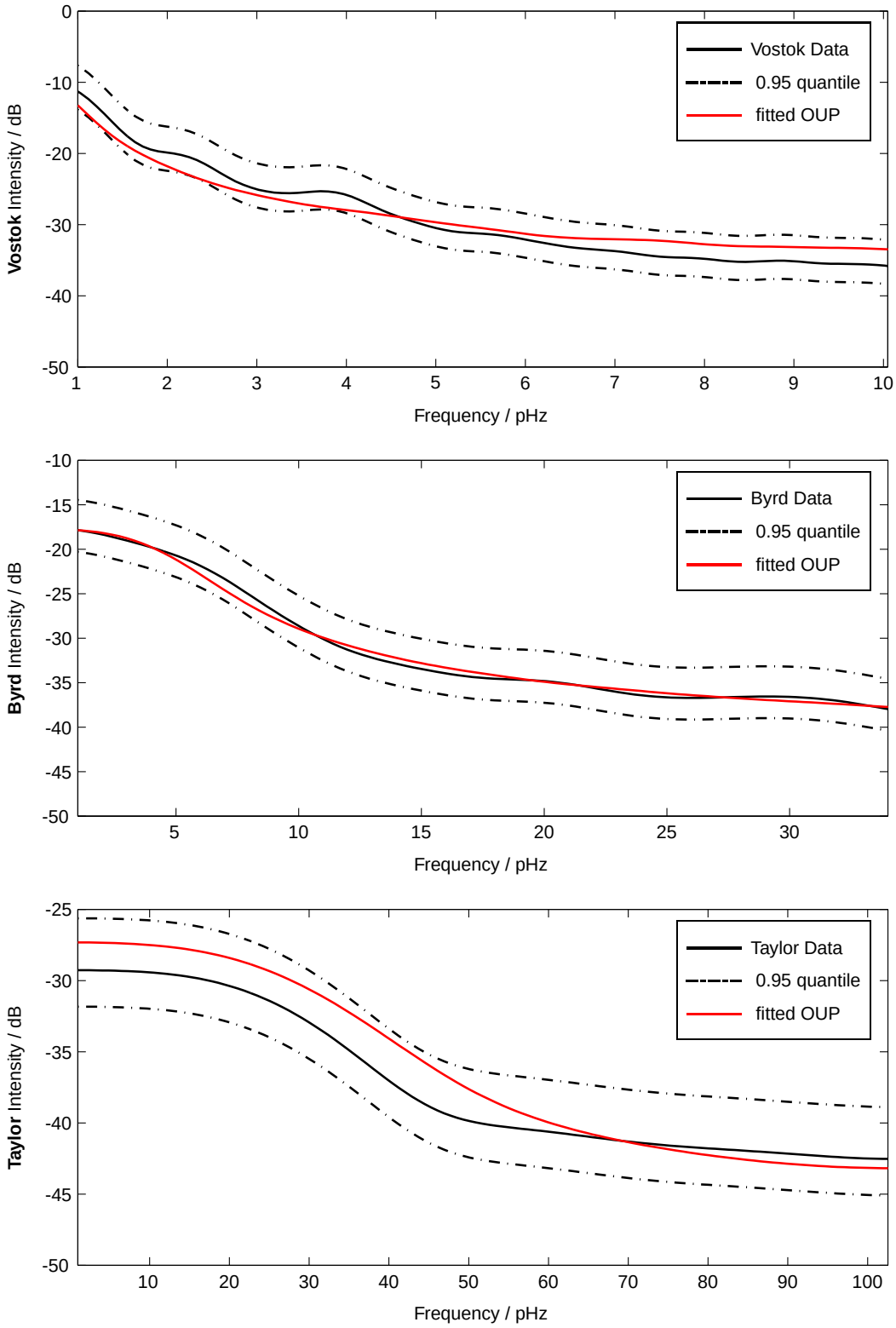


Figure 14: *Power Spectrum for OUP fitted on Ice Core Data*
Spectral plots of time series data and the corresponding fitted OUP using a Yule-Walker algorithm. Despite the small number of data points the derived OUP are within the 0.95 confidential interval.

4.4 Results

4.4.1 Sample Paths

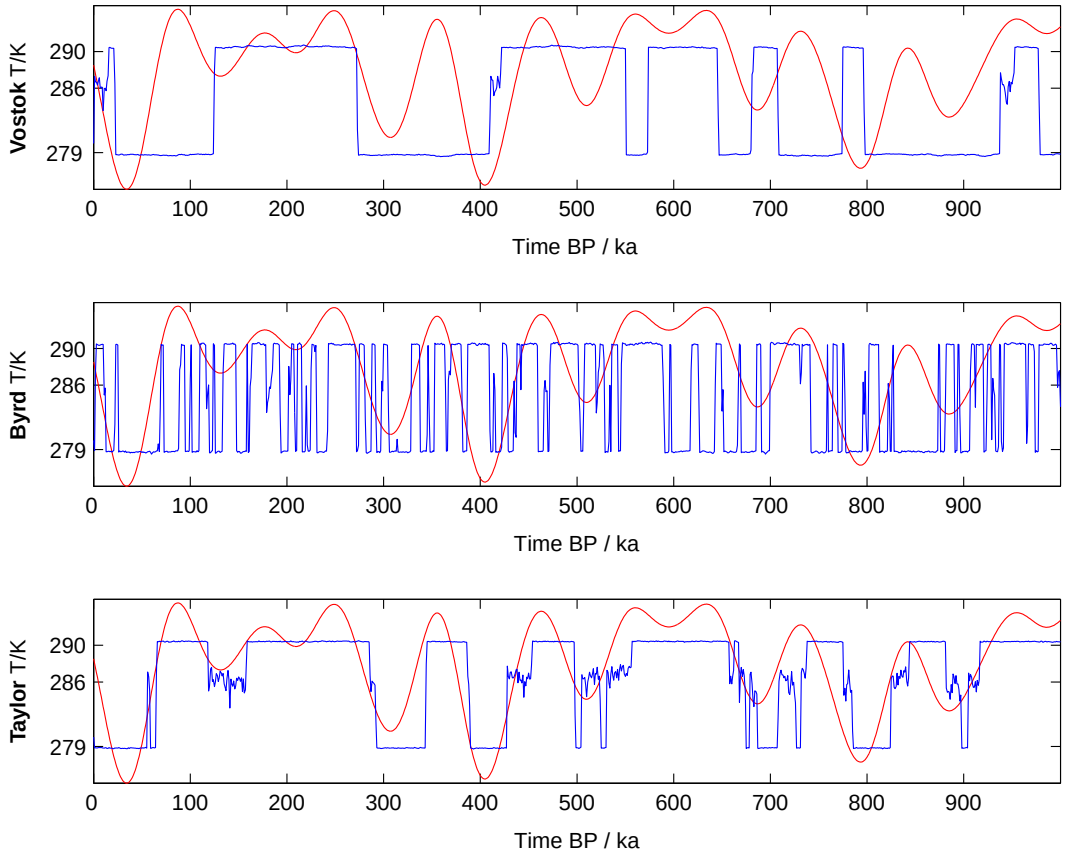


Figure 15: *Sample Paths for Three Different EBM*
Sample paths for temperature (blue) and insolation (red) of three models based on different ice core data. Since we already observed different fluctuations among the original data (Fig. 13), which resulted in inhomogeneous OUP parameters (Table 3), it comes as no surprise to find strongly differing sample paths.

The plots in Fig.15 show a single realization of the temperature process (blue line) and insolation (red line) for each model. We obtain a first intuitive impression of the models characteristics:

- There are at least two stable equilibria for the temperature, coinciding with the deterministic ones.
- The system can perform jumps between these equilibria.
- The models differ strongly regarding decorrelation times of T_t and strength of correlation between T_t and S_t.

The sample paths suggest a correlation between solar forcing and temperature depending on the strength of the SP. The coupling of solar and stochastic forcing with the temperature can be explained on a heuristic level, i.e., ignoring the systems inertia, by looking at schematic potential plots (Fig.16).

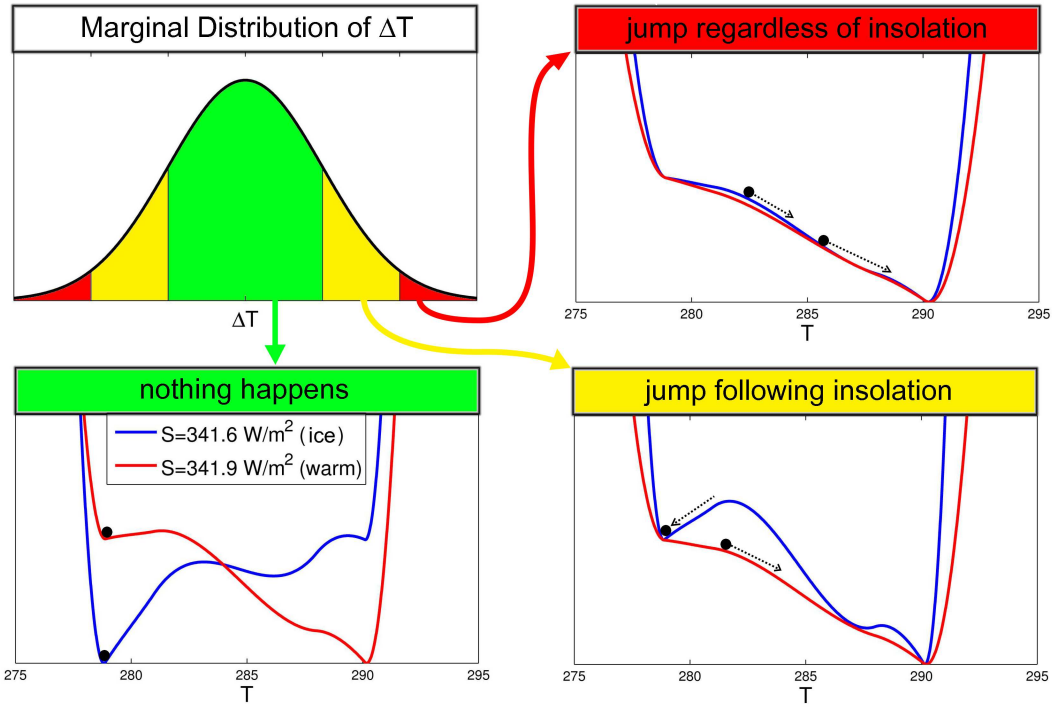


Figure 16: *Heuristic Potentials for a Stochastic EBM*
Heuristical visualization of the interaction between solar and stochastic forcing. Depending on the value of the stochastic forcing and the insolation, which is responsible for the different potential structures, the system stays in its current state, or jumps to another steady state.

Exemplary we consider positive fluctuations of ΔT_t where the system, which is illustrated by a black ball, occupies an ice-age. Positive fluctuations of insolation S_t and offset temperature ΔT_t cause the potential to tilt towards the warm stable state. Since the insolation variations are not capable of causing crossings between ice- and warm-ages on their own, we have to take a closer look at ΔT. Qualitatively the values of ΔT can be broken down into three categories:

- a area close to the mean value (green)
- a one-sided area further away from the mean, i.e., a “weak” extremum (yellow)
- a one-sided area far away from the mean, i.e., a “strong” extremum (red)

While ΔT occupies the first area (green) the system stays in its current state insusceptible to the fluctuations of the solar forcing, which corresponds to the deterministic dynamics. When ΔT adopts a weak extremum (yellow) the system follows the course of insolation, e.g. it jumps from an ice-age to a warm-age if and only if the insolation is strong at that point in time. In this case the stochastic forcing amplifies the solar forcing but is not strong enough to provoke state changes on its own. This changes when ΔT reaches a strong extremum (red): the system jumps regardless of the insolation level, e.g. it jumps from an ice- to a warm-age even if the insolation is minimal. In this case the stochastic forcing is strong enough to cause crossings between stable states even against the effect of the solar forcing. Since the OUP is a Gaussian process, the probability for extremes corresponds to $e^{-(\Delta T_t - \langle \Delta T_t \rangle)^2}$. Therefore a system with small fluctuations in ΔT_t, e.g. Taylor with var $(\Delta T_t) \approx 10^{-3} K^2$, has a much higher probability for jumps, which follow the insolation, than for jumps regardless of insolation. However, for a system with strongly varying ΔT, e.g. Byrd or Vostok with var $(\Delta T_t) \approx 10^{-2} K^2$, crossings of the second kind become more common. This behavior can be observed for the sample paths in Fig. 15.

Both Vostok and Byrd do not follow the insolation closely, yet they exhibit fundamentally different sample paths. We recall that both sets of time series data have almost the same

variance, but strongly differing decorrelation times (Table 2). The small decorrelation time of the Byrd system generates a rapidly fluctuating sample path, whereas the large decorrelation time yields a more ponderous Vostok model. Note that these effects are so dominant because the driving noise processes are very strong. Elsewise the insolation would play a bigger role and systems with large decorrelation times would "*miss*" some jumps at extremal insolation levels. Fig. 16 points out that the system is unaffected by the time a SP spends in the green area, i.e., it is close to its mean value. However, a short time span within the area of a weak extremum (yellow) at the time of an insolation extremum is suffice to initiate a jump to a new equilibrium. Naturally, for rapidly fluctuating SPs it is more likely to occupy a weak extremum at least once during a short time interval of extremal insolation. Following this line of thought we can consider the fast but weak fluctuations in the Taylor model as a "*scanner*" for jump opportunities. The combination of a small decorrelation time and weak variance leads to a system with high correlation between temperature and solar forcing. This heuristically derived idea is supported by statistical evidence in the upcoming sections 4.4.2 and 4.4.3.

4.4.2 Marginal Distributions at Local Extrema of the Solar Forcing

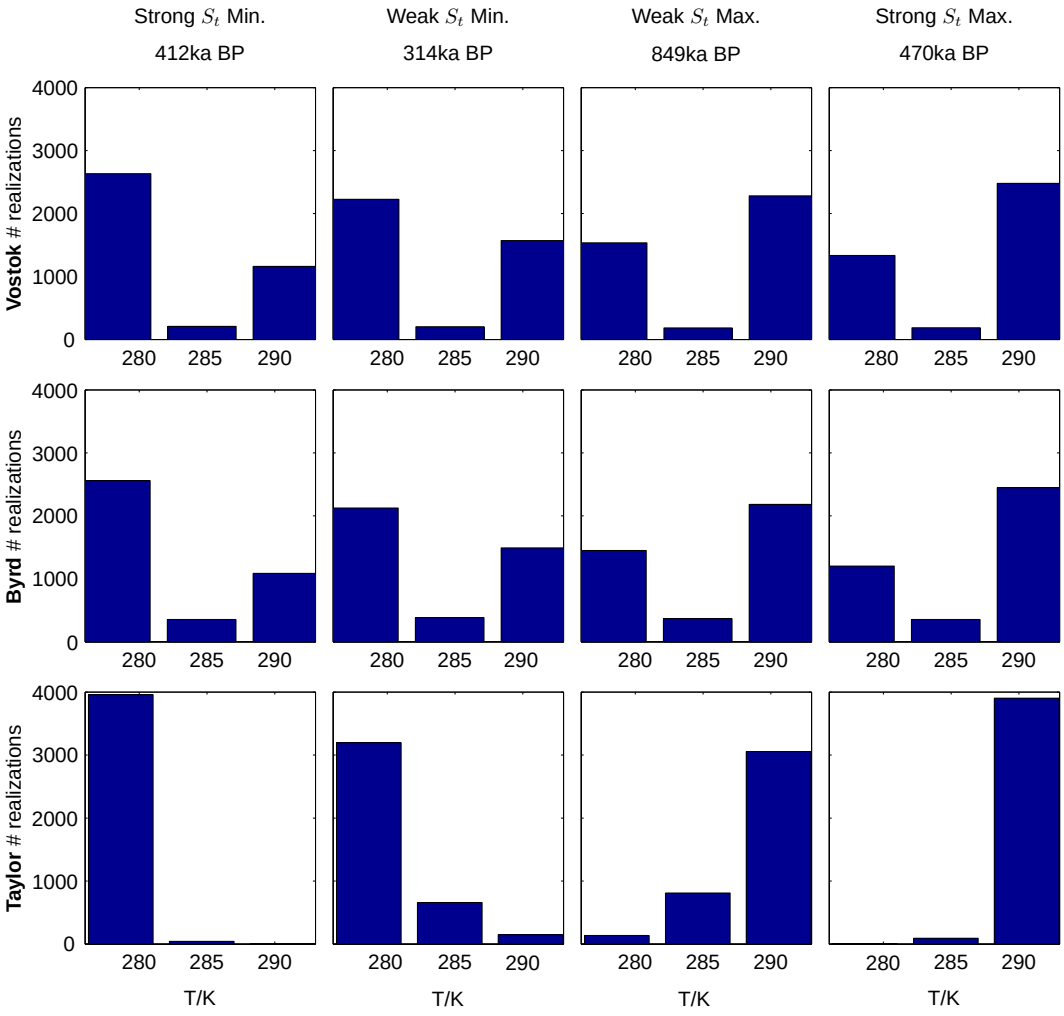


Figure 17: *Marginal Distributions for Three EBM*
Marginal distributions calculated from 4.000 realizations at local extrema of the solar forcing S_t. The plots support the impression, that the Taylor model closely follows the insolation, whereas Vostok and Byrd systems exhibit smaller solar influences on the temperature state.

Following the intuition gained by the heuristic consideration of sample paths and potentials, we take a first step to describe statistical characteristics of the models by analyzing marginal distributions of T_t at local extrema of the solar forcing S_t. We expect the Taylor model to follow the course of insolation and show one-sided distributions, whereas the Byrd and Vostok models should exhibit closer to uniform distributions. In Fig.17 the temperature for 4000 realizations of each model has been divided into three intervals corresponding to ice-ages and warm-ages and to a intermediate state.

4.4.3 Coherence

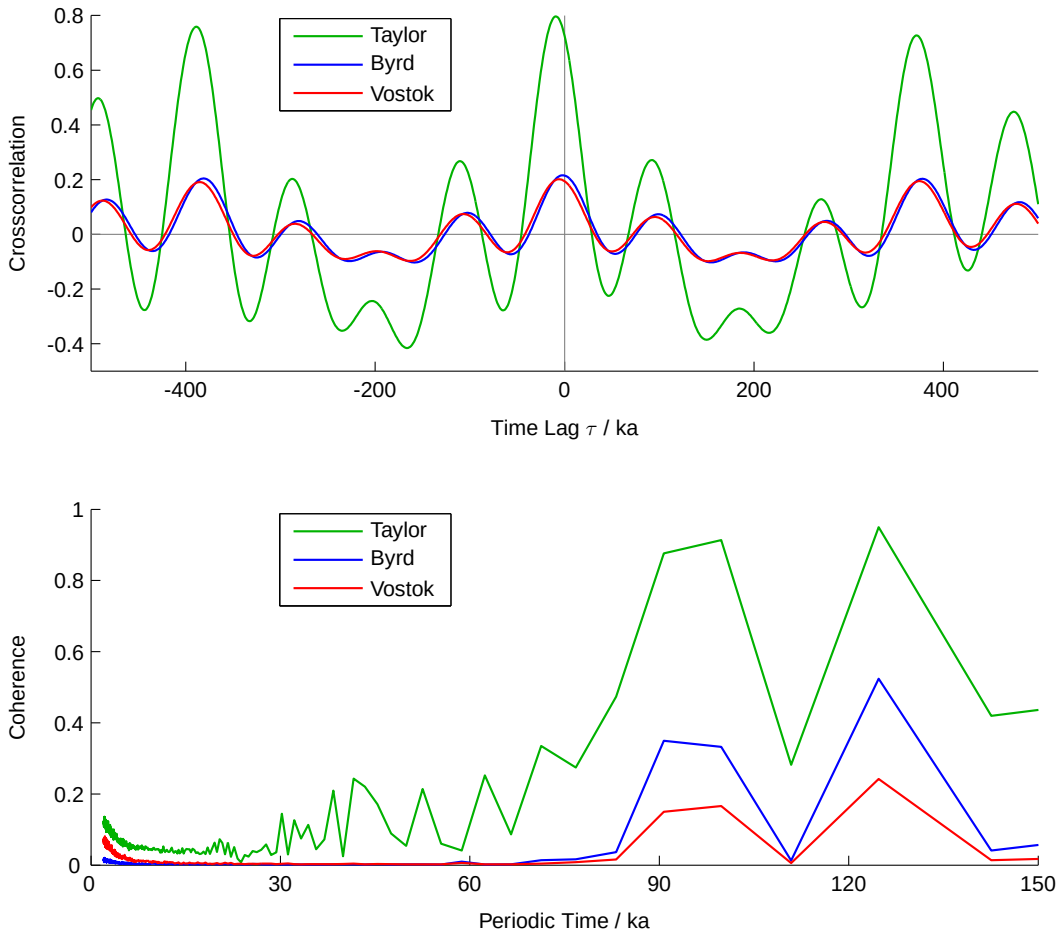


Figure 18: *Crosscorrelation and Coherence for Three EBM*
The upper plot shows the time lagged crosscorrelation between temperature and insolation. Since the crosscorrelation is maximal at $\tau = -8ka$, the insolation precedes the temperature. The deterministicity of the insolation signal then (heuristically) allows to consider the coherence as a measure for causality. As expected the Taylor system has hight crosscorrelation and coherence values than Byrd and Vostok. The maxima of the coherence correspond to the spectral peaks of the insolation, further supporting the causal connection between insolation and temperature in the Taylor model.

Obviously neither a single sample path nor marginal distributions at fixed points in time are sufficient to capture the characteristics of a stochastic dynamic system. Therefore we analyze the crosscorrelation and coherence between insolation and temperature based on the 4000 realizations of each model. The impression gained thus far are verified by the crosscorrelation (Fig.18), which is much higher in the Taylor system than it is in Byrd and Vostok systems. A maximal crosscorrelation of 0.8 at a time lag of 8ka and the peaks at time lags of about 400ka, which relates to the eccentric cycle (see [127]), suggests that the Taylor system is dominated by its insolation. The same periodic structure can be observed for the Byrd and Vostok model but with a much smaller amplitude. In these models a maximal crosscorrelation of about 0.2 indicates only slight influences of the insolation on the temperature but a dominance of the stochastic forcing.

The periodic crosscorrelation structure advises further analysis in the frequency domain. To this end we define the (quadratic) spectral coherence of two signals x, y by

$$C_{xy}^2 = \frac{|G_{xy}|^2}{G_{xx}G_{yy}},$$

where $G_{xy}(f) = E\left[X^*(f)Y(f)\right]$ is the cross-spectral density of the Fourier transformed signals X, Y. Note that for ergodic, linear systems the coherence is a measure for the causality between the input x and the output y. Although we are not dealing with a linear system, we can identify the insolation S_t as a deterministic input signal, which precedes the output signal T_t in the sense of the aforementioned time lag for maximal crosscorreletation. Therefore the causal interpretation of high coherence values is sensible but not rigorous. Denoting by $\mathcal{S}$ and $\mathcal{T}$ the Fourier transforms of insolation and temperature, yields

$$\begin{aligned} C_{\mathcal{ST}}^2 &= \frac{|E\left[\mathcal{S}^*\mathcal{T}\right]|^2}{E\left[|\mathcal{S}|^2\right]E\left[|\mathcal{T}|^2\right]} = \frac{|\mathcal{S}|^2|E\left[\mathcal{T}\right]|^2}{|\mathcal{S}|^2E\left[|\mathcal{T}|^2\right]} \\ &= \frac{|E\left[\mathcal{T}\right]|^2}{E\left[|\mathcal{T}|^2\right]} = \left(\frac{\mathrm{var}(\mathcal{T})}{|E\left[\mathcal{T}\right]|^2} + 1\right)^{-1}, \end{aligned}$$

for $E\left[\mathcal{T}\right] \neq 0$. Therefore the coherence between S and T does not depend on the insolation S. This counterintuitive result derives from the fact that S_t is a deterministic signal. As a consequence the coherence $C_{\mathcal{ST}} \in (0, 1]$ can be interpreted as a measure for the deterministicity of the temperature T: a system with no fluctuations has $\mathrm{var}(\mathcal{T}) = 0$ and therefore $C_T^2 = 1$. However, if the normalized variance is large, C_T is close to zero. The advantage of the Fourier analysis becomes obvious in Fig.18, where we can observe the impact of the stochastic forcing on each individual frequency. All models show peaks at periodic times of $95ka$ and $125ka$, which coincides with the leading terms in the spectrum of the eccentric cycle [127]. As expected the peaks of the Taylor model are close to one, whereas Byrd and Vostok models produce significant lower values. In summary, the results of the spectral analysis involving crosscorrelation and coherence supports the previously gathered insights: the Taylor model exhibits a strong (causal) connection between insolation and temperature. On the other hand the Byrd and Vostok systems are largely dominated by their strong stochastic forcing.

4.5 Conclusion

We derived three stochastic energy budget models based on three different sets of ice core data. Although the occurring SPs are based on physical data for each model, the emerging systems differ fundamentally. While the Taylor system yields the most realistic results, there is no solid ground to utterly dismiss the other data sets. The underlying problem is the discrepancy of spatial dimensions between data sets and model space. Ideally we would need data that allows the derivation of global mean fluctuations of ΔT rather than spatially pointwise ice core drillings. Another factor are feedback effects between CO_2 levels and temperature, which are included in the data but not accounted for in the model. A third obstacle is the large amount of data points one would need to determine the SPs with high accuracy and confidence. Of course we cannot expect such an abstract and simplified model to yield realistic results. We would like to emphasize that it is not the aim of this section to study the question whether or not glacial cycles can be explained by the eccentricity. In fact the diversity of the three systems shows that this basic model is inapt to produce any solid arguments on this topic. I further emphasizes the importance of choosing stochastic terms very carefully, even in the case of such a simple model. After all, if the three EBM models differ that strongly although their stochastics were derived from physical data, how can one expect realistic or even sensible results from models with arbitrarily chosen stochastic terms? In particular, this puts uncritically used additive white noise approaches into question, regardless of the models complexity.

In this section we presented a methodology to deduce physically based stochastic terms. To this end, starting from a deterministic framework, we identify a physical variable or subgrid process that is not accurately described by the model. Furthermore statistical information on this variable, e.g., sets of data or results from a high-resolution model, is necessary to derive the characteristics of a suitable stochastic formulation. The spectral fitting via Yule-Walker or Burg algorithms (see Section 4.3.3) onto *suitable* physical data then provides a powerful tool to implement physically meaningful stochastic formulations for unresolved or subgrid processes into a model.

5 Three-Dimensional Primitive Equations

The analysis of weather and climate is inextricably linked with the study of fluid dynamics, first and foremost regarding the atmosphere and ocean. At the core of this physical field are the Navier-Stokes equations (NSE), consisting of momentum, continuity and thermodynamic equations. They were formulated independently by Navier [128] and Stokes [129] in the first half of the 19th century based on the Euler equations governing inviscid flow [130], which were published in 1757. Despite their immense importance and mathematical elegance the theoretical understanding of these equations is still incomplete. In particular the existence of smooth solutions in space dimension three is still an open question and one of the "*Millennium Prize Problems*" of the Clay Mathematics Institute [131]. In two dimensions a positive answer was given by Ladythenskaya [132] in 1969 and there has been some success proving the existence of "*local*" solution in three dimensions, i.e., solutions that exist up to a finite (and possible small) "*blow up time*" $T < \infty$, which depends on the smallness of initial conditions.

The primitive equations (PE) are based on the so-called hydrostatic approximation (5.6), which simplifies the equation for the conservation of vertical momentum assuming equilibrium (in the vertical) between the pressure gradient force and the gravitational force. For many meteorological application this is a good approximation, since the vertical acceleration is very small compared to the gravitational acceleration [133]. While the PE were formulated in 1922 by Richardson [134], it took until 1992 for the first rigorous mathematical framework to be developed by Lions, Temam and Wang [135]. We closely follow this work during Section 5.1 and 5.2. In order to use recent results in a stochastic environment [7] however, we use an abstract framework based on infinite-dimensional Hilbert spaces and operators in Section 5.3, which follows the review paper by Temam and Ziane [136]. Although, from a physical point of view, the PE are of lower complexity then the Navier-Stokes equations, this behavior does not transfer to their mathematical treatment. Quite the contrary, the PE are technically even more evolved then the NSE: in Section 5.2.4 we point out that the vertical velocity can be considered as a function of the divergence of the horizontal velocity. This leads to a nonlinear advection term of the form

PE: (*first-order derivative of the velocity*) $\times$ (*first-order derivative of the velocity*)

whereas for the NSE the corresponding term has the form

NSE: (*velocity*) $\times$ (*first-order derivative of the velocity*).

It is precisely this additional derivative, that is investigated in the remainder of this work. Using Sobolev estimates it can be controlled to some extend via suitable L^p-norms (see Lemma 5.20). Since the exponents of these norms depend on the spatial dimension, solid results regarding existence and smoothness for space dimension two were soon available [137], whereas the three-dimensional case proved to be more defiant. Analogous with the NSE there were some results showing local existence of strong solution for the three-dimensional PE [136], but it was only in the last decade, that finally the global existence of smooth solution with physical boundary conditions in space dimension three was proved by Cao and Titi in 2007 [138].

Since the PE are the core of virtually every GCM, which are limited by their finite grid resolution, they are a subject of various stochastic parameterization approaches in order to account for unresolved subgrid processes or other sources of uncertainty. Consequently there has been a sharply growing interest among the meteorological and geophysical society [139, 140, 16, 141, 142, 11, 143, 144] to incorporate stochastic processes into the equations. However the mathematical rigorous analysis of the stochastic PE is lagging behind compared to the deterministic framework or even the results for stochastic NSE. Naturally there are solid results for space dimension two [145, 146, 147], and for the additive white noise case with mathematically idealized boundary conditions for three dimensions [148]. Unfortunately these results do not apply to physically reasonable cases. Recently Debussche et al. [7] proved the local existence for the three-dimensional PE

with physical boundary conditions and very general multiplicative noise using an abstract operator framework. While their work is concentrated on abstract fluid models, and therefore assumes incompressibility of the medium, we show that their results hold true for the global, compressible atmosphere.

5.1 General Equations of the Atmosphere

In this section we introduce the general equations of the atmosphere. Since we are interested in the equations of motion for the three-dimensional system, it is sensible to describe these equations in a non-inertial coordinate system rotating with the earth. In the Euclidean space the atmosphere is then characterized by the three-dimensional velocity field V_3, the pressure p, the density ρ and the temperature T.

Momentum equations

$$\begin{aligned}\frac{dV_3}{dt} &= \text{pressure gradient + gravity + Coriolis force +dissipative force} \\ &= -\frac{1}{\rho}\nabla_3 p + G - 2\Omega \times V_3 + D. \end{aligned} \tag{5.1}$$

Continuity equation

$$\frac{d\rho}{dt} + \rho \,\text{div}_3 V_3 = 0. \tag{5.2}$$

The first law of thermodynamics

$$c_p \frac{dT}{dt} - \frac{RT}{p}\frac{dp}{dt} = \frac{dQ}{dt}. \tag{5.3}$$

The equation of state

$$p = \rho R T, \tag{5.4}$$

where the viscosity term D in (5.1) is specified below. The heat flux per unit density in a unit time interval dQ/dt represents molecular, turbulent and solar heating, including the albedo of the earth, as well as heating due to evaporation. These equation are transformed into spherical coordinates, which are defined as follows.

Definition 5.1. Coordinate System
The spherical coordinate system consists of

$$\begin{aligned} \Theta &\in [0, \pi] && \text{(colatitude)} \\ \varphi &\in [0, 2\pi] && \text{(longitude)} \\ r &\in [0, \infty) && \text{(radial distance)} \end{aligned}$$

with unit vectors e_Θ, e_φ, e_z. Define furthermore the sea level radius a and the height above sea level $z = r - a$.

We obtain the following relations between euclidean and spherical coordinates

$$\vec{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r\cos\Theta\sin\varphi \\ r\sin\Theta\sin\varphi \\ r\cos\varphi \end{pmatrix},$$

which yields

$$
\begin{aligned}
e_\Theta &= \frac{\frac{\partial}{\partial\Theta}\vec{r}}{\left|\frac{\partial}{\partial\Theta}\vec{r}\right|} = \begin{pmatrix} -\sin\Theta \\ \cos\Theta \\ 0 \end{pmatrix} \\
e_\varphi &= \frac{\frac{\partial}{\partial\varphi}\vec{r}}{\left|\frac{\partial}{\partial\varphi}\vec{r}\right|} = \begin{pmatrix} \cos\Theta\cos\varphi \\ \sin\Theta\cos\varphi \\ -\sin\varphi \end{pmatrix} \\
e_z &= \frac{\frac{\partial}{\partial r}\vec{r}}{\left|\frac{\partial}{\partial r}\vec{r}\right|} = \begin{pmatrix} \cos\Theta\sin\varphi \\ \sin\Theta\sin\varphi \\ \cos\varphi \end{pmatrix}.
\end{aligned}
$$

Using this representation, the derivatives of the spherical unit vectors can be calculated. Then the velocity vector can be written as

$$V_3 = v_\Theta e_\Theta + v_\varphi e_\varphi + v_r e_z = r\dot{\Theta}e_\Theta + r\sin\Theta\ \dot{\varphi}e_\varphi + \dot{r}e_z$$

and we obtain

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \frac{v_\Theta}{r}\frac{\partial}{\partial\Theta} + \frac{v_\varphi}{r\sin\Theta}\frac{\partial}{\partial\varphi} + v_r\frac{\partial}{\partial r}.$$

Using $D := (D_\Theta, D_\varphi, D_r)$ finally allows us to translate the general equations (5.1)-(5.4) into the spherical coordinate system, yielding

$$
\begin{aligned}
&\frac{dv_\Theta}{dt} + \frac{1}{r}\left(v_r v_\Theta - v_\varphi^2\cot\Theta\right) = -\frac{1}{\rho r}\frac{\partial}{\partial\Theta}p + 2\Omega\cos\Theta v_\varphi + D_\Theta \\
&\frac{dv_\varphi}{dt} + \frac{1}{r}\left(v_r v_\varphi + v_\Theta v_\varphi\cot\Theta\right) = -\frac{1}{\rho r\sin\Theta}\frac{\partial}{\partial\varphi}p - 2\Omega\cos\Theta v_\Theta - 2\Omega\sin\Theta v_r + D_\varphi \\
&\frac{dv_r}{dt} + \frac{1}{r}\left(v_\Theta^2 - v_\varphi^2\right) = -\frac{1}{\rho}\frac{\partial}{\partial r}p - g + 2\Omega\sin\Theta v_\varphi + D_r \\
&\frac{d\rho}{dt} + \rho\left(\frac{1}{r\sin\Theta}\frac{\partial}{\partial\Theta}(v_\Theta\sin\Theta) + \frac{1}{r\sin\Theta}\frac{\partial}{\partial\varphi}v_\varphi + \frac{1}{r^2}\frac{\partial}{\partial r}(r^2 v_r)\right) = 0 \\
&c_p\frac{dT}{dt} - \frac{RT}{p}\frac{dp}{dt} = \frac{dQ}{dt} \\
&p = \rho RT.
\end{aligned}
$$

5.2 The Deterministic 3d Primitive Equations

In order to derive the primitive equations from the general equations of the atmosphere we identify terms of small magnitude via scale analysis, and approximate or omit them. The following scales, which are strongly based on observations and not so much on a theoretical deduction, are valid for midlatitude synoptic systems [133].

Variable	Scale	Unit	Description
v_Θ, v_φ	10	ms^{-1}	horizontal velocity
v_r	10^{-2}	ms^{-1}	vertical velocity
L	10^6	m	horizontal length
H	10^4	m	vertical hight
t	10^5	s	synoptic time
$\frac{\partial}{\partial\Theta}, \frac{\partial}{\partial\varphi}$	10^{-6}	m^{-1}	horizontal derivative
$\frac{\partial}{\partial r}$	10^{-4}	m^{-1}	vertical derivative
$\frac{\partial}{\partial t}$	10^{-5}	s^{-1}	time derivative
ρ	1	$kg\ m^{-3}$	density
p	10^5	Pa	pressure
$2\Omega\sin\Theta$	10^{-4}	s^{-1}	Coriolis term
g	10	ms^{-2}	gravity acceleration
r	10^7	m	earth radius

Table 4: Typical scales for the daily perturbations in midlatitude synoptic systems.

Note that the typical scales for the derivatives correspond to the inverse length and time scales. The most notably result is the "*hydrostatic approximation*", which reduces the vertical equation of motion

$$\underbrace{\frac{\partial}{\partial t}v_r}_{10^{-7}} + \underbrace{v_\Theta\frac{\partial}{\partial\Theta}v_r}_{10^{-7}} + \underbrace{v_\varphi\frac{\partial}{\partial\varphi}v_r}_{10^{-7}} + \underbrace{v_r\frac{\partial}{\partial r}v_r}_{10^{-8}} + \underbrace{\frac{1}{r}\left(v_\Theta^2 - v_\varphi^2\right)}_{10^{-5}} = \underbrace{-\frac{1}{\rho}\frac{\partial}{\partial r}p}_{10} - \underbrace{g}_{10} + \underbrace{2\Omega\sin\Theta v_\varphi}_{10^{-3}} + \underbrace{D_r}_{10^{-4}} \tag{5.5}$$

to the clearly arranged form

$$\frac{\partial}{\partial r}p = -\rho g. \tag{5.6}$$

This approximation exhibits a high accuracy, due to the gap of four order of magnitudes between the resolved and omitted scales. Supplementing this heuristic derivation of the hydrostatic approximation we refer to [149] for a mathematical justification, which is based "*on the fact that the ratio [ϵ^{-1}] between horizontal and vertical scales leads to very different sizes for the horizontal and vertical eddies*" [149, p. 847]. The authors show that, for small ϵ, weak solutions of the NSE converge to weak solutions of a system with hydrostatic pressure. Note that hydrostatic does not imply small vertical velocities but rather vertical accelerations, which are small compared to the gravitational acceleration.
Since the radius of the earth is large compared to the height of the atmosphere we further replace $r = a + z$, where z is the altitude above sea level, by the constant earth radius a in terms were r is not differentiated. This is known as "*shallow atmosphere approximation*" and, combined with scale analysis for the remaining equations analogous with (5.5), finally yields:

Definition 5.2. The PE of the Atmosphere in Spherical Coordinates

$$\begin{aligned}
&\frac{dv_\Theta}{dt} - \frac{v_\varphi^2}{a}\cot\Theta = -\frac{1}{\rho a}\frac{\partial}{\partial\Theta}p + 2\Omega\cos\Theta v_\varphi + D_\Theta \\
&\frac{dv_\varphi}{dt} + \frac{v_\Theta v_\varphi}{a}\cot\Theta = -\frac{1}{\rho a\sin\Theta}\frac{\partial}{\partial\varphi}p - 2\Omega\cos\Theta v_\Theta + D_\varphi \\
&\frac{\partial}{\partial r}p = -\rho g \\
&\frac{d\rho}{dt} + \rho\left(\frac{1}{a\sin\Theta}\frac{\partial}{\partial\Theta}(v_\Theta\sin\Theta) + \frac{1}{a\sin\Theta}\frac{\partial}{\partial\varphi}v_\varphi + \frac{\partial}{\partial r}v_r\right) = 0 \\
&c_p\frac{dT}{dt} - \frac{RT}{p}\frac{dp}{dt} = \frac{dQ}{dt} \\
&p = \rho RT.
\end{aligned}$$

Where the material differentiation is now given by

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \frac{v_\Theta}{a}\frac{\partial}{\partial\Theta} + \frac{v_\varphi}{a\sin\Theta}\frac{\partial}{\partial\varphi} + v_r\frac{\partial}{\partial r}.$$

Note that of the three-dimensional, geometric terms

$$\frac{1}{r}\left(v_r v_\Theta - v_\varphi^2\cot\Theta\right) \quad \text{and} \quad \frac{1}{r}\left(v_r v_\varphi + v_\Theta v_\varphi\cot\Theta\right),$$

only the horizontal contributions survived:

$$-\frac{v_\varphi^2}{a}\cot\Theta \quad \text{and} \quad \frac{v_\Theta v_\varphi}{a}\cot\Theta.$$

Therefore we effectively reduced the three-dimensional geometry of the atmosphere to the two-dimensional surface of the sphere with radius a.

5.2.1 Pressure Coordinates

The hydrostatic approximation (5.6) implies that the pressure p is decreasing monotonically with respect to the vertical coordinate z. This motivates the idea to use pressure as vertical coordinate and permits the following transformation

$$(t,\Theta,\varphi,z) \to (t^*,\Theta^*,\varphi^*,p = p\,(t,\Theta,\varphi,z))$$

with

$$t^* = t, \quad \Theta^* = \Theta, \quad \varphi^* = \varphi.$$

Since the inverse transformation is given by

$$(t,\Theta,\varphi,z) \to (t,\Theta,\varphi,z = z\,(t^*,\Theta^*,\varphi^*,p))$$

we have

$$p = p\,(t,\Theta,\varphi,z\,(t^*,\Theta^*,\varphi^*,p))\,. \tag{5.7}$$

While z was defined on $[0,\infty)$, the pressure p lives on the interval $[p_s,0]$, where p_s is the (unknown) surface pressure and 0 the pressure in the (infinitely) high atmosphere. Since there are technical difficulties for $p=0$, which are for instance studied in [150], we restrict p to the interval $[p_0,P]$ with an approximated surface pressure P and a small $p_0 > 0$.

Differentiating (5.7) with respect to p, Θ^*, φ^*, and using the hydrostatic approximation (5.6) leads to

$$\begin{aligned}
\frac{\partial}{\partial p}(gz) &= -\frac{1}{\rho} \\
\frac{\partial}{\partial \Theta^*}(gz) &= \frac{1}{\rho}\frac{\partial}{\partial \Theta}p \\
\frac{\partial}{\partial \varphi^*}(gz) &= \frac{1}{\rho}\frac{\partial}{\partial \varphi}p.
\end{aligned}$$

Furthermore we have

$$\begin{aligned}
\frac{\partial}{\partial z} &= \frac{\partial}{\partial z}p\frac{\partial}{\partial p} \\
\frac{\partial}{\partial \Theta} &= \frac{\partial}{\partial \Theta^*} + \frac{\partial}{\partial \Theta}p\frac{\partial}{\partial p} \\
\frac{\partial}{\partial \varphi} &= \frac{\partial}{\partial \varphi^*} + \frac{\partial}{\partial \varphi}p\frac{\partial}{\partial p}.
\end{aligned}$$

The material differentiation for a function F in the new coordinate system is given by

$$\begin{aligned}
\dot{F} = \frac{d}{dt^*}F &= \left(\frac{\partial}{\partial t^*} + \dot{\Theta}^*\frac{\partial}{\partial \Theta^*} + \dot{\varphi}^*\frac{\partial}{\partial \varphi^*} + \dot{p}\frac{\partial}{\partial p}\right)F \\
&= \left(\frac{\partial}{\partial t^*} + \dot{\Theta}\frac{\partial}{\partial \Theta} + \dot{\varphi}\frac{\partial}{\partial \varphi} + \dot{p}\frac{\partial}{\partial p}\right)F.
\end{aligned}$$

Using these expression we can deduce the horizontal (two-dimensional) gradient of a scalar function T and the horizontal (two-dimensional) divergence of a vector u, where the indices p and z denote the coordinate system of the nabla operator:

$$\begin{aligned}
\nabla_p T &= \nabla_z T + \frac{\partial}{\partial p}T\frac{\partial}{\partial z}p\nabla_p z \\
\nabla_p \cdot u &= \nabla_z \cdot u + \frac{\partial}{\partial p}u\frac{\partial}{\partial z}p \cdot \nabla_p z
\end{aligned}$$

Defining the geopotential $\Phi = gz$ then yields

$$-\frac{1}{\rho}\nabla_z p = -\nabla_p \Phi.$$

While the horizontal momentum equations of the PE take the form

$$\begin{aligned}
\frac{d}{dt^*}v_\Theta - \frac{v_\varphi^2}{a}\cot\Theta^* &= -\frac{1}{a}\frac{\partial}{\partial \Theta^*}\Phi + 2\Omega\cos\Theta^* v_\varphi + D_\Theta \\
\frac{d}{dt^*}v_\varphi + \frac{v_\Theta v_\varphi}{a}\cot\Theta^* &= -\frac{1}{a\sin\Theta^*}\frac{\partial}{\partial \varphi^*}\Phi - 2\Omega\cos\Theta^* v_\Theta + D_\varphi
\end{aligned}$$

we combine the hydrostatic approximation and the equation of state to obtain

$$\frac{\partial}{\partial p}\Phi + \frac{RT}{p} = 0.$$

Regarding the continuity equation we substitute $\rho = \frac{\partial}{\partial z}p$ and use the product rule to get

$$\begin{aligned}
\frac{d}{dt}\left(\frac{\partial}{\partial z}p\right) &= \frac{\partial}{\partial t}\left(\frac{\partial}{\partial z}p\right) + v_\Theta\frac{\partial}{\partial \Theta}\left(\frac{\partial}{\partial z}p\right) + \frac{v_\varphi}{a\sin\Theta}\frac{\partial}{\partial \varphi}\left(\frac{\partial}{\partial z}p\right) + v_r\frac{\partial}{\partial r}\left(\frac{\partial}{\partial z}p\right) \\
&= \frac{\partial}{\partial z}\left(\frac{dp}{dt}\right) - \frac{\partial}{\partial z}v_\Theta\frac{1}{a}\frac{\partial}{\partial \Theta}p - \frac{\partial}{\partial z}v_\varphi\frac{1}{a\sin\Theta}\frac{\partial}{\partial \varphi}p - \frac{\partial}{\partial z}v_r\frac{\partial}{\partial r}p.
\end{aligned}$$

Some basic calculations using the transformations of the various differential operators mentioned above lead to

$$\frac{\partial}{\partial p}\left(\frac{dp}{dt}\right) + \frac{1}{a\sin\Theta^*}\frac{\partial}{\partial\Theta^*}\left(v_\Theta \sin\Theta^* + \frac{\partial}{\partial\varphi^*}v_\varphi\right) = 0.$$

Note that the total derivative is independent of the coordinate system allowing us to define a (generalized) vertical velocity

$$\omega = \frac{d}{dt}p = \frac{d}{dt^*}p = \dot{p}.$$

Omitting the asterisk gives us

Lemma 5.3. PE of the Atmosphere in p-Coordinates
The PE in the coordinate system (t, Θ, φ, p) are given by

$$\begin{aligned}
&\frac{d}{dt}v_\Theta - \frac{v_\varphi^2}{a}\cot\Theta = -\frac{1}{a}\frac{\partial}{\partial\Theta}\Phi + 2\Omega\cos\Theta v_\varphi + D_\Theta\\
&\frac{d}{dt}v_\varphi + \frac{v_\Theta v_\varphi}{a}\cot\Theta = -\frac{1}{a\sin\Theta}\frac{\partial}{\partial\varphi}\Phi - 2\Omega\cos\Theta v_\Theta + D_\varphi\\
&\frac{\partial}{\partial p}\Phi + \frac{RT}{p} = 0\\
&\frac{\partial}{\partial p}\omega + \frac{1}{a\sin\Theta}\frac{\partial}{\partial\Theta}\left(v_\Theta\sin\Theta + \frac{\partial}{\partial\varphi}v_\varphi\right) = 0\\
&c_p\frac{dT}{dt} - \frac{RT}{p}\omega = \frac{dQ}{dt},
\end{aligned}$$

where

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \frac{v_\Theta}{a}\frac{\partial}{\partial\Theta} + \frac{v_\varphi}{a\sin\Theta}\frac{\partial}{\partial\varphi} + \omega\frac{\partial}{\partial p}$$

and

$$v_r = \frac{dr}{dt} = \frac{dz}{dt} = \frac{\partial}{\partial t}z + \frac{v_\Theta}{a}\frac{\partial}{\partial\Theta}z + \frac{v_\varphi}{a\sin\Theta}\frac{\partial}{\partial\varphi}z - \frac{\omega}{\rho g}.$$

Remark 5.4. Quasi-Incompressibility
One of the most notable advantages of the p-coordinate PE is their "quasi- incompressibility": although the atmosphere is evidently a highly compressible fluid, the continuity equation in p-coordinates takes the same form as that of an incompressible fluid. This proves to be a crucial condition for the existence of solutions for stochastic PE.

5.2.2 Approximations involving Temperature and Diffusion

In order to make the viscous PE more accessible from a mathematical point of view we give a simplification of the term $RT\omega/p$ and characterize diffusion and heating terms $D_\Theta, D_\varphi, dQ/dt$, which remained unspecified so far. As we are interested in the large scale behavior of the atmosphere we assume that the surface of the earth is isobaric with a surface pressure of P. The upper boundary of the atmosphere is given by the isobaric surface $p = p_0 > 0$. Let us define vertical distributions $\bar{T}, \bar{\Phi}$ of temperature and geopotential in order to derive equations for the difference processes

$$T' := T - \bar{T} \quad \text{and} \quad \Phi' := \Phi - \bar{\Phi}.$$

This decomposition allows us to identify terms of small magnitude. To this end define a vertical temperature distribution $\bar{T} = \bar{T}(p) \in C^\infty([p_0, P])$ such that

$$R\left(\frac{R\bar{T}}{c_p} - p\frac{\partial}{\partial p}\bar{T}\right) = \text{constant} =: C^2, \tag{5.8}$$

which can be understood as the climate mean temperature on isobaric surfaces. As the next step we define a vertical geopotential distribution $\bar{\Phi} = \bar{\Phi}(p)$ by

$$\frac{R\bar{T}}{p} - p\frac{\partial}{\partial p}\bar{\Phi} = 0. \tag{5.9}$$

Furthermore the difference processes

$$T' = T - \bar{T} \quad \text{and} \quad \Phi' = \Phi - \bar{\Phi}$$

satisfy

$$\begin{aligned}
&\frac{\partial}{\partial\Theta}\Phi = \frac{\partial}{\partial\Theta}\Phi' \\
&\frac{\partial}{\partial\varphi}\Phi = \frac{\partial}{\partial\varphi}\Phi' \\
&\frac{\partial}{\partial p}\Phi + \frac{RT}{p} = \frac{\partial}{\partial p}(\Phi' + \bar{\Phi}) + \frac{R}{p}(T' + \bar{T}) = \frac{\partial}{\partial p}\Phi' + \frac{RT'}{p}.
\end{aligned}$$

These facts allow us to rewrite the PE with exception of the first law of thermodynamics, which is treated using

$$\frac{\partial}{\partial t}\bar{T} = \frac{\partial}{\partial\Theta}\bar{T} = \frac{\partial}{\partial\varphi}\bar{T} = 0.$$

This implies

$$\frac{d\bar{T}}{dt} = \omega\frac{\partial}{\partial p}\bar{T}$$

and therefore by definition of $\bar{T}$

$$\frac{d\bar{T}}{dt} = \frac{R\bar{T}}{c_p p}\omega - \frac{C^2}{pR}\omega.$$

For the thermodynamic part of the PE we obtain

$$\frac{dQ}{dt} = c_p\frac{dT}{dt} - \frac{RT}{p}\omega = c_p\frac{dT'}{dt} + c_p\frac{d\bar{T}}{dt} - \frac{RT}{p}\omega = c_p\frac{dT'}{dt} - \frac{c_p C^2}{pR}\omega + (\bar{T} - T)\frac{R\omega}{p}$$

In order to simplify the PE we omit the last term. This is in fact the first approximation to the PE we made since the transformation to pressure coordinates. To give a (brief) justification for this approach note that $|T - \bar{T}|$ is small compared to $p/R\omega$ [151]. This leads to

$$|T - \bar{T}|\frac{R\omega}{p} = |T - \bar{T}|\left(\frac{p}{R\omega}\right)^{-1} << 1.$$

Variations of this approximation enjoy great popularity among the mathematical papers on the PE. It is for instance used in the review paper on PE by Temam and Ziane [136], where the authors refer to [150] for a more elaborate approximation. Furthermore they state that the occurring mathematical problems are of technical nature so that "*with addition precautions, and using the maximum principle for the temperature as in [152], we could keep the exact term* $RT\omega/c_p p$" [136, p. 565]. Since our focus lies in a stochastic treatment of the PE, we accept

$$(\bar{T} - T)\frac{R\omega}{p} \approx 0$$

as a good first order approximation to avoid additional (albeit technical) complications. This gives us a modified version of the PE

$$\begin{aligned}
&\frac{d}{dt}v_\Theta - \frac{v_\varphi^2}{a}\cot\Theta = -\frac{1}{a}\frac{\partial}{\partial\Theta}\Phi' + 2\Omega\cos\Theta v_\varphi + D_\Theta\\
&\frac{d}{dt}v_\varphi + \frac{v_\Theta v_\varphi}{a}\cot\Theta = -\frac{1}{a\sin\Theta}\frac{\partial}{\partial\varphi}\Phi' - 2\Omega\cos\Theta v_\Theta + D_\varphi\\
&\frac{\partial}{\partial p}\Phi' + \frac{RT'}{p} = 0\\
&\frac{\partial}{\partial p}\omega + \frac{1}{a\sin\Theta}\frac{\partial}{\partial\Theta}\Big(v_\Theta\sin\Theta + \frac{\partial}{\partial\varphi}v_\varphi\Big) = 0\\
&c_p\frac{dT'}{dt} - \frac{c_pC^2}{pR}\omega = \frac{dQ}{dt},
\end{aligned}$$

which leaves the diffusion and heating terms to be specified. To allow for a more concise representation of the PE we introduce the following differential operators on the two-dimensional sphere S_a^2. The covariant derivative of a vector field $u = u_\Theta e_\Theta + u_\varphi e_\varphi$ and a scalar function T respectively, w.r.t. $v = v_\Theta e_\Theta + v_\varphi e_\varphi$, is given by

$$\begin{aligned}
\nabla_v u = &\Big(\frac{v_\Theta}{a}\frac{\partial}{\partial\Theta}u_\Theta + \frac{v_\varphi}{a\sin\Theta}\frac{\partial}{\partial\varphi}u_\Theta - \frac{v_\varphi u_\varphi}{a}\cot\Theta\Big)e_\Theta\\
&+\Big(\frac{v_\Theta}{a}\frac{\partial}{\partial\Theta}u_\varphi + \frac{v_\varphi}{a\sin\Theta}\frac{\partial}{\partial\varphi}u_\varphi - \frac{v_\varphi u_\Theta}{a}\cot\Theta\Big)e_\varphi\\
\nabla_v T = &\frac{v_\Theta}{a}\frac{\partial}{\partial\Theta}T + \frac{v_\varphi}{a\sin\Theta}\frac{\partial}{\partial\varphi}T.
\end{aligned}$$

The horizontal gradient of a scalar function T and the horizontal divergence of a vector field $u = u_\Theta e_\Theta + u_\varphi e_\varphi$ have the form

$$\begin{aligned}
\nabla T &= \frac{1}{a}\frac{\partial}{\partial\Theta}T\,e_\Theta + \frac{1}{a\sin\Theta}\frac{\partial}{\partial\varphi}T\,e_\varphi\\
\nabla\cdot u &= \frac{1}{a\sin\Theta}\Big(\frac{\partial}{\partial\Theta}(u_\Theta\sin\Theta) + \frac{\partial}{\partial\varphi}u_\varphi\Big).
\end{aligned}$$

This leaves the horizontal Laplace-Beltrami operators acting on a scalar function T and a vector field $u = u_\Theta e_\Theta + u_\varphi e_\varphi$, respectively. They are defined by

$$\begin{aligned}
\Delta T &= \nabla\cdot(\nabla\,T) = \frac{1}{a^2\sin\Theta}\left[\frac{\partial}{\partial\Theta}\Big(\sin\Theta\frac{\partial}{\partial\Theta}T\Big) + \frac{1}{\sin\Theta}\frac{\partial^2}{\partial\varphi^2}T\right]\\
\Delta u &= \Delta\left(\frac{u_\Theta}{a}\frac{\partial}{\partial\Theta} + \frac{u_\varphi}{a\sin\Theta}\frac{\partial}{\partial\varphi}\right)\\
&= \Big(\Delta u_\Theta - \frac{2\cos\Theta}{a^2\sin^2\Theta}\frac{\partial}{\partial\varphi}u_\varphi - \frac{u_\Theta}{a^2\sin^2\Theta}\Big)e_\Theta\\
&\quad + \Big(\Delta u_\varphi - \frac{2\cos\Theta}{a^2\sin^2\Theta}\frac{\partial}{\partial\varphi}u_\Theta - \frac{u_\varphi}{a^2\sin^2\Theta}\Big)e_\varphi.
\end{aligned}$$

We can now introduce the following viscosity terms D_v, D_T for $v = (v_\Theta, v_\varphi)$

$$\begin{aligned}
D_v &= (D_\Theta, D_\varphi) = \mu_v\Delta v + \nu_v\frac{\partial}{\partial p}\left[\Big(\frac{gp}{R\bar{T}}\Big)^2\frac{\partial}{\partial p}v\right]\\
D_T &= \mu_T\Delta T' + \nu_T\frac{\partial}{\partial p}\left[\Big(\frac{gp}{R\bar{T}}\Big)^2\frac{\partial}{\partial p}T'\right],
\end{aligned}$$

which are extensively discussed in [153, 154]. Note that T was approximated by $\bar{T}$, since the coefficients ν_v, ν_T are small. The heating term is specified by

$$\frac{1}{c_p}\frac{dQ}{dt} = D_T + F_T,$$

where F_T denotes the diabatic heating, consisting of radiative heating, which includes solar heating and the albedo effect, as well as evaporative heating. We can now state

Lemma 5.5. The Viscous PE
The PE of the atmosphere with horizontal viscosity and heating terms in the coordinate system (t, Θ, φ, p) *are given by*

$$\begin{aligned}
&\frac{\partial}{\partial t}v + \nabla_v v + \omega\frac{\partial}{\partial p}v + 2\Omega\cos\Theta\ k \times v + \nabla\Phi' - D_v = F_v \\
&\frac{\partial}{\partial p}\Phi' + \frac{RT'}{p} = 0 \\
&\nabla \cdot v + \frac{\partial}{\partial p}\omega = 0 \\
&\frac{\partial}{\partial t}T' + \nabla_v T' + \omega\frac{\partial}{\partial p}T' - \frac{C^2}{pR}\omega - D_T = F_T,
\end{aligned}$$

where F_T *stands for diabatic heating, while the term* F_v*, which vanishes in reality, was added to allow for a greater mathematical generality. The diffusion terms are defined by*

$$\begin{aligned}
D_v &= \mu_v\Delta v + \nu_v\frac{\partial}{\partial p}\left[\left(\frac{gp}{R\bar{T}}\right)^2\frac{\partial}{\partial p}v\right] \\
D_T &= \mu_T\Delta T' + \nu_T\frac{\partial}{\partial p}\left[\left(\frac{gp}{R\bar{T}}\right)^2\frac{\partial}{\partial p}T'\right].
\end{aligned}$$

Furthermore $T' = T - \bar{T}$ *and* $\Phi' = \Phi - \bar{\Phi}$ *are the difference processes characterized in* (5.8) *and* (5.9)*, respectively.*

Remark 5.6. A Note on Vertical Viscosity
The vertical diffusion term D_r *vanished due to the hydrostatic approximation. As a result of the coordinate transformations, that were employed to deal with the compressibility of the atmosphere, it is not obvious how one could introduce effects of vertical viscosity into the PE. This setting is extensively studied [135, Ch. 4-7] using the ansatz of a "weaker" hydrostatic approximation*

$$\frac{\partial}{\partial z}p = -\rho g + \rho D_r,$$

where D_r *is small. This is used to make the transformation to pressure coordinates omitting any arising terms of small order of magnitude. While we won't pursue this topic any further in this thesis it might be interesting for future extensions.*

5.2.3 Boundary Conditions

The horizontal boundary conditions Γ_h are obviously given by periodicity due to the spherical geometry of our setting. Regarding the top of the atmosphere Γ_u we assume

$$\frac{\partial}{\partial p}v = 0, \quad \omega = 0, \quad \frac{\partial}{\partial p}T' = 0. \tag{5.10}$$

The lower boundary however is more complex, since it contains the interaction between the surface of the earth and its atmosphere. Since the surface is distinctively inhomogeneous there is a multitude of possible formulation to describe these interaction. A widely accepted characterization in meteorological applications is as follows [155, 154, 136].

Above land surface Γ_l

$$\begin{aligned}
&v = 0, \quad \omega = 0 \\
&\nu_T\frac{\partial}{\partial p}T = -\alpha_T(T - T_l)
\end{aligned} \tag{5.11}$$

Above oceanic surface Γ_o

$$\begin{aligned} \nu_v \left(\frac{gp}{R\bar{T}}\right)^2 \frac{\partial}{\partial p} v &= \tau_v - \alpha_v(v - v_o), \quad \omega = 0 \\ \nu_T \left(\frac{gp}{R\bar{T}}\right)^2 \frac{\partial}{\partial p} T' &= -\alpha_T(T' - T_o). \end{aligned} \tag{5.12}$$

We assume that the temperature on land surface T_l, and the velocity and temperature of the ocean v_o and T_o are given $C^\infty(S_a^2)$ functions. Furthermore, for the sake of simplicity we assume that τ_v and the coefficients $\alpha_v > 0$ and $\alpha_T > 0$ are known constants and refer to Lions et al., who state that "the non-constant [case] can be treated in exactly the same way" [135, p. 250]. Although there exist a multitude of viable formulations for the boundary conditions, the results of this section hold true for any alternative formulation based on first order linear differential equations of horizontal velocity and temperature, since we have to homogenize the boundary conditions for symmetry reasons later on, see Lemma 5.11.

5.2.4 Prognostic and Diagnostic Variables

In order to take a closer look at the generalized vertical velocity we integrate the continuity equation in Lemma 5.5

$$\nabla \cdot v + \frac{\partial}{\partial p}\omega = 0.$$

The boundary conditions on ω then yield

$$\omega = \omega(v) = -\int_{p_0}^{p} \nabla \cdot v \, dp' \tag{5.13}$$

with

$$\int_{p_0}^{P} \nabla \cdot v \, dp = 0.$$

Therefore ω is a function completely defined by the horizontal velocity v. The geopotential is obtained in an analog way based on the second equation of the PE Lemma 5.5

$$\Phi' = \Phi'(p, T') = \Phi'_s + \int_p^P \frac{RT'}{p'} dp', \tag{5.14}$$

where Φ_s is the geopotential at the surface $p = P$, which equals g times the height of the isobar $p = P$. We can now group the unknown functions into two sets: the prognostic variables $U = (v, T)$ for which we define the initial value problem, and the diagnostic variables ω and Φ, which can be determined, at each instant in time by evaluating the prognostic variables using the PE and the boundary conditions. This reduction of (prognostic) variables is a distinct advantage of the PE compared to the Navier-Stokes equations. However, we encounter the aforementioned disadvantage of this technique when trying to control the advection term $(U \cdot \nabla)U$, where we have to deal with additional horizontal derivatives on one hand and vertical averaging (integration) on the other, see Lemma 5.20. This leads to the necessity of so-called "*anisotrop Sobolev estimates*" in the proof of Lemma 5.20.

These considerations together with the boundary conditions (5.10), (5.11), (5.12) and the viscous PE (Lemma 5.5) conclude the classical deterministic framework. The next step is the derivation of an abstract framework using infinite-dimensional Hilbert spaces and operators, which finally allows us to use the results of Debussche et al [7] to conclude the existence of solutions of the PE in a stochastic framework.

5.3 An Abstract Operator Framework for the PE

Before diving straight into the mathematical construction of an operator framework, we motivate this approach by means of an illustrative example, showcasing the underlying idea of the upcoming proof of existence.

5.3.1 Variational Formulation of the Stokes-Equation - A Motivation

The Stokes equations are the stationary linearized form of the full Navier-Stokes equations. They are therefore far less complex than the PE but exhibit a similar fundamental structure. We refer to [156] for rigorous proofs of the following conceptual statements. Let $\mathcal{M}$ be an open bounded set in $\mathbb{R}^n$ with boundary $\partial\mathcal{M}$, let $f \in L^2(\mathcal{M})$ be a given vector function in $\mathcal{M}$ and let $\nu > 0$ be the constant viscosity coefficient. We can then state

Proposition 5.7. The Stokes Problem
Find a vector function u (the velocity) and a scalar function p (the pressure) defined in $\mathcal{M}$, which satisfy

$$\begin{aligned} -\nu\Delta u + \nabla p &= f \quad \textit{in} \quad \mathcal{M} \\ \nabla \cdot u &= 0 \quad \textit{in} \quad \mathcal{M} \\ u &= 0 \quad \textit{on} \quad \partial\mathcal{M}. \end{aligned}$$

We introduce the concept of Hilbert spaces, which can be regarded as a natural generalization of the Euclidean space for systems with any finite or infinite number of dimensions.

Definition 5.8. Hilbert Space
A Hilbert space H is a vector space endowed with an inner product $\langle\cdot,\cdot\rangle$ such that the norm $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$ turns H into a complete metric space, i.e., a metric space in which every Cauchy sequence is convergent.

Let us define the Hilbert space

$$V = \left\{ u \in H^1(\mathcal{M}) \;\middle|\; \nabla \cdot u = 0 \right\}$$

endowed with the inner product

$$((u,v)) = \sum_{i=1}^{n} \left(\nabla u_i, \nabla v_i\right), \quad u, v \in V.$$

Note that $H^1(\mathcal{M})$ can be understood as a space of suitable smooth functions on $\mathcal{M}$ for this conceptual argument and are rigorously defined in the next section. The crucial point is that V consists only of divergence free vector fields. Taking the scalar product of the first equation with a $v \in V$ yields

$$-\left(\nu\Delta u, v\right) + \left(\nabla p, v\right) = \left(f, v\right)$$

and, integrating by parts, we obtain

$$\begin{aligned} -\nu\left(\Delta u, v\right) &= \nu \sum_{i=1}^{n} \left(\nabla u_i, \nabla v_i\right) = \nu\left((u,v)\right) \\ \left(\nabla p, v\right) &= \left(p, \nabla \cdot v\right) = 0 \quad \text{by definition of } V. \end{aligned}$$

For this calculation we used the characteristics of V as well as the boundary conditions on $\partial\mathcal{M}$. We can now state the

Proposition 5.9. Variational formulation of the Stokes problem
Find a $u \in V$ satisfying

$$\nu\left((u,v)\right) = \left(f, v\right) \quad \textit{for all } v \in V.$$

Following up on the idea presented here by using elaborate topological arguments one can show (see Lemma 2.1 in [156]) that the variational formulation is equivalent to the Stokes problem in a weak sense. Therefore we have reduced the classical problem (Prop. 5.7) to the task of finding only u, while the existence of p follows from the fact

$$(w, v) = 0 \quad \text{for all } v \in V \quad \Rightarrow \quad \exists p \quad \text{with } w = \nabla p$$

for $w = \nu\Delta u - f$ (see Prop.1.1 and Prop.1.2. in [156]). The existence of an u satisfying Proposition 5.9 can be approached by the abstract arguments of various fix-point theorems. In the case of a finite-dimensional Euclidean space Brouwers Theorem [157] states that every continuous function f from the closed n-dimensional Ball B^n into itself has a fix point, i.e.

$$\forall f \in C(B^n, B^n) : \exists x \in B^n : f(x) = x.$$

The extension to more general sets $\mathcal{M}$ in Banach spaces was given by Schauder and Leray [158] and is also referred to as the "*Leray-Schauder principle*". Naturally the main issue in the application of these statements is the implied boundedness of the function f. This is particularly true for the PE due to their highly nonlinear advection term.

5.3.2 The Mathematical Framework

During the construction of an abstract framework for the PE we follow the notation of [136], starting with the definition of the domains

$$\mathcal{M} = \mathcal{M}_0 \times (p_0, P) = \left\{(\Theta, \varphi, p) \;\middle|\; \Theta \in (0, \pi), \varphi \in (0, 2\pi), p \in (p_0, P)\right\}.$$

The boundary of $\mathcal{M}$ is then characterized using the definitions of the previous section

$$\begin{aligned} \Gamma_h &= \partial\mathcal{M}_0 \times (p_0, P) && \text{(horizontal)} \\ \Gamma_u &= \mathcal{M}_0 \times \{p_0\} && \text{(top)} \\ \Gamma_l \cup \Gamma_o &= \mathcal{M}_0 \times \{P\}. && \text{(surface)} \end{aligned}$$

Let $L^p(\mathcal{M})$ denote the space of p-integrable functions on $\mathcal{M}$ for all $1 \leq p < \infty$. Define the Sobolev spaces $W^{k,p}(\mathcal{M})$ for $k \in \mathbb{N}$ and $1 \leq p < \infty$ by

$$W^{k,p}(\mathcal{M}) = \left\{u \in L^p(\mathcal{M}) \middle|\; D^\alpha u \in L^p(\mathcal{M}) \text{ for all } |\alpha| \leq k\right\}.$$

Here α is a multi-index and $D^\alpha u$ denotes the weak partial derivative of u. Therefore $W^{k,p}$ is the space of all L^p-functions on $\mathcal{M}$ whose weak partial derivatives of up to order k are p-integrable as well. The following notation is motivated by the fact that $W^{k,p}$ is a Hilbert space in the case of $p = 2$,

$$H^k(\mathcal{M}) = W^{k,2}(\mathcal{M}), \quad k \in \mathbb{N}.$$

Using this mathematical groundwork we introduce spaces

$$H = H_v \times H_T \quad \text{and} \quad V = V_v \times V_T$$

with

$$H_v = \left\{ v(\Theta, \varphi, p) \in L^2(\mathcal{M})^2 \;\middle|\; v \text{ periodic in } \Theta \text{ and } \varphi, \quad \nabla \cdot \int_{p_0}^{P} v \, dp = 0 \right\}$$

$$H_T = \left\{ T(\Theta, \varphi, p) \in L^2(\mathcal{M}) \;\middle|\; T \text{ periodic in } \Theta \text{ and } \varphi \right\}$$

and

$$V_v = \left\{ v(\Theta, \varphi, p) \in H^1(\mathcal{M})^2 \;\middle|\; v \text{ periodic in } \Theta \text{ and } \varphi, \quad \nabla \cdot \int_{p_0}^{P} v \, dp = 0, \quad v = 0 \text{ on } \Gamma_l \right\}$$

$$V_T = \left\{ T(\Theta, \varphi, p) \in H^1(\mathcal{M}) \;\middle|\; T \text{ periodic in } \Theta \text{ and } \varphi \right\}.$$

We endow H with the classical L^2 inner product

$$\langle U, \tilde{U} \rangle = \int_{\mathcal{M}} U \cdot \tilde{U} \, d\mathcal{M}, \quad U, \tilde{U} \in H$$

and denote the corresponding norm by $\|\cdot\|_H$. We define the inner product of V by

$$\big((U, \tilde{U})\big) = \big((v, \tilde{v})\big)_v + \big((T, \tilde{T})\big)_T, \quad U = (v, T), \tilde{U} = (\tilde{v}, \tilde{T}) \in V$$

with

$$\begin{aligned}
\big((v, \tilde{v})\big)_v &= \int_{\mathcal{M}} \mu_v \nabla v \cdot \nabla \tilde{v} + \nu_v \left(\frac{gp}{R\overline{T}}\right)^2 \frac{\partial}{\partial p} v \frac{\partial}{\partial p} \tilde{v} \, d\mathcal{M} - \int_{\Gamma_o} \alpha_v v \cdot \tilde{v} \, d\Gamma_o \\
\big((T, \tilde{T})\big)_T &= \int_{\mathcal{M}} \mu_T \nabla T \cdot \nabla \tilde{T} + \nu_T \left(\frac{gp}{R\overline{T}}\right)^2 \frac{\partial}{\partial p} T \frac{\partial}{\partial p} \tilde{T} \, d\mathcal{M} \\
&\qquad - \int_{\Gamma_o} \alpha_T T \cdot \tilde{T} \, d\Gamma_o - \int_{\Gamma_l} \alpha_T \left(\frac{gp}{R\overline{T}}\right)^2 T \cdot \tilde{T} \, d\Gamma_l.
\end{aligned}$$

Finally we define a space $V_{(2)}$ as the closure of V in the $H^2(\mathcal{M})^3$ norm and equip it with the classical $H^2(\mathcal{M})$ inner product. While H can be considered as the space containing the basic geometry of our problem, the subspaces $V_{(2)} \subset V \subset H$ additionally demand the boundedness of derivatives of the vector field (in a Sobolev sense).

Remark 5.10. Meteorological Interpretation of H- and V-norms
Considering the typical bilinear term

$$v \cdot v \quad (H - \textit{norm}) \quad \textit{and} \quad \nabla v \cdot \nabla v \quad (V - \textit{norm}),$$

we can identify the H-norm with the kinetic energy and the V-norm with the entropy of the system. Furthermore Lemma 5.18 below shows a close connection between diffusion terms and entropy norm $\|\cdot\|_V$.

5.3.3 The PE in Operator form

We are now in a position to give an operator representation for the PE. To this end let U be a square integrable vector field with horizontal periodicity, i.e.

$$U \in \left\{ u \in L^2(\mathcal{M})^3 \;\middle|\; u \text{ periodic in } \Theta \text{ and } \varphi \right\}.$$

Define the Leray projection P_H as the orthogonal projection of U onto the space of square-integrable, horizontal periodic, divergence free vector fields H, i.e., $P_H U \in H$. We define the linear operator A by merging all linear terms of second order derivatives, i.e.

$$AU = P_H \begin{pmatrix} -\mu_v \Delta v - \nu_v \frac{\partial}{\partial p} \left[\left(\frac{gp}{RT}\right)^2 \frac{\partial}{\partial p} v \right] \\ -\mu_T \Delta T' - \nu_T \frac{\partial}{\partial p} \left[\left(\frac{gp}{RT}\right)^2 \frac{\partial}{\partial p} T' \right] \end{pmatrix}, \quad U = (v, T) \in \mathcal{D}(A) \tag{5.15}$$

with

$$\mathcal{D}(A) = \left\{ U = (v, T) \in V_{(2)} \;\middle|\; \begin{array}{rl} \frac{\partial}{\partial p} v = 0, \quad \frac{\partial}{\partial p} T' = 0 & \text{on } \Gamma_u \\ \nu_T \frac{\partial}{\partial p} T' = -\alpha_T (T' - T_l) & \text{on } \Gamma_l \\ \left.\begin{array}{r} \nu_v \left(\frac{gp}{RT}\right)^2 \frac{\partial}{\partial p} v = \tau_v - \alpha_v (v - v_o) \\ \nu_T \left(\frac{gp}{RT}\right)^2 \frac{\partial}{\partial p} T' = -\alpha_T (T' - T_o) \end{array}\right\} & \text{on } \Gamma_o \end{array} \right\}.$$

While the spaces $V_{(2)} \subset V \subset H$ primarily contain information regarding differentiability and geometry, the boundary conditions are accounted for in the definition of $\mathcal{D}(A)$. In the next step we turn to the bilinear advection term and define

$$B(U, \tilde{U}) = P_H \begin{pmatrix} (\nabla_v v)\, \tilde{v} + \omega(v) \frac{\partial}{\partial p} \tilde{v} \\ (\nabla_v T')\, \tilde{T}' + \omega(v) \frac{\partial}{\partial p} \tilde{T}' \end{pmatrix}, \quad U = (v, T), \tilde{U} = (\tilde{v}, \tilde{T}) \in \mathcal{D}(A). \tag{5.16}$$

Note that we often use the abbreviatory notation $B(U) := B(U,U)$. Recall the diagnostic nature of the generalized vertical velocity $\omega(v)$ and the geopotential Φ'

$$\omega = \omega(v) = -\int_{p_0}^{p} \nabla \cdot v \, dp' \tag{5.13}$$

$$\Phi' = \Phi'(p,T') = \Phi'_s + \int_p^P \frac{RT'}{p'} dp', \tag{5.14}$$

which are combined in the pressure operator

$$A_p U = P_H \begin{pmatrix} \int_p^P \frac{RT'}{p'} dp' \\ -\frac{C^2}{pR}\omega(v) \end{pmatrix}, \quad U = (v,T) \in V.$$

While the Coriolis term is captured in

$$CU = P_H \begin{pmatrix} 2\cos\Theta k \times v \\ 0 \end{pmatrix}, \quad U = (v,T) \in V,$$

we merge the forcing terms

$$F_U = P_H \begin{pmatrix} F_v \\ F_T \end{pmatrix}$$

and finally set

$$F(U) = A_p U + CU + F_U. \tag{5.17}$$

The viscous PE than have the clearly arranged form

$$dU + (AU + B(U,U) + F(U))\, dt = 0. \tag{5.18}$$

Before we conclude the discussion of the deterministic setting we homogenize the boundary conditions, which is important for the symmetry of the linear operator A (see Lemma 5.18 and Remark 5.19).

Lemma 5.11. Homogenization of the Boundary Conditions A
Without loss of generality we may assume homogeneous boundary conditions, i.e.

$$\begin{aligned} \nu_T \frac{\partial}{\partial p} T &= -\alpha_T T \quad \textit{on } \Gamma_l \\ \nu_v \left(\frac{gp}{R\bar{T}}\right)^2 \frac{\partial}{\partial p} v &= -\alpha_v v \quad \textit{on } \Gamma_o \\ \nu_T \left(\frac{gp}{R\bar{T}}\right)^2 \frac{\partial}{\partial p} T &= -\alpha_T T \quad \textit{on } \Gamma_o \end{aligned}$$

Proof. *Lemma 5.11*
Since the reduction of the system to homogeneous boundary conditions is a common technique, we present the main idea and refer to [136, p. 574ff.] for a rigorous proof. Let U^* be the solution of the deterministic linear system, i.e.,

$$\frac{dU^*}{dt} + AU^* + F(U^*) = 0, \quad U(0) = U_0 \tag{\#}$$

using the same boundary conditions (5.10), (5.11), (5.12) as the fully nonlinear system. The existence and uniqueness of this system follows from the Lax-Milgram theorem, which gives us the existence of weak solutions for very general systems, see Section 4.4.1

"*Weak formulation of the Stokes problem*" [136, p. 640ff.]. Furthermore it is shown that the solution belongs to $H^2(\mathcal{M})$, which ultimately allows to control U^* as follows

$$\sup_{t\in[0,T]} \|U_t^*\|_V^2 + \int_0^T \|U^*\|_{H^2(\mathcal{M})}^2 ds \leq k_2, \tag{\#\#}$$

with a bounded expression k_2, which is governed by the initial values and the vector field of the linear stationary system, see (3.18), (3.22) and (3.23) in [136, p. 576f.]. We can now use the decomposition $U = U^* + U'$ on the (fully nonlinear) PE and obtain

$$\frac{dU'}{dt} + AU' + B(U',U') + B(U^*,U') + B(U',U^*) + B(U^*,U^*) + F(U') = 0, \quad U'(0) = 0,$$

where we used (#) and the (bi-)linearity of the involved terms. The arising bilinear terms containing U^* can be controlled easily using (##), see [136, p. 578ff.]. Since the boundary conditions are in fact a linear PDE, we conclude that $U' = (v', T')$ satisfies homogeneous boundary conditions. Exemplarily the temperature over land surface obeys

$$\nu_T \frac{\partial}{\partial p} T' = \nu_T \frac{\partial}{\partial p} T - \nu_T \frac{\partial}{\partial p} T^* = -\alpha_T (T - T_l) + \alpha_T (T^* - T_l) = -\alpha_T (T - T^*) = -\alpha_T T'.$$

□

As a consequence we obtain for the homogenized system

$$\mathcal{D}(A) = \left\{ U = (v,T) \in V_{(2)} \;\middle|\; \begin{array}{rl} \frac{\partial}{\partial p} v = 0, \quad \frac{\partial}{\partial p} T' = 0 & \text{on } \Gamma_u \\ \nu_T \frac{\partial}{\partial p} T' = -\alpha_T T' & \text{on } \Gamma_l \\ \left.\begin{array}{r} \nu_v \left(\frac{gp}{RT}\right)^2 \frac{\partial}{\partial p} v = \alpha_v v \\ \nu_T \left(\frac{gp}{RT}\right)^2 \frac{\partial}{\partial p} T' = -\alpha_T T \end{array}\right\} & \text{on } \Gamma_o \end{array} \right\}. \tag{5.19}$$

5.4 The Stochastic 3d Primitive Equations

We are finally in a position to introduce stochasticity into the PE by stating the stochastic Itô evolution system, which is the subject of investigation for the remainder of this work

$$dU + (AU + B(U,U) + F(U))\, dt = \sigma(U) dW_t, \quad U(0) = \xi \tag{5.20}$$

where the stochastic term is given by

$$\sigma(U) dW_t = \begin{pmatrix} \sigma_v(v,T) dW_t^1 \\ \sigma_T(v,T) dW_t^2 \end{pmatrix}.$$

Here $W_t^1 = (W_t^{1,1}, W_t^{1,2})$ and $W_t^{1,1}, W_t^{1,2}, W_t^2$ are independent white noise processes defined below. Although we have to constrain the coefficient σ to abide certain growth conditions (see (5.24) and (5.25)), which are needed to guarantee the existence of solutions, this formulation allows for a wide variety of stochastic influences, e.g., white, colored, additive, multiplicative, linear and nonlinear noise. In order to prove the existence of solutions we aim to use the results of Debussche et al. [7], where criteria for the existence of local martingale (Definition 5.13) and unique maximal pathwise solutions (Definition 5.14) are given for an abstract operator evolution equation. Therefore we verify that the operators defined for the PE satisfy the assumptions specified therein, starting with some necessary definitions and the statement of the main thesis.

5.4.1 The Stochastic Framework

Before we can define various solutions of stochastic evolutions systems, let alone give criteria for the existence of such solutions for (5.20), we have to characterize the mathematical structure of the stochastic terms.

Definition 5.12. Stochastic Basis
$S = (\Omega, \mathcal{F}, \{\mathcal{F}\}_{t\geq 0}, P, \{W^k\}_{k\geq 1})$ *is called a stochastic basis if* $(\Omega, \mathcal{F}, P)$ *is a probability space and* $\{W^k\}_{k\geq 1}$ *a sequence of independent standard 1d Brownian motions adapted to the filtration* $\{\mathcal{F}_t\}_{t\geq 0}$*. To avoid unnecessary complications we assume that the filtration* $\mathcal{F}_t$ *is complete and right-continuous.*

We fix a stochastic basis $S = (\Omega, \mathcal{F}, \{\mathcal{F}\}_{t\geq 0}, P, \{W^k\}_{k\geq 1})$ and a separable Hilbert space G with an orthonormal basis $\{g_k\}_{k\geq 1}$. We can now define the "*cylindrical Brownian motion (evolving over G)*"

$$W = \sum_{k\geq 1} W^k g_k.$$

Let X be a separable Hilbert space with the associated inner product $\langle\cdot,\cdot\rangle_X$ and norm $|\cdot|_X$. Denote the set of all operators from G to X by $\mathcal{L}(G, X)$. Then the collection of all "*Hilbert-Schmidt operators*" from G to X is given by

$$HS(G, X) = \left\{ A \in \mathcal{L}(G, X) \;\middle|\; \sum_{k\geq 1} |Ag_k|_X^2 < \infty \right\}.$$

By endowing this collection with the inner product

$$\langle A, B\rangle_{HS(G,X)} = \sum_{k\geq 1} \langle Ag_k, Bg_k\rangle_X, \quad A, B \in HS(G, X)$$

$HS(G, X)$ is itself a Hilbert space. Furthermore define a space $G_0 \supset G$ via

$$G_0 = \left\{ v = \sum_{k\geq 1} v_k g_k \;\middle|\; \sum_{k\geq 1} v_k^2 k^2 < \infty \right\}$$

and endow it with the norm

$$|v|_{G_0}^2 = \sum_{k\geq 1} \frac{v_k^2}{k^2}.$$

Then it is a well known fact that W hast almost surely continuous sample paths in G_0, i.e., $W(\omega) \in C([0, T], G_0)$ for almost every $\omega \in \Omega$, see for instance [159]. Based on this setting one can obtain the well-posedness of (5.20) and the occurring stochastic Itô integrals

$$\int_0^t G\, dW_t = \sum_{k\geq 1} \int_0^t G\, dW_t^k,$$

for all X-valued predictable processes $G \in L^2(\Omega, L^2_{loc}([0, \infty), HS(G_0, X)))$, see for instance [160, Ch. 2.2 and 2.3]. Let us specify two types of solutions for stochastic evolution systems, which are necessary to state the main result of this section.

Definition 5.13. Local and Global Martingale Solutions

1. *A triple* (S, U, τ) *is a local Martingale solution of*

$$dU + (AU + B(U,U) + F(U))\,dt = \sigma(U)dW_t, \quad U(0) = \xi \tag{5.20}$$

if

- $S = (\Omega, \mathcal{F}, \{\mathcal{F}\}_{t\geq 0}, P, \{W^k\}_{k\geq 1})$ *is a stochastic basis*
- τ *is a stopping time relative to* $\mathcal{F}_t$
- $U(\cdot) = U(\cdot \wedge \tau) : \Omega \times [0,\infty) \to V$ *is a* $\mathcal{F}_t$ *adapted process such that*

$$\begin{aligned} &U(\cdot \wedge \tau) \in L^2(\Omega; C([0,\infty); V) \\ &U\mathbb{1}_{t\leq\tau} \in L^2_{loc}(\Omega; C([0,\infty); \mathcal{D}(A)), \end{aligned} \tag{5.21}$$

the law μ_0 *of* $U(0)$ *satisfies*

$$\int_V \|U\|_V^q \; d\mu_0(U) < \infty \tag{5.22}$$

with $q \geq 8$*, and* U *satisfies*

$$U(t\wedge\tau) = U(0) - \int_0^{t\wedge\tau} AU + B(U) + F(U)\; ds + \int_0^{t\wedge\tau} \sigma(U)\; dW, \tag{5.23}$$

with equality understood in H*.*

2. *A Martingale solution* (S, U, τ) *is global if* $\tau = \infty$ *P-a.s.*

Definition 5.14. Local, Maximal and Global Pathwise Solutions and Uniqueness
Let $S = (\Omega, \mathcal{F}, \{\mathcal{F}\}_{t\geq 0}, P, \{W^k\}_{k\geq 1})$ *be a fixed stochastic basis and let* $\xi \in L^2(\Omega; V)$ *be a* $\mathcal{F}_0$*-measurable,* V*-valued random variable.*

1. *A pair* (U, τ) *is a local pathwise solution of* (5.20) *if* τ *is a strictly positive stopping time and* $U(\cdot \wedge \tau)$ *is an* $\mathcal{F}_t$*-adapted process in* V *satisfying* (5.21) *and* (5.23)*.*

2. *A pathwise solutions* (U, τ) *is (pathwise) unique up to a stopping time* $\tau > 0$ *if for any two pathwise solutions* (U^1, τ) *and* (U^2, τ) *which coincide at* $t = 0$ *on a subset* $\Omega \supset \bar{\Omega} = \{(U^1(0) = U^2(0)\}$ *it holds true that*

$$P\left(\mathbb{1}_{\bar{\Omega}}\left(U^1(t\wedge\tau) - U^2(t\wedge\tau)\right) = 0;\ \forall t \geq 0\right) = 1.$$

3. *Let* ζ *be a (possibly infinite) stopping time and assume that* U *is a predictable process in* H*. Then the pair* (U, ζ) *is a maximal pathwise solution if there exists a strictly increasing sequence of stopping times*$\{\tau_n\}_{n\geq 1}$ *converging to* ζ*, where* (U, τ_n) *is a local pathwise solution for every* n *and*

$$\sup_{t\in[0,\zeta]} \|U\|_V^2 + \int_0^\zeta \|AU\|_H^2 \; ds = \infty$$

P-a.s. on the set $\{\zeta < \infty\}$*.*

4. *A maximal pathwise solution* (U, ζ) *is global if* $\zeta = \infty$ *P-a.s.*

Definition 5.15. Boundedness and Lipschitz Continuity
Let X, Y *be Banach spaces.*

1. *We denote by* $Bnd(X, Y)$ *the set of all continuous mappings*

$$\Psi : [0,\infty) \times X \to Y$$

satisfying

$$\|\Psi(x,t)\|_Y \leq c\,(1 + \|x\|_X), \quad x \in X,\ t \geq 0$$

with a constant c *independent of* x *and* t*.*

2. *We define* $Lip(X, Y)$ *to be the set of all* $\Psi \in Bnd(X, Y)$ *such that*

$$\|\Psi(x,t) - \Psi(y,t)\|_Y \leq c\,(\|x - y\|_X), \quad x, y \in X,\ t \geq 0.$$

5.4.2 Existence of Solutions of the Stochastic 3d Primitive Equations

After this preliminary work we may now state the main result, which guarantees the existence of solutions of (5.20) for suitable constrained coefficients σ. In order to prove this statement we aim to use Theorem 2.1. in [7], which yields the existence of solutions for an abstract evolution system in operator form. We then show in Lemma 5.18,5.20 and 5.21 that the the operators A, B and F of the PE, defined in Section 5.3.3, satisfy the necessary assumptions.

Theorem 5.16. Existence of Solutions for the Stochastic 3d PE
Let $\xi = (v_0, T_0)$ take values in V and let μ_0 satisfy (5.22) *with $q \geq 8$. Regarding the forcing terms we assume*

$$F_v, F_T \in L^2_{loc}\left([0,\infty), L^2(\mathcal{M})\right).$$

Furthermore suppose that σ satisfies

$$\sigma \in Bnd(H, L_2(G,H)) \cap Bnd(V, L_2(G,V)) \cap Bnd(\mathcal{D}(A), L_2(G, \mathcal{D}(A))). \tag{5.24}$$

Then:

1. *There exists a local martingale solution of the stochastic three-dimensional PE* (5.20)*.*
2. *If σ additionally satisfies*

$$\sigma \in Lip(H, L_2(G,H)) \cap Lip(V, L_2(G,V)) \cap Lip(\mathcal{D}(A), L_2(G, \mathcal{D}(A))), \tag{5.25}$$

and the initial conditions satisfy (5.22) *for all $q \geq 2$, then there exists a unique maximal, pathwise solution of the stochastic three-dimensional PE* (5.20)*.*

Remark 5.17. Uniqueness of Pathwise Solutions
We would like to emphasize, that Theorem 5.16 not only yields the existence of pathwise solution for suitable "tame" noise, but also guarantees the uniqueness of this solution. This in turn is a necessary prerequisite for the construction of convergent numerical schemes, which is the subject of our study in the following Sections 6 and 7.

The assumptions on the linear operator A, particularly the existence of a complete orthonormal basis of eigenvectors for the Hilbert space H, which is used for a Galerkin approximation, are a direct consequence from the following Lemma.

Lemma 5.18. Characteristics of the Linear Operator A
For all $U \in \mathcal{D}(A)$ and $\tilde{U} \in V$ we have

$$\left\langle AU, \tilde{U} \right\rangle = \left((U, \tilde{U})\right).$$

This implies that A is a symmetric operator. Furthermore we have

$$\|AU\|_H \cong ||U||_{H^2}.$$

Proof. *Lemma 5.18*
Since H is equipped with the classical L^2 inner product we have

$$\begin{aligned}
\left\langle AU, \tilde{U} \right\rangle &= \int_{\mathcal{M}} AU \cdot \tilde{U} \, d\mathcal{M} \\
&= -\int_{\mathcal{M}} \mu_v \Delta v \cdot \tilde{v} \, d\mathcal{M} - \int_{\mathcal{M}} \nu_v \frac{\partial}{\partial p}\left[\left(\frac{gp}{R\bar{T}}\right)^2 \frac{\partial}{\partial p} v\right] \cdot \tilde{v} \, d\mathcal{M} \\
&\quad - \int_{\mathcal{M}} \mu_T \Delta T \cdot \tilde{T} \, d\mathcal{M} - \int_{\mathcal{M}} \nu_T \frac{\partial}{\partial p}\left[\left(\frac{gp}{R\bar{T}}\right)^2 \frac{\partial}{\partial p} T\right] \cdot \tilde{T} \, d\mathcal{M} \\
&=: J_v^1 + J_v^2 + J_T^1 + J_T^2
\end{aligned}$$

These terms are treated by splitting the integral in horizontal and vertical directions followed by integration by parts using the homogeneous boundary conditions, i.e., the characteristics of the spaces $\mathcal{D}(A)$ and V. The horizontal derivatives $J^1_{v,T}$ satisfy

$$\begin{aligned}
J^1_v + J^1_T &= -\int_{\mathcal{M}} \mu_v \Delta v \cdot \tilde{v} + \mu_T \Delta T \cdot \tilde{T}\, d\mathcal{M} \\
&= -\int_P^{p_0} \int_{\mathcal{M}_0} \mu_v \Delta v \cdot \tilde{v} + \mu_T \Delta T \cdot \tilde{T}\, d\mathcal{M}_0\, dp \\
&= -\int_P^{p_0} \underbrace{\left[\mu_v \nabla v \cdot \tilde{v} + \mu_T \nabla T \cdot \tilde{T}\right]_{\partial\mathcal{M}_0}}_{=0 \text{ due to horizontal periodicity}} dp \\
&\qquad + \int_P^{p_0} \int_{\mathcal{M}_0} \mu_v \nabla v \cdot \nabla \tilde{v} + \mu_T \nabla T \cdot \nabla \tilde{T}\, d\mathcal{M}_0\, dp \\
&= \int_{\mathcal{M}} \mu_v \nabla v \cdot \nabla \tilde{v} + \mu_T \nabla T \cdot \nabla \tilde{T}\, d\mathcal{M}.
\end{aligned}$$

Regarding the vertical derivatives $J^2_{v,T}$ we get

$$\begin{aligned}
J^2_v &= -\int_{\mathcal{M}} \nu_v \frac{\partial}{\partial p}\left[\left(\frac{gp}{R\bar{T}}\right)^2 \frac{\partial}{\partial p} v\right] \cdot \tilde{v}\, d\mathcal{M} \\
&= -\int_{\mathcal{M}_0} \int_P^{p_0} \nu_v \frac{\partial}{\partial p}\left[\left(\frac{gp}{R\bar{T}}\right)^2 \frac{\partial}{\partial p} v\right] \cdot \tilde{v}\, d\mathcal{M}_0\, dp \\
&= -\int_{\mathcal{M}_0} \left[\nu_v \left(\frac{gp}{R\bar{T}}\right)^2 \frac{\partial}{\partial p} v \cdot \tilde{v}\right]_P^{p_0} d\mathcal{M}_0 + \int_{\mathcal{M}_0} \int_P^{p_0} \nu_v \left(\frac{gp}{R\bar{T}}\right)^2 \frac{\partial}{\partial p} v \cdot \frac{\partial}{\partial p}\tilde{v}\, d\mathcal{M}_0\, dp \\
&= -\int_{\Gamma_u} \nu_v \left(\frac{gp}{R\bar{T}}\right)^2 \frac{\partial}{\partial p} v \cdot \tilde{v}\, d\Gamma_u + \int_{\Gamma_o} \nu_v \left(\frac{gp}{R\bar{T}}\right)^2 \frac{\partial}{\partial p} v \cdot \tilde{v}\, d\Gamma_o \\
&\qquad + \int_{\Gamma_l} \nu_v \left(\frac{gp}{R\bar{T}}\right)^2 \frac{\partial}{\partial p} v \cdot \tilde{v}\, d\Gamma_l + \int_{\mathcal{M}} \nu_v \left(\frac{gp}{R\bar{T}}\right)^2 \frac{\partial}{\partial p} v \cdot \frac{\partial}{\partial p}\tilde{v}\, d\mathcal{M}
\end{aligned}$$

Note that the homogeneous boundary conditions (5.19) lead to

$$\begin{aligned}
\frac{\partial}{\partial p} v &= 0 \qquad \text{on} \quad \Gamma_u \\
\nu_v \left(\frac{gp}{R\bar{T}}\right)^2 \frac{\partial}{\partial p} v &= -\alpha_v v \quad \text{on} \quad \Gamma_o.
\end{aligned}$$

Furthermore we have $\tilde{v} = 0$ on Γ_l by definition of V, yielding

$$J^2_v = -\int_{\Gamma_o} \alpha_v v \cdot \tilde{v}\, d\Gamma_o + \int_{\mathcal{M}} \nu_v \left(\frac{gp}{R\bar{T}}\right)^2 \frac{\partial}{\partial p} v \cdot \frac{\partial}{\partial p}\tilde{v}\, d\mathcal{M}.$$

In an analog way but using the slightly different boundary conditions for T we obtain

$$\begin{aligned}
J^2_T = &-\int_{\Gamma_o} \alpha_T T \cdot \tilde{T}\, d\Gamma_o - \int_{\Gamma_l} \alpha_T \left(\frac{gp}{R\bar{T}}\right)^2 T \cdot \tilde{T}\, d\Gamma_l \\
&+ \int_{\mathcal{M}} \nu_T \left(\frac{gp}{R\bar{T}}\right)^2 \frac{\partial}{\partial p} T \cdot \frac{\partial}{\partial p}\tilde{T}\, d\mathcal{M},
\end{aligned}$$

yielding

$$\begin{aligned}
\left\langle AU, \tilde{U} \right\rangle = &\int_{\mathcal{M}} \mu_v \nabla v \cdot \nabla \tilde{v} + \nu_v \left(\frac{gp}{R\bar{T}}\right)^2 \frac{\partial}{\partial p} v \cdot \frac{\partial}{\partial p}\tilde{v}\, d\mathcal{M} - \int_{\Gamma_o} \alpha_v v \cdot \tilde{v}\, d\Gamma_o \\
&+ \int_{\mathcal{M}} \mu_T \nabla T \cdot \nabla \tilde{T} + \nu_T \left(\frac{gp}{R\bar{T}}\right)^2 \frac{\partial}{\partial p} T \cdot \frac{\partial}{\partial p}\tilde{T}\, d\mathcal{M} \\
&- \int_{\Gamma_o} \alpha_T T \cdot \tilde{T}\, d\Gamma_o - \int_{\Gamma_l} \alpha_T \left(\frac{gp}{R\bar{T}}\right)^2 T \cdot \tilde{T}\, d\Gamma_l \\
= &\left((U, \tilde{U})\right).
\end{aligned}$$

The second statement $\|AU\|_H \cong ||U||_{H^2}$ follows from the same line of thought, but requires some additional technical considerations. For a rigorous proof see for instance [136, Ch. 4]. □

Remark 5.19. The Necessity of Homogeneous Boundary Conditions
The necessity for homogeneous boundary conditions derives from the fact, that $\big((U,\tilde{U})\big)$ is an inner product and therefore has to be symmetric. Using inhomogeneous boundary conditions would however lead to terms of the form

$$-\int_{\Gamma_o} \alpha_v (v - v_o)\cdot \tilde{v}\; d\Gamma_o,$$

when calculating $\langle AU, U\rangle$. Obviously the symmetry would be broken for $v_o \neq 0$ falsifying Lemma 5.18. For a slightly different approach to this problem we refer to [135, p. 264], where smooth cut-off functions in a small area around the boundaries are used to enforce homogeneity.

Lemma 5.20. Characteristics of the Bilinear Advection Operator B
The trilinear form

$$\big\langle B(U,U^{\#}),U^b\big\rangle$$

is continuous on $V\times\mathcal{D}(A)\times V$ and on $\mathcal{D}(A)\times\mathcal{D}(A)\times H$ with

$$\begin{aligned}
&\big\langle B(U,U^{\#}),U^{\#}\big\rangle = 0, && \forall U\in V,\ U^{\#}\in\mathcal{D}(A)\\
&|\big\langle B(U,U^{\#}),U^b\big\rangle| \le c_0\,\|U\|_V \big\|AU^{\#}\big\|_H \big\|U^b\big\|_V, && \forall U,U^b\in V,\ U^{\#}\in\mathcal{D}(A)\\
&|\big\langle B(U,U^{\#}),U^b\big\rangle| \le c_0\,\|U\|_V^{\frac12}\|AU\|_H^{\frac12}\big\|U^{\#}\big\|_V^{\frac12}\big\|AU^{\#}\big\|_H^{\frac12}\big\|U^b\big\|_H, && \forall U^b\in H,\ U,U^{\#}\in\mathcal{D}(A)
\end{aligned}$$

Proof. *Lemma 5.20*
The advection term can be considered as the key element of the PE, due to the interdependence of nonlinearity and horizontal as well as vertical derivatives. This is particularly true for terms involving the vertical velocity, e.g.

$$\omega(v)\frac{\partial}{\partial p}v = -\int_{p_0}^{p}\nabla\cdot v\; dp'\;\frac{\partial}{\partial p}v.$$

As we mentioned earlier during the discussion on diagnostic variables, it is precisely the appearance of this additional horizontal derivative that makes the technical treatment of the advection term so challenging. Consequently these terms have been intensively studied and proofs of the statement can be found in literature, see for instance Lemma 2.1 and Lemma 3.1 in [136]. However, we sketch the proof for the third inequality by example of the above equation in order to present the basic ideas. We have to estimate the expression

$$\left|\int_{\mathcal{M}}\omega(v)\frac{\partial}{\partial p}v^{\#}v^b d\mathcal{M}\right| = \left|\int_{\mathcal{M}_0}\int_{p_0}^{P}\omega(v)\frac{\partial}{\partial p}v^{\#}v^b dp\; d\mathcal{M}_0\right|$$

using Hölder's inequality separately on the horizontal and vertical integral. Denote by $\|\cdot\|_{L_p^q}$ the L^q-norm in vertical direction and by $\|\cdot\|_{L^q_{\mathcal{M}_0}}$ the L^q-norm in horizontal directions. Note that

$$\|\omega(v)\|_{L_p^\infty} = \left\|\int_{p_0}^{p}\nabla\cdot v\; dp'\right\|_{L_p^\infty} \le (P-p_0)^{\frac12}\,\|\nabla\cdot v\,\|_{L_p^2},$$

which leads to

$$\begin{aligned}
\left|\int_{\mathcal{M}} \omega(v)\frac{\partial}{\partial p}v^{\#}v^{b}d\mathcal{M}\right| &= \left|\int_{\mathcal{M}_0}\int_{p_0}^{P} \omega(v)\frac{\partial}{\partial p}v^{\#}v^{b}\, dp\, d\mathcal{M}_0\right| \\
&\leq \int_{\mathcal{M}_0} \|\omega(v)\|_{L^{\infty}_{p}} \left\|\frac{\partial}{\partial p}v^{\#}\right\|_{L^2_p} \left\|v^{b}\right\|_{L^2_p} d\mathcal{M}_0 \\
&\leq \int_{\mathcal{M}_0} \|\omega(v)\|_{L^{\infty}_{p}} \left\|\frac{\partial}{\partial p}v^{\#}\right\|_{L^2_p} \left\|v^{b}\right\|_{L^2_p} d\mathcal{M}_0 \\
&\leq (P-p_0)^{\frac{1}{2}} \int_{\mathcal{M}_0} \|\nabla\cdot v\|_{L^2_p} \left\|\frac{\partial}{\partial p}v^{\#}\right\|_{L^2_p} \left\|v^{b}\right\|_{L^2_p} d\mathcal{M}_0 \\
&\leq (P-p_0)^{\frac{1}{2}} \|\nabla\cdot v\|_{L^4_{\mathcal{M}_0}L^2_p} \left\|\frac{\partial}{\partial p}v^{\#}\right\|_{L^4_{\mathcal{M}_0}L^2_p} \left\|v^{b}\right\|_{L^2_{\mathcal{M}_0}L^2_p}.
\end{aligned}$$

For the last term we have

$$\|\cdot\|_{L^2_{\mathcal{M}_0}L^2_p} = \|\cdot\|_{L^2_{\mathcal{M}}}.$$

Note that the horizontal norms act on space dimension two and therefore satisfy the following Sobolev inequality

$$\|\cdot\|_{L^4_{\mathcal{M}_0}} \leq c(\mathcal{M}_0)\, \|\cdot\|^{\frac{1}{2}}_{L^2_{\mathcal{M}_0}} \|\cdot\|^{\frac{1}{2}}_{H^1_{\mathcal{M}_0}},$$

see for instance equation (3.36) in [136, p. 582]. After some technical considerations [136, p. 582f.] this yields

$$\|\cdot\|_{L^4_{\mathcal{M}_0}L^2_p} \leq c(\mathcal{M}_0)\, \|\cdot\|^{\frac{1}{2}}_{L^2_{\mathcal{M}}} \|\cdot\|^{\frac{1}{2}}_{H^1_{\mathcal{M}}}.$$

Denoting $c_0 = c(\mathcal{M}_0)(P-p_0)^{\frac{1}{2}}$ we obtain

$$\begin{aligned}
\left|\int_{\mathcal{M}} \omega(v)\frac{\partial}{\partial p}v^{\#}v^{b}d\mathcal{M}\right| &\leq c_0 \|\nabla\cdot v\|^{\frac{1}{2}}_{L^2_{\mathcal{M}}} \|\nabla\cdot v\|^{\frac{1}{2}}_{H^1_{\mathcal{M}}} \|\frac{\partial}{\partial p}v^{\#}\|^{\frac{1}{2}}_{L^2_{\mathcal{M}}} \|\frac{\partial}{\partial p}v^{\#}\|^{\frac{1}{2}}_{H^1_{\mathcal{M}}} \left\|v^{b}\right\|_{L^2_{\mathcal{M}}} \\
&\leq \tilde{c_0} \|v\|^{\frac{1}{2}}_{H^1_{\mathcal{M}}} \|v\|^{\frac{1}{2}}_{H^2_{\mathcal{M}}} \|v^{\#}\|^{\frac{1}{2}}_{H^1_{\mathcal{M}}} \|v^{\#}\|^{\frac{1}{2}}_{H^2_{\mathcal{M}}} \left\|v^{b}\right\|_{L^2_{\mathcal{M}}},
\end{aligned}$$

where we used the definition of the H^1 and H^2 norms in the last step. Along the lines of Lemma 5.18 one can conclude that the norms $\|U\|_V$ and $\|AU\|_H$ are equivalent to $\|U\|_{H^1_{\mathcal{M}}}$ and $\|U\|_{H^2_{\mathcal{M}}}$ respectively. □

Lemma 5.21. Characteristics of the Forcing Operator F
Let the forcing terms F_v and F_T appearing in the definition of the viscous PE (Lemma 5.5) satisfy

$$F_v, F_T \in L^2_{loc}\left([0,\infty), L^2(\mathcal{M})\right)$$

Then we have

$$F \in Bnd(V,H) \quad \textit{and} \quad F \in Lip(V,H).$$

Proof. *Lemma 5.21*
Using the fact that the norm $\|\cdot\|_V$ is equivalent to $\|\cdot\|_{H^1_{\mathcal{M}}}$, the statement follows directly by definition (5.17) of F. □

5.5 Conclusion

During this section we derived an abstract formulation of the three-dimensional PE both in a deterministic and a stochastic framework. This approach allows us to use the fundamentally important results on existence and uniqueness of solutions (Theorem 5.16), which is the basis for the numerical consideration in the remainder of this work.

Starting from general equations of the atmosphere (Section 5.1), we obtained the viscous PE using the concepts of coordinate transformation and scale analysis. The first aspect includes the transformation of the system from Cartesian into spherical coordinates (Section 5.1). The subsequent transformation into pressure coordinates (Section 5.2.1) allowed us to treat the system as an incompressible fluid from a mathematical point of view (Remark 5.4). The importance of this property is demonstrated by the conceptual example of Stokes equations (Section 5.3.1). The hydrostatic and shallow atmosphere approximation (Section 5.2) together with diffusion parameterization and an approximation on vertical temperature distributions (Section 5.2.2) produces mathematically more accessible equations while preserving a high accuracy on synoptic scales (5.5). The resulting viscous PE (Lemma 5.5) were completed by a setting of boundary conditions (Section 5.2.3) allowing the simultaneous treatment of land and oceanic surfaces.

In order to take account of the growing complexity of the equations in a stochastic setting we defined an abstract operator framework, which leads to the clearly arranged representation of the viscous PE

$$dU + (AU + B(U,U) + F(U))\,dt = 0. \tag{5.18}$$

In Section 5.4 we introduced the stochastic PE

$$dU + (AU + B(U,U) + F(U))\,dt = \sigma(U)dW_t. \tag{5.20}$$

The homogenization of boundary conditions (Lemma 5.11) enabled us to utilize recent results of Debussche et al [7], which guarantee local existence and uniqueness of pathwise solutions for suitable bounded initial conditions and forcing terms F, σ (Theorem 5.16). Drawing on these findings we derive a Galerkin scheme for the discretization of spatial and stochastic dimensions in the following Section 6. Furthermore, under additional assumptions, we discuss an approach for a complete numerical scheme, i.e., including temporal discretization, in Section 7.

6 Galerkin Approximation for an Abstract Fluid Model

In the last section we derived local existence and uniqueness of pathwise solutions for the PE (Theorem 5.16). As a result, we now in a position to start the development of a numerical scheme for these equations aiming at quantitative convergence rate estimates. While the numerical framework for SODE is well developed (see for instance [44]), the "*state of development of numerical schemes for SPDEs compares with that for SODEs in the early 1970s and most of the numerical schemes that have been proposed to date have a low order of convergence, especially in terms of an overall computational effort*" [29]. Most higher order schemes for SODE are derived by iterated application of the Itô formula, yielding so-called *stochastic Taylor expansions*. Unfortunately, we have no access to an Itô formula in a general, infinite-dimensional environment for SPDE, which is the main obstacle for the development of an extensive numerical framework. The theoretical understanding of SPDE is further complicated by the multitude of different solution-concepts and function spaces. We refer to [161] and [160] for a comprehensive treatment of this topic. During the last 15 years the interest in the numerical treatment of SPDE has grown, which resulted in a variety of different approaches. Most notably are the linear-implicit Euler scheme and the Crank-Nicolson scheme, which combine a spatial discretization via finite elements or Galerkin approximations with time discretization methods for stiff SODE. For an extensive overview and list of publications we refer to [29, Ch. 3.2]. In 1999, Gyöngy [162, 163] showed that these schemes converge with order $\frac{1}{2}$ in space and $\frac{1}{4}$ in time, resulting in the overall convergence rate of $\frac{1}{6}$ with respect to the computational effort. One year later, Davie and Gaines published a highly esteemed paper [164], where they studied the linear system

$$dU_t = \frac{\partial^2}{\partial x^2} U_t dt + dW_t,$$

which can be considered as one of the most basic SPDE. Their results state that every numerical scheme of this type, which uses only equidistant values of the noise W_t, has a maximal overall convergence order of $\frac{1}{6}$. However they stated that "*it is still reasonable to hope that by generating suitable linear functionals of the noise (which will be normally distributed) one may be able to improve the order of approximation*". It was not until 2009 that this thought was followed up by Jentzen and Kloeden [165], who considered the stochastic evolution equation

$$dU_t = (AU_t + F(U_t))dt + dW_t, \quad U(0) = U_0,$$

defined on a Hilbert space H with an unbounded linear operator A and a smooth function F. This equation is interpreted in the mild sense, i.e.

$$U_t = e^{At}U_0 + \int_0^t e^{A(t-s)}F(U_t)dt + \int_0^t e^{A(t-s)}dW_t.$$

The mild solution is then spatially approximated by a Galerkin scheme using the eigenvectors of A as an orthonormal basis for the projection of operators and processes onto finite-dimensional subspaces. For the temporal discretization, an approach analogous with the Euler-Maruyama scheme for SODE is used, leading to the denotation "*exponential Euler scheme*". This scheme has a temporal convergence order of $1 - \epsilon$ for an arbitrary small $\epsilon > 0$. The spatial convergence order depends on the specifics of the nonlinear operator F. Hence, using the functional of the noise instead of a equidistant evaluation leads to a remarkable improvement in the temporal convergence rate. While this is already valuable in itself, the approach with mild solutions bears the potential to derive even higher order schemes using Taylor expansions without the need of an Itô formula [166, 167]. However, these expansions require bounded derivatives of the nonlinear terms, which is not satisfied by the system studied in this work. Nevertheless, the framework of mild solutions provides an elegant way to treat the linear operator A and allows us to utilize existing bounds on the nonlinear terms (see Assumption 6.4 and 6.5).

We aim to discretize the state space using a Galerkin approximation based on the eigenfunctions of the linear operator A. The noise term can be controlled in a similar way under

some quantitative assumption on the operator σ. It turns out that we have to postulate an additional a priori estimate regarding the boundedness of the numerical solution (see Assumption 7.3) in order to prove the convergence for the time-discretization model. This is a common issue for explicit schemes in the context of non-Lipschitz continuous differential equations, particularly in the case of infinite-dimensions. We refer to Section 7 for a more elaborate discussion of this subject, where we focus on the construction of a time-discretized numerical scheme. In the following we draw our attention on rigorous Galerkin approximations for spatial and stochastic dimensions, which are valid without additional assumptions.

While the idea of using Galerkin approximations can certainly be considered a standard approach when dealing with PDE, it exhibits some nontrivial issues in the context of stochastic non-Lipschitz systems:

- The presence of noise leads to more irregular paths
- Non-Lipschitz operators make it hard to control the growth of the solution
- The infinite-dimensional framework prohibits the use of many standard techniques and results

Each of these aspects would not be critical on its own but in combination they are the source for serious technical obstacles that are discussed in more detail during the outline of the proof for Theorem 6.10.

6.1 Setting and Assumptions

Our aim is to derive a numerical scheme to solve the viscous PE. The aforementioned work by Debussche et al. [7] provides the existence and uniqueness of solutions for a more general abstract fluid model, which includes the viscous PE, see Theorem 5.16. Therefore we consider the abstract evolution equation

$$dU = (AU + B(U,U) + F(U))\,dt + \sigma(U)dW, \quad U_0 = \xi. \tag{6.1}$$

Note that this equation corresponds to the abstract PE 5.20 with inverted signs of the operators A, B and F. Since the assumptions on B and F involve only absolute values, the change of sign is trivial. Regarding the linear operator A this is purely a matter of convenience for the later technical calculations, and is balanced by the sign of the eigenvalues in Assumption 6.3. Let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathcal{F}_t)_{t\in[0,T]}$ for a $0 < T \leq \tau$, where the stopping time τ is specified in Corollary 6.9 below. Let $(H, \langle\cdot,\cdot\rangle_H, \|\cdot\|_H)$ and $(G, \langle\cdot,\cdot\rangle_G, \|\cdot\|_G)$ be two real valued Hilbert spaces. Denote by $V_r := \mathcal{D}((-A)^r)$ the real valued Hilbert spaces of domains of fractional powers of the linear operator $(-A)$ defined below, equipped with the norm $\|v\|_{V_r} := \|(-A)^r v\|_H$ for all $v \in V_r$ and all $r \in [0,\infty)$, where we use the abbreviation $V := V_{1/2}$. In order to specify the stochastic framework, we need the following definition for infinite-dimensional Brownian motions.

Definition 6.1. Q-Wiener Process
A stochastic process $\{W_t\}_{t\geq 0}$ on a probability space $(\Omega, \mathcal{F}, P)$ with values in a Hilbert space U is called a Q-Wiener process, if

- $W(0) = 0$,
- *W has P-a.s. continuous sample paths,*
- *the increments of W are independent and have the following Gaussian distribution*

$$W_t - W_s \sim \mathcal{N}(0, (t-s)Q)$$

for all $0 \leq s \leq t$.

Assumption 6.2. Stochastic Framework
Let $Q : G \to G$ be a trace class operator, i.e.

$$Tr(Q) = \sum_k \langle Qe_k, e_k\rangle < \infty$$

for all orthonormal bases of G. Let W_t be a standard Q-Wiener process with respect to the filtration $(\mathcal{F}_t)_{t\in[0,T]}$. For a finite or countable set $\mathcal{J}$ let $(g_j)_{j\mathcal{J}} \subset G$ be a orthonormal basis of eigenfunctions of Q with the corresponding eigenvalues $(\mu_j)_{j\mathcal{J}} \subset [0,\infty)$. Such a basis exists since Q is a trace class operator, see Prop. 2.1.5 in [160]. Then we have for all $u \in G$

$$Qu = \sum_{j\in\mathcal{J}} \mu_j \langle g_j, u\rangle_G \; g_j.$$

Denote by $(G_0, \langle\cdot,\cdot\rangle_{G_0}, \|\cdot\|_{G_0})$ the separable real valued Hilbert space $G_0 := Q^{\frac{1}{2}}(G)$ with $\langle u, v\rangle_{G_0} = \langle Q^{-1/2}u, Q^{-1/2}v\rangle_G$ for all $u, v \in G_0$ (see for instance [160, Ch. 2.3.2]). Here for an arbitrary linear operator $S \in L(G)$ we denote by $S^{-1} : im(S) \subset G \to G$ the pseudo inverse of S (see [160, Appx C]). Then we have $W_t : [0,T]\times\Omega \to G_0$ with

$$W_t(\omega) = \sum_{j\in\mathcal{J},\ \mu_j\neq 0} \langle g_j, W_t(\omega)\rangle_G \; g_j = \sum_{j\in\mathcal{J},\ \mu_j\neq 0} \sqrt{\mu_j} g_j \beta_t^j(\omega),$$

where $\beta^j : [0,T]\times\Omega \to \mathbb{R},\ j \in \mathcal{J}, \mu_j \neq 0$ is a family of independent standard Brownian motions.

Assumption 6.3. Linear Operator A
Let $\mathcal{I}$ be a finite or countable set and let $(\phi_i)_{i\mathcal{I}} \subset H$ be a orthonormal basis of H with the corresponding increasing sequence of eigenvalues $(\lambda_i)_{i\mathcal{I}} \subset (0,\infty)$ satisfying $\inf_{i\in\mathcal{I}} \lambda_i \in (0,\infty)$. Let $A : \mathcal{D}(A) \subset H \to H$ be a linear operator with

$$Av = \sum_{i\in\mathcal{I}} -\lambda_i \langle v, \phi_i\rangle_H \phi_i$$

for every $v \in \mathcal{D}(A)$ with $\mathcal{D}(A) = \{w \in H|\ \sum_{i\in\mathcal{I}} |\lambda_i|^2 |\langle v, \phi_i\rangle_H|^2 < \infty\}$.

Assumption 6.4. Drift Term F
Let $F : V \to H$ be a continuous mapping with

$$\forall u, v \in V \left\| (-A)^{-\vartheta}\Big(F(u) - F(v)\Big)\right\|_H \leq c \|u - v\|_H \qquad \text{(Lipschitz)}$$

$$\forall u \in V \|F(u)\|_H \leq c(1 + \|u\|_V) \qquad \text{(Bounded)}$$

for a $\vartheta \in [\frac{1}{2}, 1)$ and $c \geq 0$.

Assumption 6.5. Bilinear term B
Let the form $B : H \times H \to H$ be bilinear continuous with

$$|\langle B(u,v), v\rangle| = 0 \qquad \forall u \in V, v \in \mathcal{D}(-A) \qquad (6.2)$$

$$|\langle B(u,v), w\rangle| \leq c_0 \|u\|_H \|v\|_V^{\frac{1}{2}} \|Av\|_H^{\frac{1}{2}} \|w\|_V^{\frac{1}{2}} \|Aw\|_H^{\frac{1}{2}} \qquad \forall u \in H,\ v, w \in \mathcal{D}(-A) \qquad (6.3)$$

$$|\langle B(u,v), w\rangle| \leq c_0 \|u\|_V^{\frac{1}{2}} \|Au\|_H^{\frac{1}{2}} \|v\|_V^{\frac{1}{2}} \|Av\|_H^{\frac{1}{2}} \|w\|_H \qquad \forall w \in H,\ u, v \in \mathcal{D}(-A). \qquad (6.4)$$

Remark 6.6. PE and Equation (6.4)
Note that, in the context of PE, (6.3) *implies* (6.4) *via partial integration, essentially combining the strategies of proof used in Lemma 5.18 and 5.20. An independent proof for an advection operator on a smooth horizontal vector field with vanishing vertically integrated divergence is for instance given in [168].*

Assumption 6.7. Diffusion term σ
Let $\sigma : V \to H$ be a continuous mapping with

$$\forall u, v \in V \left\| (-A)^{-\vartheta/2}\Big(\sigma(u) - \sigma(v)\Big) \right\|_{HS(G_0,H)} \leq c \left\| u - v \right\|_H \qquad \text{(Lipschitz)}$$

$$\forall u \in V \left\| (-A)^{-\vartheta/2}\sigma(u)Q^{-\alpha} \right\|_{HS(G_0,H)} \leq c\left(1 + \left\| u \right\|_V\right) \qquad (Q\text{-bounded})$$

$$\forall u \in V \left\| \sigma(u) \right\|_{HS(G_0,V)} \leq c\left(1 + \left\| u \right\|_V\right) \qquad (V\text{-bounded})$$

for an $\alpha \in (0, \infty)$ and $\vartheta \in [\frac{1}{2}, 1), c \geq 0$ from Assumption 6.4. Furthermore let the assumptions of Theorem 5.16 be satisfied, i.e.

$$\sigma \in Bnd(H, L_2(G, H)) \cap Bnd(V, L_2(G, V)) \cap Bnd(\mathcal{D}(-A), L_2(G, \mathcal{D}(-A)))$$
$$\sigma \in Lip(H, L_2(G, H)) \cap Lip(V, L_2(G, V)) \cap Lip(\mathcal{D}(-A), L_2(G, \mathcal{D}(-A))).$$

Assumption 6.8. Initial Value
Let the initial condition ξ be a $\mathcal{F}_0$-measurable V-valued random variable such that

$$\xi \in L^2(\Omega, V).$$

Corollary 6.9. Existence of Local Mild Solutions for the PE
Let L be the solution of the linear deterministic system

$$\frac{d}{dt}L_t = AL_t, \quad L_0 = \xi_0. \tag{6.5}$$

Define the stopping time

$$\tau = \inf_{s>0}\left\{ \left\| U - L \right\|_V \geq \frac{1}{64c_0} \right\},$$

where c_0 is the constant occurring in Assumption 6.5 on the bilinear form. Then there exists an (up to modification) unique mild solution of (6.1) *satisfying*

$$U_t = e^{At}U_0 + \int_0^t e^{A(t-s)}\big(F(U_s) + B(U_s)\big)ds + \int_0^t e^{A(t-s)}\sigma(U_s)dW_s, \tag{6.6}$$

for all $t \in [0, \tau]$.

Proof. *Corollary 6.9*
The existence of local, unique, pathwise solution up to the stopping time τ is shown in [7], which we used in the previous section to obtain the existence of a solution for the PE. A well known result (see Theorem 6.5 in [161, p. 156]) states, that this implies the existence of a mild solution. □

Although a formal proof of this fact is given in [161] we would like to point out the idea behind mild solutions by informally using Itô's formula, which is not guaranteed to hold in this infinite-dimensional environment. To this end define the function $f(x,t) = xe^{-At}$. Then Itô's formula yields

$$\begin{aligned} df(U_t,t) &= \left(\frac{\partial}{\partial t}f(U_t,t) + \frac{\partial^2}{\partial x^2}f(U_t,t)\right)dt + \frac{\partial}{\partial x}f(U_t,t)dU_t \\ &= -AU_te^{-At}dt + e^{-At}\left(AU_t + B(U_t) + F(U_t)\right)dt + e^{-At}\sigma(U_t)dW_t \\ &= e^{-At}\left(B(U_t) + F(U_t)\right)dt + e^{-At}\sigma(U_t)dW_t \\ \Rightarrow \quad f(U_t,t) &= f(U_0,0) + \int_0^t e^{-As}\left(B(U_s) + F(U_s)\right)ds + \int_0^t e^{-As}\sigma(U_s)dW_s \\ \Rightarrow \quad e^{-At}U_t &= U_0 + \int_0^t e^{-As}\left(B(U_s) + F(U_s)\right)ds + \int_0^t e^{-As}\sigma(U_s)dW_s \end{aligned}$$

Mild solutions obviously allow us to substitute the linear additive term AU_t by multiplicative exponential terms e^{AU_t}. Although, at first sight, this seems to complicate matters, due to the more interwoven mathematical structure, the advantages of this formulation become obvious considering the strictly negative, unbounded eigenvalues of A. The unboundedness of eigenvalues in an infinite-dimensional environment makes it hard to control the growth of the term AU_t. The corresponding term e^{AU_t} for mild solutions is strictly bounded due to the elementary characteristics of the exponential function. We utilize this behavior in the technical Lemma 6.13 stated below.

6.2 Galerkin Approximation

In this section we formulate a finite-dimensional Galerkin approximation of the full solution (6.6). To this end let $(\mathcal{I}_N)_{N\in\mathbb{N}}$ and $(\mathcal{J}_K)_{K\in\mathbb{N}}$ be sequences of finite subsets of $\mathcal{I}$ and $\mathcal{J}$. Define the linear projection operator $P_N : H \to H$, $N \in \mathbb{N}$, by

$$P_N v := \sum_{i\in\mathcal{I}_N} \langle \phi_i, v\rangle_H \phi_i,$$

for all $v \in H$ and $N \in \mathbb{N}$. Furthermore define for $U \in V$ the projected operators

$$F^N(U) := P_N F(U), \quad B^N(U) := P_N B(U), \quad \sigma^N(U) := P_N \sigma(U)$$

and let the Wiener processes $W^K : [0,T] \times \Omega \to G_0$, $K \in \mathbb{N}$, be given by

$$W_t^K(\omega) := \sum_{j\in\mathcal{J}_K, \mu_j \le 0} \langle g_j, W_t(\omega)\rangle_G g_j, \tag{6.7}$$

for all $t \in [0,T]$, $\omega \in \Omega$ and all $K \in \mathbb{N}$. [7] yields the existence of an up to modifications unique mild solution of the finite-dimensional system

$$\begin{aligned} U_t^{N,K} = e^{At}P_N\xi &+ \int_0^t e^{A(t-s)}\big(F^N(U_s^{N,K}) + B^N(U_s^{N,K})\big)ds \\ &+ \int_0^t e^{A(t-s)}\sigma^N(U_s^{N,K})dW_s^K, \end{aligned} \tag{6.8}$$

P-a.s. for all $t \in [0,T]$ and $N, K \in \mathbb{N}$. We are now in a position to state the main result of this section.

Theorem 6.10. Convergence Rates of the Galerkin Approximation
Let Assumptions 6.2-6.8 be satisfied. Then there exists a constant $c < \infty$ independent of N, K, α, β with

$$E\left[\left\|U_t - V_t^{N,K}\right\|_H^2\right]^{\frac{1}{2}} \le e^{c\beta^3} g(\beta^{-1}, N)\left(\left(\inf_{i\in\mathcal{I}\setminus\mathcal{I}_N}\lambda_i\right)^{-\frac{1}{2}} + \left(\sup_{j\in\mathcal{J}\setminus\mathcal{J}_K}\mu_j\right)^{\alpha}\right),$$

where

$$g(\beta^{-1}, N) = \exp\left(\frac{1}{\beta}\left(1 - e^{-\lambda_1 t} + \sum_{j=2}^{N}\left(\frac{\lambda_{j-1}}{\lambda_j}\right)^{\frac{\lambda_{j-1}}{\lambda_j - \lambda_{j-1}}}\left(1 - \frac{\lambda_{j-1}}{\lambda_j}\right)\right)\right)$$

for all $N, K \in \mathbb{N}$ and all $\beta > 0$.

Remark 6.11. The Parameter β
Galerkin estimates for SPDE with Lipschitz coefficients usually have a similar structure, but satisfy

$$e^{c\beta^3} g(\beta^{-1}, N) < \tilde{c},$$

with a constant $\tilde{c} < \infty$ independent of N, K, α, β. In the present system this holds true only if the eigenvalues of A grow fast enough, i.e., the sum in $g(\beta^{-1}, N)$ converges for $N \to \infty$. Unfortunately the PE do not satisfy this assumption, which is discussed in Section 6.4.1. However, we may choose $\beta > 0$ freely and therefore achieve a convergence rating arbitrarily close to $(\inf_{i \in \mathcal{I} \setminus \mathcal{I}_N} \lambda_i)^{-\frac{1}{2}}$ in the spacial-dimension. This variable error term arises due to the ill-behaved bilinear advection term, which leads to an integral of the form

$$\int_0^t \left\| -Ae^{As} \right\|_{L(H)}^{\frac{3}{4}} \|AU_s\|_H^{\frac{1}{2}} \, ds$$

in the proof of Theorem 6.10. It turns out that $\|AU_s\|_H^2$ is P-a.s. integrable, see Lemma 6.18. We apply Hölder's inequality:

$$\int_0^t \left\| -Ae^{As} \right\|_{L(H)}^{\frac{3}{4}} \|AU_s\|_H^{\frac{1}{2}} \, ds \leq \left(\int_0^t \left\| -Ae^{As} \right\|_{L(H)} ds \right)^{\frac{3}{4}} \left(\int_0^t \|AU_s\|_H^2 \, ds \right)^{\frac{1}{4}}.$$

Furthermore, based on estimates of the exponential function (see Lemma 6.18), we have

$$\left\| -Ae^{As} \right\|_{L(H)} \leq \frac{1}{s}.$$

This approximation would lead to a divergent integral. Therefore, we have to use a more delicate estimate for the operator norm. In fact we derive the function $g(\beta^{-1}, N)$ as the exact solution of the problematic integral. Introducing the parameter β allows us to minimize the undesired contribution of the sum of eigenvalues in g on the convergence rate at the cost of a greater constant $e^{c\beta^3}$. In particular this yields convergence in the spatial-dimensions if the sum of eigenvectors in g is bound by a polynomial of the maximal eigenvalue λ_N. It is a typical side effect of Gronwall-type arguments, which are applied in the upcoming proof, to yield very pessimistic constants. Since the present case is no exception, the importance of the estimated value of the resulting constant $e^{c\beta^3}$ should not be overrated. To summarize: Theorem 6.10 guarantees the convergence of a Galerkin approximation, but concrete numerical experiments are needed for reasonable statements regarding the actual error for a specific finite number of dimensions N.

6.3 Technical Lemmata

For the upcoming proof of Theorem 6.10 we have to make some technical preparations.

Corollary 6.12. Anticommutativity of the Trilinear Form
For all $u \in H$, $v, w \in \mathcal{D}(-A)$ we have

$$\langle B(u, v), w \rangle = -\langle B(u, w), v \rangle$$

Proof. *Corollary 6.12*

$$\begin{aligned}
\langle B(u, v), w \rangle &= \langle B(u, w), w \rangle + \langle B(u, v - w), w \rangle \\
&= \langle \langle B(u, w), w \rangle + B(u, v - w), w - v \rangle + \langle B(u, v - w), v \rangle \\
&= \underbrace{\langle B(u, w), w \rangle - \langle B(u, v - w), v - w \rangle + \langle B(u, v), v \rangle}_{=0 \text{ due to (6.2)}} - \langle B(u, w), v \rangle
\end{aligned}$$

□

Lemma 6.13. Estimates on the Propagator
For $r \in [0,1)$ and all $t > 0$

$$\begin{aligned}\left\|(-A)^r e^{At}\right\|_{L(H)} &\leq t^{-r}\\ \left\|(-A)^{-r}\left(1-e^{At}\right)\right\|_{L(H)} &\leq t^{r}\end{aligned}$$

Proof. *Lemma 6.13*
Since A is a diagonal matrix with negative eigenvalues, basic exponential inequalities yield

$$\begin{aligned}\left\|(-A)^r e^{At}\right\|_{L(H)} &= t^{-r}\left\|(-At)^r e^{At}\right\|_{L(H)}\\ &\leq t^{-r}\sup_{x>0} x^r e^{-x} \leq t^{-r}\\ \left\|(-A)^{-r}\left(1-e^{At}\right)\right\|_{L(H)} &= t^{r}\left\|(-At)^{-r}\left(1-e^{At}\right)\right\|_{L(H)}\\ &\leq t^{r}\sup_{x>0} x^{-r}(1-e^{-x}) \leq t^{r}.\end{aligned}$$

□

Lemma 6.14. A "Mild" Itô Formula
Let $X : [t_0,T]\times\Omega \to V_M$ be a stochastic-process in a M-dimensional setting, $M \in \mathbb{N}$, with

$$X_t^M = e^{A(t-t_0)}X_{t_0}^M + \int_{t_0}^t e^{A(t-s)}\left(F^M(X_s^M)+B^M(X_s^M)\right)ds + \int_{t_0}^t e^{A(t-s)}\sigma^M(X_s^M)dW_s^M$$

P-a.s. for all $t \in [t_0,T]$. Then, for a twice continuously differentiable function $\varphi : V_M \to H_M$, we have

$$\begin{aligned}\varphi(X_t^M) = \varphi(e^{A(t-t_0)}X_{t_0}^M) &+ \int_{t_0}^t \varphi'(e^{A(t-s)}X_s^M)e^{A(t-s)}\left(F^M(X_s^M)+B^M(X_s^M)\right)ds\\ &+ \int_{t_0}^t \varphi'(e^{A(t-s)}X_s^M)e^{A(t-s)}\sigma^M(X_s^M)dW_s^M\\ &+ \frac{1}{2}\sum_{j\in\mathcal{J}_M}\int_{t_0}^t \varphi''(e^{A(t-s)}X_s^M)\langle e^{A(t-s)}\sigma(X_s^M)g_j, e^{A(t-s)}\sigma(X_s^M)g_j\rangle ds.\end{aligned}$$

A more general version of this lemma in an infinite-dimensional setting is given in [169]. Unfortunately the bilinear term in the present system does not satisfy the Lipschitz condition required in [169]. Therefore we restrict the statement on finite-dimensional subspaces to ensure the well posedness of all occurring integrals and give a brief outline of the proof given in [169] to present the main idea.

Proof. *Lemma: 6.14*
The central difficulty is the fact, that the solution process X^M is, in general, no semimartingale since the integrands depend on $e^{A(t-s)}$ and therefore on t. As a consequence we cannot apply Itô's formula in a direct manner. In order to circumvent this issue, we fix the time variable t and define a family of stochastic processes $Y^t : [t_0,t]\times\Omega \to V_M$, $t \in [t_0,T]$, by

$$Y_u^t = e^{A(t-t_0)}X_{t_0}^M + \int_{t_0}^u e^{A(t-s)}\left(F^M(X_s^M)+B^M(X_s^M)\right)ds + \int_{t_0}^u e^{A(t-s)}\sigma^M(X_s^M)dW_s^M.$$

for all $u \in [t_0, t]$ P-a.s. and all $t \in [t_0, T]$. Using Itô's formula and noting that $Y_t^t = X_t$ and $Y_{t_0}^t = e^{A(t-t_0)}X_t$ P-a.s. for all $t \in [t_0, T]$ then yields

$$\begin{aligned}\varphi(Y_u^t) = \varphi(Y_{t_0}^t) &+ \int_{t_0}^u \varphi'(Y_s^t)e^{A(t-s)}\big(F^M(X_s^M) + B^M(X_s^M)\big)ds \\ &+ \int_{t_0}^u \varphi'(Y_s^t)e^{A(t-s)}\sigma^M(X_s^M)dW_s^M \\ &+ \frac{1}{2}\sum_{j\in\mathcal{J}_M}\int_{t_0}^u \varphi''(Y_s^t)\langle e^{A(t-s)}\sigma(X_s^M)g_j, e^{A(t-s)}\sigma(X_s^M)g_j\rangle ds.\end{aligned}$$

Furthermore a straightforward calculation gives us

$$Y_s^t = e^{A(t-s)}X_s,$$

for all $s \in [t_0, T]$ P-a.s. and all $t \in [t_0, T]$, which directly leads to Lemma 6.14. For a rigorous proof we refer to the more general case in [169]. □

Corollary 6.15. Squared Norm and Itô
Using the setting of Lemma 6.14 we have for the squared norm of H_M

$$\begin{aligned}\left\|X_t^M\right\|_H^2 = \left\|e^{A(t-t_0)}X_{t_0}^M\right\|_H^2 &+ 2\int_{t_0}^t \langle e^{A(t-s)}X_s^M, e^{A(t-s)}\big(F^M(X_s^M) + B^M(X_s^M)\big)\rangle ds \\ &+ 2\int_{t_0}^t \langle e^{A(t-s)}X_s^M, e^{A(t-s)}\sigma^M(X_s^M)\rangle dW_s^M \\ &+ \int_{t_0}^t \left\|e^{A(t-s)}\sigma(X_s^M)\right\|_{HS(G_0,H)} ds\end{aligned}$$

Lemma 6.16. A Stochastic Gronwall Lemma
Let Ψ, Φ, χ and G be real-valued stochastic processes satisfying

$$\Psi_t \leq \Psi_0 + \int_0^t \Phi_s\Psi_s ds + \int_0^t \chi_s ds + \int_0^t G_s dW_s,$$

P-a.s.. Then we have

$$E\left[e^{-\int_0^t \Phi_s ds}\Psi_t\right] \leq E\left[\Psi_0\right] + E\left[\int_0^t \chi_s e^{-\int_0^t \Phi_u du} ds\right].$$

Proof. *Lemma: 6.16*
For a convenient notation we define

$$h_t := e^{-\int_0^t \Phi_s ds}.$$

Then Itô's lemma and a differential formulation of the assumption yield

$$\begin{aligned}d(h_t\Psi_t) &= \Psi_t dh_t + h_t d\Psi_t \\ &= -\Phi_t\Psi_t h_t dt + h_t d\Psi_t \\ &\leq -\Phi_t\Psi_t h_t dt + h_t\Phi_t\Psi_t dt + h_t\chi_t dt + h_t G_t dW_t \\ &= h_t\chi_t dt + h_t G_t dW_t \\ \Rightarrow \qquad h_t\Psi_t &\leq h_0\Psi_0 + \int_0^t \chi_s h_s ds + \int_0^t G_s h_s dW_s\end{aligned}$$

Since h_0 equals one and the stochastic integral is a martingale, taking the expectation value gives us

$$E\left[h_t\Psi_t\right] \leq E\left[\Psi_0\right] + E\left[\int_0^t \chi_s h_s ds\right].$$

□

Corollary 6.17. A version of Young's inequality
Let $a, b > 0$ and $1/p + 1/q = 1$. Then, for all $\beta > 0$

$$ab \leq \frac{a^p}{\beta} + \left(\frac{\beta}{p}\right)^{q/p} \frac{b^q}{q}.$$

Proof. *Corollary: 6.17*
Using the classical version of Young's inequality gives us

$$\begin{aligned} ab &= a\left(\frac{p}{\beta}\right)^{1/p}\left(\frac{\beta}{p}\right)^{1/p} b \\ &\leq \left(a\left(\frac{p}{\beta}\right)^{1/p}\right)^p \frac{1}{p} + \left(\left(\frac{\beta}{p}\right)^{1/p} b\right)^q \frac{1}{q} \\ &= \frac{a^p}{\beta} + \left(\frac{\beta}{p}\right)^{q/p} \frac{b^q}{q}. \end{aligned}$$

□

Lemma 6.18. P-a.s. Bounds on Finite-Dimensional Solutions
Let $V_t^{N,K}$ be a solution of the finite-dimensional mild system (6.8)*. Then there exists a constant $k > 0$ with*

$$\sup_{s\in[0,t]} \left\|V_s^{N,K}\right\|_V^4 + \int_0^t \left\|V_s^{N,K}\right\|_V^2 \left\|AV_s^{N,K}\right\|_H^2 ds \leq k$$

P-a.s. for all $t \in [0,T]$ and all $N, M, K \in \mathbb{N}$.

Proof. *Lemma 6.18*
Since this Lemma is one of the central statements of the work by Debussche et al. [7] with a rather arduous proof, we present the main ideas but omit the technical details. As a first step we introduce a smooth, Lipschitz continuous cutoff-function $\Theta^M : V \to [0,\infty)$ with

$$\Theta^M(u) := \begin{cases} 1, & \left\|P_M(u - L_s^M)\right\|_V \leq \frac{1}{64c_0} \\ 0, & \left\|P_M(u - L_s^M)\right\|_V \geq \frac{2}{64c_0} \end{cases}$$

for all $u \in V$ and $M \in \mathbb{N}$. The constant c_0 is defined through the estimates on the bilinear form in Assumption 6.5. Following the proof of Proposition 5.1 in [7, p. 16] the local mild solution $V_t^{N,K}$ of the finite-dimensional setting (6.8) is a global solutions of the "*tamed*" system

$$\begin{aligned} V_t^{N,K} = e^{At}P_N\xi &+ \int_0^t e^{A(t-s)}\left(F^N(V_s^{N,K}) + \Theta^N(V_s^{N,K})B^N(V_s^{N,K})\right) ds \\ &+ \int_0^t e^{A(t-s)}\sigma^N(V_s^{N,K}) dW_s^K \end{aligned} \tag{6.9}$$

Note that, due to the characteristics of the stopping time τ in Corollary 6.9 and by definition of Θ^M, the solutions of the tamed systems are identical to the corresponding finite-dimensional solutions of (6.8) for all $t \in [0,T]$. Define a sequence of stopping times

$$\tau_n := \inf_{t\geq 0}\left\{\sup_{s\in[0,t]} \left\|V_s^{N,K}\right\|_V^4 + \int_0^t \left\|V_s^{N,K}\right\|_V^2 \left\|AV_s^{N,K}\right\|_H^2 \geq n\right\},$$

with $n \in \mathbb{N}$. This is an increasing sequence satisfying $\lim_{n\to\infty} \tau_n = \infty$ since $V_t^{N,K}$ is a global solutions of (6.9), see Lemma 3.1 in [7]. Therefore there is a $n_0 \in \mathbb{N}$ with $\tau < \tau_{n_0}$ and

$$\sup_{s\in[0,t]} \left\|V_s^{N,K}\right\|_V^4 + \int_0^t \left\|V_s^{N,K}\right\|_V^2 \left\|AV_s^{N,K}\right\|_H^2 + \left\|U_s^{M,M}\right\|_V^2 \left\|AU_s^{M,M}\right\|_H^2 ds \leq n_0$$

P-a.s. for all $t \in [0,T]$. □

6.4 Proof of Theorem 6.10

Due to the ill-behaved bilinear term B, we cannot use the standard approach to derive convergence rates for systems with smooth coefficients, i.e., Taylor expansions and direct mean square estimates for each operator. We aim to use Itô's formula to exploit the assumptions on B, which requires a finite-dimensional setting, see Lemma 6.14. Therefore we project the full system on a finite-dimensional subspace and apply a result by Debussche et al [7] used in the proof of existence of solution, yielding L^2-convergence of these projections. Using Itô's formula on the projected discretization error allows us to derive uniform estimates, i.e., error estimate that do not depend on the dimensions of the projection. Finally, the Gronwall Lemma 6.16 and some elaborate technical calculations complete the proof. Indeed, as a result of [7, Ch. 7.2] we have

$$U^{M,M} \to U \text{ for } M \to \infty \text{ P-a.s. in } L^2([0,T],V).$$

Therefore we can use uniform estimates of the finite-dimensional error $\|U_t^{M,M} - V_t^{N,K}\|_H^2$ in order to proof the theorem. We have

$$\begin{aligned}
&\left\|U_t - V_t^{N,K}\right\|_H^2 \\
&\quad\le \sup_{M>N} \left\|U_t^{M,M} - V_t^{N,K}\right\|_H^2 \\
&\quad\le \sup_{M>N} \left(2\left\|(1-P_N)U_t^{M,M}\right\|_H^2 + 2\left\|P_N U_t^{M,M} - V_t^{N,K}\right\|_H^2\right) \\
&\quad\le \sup_{M>N} \left(2\left\|(-A)^{\frac{1}{2}}U_t^{M,M}\right\|_H^2 \left\|(-A)^{-1/2}(1-P_N)\right\|_{L(H)}^2 + 2\left\|P_N U_t^{M,M} - V_t^{N,K}\right\|_H^2\right) \\
&\quad\le \sup_{M>N} \left(2\left\|U_t^{M,M}\right\|_V^2 \left(\inf_{i\in\mathcal{I}\setminus\mathcal{I}_N}\lambda_i\right)^{-1} + 2\left\|P_N U_t^{M,M} - V_t^{N,K}\right\|_H^2\right) \\
&\quad\le \sup_{M>N} \left(2\sup_{t\in[0,T]}\left\|U_t^{M,M}\right\|_V^2 \left(\inf_{i\in\mathcal{I}\setminus\mathcal{I}_N}\lambda_i\right)^{-1} + 2\left\|P_N U_t^{M,M} - V_t^{N,K}\right\|_H^2\right). \qquad (6.10)
\end{aligned}$$

The second term can be estimated using the mild Itô formula for the squared H-norm. Corollary 6.15 yields for an arbitrary $M > N$

$$\left\|P_N U_t^{M,M} - V_t^{N,K}\right\|_H^2 = (U_0) + (F) + (B) + (W^K) + (W^{M-K}) + (\text{Itô K}) + (\text{Itô M-K}) \qquad (6.11)$$

with

$$\begin{aligned}
(U_0) &= \left\|e^{At}P_N\big(U_0^{M,M} - V_0^{N,K}\big)\right\|_H^2 \\
(F) &= 2\int_0^t \left\langle e^{A(t-s)}P_N(U_s^{M,M} - V_s^{N,K}), e^{A(t-s)}P_N\Big(F^M(U_s^{M,M}) - F^N(V_s^{N,K})\Big)\right\rangle ds \\
(B) &= 2\int_0^t \left\langle e^{A(t-s)}P_N(U_s^{M,M} - V_s^{N,K}), e^{A(t-s)}P_N\Big(B^M(U_s^{M,M}) - B^N(V_s^{N,K})\Big)\right\rangle ds \\
(W^K) &= 2\int_0^t \left\langle e^{A(t-s)}P_N(U_s^{M,M} - V_s^{N,K}), e^{A(t-s)}P_N\Big(\sigma^M(U_s^{M,M}) - \sigma^N(V_s^{N,K})\Big)\right\rangle dW_s^K \\
(W^{M-K}) &= 2\int_0^t \left\langle e^{A(t-s)}P_N(U_s^{M,M} - V_s^{N,K}), e^{A(t-s)}P_N\sigma^M(U_s^{M,M})\right\rangle d(W_s^M - W_s^K) \\
(\text{Itô K}) &= \sum_{j\in J_K,\ \mu_j\neq 0} \mu_j \int_0^t \left\|e^{A(t-s)}P_N\Big(\sigma^M(U_s^{M,M}) - \sigma^N(V_s^{N,K})\Big)g_j\right\|_H^2 ds \\
(\text{Itô M-K}) &= \sum_{j\in J_M\setminus J_K,\ \mu_j\neq 0} \mu_j \int_0^t \left\|e^{A(t-s)}P_N\sigma^M(U_s^{M,M})g_j\right\|_H^2 ds
\end{aligned}$$

We aim to use the stochastic Gronwall Lemma 6.16. To this end we successively estimate the terms above. Regarding the initial values (U_0) we have

$$\begin{aligned}(U_0) &= \left\| e^{At} P_N \big(U_0^{M,M} - V_0^{N,K}\big) \right\|_H^2 \\ &= \left\| e^{At} P_N \big(P_M \xi - P_N \xi\big) \right\|_H^2 \\ &= \left\| e^{At} \big(P_N \xi - P_N \xi\big) \right\|_H^2 = 0 \end{aligned} \tag{6.12}$$

The nonlinear drift term (F) can be estimated using $\|P_N\|_{L(H)} \leq 1$ and $\left\|e^{A(t-s)}\right\|_{L(H)} \leq 1$ for all $N \in \mathbb{N}$ and $0 \leq s \leq t$.

$$\begin{aligned}\frac{(F)}{2} &= \int_0^t \left\langle e^{A(t-s)} P_N (U_s^{M,M} - V_s^{N,K}), e^{A(t-s)} P_N \Big(F^M(U_s^{M,M}) - F^N(V_s^{N,K}) \Big) \right\rangle ds \\ &\leq \int_0^t \left\| e^{A(t-s)} P_N (U_s^{M,M} - V_s^{N,K}) \right\|_H \left\| e^{A(t-s)} P_N \Big(F^M(U_s^{M,M}) - F^N(V_s^{N,K}) \Big) \right\|_H ds \\ &\leq \int_0^t \left\| P_N (U_s^{M,M} - V_s^{N,K}) \right\|_H \left\| e^{A(t-s)} \Big(F(U_s^{M,M}) - F(V_s^{N,K}) \Big) \right\|_H ds \\ &\leq \int_0^t \left\| P_N (U_s^{M,M} - V_s^{N,K}) \right\|_H \left\| (-A)^{\vartheta} e^{A(t-s)} \right\|_{L(H)} \\ &\qquad \times \left\| (-A)^{-\vartheta} \Big(F(U_s^{M,M}) - F(V_s^{N,K}) \Big) \right\|_H ds \end{aligned}$$

Lemma 6.13, Assumption 6.4 and the basic inequality $ab \leq \frac{a^2}{2} + \frac{b^2}{2}$ yield

$$\begin{aligned}\frac{(F)}{2} &\leq \int_0^t c(t-s)^{-\vartheta} \left\| P_N (U_s^{M,M} - V_s^{N,K}) \right\|_H \left\| U_s^{M,M} - V_s^{N,K} \right\|_H ds \\ &\leq \int_0^t c(t-s)^{-\vartheta} \left\| P_N (U_s^{M,M} - V_s^{N,K}) \right\|_H \\ &\qquad \times \Big(\left\| (1 - P_N) U_s^{M,M} \right\|_H + \left\| P_N (U_s^{M,M} - V_s^{N,K}) \right\|_H \Big) ds \\ &\leq \int_0^t c(t-s)^{-\vartheta} \Bigg[\left\| P_N (U_s^{M,M} - V_s^{N,K}) \right\|_H^2 \\ &\qquad + \left\| U_s^{M,M} \right\|_V \Big(\inf_{i \in \mathcal{I} \setminus \mathcal{I}_N} \lambda_i \Big)^{-1/2} \left\| P_N (U_s^{M,M} - V_s^{N,K}) \right\|_H \Bigg] ds \\ &\leq \int_0^t c(t-s)^{-\vartheta} \Big(\frac{1}{2} \left\| U_s^{M,M} \right\|_V^2 \Big(\inf_{i \in \mathcal{I} \setminus \mathcal{I}_N} \lambda_i \Big)^{-1} + \frac{3}{2} \left\| P_N (U_s^{M,M} - V_s^{N,K}) \right\|_H^2 \Big) ds \\ &\leq \int_0^t c(t-s)^{-\vartheta} \Bigg[2 \sup_{t \in [0,T]} \left\| U_t^{M,M} \right\|_V^2 \Big(\inf_{i \in \mathcal{I} \setminus \mathcal{I}_N} \lambda_i \Big)^{-1} \\ &\qquad + 2 \left\| P_N (U_s^{M,M} - V_s^{N,K}) \right\|_H^2 \Bigg] ds \end{aligned} \tag{6.13}$$

For the bilinear term (B) we obtain

$$\begin{aligned}\frac{(B)}{2} &= \int_0^t \left\langle e^{A(t-s)} P_N (U_s^{M,M} - V_s^{N,K}), e^{A(t-s)} P_N \Big(B^M(U_s^{M,M}) - B^N(V_s^{N,K}) \Big) \right\rangle ds \\ &= \int_0^t \left\langle e^{A(t-s)} P_N (U_s^{M,M} - V_s^{N,K}), e^{A(t-s)} P_N \Big(B(U_s^{M,M}) - B(V_s^{N,K}) \Big) \right\rangle ds \\ &= \int_0^t \left\langle e^{A(t-s)} P_N (U_s^{M,M} - V_s^{N,K}), e^{A(t-s)} P_N B\big(U_s^{M,M} - V_s^{N,K}, U_s^{M,M}\big) \right\rangle ds \\ &\qquad + \int_0^t \left\langle e^{A(t-s)} P_N (U_s^{M,M} - V_s^{N,K}), e^{A(t-s)} P_N B\big(V_s^{N,K}, U_s^{M,M} - V_s^{N,K}\big) \right\rangle ds \\ &= \int_0^t \left\langle e^{2A(t-s)} P_N (U_s^{M,M} - V_s^{N,K}), B\big(U_s^{M,M} - V_s^{N,K}, U_s^{M,M}\big) \right\rangle ds \\ &\qquad + \int_0^t \left\langle e^{2A(t-s)} P_N (U_s^{M,M} - V_s^{N,K}), B\big(V_s^{N,K}, U_s^{M,M} - V_s^{N,K}\big) \right\rangle ds \end{aligned}$$

Corollary 6.12 and Assumption 6.5 yield

$$\begin{aligned}
\frac{(B)}{2} &= \int_0^t \Big\langle e^{2A(t-s)}P_N(U_s^{M,M}-V_s^{N,K}), B\big(U_s^{M,M}-V_s^{N,K},U_s^{M,M}\big)\Big\rangle \\
&\qquad + \Big\langle U_s^{M,M}-V_s^{N,K}, B\big(V_s^{N,K}, e^{2A(t-s)}P_N(U_s^{M,M}-V_s^{N,K})\big)\Big\rangle\, ds \\
&\le \int_0^t c_0 \left\|e^{2A(t-s)}P_N(U_s^{M,M}-V_s^{N,K})\right\|_V^{\frac12} \left\|Ae^{2A(t-s)}P_N(U_s^{M,M}-V_s^{N,K})\right\|_H^{\frac12} \\
&\qquad\qquad \times \left\|U_s^{M,M}-V_s^{N,K}\right\|_H \left\|U_s^{M,M}\right\|_V^{\frac12} \left\|AU_s^{M,M}\right\|_H^{\frac12} \\
&\qquad + c_0 \left\|U_s^{M,M}-V_s^{N,K}\right\|_H \left\|V_s^{N,K}\right\|_V^{\frac12} \left\|AV_s^{N,K}\right\|_H^{\frac12} \\
&\qquad\qquad \times \left\|e^{2A(t-s)}P_N(U_s^{M,M}-V_s^{N,K})\right\|_V^{\frac12} \left\|Ae^{2A(t-s)}P_N(U_s^{M,M}-V_s^{N,K})\right\|_H^{\frac12} ds \\
&\le \int_0^t c_0 \Big(\left\|V_s^{N,K}\right\|_V^{\frac12} \left\|AV_s^{N,K}\right\|_H^{\frac12} + \left\|U_s^{M,M}\right\|_V^{\frac12} \left\|AU_s^{M,M}\right\|_H^{\frac12} \Big) \left\|U_s^{M,M}-V_s^{N,K}\right\|_H \\
&\qquad\qquad \times \left\|e^{2A(t-s)}P_N(U_s^{M,M}-V_s^{N,K})\right\|_V^{\frac12} \left\|Ae^{2A(t-s)}P_N(U_s^{M,M}-V_s^{N,K})\right\|_H^{\frac12} ds
\end{aligned}$$

Since $\|u\|_V = \langle Au,u\rangle^{\frac12} \le \|Au\|_H^{\frac12}\|u\|_H^{\frac12}$, for all $u\in V$, we have

$$\begin{aligned}
&\left\|e^{2A(t-s)}P_N(U_s^{M,M}-V_s^{N,K})\right\|_V^{\frac12} \left\|Ae^{2A(t-s)}P_N(U_s^{M,M}-V_s^{N,K})\right\|_H^{\frac12} \\
&\quad\le \left\|e^{2A(t-s)}P_N(U_s^{M,M}-V_s^{N,K})\right\|_H^{\frac14} \left\|Ae^{2A(t-s)}P_N(U_s^{M,M}-V_s^{N,K})\right\|_H^{3/4} \\
&\quad\le \left\|P_N(U_s^{M,M}-V_s^{N,K})\right\|_H^{\frac14} \left\|Ae^{2A(t-s)}P_N\right\|_{L(H)}^{3/4} \left\|P_N(U_s^{M,M}-V_s^{N,K})\right\|_H^{3/4} \\
&\quad= \left\|(-A)e^{2A(t-s)}P_N\right\|_{L(H)}^{3/4} \left\|P_N(U_s^{M,M}-V_s^{N,K})\right\|_H
\end{aligned}$$

As before, we use

$$\begin{aligned}
\left\|U_s^{M,M}-V_s^{N,K}\right\|_H &\le \left\|(1-P_N)U_s^{M,M}\right\|_H + \left\|P_N(U_s^{M,M}-V_s^{N,K})\right\|_H \\
&\le \left\|U_s^{M,M}\right\|_V \Big(\inf_{i\in\mathcal{I}\setminus\mathcal{I}_N}\lambda_i\Big)^{-1/2} + \left\|P_N(U_s^{M,M}-V_s^{N,K})\right\|_H,
\end{aligned}$$

and $ab \le \frac{a^2}{2}+\frac{b^2}{2}$, which leads to

$$\begin{aligned}
\frac{(B)}{2} &\le \int_0^t c_0 \Big(\left\|V_s^{N,K}\right\|_V^{\frac12} \left\|AV_s^{N,K}\right\|_H^{\frac12} + \left\|U_s^{M,M}\right\|_V^{\frac12} \left\|AU_s^{M,M}\right\|_H^{\frac12} \Big) \\
&\qquad \times \left\|(-A)e^{2A(t-s)}P_N\right\|_{L(H)}^{\frac34} \left\|P_N(U_s^{M,M}-V_s^{N,K})\right\|_H \\
&\qquad \times \Big(\left\|U_s^{M,M}\right\|_V \Big(\inf_{i\in\mathcal{I}\setminus\mathcal{I}_N}\lambda_i\Big)^{-1/2} + \left\|P_N(U_s^{M,M}-V_s^{N,K})\right\|_H \Big) ds \\
&\le \int_0^t c_0 \Big(\left\|V_s^{N,K}\right\|_V^{\frac12} \left\|AV_s^{N,K}\right\|_H^{\frac12} + \left\|U_s^{M,M}\right\|_V^{\frac12} \left\|AU_s^{M,M}\right\|_H^{\frac12} \Big) \left\|(-A)e^{2A(t-s)}P_N\right\|_{L(H)}^{\frac34} \\
&\qquad \times \Big(\frac12 \left\|U_s^{M,M}\right\|_V^2 \Big(\inf_{i\in\mathcal{I}\setminus\mathcal{I}_N}\lambda_i\Big)^{-1} + \frac32 \left\|P_N(U_s^{M,M}-V_s^{N,K})\right\|_H^2 \Big) ds \\
&\le \int_0^t c_0 \Big(\left\|V_s^{N,K}\right\|_V^{\frac12} \left\|AV_s^{N,K}\right\|_H^{\frac12} + \left\|U_s^{M,M}\right\|_V^{\frac12} \left\|AU_s^{M,M}\right\|_H^{\frac12} \Big) \left\|(-A)e^{2A(t-s)}P_N\right\|_{L(H)}^{\frac34} \\
&\qquad \times \Big(2\sup_{t\in[0,T]} \left\|U_t^{M,M}\right\|_V^2 \Big(\inf_{i\in\mathcal{I}\setminus\mathcal{I}_N}\lambda_i\Big)^{-1} + 2\left\|P_N(U_s^{M,M}-V_s^{N,K})\right\|_H^2 \Big) ds \qquad (6.14)
\end{aligned}$$

Since the stochastic integrals (W^K) and (W^{M-K}) vanish when taking the expectation value for the L^2-error, we draw our attention to the Itô terms by using Assumption 6.7.

$$
\begin{aligned}
(\text{Itô K}) &= \sum_{j\in J_K,\ \mu_j\neq 0} \mu_j \int_0^t \left\| e^{A(t-s)} P_N \Big(\sigma^M(U_s^{M,M}) - \sigma^N(V_s^{N,K}) \Big) g_j \right\|_H^2 ds \\
&\leq \sum_{j\in J,\ \mu_j\neq 0} \mu_j \int_0^t \left\| e^{A(t-s)} P_N \Big(\sigma^M(U_s^{M,M}) - \sigma^N(V_s^{N,K}) \Big) g_j \right\|_H^2 ds \\
&= \int_0^t \left\| e^{A(t-s)} P_N \Big(\sigma^M(U_s^{M,M}) - \sigma^N(V_s^{N,K}) \Big) \right\|_{HS(G_0,H)}^2 ds \\
&\leq \int_0^t c \left\| e^{A(t-s)} (-A)^{\vartheta/2} \right\|_{L(H)}^2 \left\| (-A)^{-\vartheta/2} \Big(\sigma^M(U_s^{M,M}) - \sigma^N(V_s^{N,K}) \Big) \right\|_{HS(G_0,H)}^2 ds \\
&\leq \int_0^t c(t-s)^{-\vartheta} \left\| U_s^{M,M} - V_s^{N,K} \right\|_H^2 ds \\
&\leq \int_0^t c(t-s)^{-\vartheta} \Big(2 \left\| U_s^{M,M} \right\|_V^2 \Big(\inf_{i\in\mathcal{I}\setminus\mathcal{I}_N} \lambda_i \Big)^{-1} + 2 \left\| P_N(U_s^{M,M} - V_s^{N,K}) \right\|_H^2 \Big) ds \\
&\leq \int_0^t c(t-s)^{-\vartheta} \Big(2 \sup_{t\in[0,T]} \left\| U_t^{M,M} \right\|_V^2 \Big(\inf_{i\in\mathcal{I}\setminus\mathcal{I}_N} \lambda_i \Big)^{-1} + 2 \left\| P_N(U_s^{M,M} - V_s^{N,K}) \right\|_H^2 \Big) ds
\end{aligned}
\tag{6.15}
$$

The second Itô term deals with the error due to the reduction of stochastic dimensions.

$$
\begin{aligned}
(\text{Itô M-K}) &= \sum_{\substack{j\in J_M\setminus J_K \\ \mu_j\neq 0}} \mu_j \int_0^t \left\| e^{A(t-s)} P_N \sigma^M(U_s^{M,M}) g_j \right\|_H^2 ds \\
&\leq \sum_{\substack{j\in\mathcal{J}\setminus\mathcal{J}_K \\ \mu_j\neq 0}} \mu_j \int_0^t \left\| e^{A(t-s)} P_N \sigma^M(U_s^{M,M}) g_j \right\|_H^2 ds \\
&= \sum_{\substack{j\in\mathcal{J}\setminus\mathcal{J}_K \\ \mu_j\neq 0}} \mu_j \int_0^t \left\| e^{A(t-s)} P_N \sigma^M(U_s^{M,M}) Q^{-\alpha} Q^{\alpha} g_j \right\|_H^2 ds \\
&= \sum_{\substack{j\in\mathcal{J}\setminus\mathcal{J}_K \\ \mu_j\neq 0}} \mu_j \int_0^t \left\| e^{A(t-s)} P_N \sigma^M(U_s^{M,M}) Q^{-\alpha} \mu_j^{\alpha} g_j \right\|_H^2 ds \\
&\leq \sup_{j\in\mathcal{J}\setminus\mathcal{J}_K} \mu_j^{2\alpha} \sum_{\substack{j\in\mathcal{J}\setminus\mathcal{J}_K \\ \mu_j\neq 0}} \mu_j \int_0^t \left\| e^{A(t-s)} (-A)^{\vartheta/2} \right\|_{L(H)}^2 \left\| (-A)^{-\vartheta/2} \sigma^M(U_s^{M,M}) Q^{-\alpha} g_j \right\|_H^2 ds \\
&\leq \sup_{j\in\mathcal{J}\setminus\mathcal{J}_K} \mu_j^{2\alpha} \sum_{j\in J,\ \mu_j\neq 0} \mu_j \int_0^t (t-s)^{-\vartheta} \left\| (-A)^{-\vartheta/2} \sigma^M(U_s^{M,M}) Q^{-\alpha} g_j \right\|_H^2 ds \\
&= \sup_{j\in\mathcal{J}\setminus\mathcal{J}_K} \mu_j^{2\alpha} \int_0^t (t-s)^{-\vartheta} \left\| (-A)^{-\vartheta/2} \sigma^M(U_s^{M,M}) Q^{-\alpha} \right\|_{HS(G_0,H)}^2 ds \\
&\leq \sup_{j\in\mathcal{J}\setminus\mathcal{J}_K} \mu_j^{2\alpha} \int_0^t c(t-s)^{-\vartheta} \Big(1 + \left\| U_s^{M,M} \right\|_V^2 \Big) ds
\end{aligned}
\tag{6.16}
$$

At this point, we have estimates for all temporal integrals in (6.11). Recall that we did not apply Itô's formula on the whole Galerkin error, but only on its finite-dimensional projection, i.e., the last term on the right hand side of

$$
\left\| U_t - V_t^{N,K} \right\|_H^2 \leq \sup_{M>N} \Big(2 \sup_{t\in[0,T]} \left\| U_t^{M,M} \right\|_V^2 \Big(\inf_{i\in\mathcal{I}\setminus\mathcal{I}_N} \lambda_i \Big)^{-1} + 2 \left\| P_N U_t^{M,M} - V_t^{N,K} \right\|_H^2 \Big).
\tag{6.10}
$$

However, we recognize terms corresponding to the entire right hand side in (6.13), (6.14) and (6.15). Using the notations

$$\Psi_s := 2 \sup_{t\in[0,T]} \left\|U_t^{M,M}\right\|_V^2 \left(\inf_{i\in\mathcal{I}\setminus\mathcal{I}_N} \lambda_i\right)^{-1} + 2\left\|P_N U_s^{M,M} - V_s^{N,K}\right\|_H^2$$

$G_s :=$ *stochastic integrands*

and combining (6.12)-(6.16), we get for arbitrary $M > N$

$$\begin{aligned}\Psi_t \le 2 \sup_{t\in[0,T]} \left\|U_t^{M,M}\right\|_V^2 \left(\inf_{i\in\mathcal{I}\setminus\mathcal{I}_N} \lambda_i\right)^{-1} &+ \int_0^t \Big(4c(t-s)^{-\vartheta} + 2c_0\Big(\left\|V_s^{N,K}\right\|_V^{\frac12} \left\|AV_s^{N,K}\right\|_H^{\frac12} \\ &+ \left\|U_s^{M,M}\right\|_V^{\frac12} \left\|AU_s^{M,M}\right\|_H^{\frac12}\Big) \left\|(-A)e^{2A(t-s)}P_N\right\|_{L(H)}^{\frac34}\Big)\Psi_s\, ds \\ &+ \int_0^t c(t-s)^{-\vartheta}\Big(1 + \left\|U_s^{M,M}\right\|_V^2\Big) \sup_{j\in\mathcal{J}\setminus\mathcal{J}_K} \mu_j^{2\alpha} ds + \int_0^t G_s dW_s.\end{aligned}$$

Note that the first summand corresponds to Ψ_0. We introduce the abbreviations

$$\begin{aligned}\Phi_s &:= 4c(t-s)^{-\vartheta} + 2c_0 \left\|(-A)e^{2A(t-s)}P_N\right\|_{L(H)}^{\frac34} \\ &\qquad \times \Big(\left\|V_s^{N,K}\right\|_V^{\frac12} \left\|AV_s^{N,K}\right\|_H^{\frac12} + \left\|U_s^{M,M}\right\|_V^{\frac12} \left\|AU_s^{M,M}\right\|_H^{\frac12} \Big) \\ \chi_s &:= c(t-s)^{-\vartheta}\Big(1 + \left\|U_s^{M,M}\right\|_V^2\Big) \sup_{j\in\mathcal{J}\setminus\mathcal{J}_K} \mu_j^{2\alpha},\end{aligned}$$

and obtain

$$\Psi_t \le \Psi_0 + \int_0^t \Phi_s\Psi_s ds + \int_0^t \chi_s ds + \int_0^t G_s dW_s.$$

We can now apply the stochastic Gronwall Lemma 6.16, which leads to

$$E\left[e^{-\int_0^t \Phi_s ds}\Psi_t\right] \le E\left[\Psi_0\right] + E\left[\int_0^t \chi_s e^{-\int_0^s \Phi_u du} ds\right]. \tag{6.17}$$

On the other hand (6.10) gives us

$$E\left[\left\|U_t - V_t^{N,K}\right\|_H^2\right] \le E\left[\Psi_t\right]. \tag{6.18}$$

Therefore we have to split $E[\Psi_t e^{-\int_0^t \Phi_s ds}]$, for instance, by finding a P-a.s. *lower* bound for $e^{-\int_0^t \Phi_s ds}$. To this end we use a stopping time argument (Lemma 6.18) to obtain a P-a.s. *upper* bound for $\int_0^t \Phi_s ds$. A version of Young's inequality, given by Corollary 6.17 with Hölder exponents $p = 4/3$, $q = 4$, gives us for all $\beta > 0$

$$\begin{aligned}\int_0^t \Phi_s ds &= 4c\int_0^t (t-s)^{-\vartheta} ds + 2c_0 \int_0^t \left\|(-A)e^{2A(t-s)}P_N\right\|_{L(H)}^{\frac34} \\ &\qquad \times \Big(\left\|V_s^{N,K}\right\|_V^{\frac12} \left\|AV_s^{N,K}\right\|_H^{\frac12} + \left\|U_s^{M,M}\right\|_V^{\frac12} \left\|AU_s^{M,M}\right\|_H^{\frac12} \Big) ds \\ &\le 4c\frac{t^{1-\vartheta}}{1-\vartheta} + \frac{2}{\beta}\int_0^t \left\|(-A)e^{2A(t-s)}P_N\right\|_{L(H)} ds \\ &\qquad + \frac{(2c_0)^4}{4}\Big(\frac{3\beta}{8}\Big)^3 \int_0^t \Big(\left\|V_s^{N,K}\right\|_V^{\frac12} \left\|AV_s^{N,K}\right\|_H^{\frac12} + \left\|U_s^{M,M}\right\|_V^{\frac12} \left\|AU_s^{M,M}\right\|_H^{\frac12} \Big)^4 ds \\ &\le 4c\frac{t^{1-\vartheta}}{1-\vartheta} + \frac{2}{\beta}\int_0^t \left\|(-A)e^{2A(t-s)}P_N\right\|_{L(H)} ds \\ &\qquad + c_0^4\beta^3 \int_0^t \left\|V_s^{N,K}\right\|_V^2 \left\|AV_s^{N,K}\right\|_H^2 + \left\|U_s^{M,M}\right\|_V^2 \left\|AU_s^{M,M}\right\|_H^2 ds,\end{aligned} \tag{6.19}$$

where $(a+b)^2 \leq 2a^2 + 2b^2$ was applied twice in the last line. Lemma 6.18 then yields

$$\int_0^t \Phi_s ds \leq \frac{2}{\beta} \int_0^t \left\| (-A) e^{2A(t-s)} P_N \right\|_{L(H)} ds + 4c \frac{t^{1-\vartheta}}{1-\vartheta} + c_0^4 k \beta^3 \tag{6.20}$$

P-a.s. for all $t \in [0,T]$. Unfortunately, a direct estimate of the remaining integral via Lemma 6.13 would lead to a divergent expression, see Remark 6.11. Therefore we need a tighter bound on

$$\int_0^t \left\| (-A) P_N e^{A(t-s)} \right\|_{L(H)} ds = \int_0^t \sup_{i \in \mathcal{I}_N} \lambda_i e^{-\lambda_i (t-s)} ds = \int_0^t \sup_{i \in \mathcal{I}_N} \lambda_i e^{-\lambda_i s} ds. \tag{6.21}$$

To this end, recall that the eigenvalues of $(-A)P_N$ are a strictly increasing sequence $\lambda_1 < \lambda_2 < ... < \lambda_N$. Furthermore, for a fixed $t \in [0,T]$, the function

$$x \mapsto x e^{-xt}$$

is maximal at $x = 1/t$. Therefore, the index of the supremum in (6.21) increases when t decreases. We divide the integral into disjunct time intervals, in each of which the index of the supremum is fixed. More precisely, for $j \in \{2, ..., N\}$, denote by t_j the time at which the index of the supremum changes from $j-1$ to j. Then t_j satisfies

$$\begin{aligned}
& \lambda_j e^{-\lambda_j t_j} = \lambda_{j-1} e^{-\lambda_{j-1} t_j} \\
\Rightarrow \quad & \frac{\lambda_j}{\lambda_{j-1}} = e^{(\lambda_j - \lambda_{j-1}) t_j} \\
\Rightarrow \quad & t_j = \frac{\ln \lambda_j - \ln \lambda_{j-1}}{\lambda_j - \lambda_{j-1}}.
\end{aligned}$$

Due to $\ln(x) < x - 1$, for $x > 1$, we have

$$\begin{aligned}
& \underbrace{(\lambda_{j+1} - \lambda_j)}_{>0} \underbrace{(\lambda_j - \lambda_{j-1})}_{>0} (t_{j+1} - t_j) \\
&= (\lambda_j - \lambda_{j-1}) \ln \frac{\lambda_{j+1}}{\lambda_j} - (\lambda_{j+1} - \lambda_j) \ln \frac{\lambda_j}{\lambda_{j-1}} \\
&< (\lambda_j - \lambda_{j-1}) \left(\frac{\lambda_{j+1}}{\lambda_j} - 1 \right) - (\lambda_{j+1} - \lambda_j) \left(\frac{\lambda_j}{\lambda_{j-1}} - 1 \right) \\
&= \lambda_{j+1} - \lambda_j - \frac{\lambda_{j-1} \lambda_{j+1}}{\lambda_j} - \lambda_{j-1} + \frac{\lambda_{j+1} \lambda_{j-1}}{\lambda_j} - \lambda_{j+1} - \lambda_{j-1} + \lambda_j \\
&= 0,
\end{aligned}$$

and hence

$$t_{j+1} < t_j.$$

Consequently, $\{t_j\}_{j=1}^N$ is a strictly decreasing sequence with

$$0 < t_N < t_{N-1} < ... < t_1.$$

This gives us the aforementioned disjunct decomposition

$$[0,t) = \bigcup_{j=1}^{N} [t_{j+1}, t_j),$$

for $t_{N+1} := 0$ and $t_1 := t$. Define the abbreviatory notation

$$S_j(s) := \lambda_j e^{-\lambda_j s},$$

and fix an arbitrary $j \in \{1, ..., N\}$. Then, for all $k \in \{0, ..., N-j-1\}$ and all $s \in [t_{j+k+1}, \infty)$, we have by definition of t_{j+k+1}

$$\frac{S_{j+k+1}(s)}{S_{j+k}(s)} = \frac{S_{j+k+1}(t_{j+k+1})}{S_{j+k}(t_{j+k+1})} \frac{e^{-\lambda_{j+k+1}(s - t_{j+k+1})}}{e^{-\lambda_{j+k}(s - t_{j+k+1})}} = e^{(\lambda_{j+k} - \lambda_{j+k+1})(s - t_{j+k+1})} \leq 1.$$

Therefore, $S_{j+k} \geq S_{j+k+1}$ on $[t_{j+k+1}, \infty)$. Since $[t_{j+1}, \infty) \subset [t_{j+k+1}, \infty)$, we get

$$\forall s \in [t_{j+1}, \infty)\ \forall k \in \{j+1, ..., N\}: \quad S_j \geq S_{j+k}. \tag{*}$$

On the other hand, for $k \in \{0, ..., j-1\}$ and all $s \in [0, t_{j-k})$, we obtain

$$\frac{S_{j-k}(s)}{S_{j-k-1}(s)} = \frac{S_{j-k}(t_{j-k})}{S_{j-k-1}(t_{j-k})} \frac{e^{-\lambda_{j-k}(s-t_{j-k})}}{e^{-\lambda_{j-k-1}(s-t_{j-k})}} = e^{(\lambda_{j-k-1}-\lambda_{j-k})(t_{j-k}-s)} \geq 1,$$

leading to $S_{j-k} \geq S_{j-k-1}$ on $[0, t_{j-k})$. Since $[0, t_j) \subset [0, t_{j-k})$ we get

$$\forall s \in [0, t_j)\ \forall k \in \{0, ..., j-1\}: \quad S_j \geq S_{j-k}. \tag{**}$$

Combining (*) and (**) yields

$$\forall s \in [t_{j+1}, t_j): \quad \sup_{i \in \mathcal{I}_N} \lambda_i e^{-\lambda_i s} = \lambda_j e^{-\lambda_j s},$$

which finally gives us

$$\begin{aligned}
&\int_0^t \left\| (-A) P_N e^{A(t-s)} \right\|_{L(H)} ds \\
&\quad = \int_0^t \sup_{i \in \mathcal{I}_N} \lambda_i e^{-\lambda_i s} ds \\
&\quad = \int_0^t \sum_{j=1}^N \mathbb{1}_{[t_{j+1}, t_j)}\ \lambda_j e^{-\lambda_j s} ds \\
&\quad = \sum_{j=1}^N \int_{t_{j+1}}^{t_j} \lambda_j e^{-\lambda_j s} ds \\
&\quad = \sum_{j=1}^N e^{-\lambda_j t_{j+1}} - e^{-\lambda_j t_j} \\
&\quad = e^{-\lambda_N t_{N+1}} - e^{-\lambda_1 t_1} + \sum_{j=2}^N e^{-\lambda_{j-1} t_j} - e^{-\lambda_j t_j} \\
&\quad = 1 - e^{-\lambda_1 t} + \sum_{j=2}^N e^{-\lambda_{j-1} t_j} \left(1 - e^{-(\lambda_j - \lambda_{j-1}) t_j}\right).
\end{aligned}$$

By definition of t_j we have

$$\begin{aligned}
\exp\left(-\lambda_{j-1} t_j\right) &= \exp\left(-\lambda_{j-1} \frac{\ln \lambda_j - \ln \lambda_{j-1}}{\lambda_j - \lambda_{j-1}}\right) \\
&= \exp\left(\frac{\lambda_{j-1}}{\lambda_j - \lambda_{j-1}}\ \ln\left(\frac{\lambda_{j-1}}{\lambda_j}\right)\right) \\
&= \left(\frac{\lambda_{j-1}}{\lambda_j}\right)^{\frac{\lambda_{j-1}}{\lambda_j - \lambda_{j-1}}} \\
\exp\left(-(\lambda_j - \lambda_{j-1}) t_j\right) &= \exp\left(-(\lambda_j - \lambda_{j-1}) \frac{\ln \lambda_j - \ln \lambda_{j-1}}{\lambda_j - \lambda_{j-1}}\right) \\
&= \exp\left(-\ln\left(\frac{\lambda_j}{\lambda_{j-1}}\right)\right) \\
&= \frac{\lambda_{j-1}}{\lambda_j},
\end{aligned}$$

and obtain

$$\int_0^t \left\| (-A) P_N e^{A(t-s)} \right\|_{L(H)} ds = 1 - e^{-\lambda_1 t} + \sum_{j=2}^N \left(\frac{\lambda_{j-1}}{\lambda_j}\right)^{\frac{\lambda_{j-1}}{\lambda_j - \lambda_{j-1}}} \left(1 - \frac{\lambda_{j-1}}{\lambda_j}\right).$$

Returning to the upper bound (6.20) this gives us

$$\int_0^t \Phi_s ds \leq \frac{2}{\beta}\left(1 - e^{-\lambda_1 t} + \sum_{j=2}^{N}\left(\frac{\lambda_{j-1}}{\lambda_j}\right)^{\frac{\lambda_{j-1}}{\lambda_j - \lambda_{j-1}}}\left(1 - \frac{\lambda_{j-1}}{\lambda_j}\right)\right) + 4c\frac{t^{1-\vartheta}}{1-\vartheta} + c_0^4 k\beta^3$$
$$=: 2\ln(g(\beta^{-1}, N)) + 2\ln(C(\beta)).$$

Introducing the notations

$$g(\beta^{-1}, N) := \exp\left(\frac{1}{\beta}\left(1 - e^{-\lambda_1 t} + \sum_{j=2}^{N}\left(\frac{\lambda_{j-1}}{\lambda_j}\right)^{\frac{\lambda_{j-1}}{\lambda_j - \lambda_{j-1}}}\left(1 - \frac{\lambda_{j-1}}{\lambda_j}\right)\right)\right)$$
$$C(\beta) := \exp\left(2c\frac{t^{1-\vartheta}}{1-\vartheta} + \frac{c_0^4 k\beta^3}{2}\right),$$

gives us the desired lower bound

$$e^{-\int_0^t \Phi_s ds} \geq C(\beta)^{-2} g(\beta^{-1}, N)^{-2} > 0,$$

P-a.s. for all $\beta > 0$. Since $\Psi_t \geq 0$ P-a.s. we have

$$C(\beta)^{-2} g(\beta^{-1}, N)^{-2} E[\Psi_t] \leq E\left[e^{-\int_0^t \Phi_s ds}\Psi_t\right],$$

and equation (6.17) yields

$$E[\Psi_t] \leq C(\beta)^2 g(\beta^{-1}, N)^2 \left(E[\Psi_0] + E\left[\int_0^t \chi_s e^{-\int_0^s \Phi_u du} ds\right]\right).$$

The first expectation value on the right hand side is given by

$$E[\Psi_0] = 2E\left[\sup_{t\in[0,T]}\left\|U_t^{M,M}\right\|_V^2\right]\left(\inf_{i\in\mathcal{I}\setminus\mathcal{I}_N}\lambda_i\right)^{-1}.$$

Since $\chi_s \geq 0$ and $\Phi_s \geq 0$, P-a.s. for all $s \in [0,t]$, we obtain for the second term

$$\begin{aligned}
&E\left[\int_0^t \chi_s e^{-\int_0^s \Phi_u du} ds\right] \\
&\quad\leq E\left[\int_0^t \chi_s ds\right] \\
&\quad= E\left[\int_0^t c(t-s)^{-\vartheta}\left(1 + \left\|U_s^{M,M}\right\|_V^2\right)\sup_{j\in\mathcal{J}\setminus\mathcal{J}_K}\mu_j^{2\alpha} ds\right] \\
&\quad\leq cE\left[1 + \sup_{t\in[0,T]}\left\|U_t^{M,M}\right\|_V^2\right]\left(\sup_{j\in\mathcal{J}\setminus\mathcal{J}_K}\mu_j\right)^{2\alpha}\int_0^t (t-s)^{-\vartheta} ds \\
&\quad= c\frac{t^{1-\vartheta}}{1-\vartheta}E\left[1 + \sup_{t\in[0,T]}\left\|U_t^{M,M}\right\|_V^2\right]\left(\sup_{j\in\mathcal{J}\setminus\mathcal{J}_K}\mu_j\right)^{2\alpha}.
\end{aligned}$$

Due to Lemma 6.18, there exists a constant $k < \infty$ with

$$E\left[\sup_{t\in[0,T]}\left\|U_t^{M,M}\right\|_V^2\right] \leq k.$$

Using equation (6.18) and absorbing all constants into a $c > 0$, independent of N, M, K and β, finally gives us

$$E\left[\left\|U_t - V_t^{N,K}\right\|_H^2\right] \leq E[\Psi_t] \leq e^{2c\beta^3} g(\beta^{-1}, N)^2 \left(\left(\inf_{i\in\mathcal{I}\setminus\mathcal{I}_N}\lambda_i\right)^{-1} + \left(\sup_{j\in\mathcal{J}\setminus\mathcal{J}_K}\mu_j\right)^{2\alpha}\right)$$

completing the proof. □

6.4.1 Theorem 6.10 and the Primitive Equations

Comparing Assumptions 6.2-6.8 from this section with the abstract framework of stochastic three-dimensional PE in Section 5.4.2 shows that the atmospheric PE on the sphere are a special case of the abstract fluid model at hand. Since the general formulation of Theorem 6.10, particularly the structure of the function g, is not very intuitive, we exemplarily discuss the results for the PE. The most complex error term is given by the eigenvalues of the linear operator A. For the sake of simplicity we consider the classical three-dimensional Laplace operator on the sphere, which satisfies

$$\lambda_j = j(j+1) \tag{6.22}$$

for all $j \in \mathbb{N}$. Then, we have for all $\beta > 0$

$$\begin{aligned}
\beta \ln g &= 1 - e^{-\lambda_1 t} + \sum_{j=2}^{N} \left(\frac{\lambda_{j-1}}{\lambda_j}\right)^{\frac{\lambda_{j-1}}{\lambda_j - \lambda_{j-1}}} \left(1 - \frac{\lambda_{j-1}}{\lambda_j}\right) \\
&= 1 - e^{-2t} + \sum_{j=2}^{N} \left(\frac{j(j-1)}{j(j+1)}\right)^{\frac{j(j-1)}{j(j+1)-j(j-1)}} \left(1 - \frac{j(j-1)}{j(j+1)}\right) \\
&= 1 - e^{-2t} + \sum_{j=2}^{N} \left(\frac{j-1}{j+1}\right)^{\frac{j-1}{j+1-j+1}} \left(1 - \frac{j-1}{j+1}\right) \\
&= 1 - e^{-2t} + \sum_{j=2}^{N} \left(1 - \frac{2}{j+1}\right)^{\frac{j-1}{2}} \left(\frac{2}{j+1}\right) \\
&= 1 - e^{-2t} + \sum_{j=2}^{N} \left(1 - \frac{2}{j+1}\right)^{\frac{j+1}{2}} \left(1 - \frac{2}{j+1}\right)^{-1} \left(\frac{2}{j+1}\right) \\
&= 1 - e^{-2t} + \sum_{j=2}^{N} \frac{2}{j-1} \left(1 - \frac{2}{j+1}\right)^{\frac{j+1}{2}}.
\end{aligned}$$

For large j, we get the following asymptotic behavior

$$2\left(1 - \frac{2}{j+1}\right)^{\frac{j+1}{2}} = 2\left(1 - \frac{1}{\left(\frac{j+1}{2}\right)^{-1},}\right)^{\frac{j+1}{2}} \approx 2e^{-1}$$

and

$$\frac{1}{j-1} \approx \frac{1}{j}.$$

Therefore, the sum does not converge for $N \to \infty$. However, it is bounded by

$$\begin{aligned}
\beta \ln g &= 1 - e^{-\lambda_1 t} + \sum_{j=2}^{N} \left(\frac{\lambda_{j-1}}{\lambda_j}\right)^{\frac{\lambda_{j-1}}{\lambda_j - \lambda_{j-1}}} \left(1 - \frac{\lambda_{j-1}}{\lambda_j}\right) \\
&= 1 - \underbrace{e^{-2t}}_{>0} + 2 \sum_{j=1}^{N-1} \frac{1}{j} \Bigg(\underbrace{1 - \frac{2}{j+2}}_{\leq 1}\Bigg)^{\frac{j+1}{2}} \\
&\leq 1 + 2\ln(N-1).
\end{aligned}$$

Using this expression, Theorem 6.10 yields the following error bound

$$\begin{aligned}
E\left[\left\|U_t - V_t^{N,K}\right\|_V^2\right]^{\frac{1}{2}} &\leq e^{c\beta^3} e^{\frac{2}{\beta}\ln(N)} \left(\left(\inf_{i \in \mathcal{I} \setminus \mathcal{I}_N} \lambda_i\right)^{-\frac{1}{2}} + \left(\sup_{j \in \mathcal{J} \setminus \mathcal{J}_K} \mu_j\right)^{\alpha}\right) \\
&\leq e^{c\epsilon^{-3}} \left(N^{-1+\epsilon} + \left(\sup_{j \in \mathcal{J} \setminus \mathcal{J}_K} N^{\frac{\epsilon}{\alpha}} \mu_j\right)^{\alpha}\right),
\end{aligned}$$

for all $\epsilon > 0$. Note that we introduced $\epsilon = \frac{1}{\beta}$ for a more intuitive representation. Regarding the stochastic dimension we recall, that Q is a trace-class operator, i.e., the sum of its eigenvalues converges. Assuming, for instance, a polynomial behavior $\mu_j = \mathcal{O}(j^{-p})$, for a $p > 1$, it is advisable to set $K = N^{\frac{1}{\alpha p}}$, so that the errors due to finite spatial and stochastic dimensions share the same mean square convergence rate, yielding

$$E\left[\left\|U_t - V_t^{N,K}\right\|_V^2\right]^{\frac{1}{2}} \leq e^{c\epsilon^{-3}} N^{-1+\epsilon}, \tag{6.23}$$

for all $\epsilon > 0$. This motivates the following definition.

Definition 6.19. Convergence Order p_-
A numerical scheme converges with order p_-, with $p > 0$, if it converges with order $p - \epsilon$ for all $\epsilon > 0$.

The complex error bound of Theorem 6.10 reduces to the clearly arranged term (6.23) for the Laplace operator. The crucial information we need to obtain this convergence result is the quadratical growth of the eigenvalues for the linear operator A.

6.5 Results

Theorem 6.10 provides an upper bound on the mean square error regarding the discretization of spatial and stochastic dimensions for an abstract fluid model (6.1). Note that mean square convergence implies strong convergence (Definition 2.34) due to Jensen's inequality. Convergence is guaranteed, if the sum

$$\sum_{j=2}^{N} \left(\frac{\lambda_{j-1}}{\lambda_j}\right)^{\frac{\lambda_{j-1}}{\lambda_j - \lambda_{j-1}}} \left(1 - \frac{\lambda_{j-1}}{\lambda_j}\right)$$

does not grow faster than $\ln(N)$. This assumption is, for instance, satisfied for spatial eigenvalues λ_N increasing at least quadratically, which includes the important case of the Laplace operator, see Section 6.4.1. For practical applications, the spatial operator A is defined by the system's physics. In the abstract setting at hand the system's stochastics are defined by σ and Q. While σ links the physical dynamics and the underlying stochastic fluctuations, the trace class operator Q describes the correlation between the different stochastic dimensions. The combination of both arises in the error bound in form of the term μ_k^α, where μ_k are the eigenvalues of Q, and α does depend on σ via Assumption 6.7. A common choice for the stochastic setting are quadratically decreasing μ_k. This yields a root mean square convergence order 1_- in the spatial and α_- in the stochastic dimensions. For most practical applications – such as the three-dimensional PE – a deterministic setting has already been studied prior to a stochastic formulation. Naturally, the convergence rates are a decisive factor for the success of a specific numerical treatment, particularly in the context of elaborate global models. The efficiency of a stochastic numerical scheme is therefore directly compared to its deterministic counterpart. Consequently, the parameter α plays a crucial role in practical applications, specifying the additional computational effort necessary to resolve stochastic dimensions. However, its value may not be obvious from Assumption 6.7 for a given system. We refer to [166, 167] for detailed discussions on this topic for various exemplary systems.

7 A Numerical Scheme for an Abstract Fluid Model

The Galerkin approximation from the previous section allows us to study time discretization methods for SPDE (6.1) in the finite-dimensional environment of SODE. Although the numerical framework for SODE has been extensively studied since the early 1970s, the vast majority of literature deals with Lipschitz continuous coefficient functions, see Definition 2.33. To the best of our knowledge the first strong convergence result with weaker assumptions was established by Hu in 1996 [170]. His results require a one-sided Lipschitz continuous drift function $f : \mathbb{R} \to \mathbb{R}$, i.e. there exists a $c > 0$ with

$$\langle x - y, f(x) - f(y) \rangle \leq c|x - y|^2,$$

for all $x, y \in \mathbb{R}$. This may be regarded as a weak formulation of Lipschitz continuity. Analog assumptions are used in more recent works by Higham et al. [171] and Hutzenthaler [172]. The more general setting by Higham et al. [171] assumes local Lipschitz continuous drift functions f. It turns out that the moments of both, the exact solution X and the numerical solution Y, have to satisfy a priori bounds of the form

$$E\left[\sup_{t\in[0,T]} X^p\right] + E\left[\sup_{t\in[0,T]} Y^p\right] < c,$$

for a $c > 0$, to achieve convergence. Higham et al. state that in "*general, it is not clear when such moment bounds can be expected to hold for explicit methods*" [171]. In the same paper it is shown that these bounds indeed hold true for an implicit numerical scheme, whereas Hutzenthaler [172] applies a "*tamed*" explicit scheme to derive such bounds under the additional assumption of polynomial growth of the drift function. Although the estimates in Assumption 6.5 on the bilinear advection term B are of similar structure, they do not follow a one-sided Lipschitz condition due to the involvement of the linear operator A. In the context of PE the difference becomes obvious, where A is essentially the Laplace operator. Therefore the advection can only be controlled by derivatives of the velocity and temperature but not by the state vector itself (see Lemma 5.20). As a consequence it is beyond the scope of this work to prove such a priori bounds. Furthermore the bilinear term B is ill-behaved to an extend that even the exact solution cannot be controlled in the above sense, i.e., only the local existence of solutions is guaranteed. Therefore we have to postulate a priori bounds for the numerical solution analogous with the existing bounds for the exact solution via Lemma 6.18, see Assumption 7.3. The reason for the specific choice of bounds, their applicability and some ideas for future improvements are discussed in Sections 7.2.2 and 7.2.4.

Definition 7.1. Numerical Scheme
The numerical scheme consists of $\mathcal{F}/\mathcal{B}(H)$-measurable mappings

$$Y_m^{N,M,K} : \Omega \longrightarrow H_N,$$

for $N, M, K \in \mathbb{N}$ and $m \in \{0, 1, ..., M\}$. These are successively defined by

$$Y_0^{N,M,K} := P_N(\xi)$$
$$Y_{m+1}^{N,M,K} := P_N e^{Ah}\left\{Y_m^{N,M,K} + hF(Y_m^{N,M,K}) + hB(Y_m^{N,M,K}) + \sigma(Y_m^{N,M,K})\Delta W_m^{N,K}\right\},$$

where $h := \frac{T}{M}$ and

$$\Delta W_m^{M,K} : \Omega \ni \omega \mapsto \left(W^K_{\frac{(m+1)T}{M}}(\omega) - W^K_{\frac{mT}{M}}\omega\right) \in G_0,$$

denote the $\mathcal{F}/\mathcal{B}(G_0)$-measurable increments of the Wiener process for all $N, M, K \in \mathbb{N}$ and $m \in \{0, 1, ..., M - 1\}$.

Definition 7.2. Interpolated Numerical Solution
In order to properly use the Itô formula, we need a time continuous representation of the numerical solution $Y_t^{N,M,K}$, see Lemma 6.14. To this end we define

$$\lfloor s \rfloor := \max\left\{ u \in \{0, h, 2h, ..., Mh = T\} \,\middle|\, u \leq s \right\}$$

for all $s \in [0,T]$. Using this notation we interpolate Definition 7.1 by

$$\begin{aligned} Y_t^{N,M,K} = e^{At} P_N \xi + \int_0^t e^{A(t-\lfloor s \rfloor)} \big(F^N(Y_{\lfloor s \rfloor}^{N,M,K}) + B^N(Y_{\lfloor s \rfloor}^{N,M,K}) \big) ds \\ + \int_0^t e^{A(t-\lfloor s \rfloor)} \sigma^N(Y_{\lfloor s \rfloor}^{N,M,K}) dW_s^K, \end{aligned}$$

P-a.s. for all $t \in [0,T]$ and $N, M, K \in \mathbb{N}$.

Assumption 7.3. A Priori Bounds of the Numerical Solution
Assume the existence of a constant $c > 0$ with

$$\left\| Y_{\lfloor t \rfloor}^{N,M,K} \right\|_V^2 \left\| A Y_{\lfloor t \rfloor}^{N,M,K} \right\|_H^2 \leq c$$

P-a.s. for all $N, M, K \in \mathbb{N}$ and all $t \in [0,T]$.

Theorem 7.4. Main Result
Let Assumptions 6.2-6.8 and 7.3 be satisfied. Then there exists a constant $c < \infty$ independent of $N, M, K, \alpha, \beta, \gamma, \delta$ with

$$\begin{aligned} E\left[\left\| U_t - Y_t^{N,M,K} \right\|_H^2 \right]^{1/2} \leq e^{c\beta^3} g(\beta^{-1}, N) \left(\lambda_{N+1}^{-\frac{1}{2}} + \left(\sup_{j \in \mathcal{J} \setminus \mathcal{J}_K} \mu_j \right)^{\alpha} \right) \\ + e^{c\gamma^3} \left(M^{-\frac{1}{2}+\frac{1}{\gamma}} + e^{c\delta} \lambda_N^{\frac{1}{2\delta}} M^{-\frac{1}{2}+\frac{1}{2\gamma}} \right) \end{aligned}$$

where

$$g(\beta^{-1}, N) = \exp\left(\frac{1}{\beta} \left(1 - e^{-\lambda_1 t} + \sum_{j=2}^{N} \left(\frac{\lambda_{j-1}}{\lambda_j} \right)^{\frac{\lambda_{j-1}}{\lambda_j - \lambda_{j-1}}} \left(1 - \frac{\lambda_{j-1}}{\lambda_j} \right) \right) \right)$$

for all $N, M, K \in \mathbb{N}$, $\beta, \gamma > 0$ and $\delta > 1$.

In order to proof Theorem 7.4 we need to estimate the interpolation error.

Lemma 7.5. Interpolation Error
There exists a $c > 0$ with

$$E\left[\left\| Y_t^{N,M,K} - Y_{\lfloor t \rfloor}^{N,M,K} \right\|_H^2 \right] \leq ch.$$

for all $t \in [0,T]$

Proof. *Lemma 7.5*
Note that the bilinear form B satisfies

$$\begin{aligned} & \|B(u,u)\|_H^2 = \langle B(u,u), B(u,u) \rangle \leq c_0 \|B(u,u)\|_H \|u\|_V \|Au\|_H \\ \Rightarrow \quad & \|B(u,u)\|_H^2 \leq c_0^2 \|u\|_V^2 \|Au\|_H^2 \end{aligned}$$

for all $u \in V$.

$$
\begin{aligned}
&E\left[\left\|Y_t^{N,M,K} - Y_{\lfloor t\rfloor}^{N,M,K}\right\|_H^2\right] \\
&\le E\left[\left\|e^{At}P_N\xi - e^{A\lfloor t\rfloor}P_N\xi\right\|_H^2\right] \\
&\qquad + E\left[\left\|\int_{\lfloor t\rfloor}^t e^{A(t-\lfloor s\rfloor)}\big(F^N(Y_{\lfloor s\rfloor}^{N,M,K}) + B^N(Y_{\lfloor s\rfloor}^{N,M,K})\big)ds\right\|_H^2\right] \\
&\qquad + E\left[\left\|\int_{\lfloor t\rfloor}^t e^{A(t-\lfloor s\rfloor)}\sigma^N(Y_{\lfloor s\rfloor}^{N,M,K})dW_s^K\right\|_H^2\right] \\
&\le E\left[\left\|e^{A\lfloor t\rfloor}\left(1 - e^{A(t-\lfloor t\rfloor)}\right)P_N\xi\right\|_H^2\right] \\
&\qquad + E\left[\left\|\int_{\lfloor t\rfloor}^t e^{A(t-\lfloor t\rfloor)}\big(F^N(Y_{\lfloor t\rfloor}^{N,M,K}) + B^N(Y_{\lfloor t\rfloor}^{N,M,K})\big)ds\right\|_H^2\right] \\
&\qquad + E\left[\int_{\lfloor t\rfloor}^t \left\|e^{A(t-\lfloor t\rfloor)}\sigma^N(Y_{\lfloor t\rfloor}^{N,M,K})\right\|_H^2 ds\right] \\
&\le E\left[\left\|(-A)^{-\frac{1}{2}}\left(1 - e^{A(t-\lfloor t\rfloor)}\right)\right\|_{L(H)}^2 \|P_N\xi\|_V^2\right] + E\left[\left\|(t-\lfloor t\rfloor)e^{A(t-\lfloor t\rfloor)}B^N(Y_{\lfloor t\rfloor}^{N,M,K})\right\|_H^2\right] \\
&\qquad + E\left[\left\|(t-\lfloor t\rfloor)e^{A(t-\lfloor t\rfloor)}F^N(Y_{\lfloor t\rfloor}^{N,M,K})\right\|_H^2\right] + E\left[(t-\lfloor t\rfloor)\left\|e^{A(t-\lfloor t\rfloor)}\sigma^N(Y_{\lfloor t\rfloor}^{N,M,K})\right\|_H^2\right] \\
&\le (t-\lfloor t\rfloor)\Big\{E\left[\|P_N\xi\|_V^2\right] + (t-\lfloor t\rfloor)E\left[\left\|e^{A(t-\lfloor t\rfloor)}B^N(Y_{\lfloor t\rfloor}^{N,M,K})\right\|_H^2\right] \\
&\qquad + (t-\lfloor t\rfloor)E\left[\left\|e^{A(t-\lfloor t\rfloor)}F^N(Y_{\lfloor t\rfloor}^{N,M,K})\right\|_H^2\right] + E\left[\left\|e^{A(t-\lfloor t\rfloor)}\sigma^N(Y_{\lfloor t\rfloor}^{N,M,K})\right\|_H^2\right]\Big\} \\
&\le h\Big\{E\left[\|\xi\|_V^2\right] + hc_0^2E\left[\left\|Y_{\lfloor t\rfloor}^{N,M,K}\right\|_V^2\left\|AY_{\lfloor t\rfloor}^{N,M,K}\right\|_H^2\right] \\
&\qquad + chE\left[1 + \left\|Y_{\lfloor t\rfloor}^{N,M,K}\right\|_V^2\right] + cE\left[1 + \left\|Y_{\lfloor t\rfloor}^{N,M,K}\right\|_V^2\right]\Big\} \\
&\le ch\Big\{E\left[\|\xi\|_V^2\right] + E\left[1 + \left\|Y_{\lfloor t\rfloor}^{N,M,K}\right\|_V^2\right] + E\left[\left\|Y_{\lfloor t\rfloor}^{N,M,K}\right\|_V^2\left\|AY_{\lfloor t\rfloor}^{N,M,K}\right\|_H^2\right]\Big\} \\
&\le ch\Big\{1 + E\left[\|\xi\|_V^2\right] + E\left[\left\|Y_{\lfloor t\rfloor}^{N,M,K}\right\|_V^2\left\|AY_{\lfloor t\rfloor}^{N,M,K}\right\|_H^2\right]\Big\} \\
&\le ch,
\end{aligned}
$$

where we used Assumption 7.3 for the last estimate. □

Corollary 7.6. A Reverse Hölder Inequality
For all $0 \le t_1 < t_2$ and all p-integrable functions $L^p \ni f : [t_1, t_2] \to \mathbb{R}$ we have

$$\left|\int_{t_1}^{t_2} f(s)ds\right|^p \le (t_2 - t_1)^{p-1}\int_{t_1}^{t_2} |f(s)|^p\, ds$$

for all $p > 1$.

Proof. *Corollary 7.6*
The reverse Hölder inequality for functions g, h on a measurable space (S, Σ, μ) with $\mu(S) > 0$ with $h(s) \ne 0$ for μ-almost all $s \in S$ and

$$\int_S |gh|d\mu < \infty, \qquad \left(\int_S |h|^{-\frac{1}{p-1}}d\mu\right)^{-(p-1)} > 0$$

states

$$\int_S |gh| d\mu \geq \left(\int_S |g|^{\frac{1}{p}} d\mu \right)^p \left(\int_S |h|^{-\frac{1}{p-1}} d\mu \right)^{-(p-1)}$$

for all $p > 1$ (see for instance [173]). Choose $g \equiv |f|^p$, $h \equiv 1$, $S = [t_1, t_2]$

$$\Rightarrow \qquad \left(\int_{t_1}^{t_2} |f| ds \right)^p \underbrace{\left(\int_{t_1}^{t_2} 1^{-\frac{1}{p-1}} ds \right)^{-(p-1)}}_{=(t_2-t_1)^{-(p-1)}} \leq \int_{t_1}^{t_2} |f|^p ds$$

□

7.1 Proof of Theorem 7.4

The first step towards the error estimate in Theorem 7.4 is given by the error bounds on spatial and stochastic discretization via Theorem 6.10:

$$E\left[\left\| U_t - Y_t^{N,M,K} \right\|_H^2 \right]^{1/2} \leq E\left[\left\| U_t - V_t^{N,K} \right\|_H^2 \right]^{1/2} + E\left[\left\| V_t^{N,K} - Y_t^{N,M,K} \right\|_H^2 \right]^{1/2}. \quad (7.1)$$

The time discretization error is obtained in analogy to the proof of Theorem 6.10, using characteristics of the semigroup operator $e^{A(t-s)}$. Applying Corollary 6.15, i.e., the mild Itô formula for the squared H-norm, on the difference process

$$\begin{aligned} V_t^{N,K} - Y_t^{N,M,K} = &\int_0^t e^{A(t-s)} F^N(V_s^{N,K}) - e^{A(t-\lfloor s \rfloor)} F^N(Y_{\lfloor s \rfloor}^{N,M,K}) ds \\ &+ \int_0^t e^{A(t-s)} B^N(V_s^{N,K}) - e^{A(t-\lfloor s \rfloor)} B^N(Y_{\lfloor s \rfloor}^{N,M,K}) ds \\ &+ \int_0^t e^{A(t-s)} \sigma^N(V_s^{N,K}) - e^{A(t-\lfloor s \rfloor)} \sigma^N(Y_{\lfloor s \rfloor}^{N,M,K}) dW_s^K \end{aligned}$$

yields

$$\left\| V_t^{N,K} - Y_t^{N,M,K} \right\|_H^2 = (F) + (B) + (W) + (\text{Itô}), \quad (7.2)$$

with

$$\begin{aligned} (F) &= 2\int_0^t \left\langle e^{A(t-s)} \left(V_s^{N,K} - Y_s^{N,M,K} \right), e^{A(t-s)} F^N(V_s^{N,K}) - e^{A(t-\lfloor s \rfloor)} F^N(Y_{\lfloor s \rfloor}^{N,M,K}) \right\rangle ds \\ (B) &= 2\int_0^t \left\langle e^{A(t-s)} \left(V_s^{N,K} - Y_s^{N,M,K} \right), e^{A(t-s)} B^N(V_s^{N,K}) - e^{A(t-\lfloor s \rfloor)} B^N(Y_{\lfloor s \rfloor}^{N,M,K}) \right\rangle ds \\ (W) &= 2\int_0^t \left\langle e^{A(t-s)} \left(V_s^{N,K} - Y_s^{N,M,K} \right), e^{A(t-s)} \sigma^N(V_s^{N,K}) - e^{A(t-\lfloor s \rfloor)} \sigma^N(Y_{\lfloor s \rfloor}^{N,M,K}) \right\rangle dW_s^K \\ (\text{Itô}) &= \sum_{j \in J_K,\ \mu_j \neq 0} \mu_j \int_0^t \left\| \left(e^{A(t-s)} \sigma^N(V_s^{N,K}) - e^{A(t-\lfloor s \rfloor)} \sigma^N(Y_{\lfloor s \rfloor}^{N,M,K}) \right) g_j \right\|_H^2 ds. \end{aligned}$$

Regarding the nonlinear drift term (F) we have

$$
\begin{aligned}
\frac{(F)}{2} &= \int_0^t \left\langle e^{A(t-s)}\left(V_s^{N,K} - Y_s^{N,M,K}\right), e^{A(t-s)}F^N(V_s^{N,K}) - e^{A(t-\lfloor s\rfloor)}F^N(Y_{\lfloor s\rfloor}^{N,M,K})\right\rangle ds \\
&\le \int_0^t \left\|e^{A(t-s)}\left(V_s^{N,K} - Y_s^{N,M,K}\right)\right\|_H \left\|e^{A(t-s)}F^N(V_s^{N,K}) - e^{A(t-\lfloor s\rfloor)}F^N(Y_{\lfloor s\rfloor}^{N,M,K})\right\|_H ds \\
&\le \int_0^t \left\|e^{A(t-s)}\left(V_s^{N,K} - Y_s^{N,M,K}\right)\right\|_H \\
&\quad \times \left(\left\|\left(e^{A(t-s)} - e^{A(t-\lfloor s\rfloor)}\right)F(V_s^{N,K})\right\|_H + \left\|e^{A(t-\lfloor s\rfloor)}\left(F(V_s^{N,K}) - F(Y_{\lfloor s\rfloor}^{N,M,K})\right)\right\|_H\right) ds \\
&\le \int_0^t \left\|V_s^{N,K} - Y_s^{N,M,K}\right\|_H \left\|e^{A(t-s)} - e^{A(t-\lfloor s\rfloor)}\right\|_{L(H)} \left\|F(V_s^{N,K})\right\|_H ds \\
&\quad + \int_0^t \left\|V_s^{N,K} - Y_s^{N,M,K}\right\|_H \left\|(-A)^{\vartheta}e^{A(t-\lfloor s\rfloor)}\right\|_{L(H)} \\
&\qquad \times \left\|(-A)^{-\vartheta}\left(F(V_s^{N,K}) - F(Y_{\lfloor s\rfloor}^{N,M,K})\right)\right\|_H ds \\
&\le \int_0^t \left\|V_s^{N,K} - Y_s^{N,M,K}\right\|_H \left\|e^{A(t-s)} - e^{A(t-\lfloor s\rfloor)}\right\|_{L(H)} c\left(1 + \left\|V_s^{N,K}\right\|_V\right) ds \\
&\quad + \int_0^t \left\|V_s^{N,K} - Y_s^{N,M,K}\right\|_H c(t-\lfloor s\rfloor)^{-\vartheta} \left\|V_s^{N,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H ds \\
&\le \int_0^t \left\|V_s^{N,K} - Y_s^{N,M,K}\right\|_H (t-s)^{-\frac{1}{4}} \\
&\qquad \times (t-s)^{\frac{1}{4}} c\left(1 + \left\|V_s^{N,K}\right\|_V\right) \left\|e^{A(t-s)} - e^{A(t-\lfloor s\rfloor)}\right\|_{L(H)} ds \\
&\quad + \int_0^t \left\|V_s^{N,K} - Y_s^{N,M,K}\right\|_H c(t-\lfloor s\rfloor)^{-\vartheta} \\
&\qquad \times \left(\left\|V_s^{N,K} - Y_s^{N,M,K}\right\|_H + \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H\right) ds \\
&\le \int_0^t \left\|V_s^{N,K} - Y_s^{N,M,K}\right\|_H^2 (t-s)^{-\frac{1}{2}} ds \\
&\quad + \int_0^t c^2\left(1 + \left\|V_s^{N,K}\right\|_V\right)^2 (t-s)^{\frac{1}{2}} \left\|e^{A(t-s)} - e^{A(t-\lfloor s\rfloor)}\right\|_{L(H)}^2 ds \\
&\quad + \int_0^t c(t-\lfloor s\rfloor)^{-\vartheta} \left\|V_s^{N,K} - Y_s^{N,M,K}\right\|_H^2 ds \\
&\quad + \int_0^t c(t-\lfloor s\rfloor)^{-\vartheta} \left\|V_s^{N,K} - Y_s^{N,M,K}\right\|_H \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H ds \\
&\le \int_0^t \left\|V_s^{N,K} - Y_s^{N,M,K}\right\|_H^2 (t-s)^{-\frac{1}{2}} ds \\
&\quad + \int_0^t c^2\left(1 + \left\|V_s^{N,K}\right\|_V\right)^2 (t-s)^{\frac{1}{2}} \\
&\qquad \times \left\|(-A)^{\frac{1}{2}}e^{A(t-s)}\right\|_{L(H)}^2 \left\|(-A)^{-\frac{1}{2}}\left(1 - e^{A(s-\lfloor s\rfloor)}\right)\right\|_{L(H)}^2 ds \\
&\quad + \int_0^t \frac{3}{2}c(t-\lfloor s\rfloor)^{-\vartheta} \left\|V_s^{N,K} - Y_s^{N,M,K}\right\|_H^2 ds \\
&\quad + \int_0^t \frac{1}{2}c(t-\lfloor s\rfloor)^{-\vartheta} \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2 ds,
\end{aligned}
$$

leading to

$$\begin{aligned}
\frac{(F)}{2} &\leq \int_0^t \left\|V_s^{N,K} - Y_s^{N,M,K}\right\|_H^2 \left((t-s)^{-\frac{1}{2}} + \frac{3}{2}c(t-\lfloor s\rfloor)^{-\vartheta}\right) ds \\
&\quad + \int_0^t \frac{1}{2}c(t-\lfloor s\rfloor)^{-\vartheta} \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2 ds \\
&\quad + c^2 h \int_0^t \left(1 + \left\|V_s^{N,K}\right\|_V\right)^2 (t-s)^{-\frac{1}{2}} ds \\
&\leq \int_0^t \left\|V_s^{N,K} - Y_s^{N,M,K}\right\|_H^2 \left((t-s)^{-\frac{1}{2}} + \frac{3}{2}c(t-\lfloor s\rfloor)^{-\vartheta}\right) ds \\
&\quad + \int_0^t \frac{1}{2}c(t-\lfloor s\rfloor)^{-\vartheta} \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2 ds \qquad (7.3) \\
&\quad + c^2 h \int_0^t \left(1 + \left\|V_s^{N,K}\right\|_V\right)^2 (t-s)^{-\frac{1}{2}} ds.
\end{aligned}$$

The bilinear term (B) is bounded by Assumption 6.5.

$$\begin{aligned}
\frac{(B)}{2} &= \int_0^t \left\langle e^{A(t-s)}\left(V_s^{N,K} - Y_s^{N,M,K}\right), e^{A(t-s)}B^N(V_s^{N,K}) - e^{A(t-\lfloor s\rfloor)}B^N(Y_{\lfloor s\rfloor}^{N,M,K})\right\rangle ds \\
&= \int_0^t \left\langle e^{A(t-s)}\left(V_s^{N,K} - Y_s^{N,M,K}\right), \left(e^{A(t-s)} - e^{A(t-\lfloor s\rfloor)}\right) B^N(V_s^{N,K})\right\rangle ds \\
&\quad + \int_0^t \left\langle e^{A(t-s)}\left(V_s^{N,K} - Y_s^{N,M,K}\right), e^{A(t-\lfloor s\rfloor)}\left(B^N(V_s^{N,K}) - B^N(Y_{\lfloor s\rfloor}^{N,M,K})\right)\right\rangle ds \\
&= \int_0^t \left\langle e^{A(t-s)}\left(e^{A(t-s)} - e^{A(t-\lfloor s\rfloor)}\right)\left(V_s^{N,K} - Y_s^{N,M,K}\right), B^N(V_s^{N,K})\right\rangle ds \\
&\quad + \int_0^t \left\langle e^{A(t-s)}e^{A(t-\lfloor s\rfloor)}\left(V_s^{N,K} - Y_s^{N,M,K}\right), B^N(V_s^{N,K} - Y_{\lfloor s\rfloor}^{N,M,K}, V_s^{N,K})\right\rangle ds \\
&\quad + \int_0^t \left\langle e^{A(t-s)}e^{A(t-\lfloor s\rfloor)}\left(V_s^{N,K} - Y_s^{N,M,K}\right), B^N(Y_{\lfloor s\rfloor}^{N,M,K}, V_s^{N,K} - Y_{\lfloor s\rfloor}^{N,M,K})\right\rangle ds \\
&= \int_0^t \left\langle e^{A(t-s)}\left(e^{A(t-s)} - e^{A(t-\lfloor s\rfloor)}\right)\left(V_s^{N,K} - Y_s^{N,M,K}\right), B^N(V_s^{N,K}, V_s^{N,K})\right\rangle ds \\
&\quad + \int_0^t \left\langle e^{A(t-s)}e^{A(t-\lfloor s\rfloor)}\left(V_s^{N,K} - Y_s^{N,M,K}\right), B^N(V_s^{N,K} - Y_{\lfloor s\rfloor}^{N,M,K}, V_s^{N,K})\right\rangle ds \\
&\quad + \int_0^t \left\langle V_s^{N,K} - Y_{\lfloor s\rfloor}^{N,M,K}), B^N(Y_{\lfloor s\rfloor}^{N,M,K}, e^{A(t-s)}e^{A(t-\lfloor s\rfloor)}\left(V_s^{N,K} - Y_s^{N,M,K}\right)\right\rangle ds \\
&\leq c_0 \int_0^t \left\|e^{A(t-s)}\left(e^{A(t-s)} - e^{A(t-\lfloor s\rfloor)}\right)\left(V_s^{N,K} - Y_s^{N,M,K}\right)\right\|_H \left\|V_s^{N,K}\right\|_V \left\|AV_s^{N,K}\right\|_H ds \\
&\quad + c_0 \int_0^t \left\|e^{A(t-s)}e^{A(t-\lfloor s\rfloor)}\left(V_s^{N,K} - Y_s^{N,M,K}\right)\right\|_V^{\frac{1}{2}} \left\|Ae^{A(t-s)}e^{A(t-\lfloor s\rfloor)}\left(V_s^{N,K} - Y_s^{N,M,K}\right)\right\|_H^{\frac{1}{2}} \\
&\qquad \times \left\|V_s^{N,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H \left\|V_s^{N,K})\right\|_V^{\frac{1}{2}} \left\|V_s^{N,K}\right\|_H^{\frac{1}{2}} ds \\
&\quad + c_0 \int_0^t \left\|e^{A(t-s)}e^{A(t-\lfloor s\rfloor)}\left(V_s^{N,K} - Y_s^{N,M,K}\right)\right\|_V^{\frac{1}{2}} \left\|Ae^{A(t-s)}e^{A(t-\lfloor s\rfloor)}\left(V_s^{N,K} - Y_s^{N,M,K}\right)\right\|_H^{\frac{1}{2}} \\
&\qquad \times \left\|V_s^{N,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H \left\|Y_{\lfloor s\rfloor}^{N,M,K})\right\|_V^{\frac{1}{2}} \left\|Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^{\frac{1}{2}} ds,
\end{aligned}$$

which gives us

$$
\begin{aligned}
\frac{(B)}{2} &\leq \frac{c_0}{2}\int_0^t \left\| \left(e^{A(t-s)} - e^{A(t-\lfloor s \rfloor)} \right) \right\|^2_{L(H)} + \left\| V_s^{N,K} - Y_s^{N,M,K} \right\|_H^2 \left\| V_s^{N,K} \right\|_V^2 \left\| AV_s^{N,K} \right\|_H^2 ds \\
&\quad + c_0 \int_0^t \left\| V_s^{N,K} - Y_s^{N,M,K} \right\|_H \left\| V_s^{N,K} - Y_{\lfloor s \rfloor}^{N,M,K} \right\|_H \\
&\qquad \times \left\| (-A) e^{A(t-s)} e^{A(t-\lfloor s \rfloor)} \right\|^{\frac{3}{4}}_{L(H)} \left\| V_s^{N,K}) \right\|_V^{\frac{1}{2}} \left\| V_s^{N,K} \right\|_H^{\frac{1}{2}} ds \\
&\quad + c_0 \int_0^t \left\| V_s^{N,K} - Y_s^{N,M,K} \right\|_H \left\| V_s^{N,K} - Y_{\lfloor s \rfloor}^{N,M,K} \right\|_H \\
&\qquad \times \left\| (-A) e^{A(t-s)} e^{A(t-\lfloor s \rfloor)} \right\|^{\frac{3}{4}}_{L(H)} \left\| Y_{\lfloor s \rfloor}^{N,M,K}) \right\|_V^{\frac{1}{2}} \left\| Y_{\lfloor s \rfloor}^{N,M,K} \right\|_H^{\frac{1}{2}} ds \\
&\leq \frac{c_0}{2} \int_0^t \left\| (-A)^{\frac{1-\epsilon}{2}} e^{A(t-s)} \right\|^2_{L(H)} \left\| (-A)^{-\frac{1-\epsilon}{2}} \left(1 - e^{A(s-\lfloor s \rfloor)} \right) \right\|^2_{L(H)} ds \\
&\quad + \frac{c_0}{2} \int_0^t \left\| V_s^{N,K} - Y_s^{N,M,K} \right\|_H^2 \left\| V_s^{N,K} \right\|_V^2 \left\| AV_s^{N,K} \right\|_H^2 ds \\
&\quad + c_0 \int_0^t \left\| V_s^{N,K} - Y_s^{N,M,K} \right\|_H \left(\left\| V_s^{N,K} - Y_s^{N,M,K} \right\|_H + \left\| Y_s^{N,M,K} - Y_{\lfloor s \rfloor}^{N,M,K} \right\|_H \right) \\
&\qquad \times \left\| (-A) e^{A(t-s)} e^{A(t-\lfloor s \rfloor)} \right\|^{\frac{3}{4}}_{L(H)} \\
&\qquad \times \left(\left\| V_s^{N,K} \right\|_V^{\frac{1}{2}} \left\| V_s^{N,K} \right\|_H^{\frac{1}{2}} + \left\| Y_{\lfloor s \rfloor}^{N,M,K}) \right\|_V^{\frac{1}{2}} \left\| Y_{\lfloor s \rfloor}^{N,M,K} \right\|_H^{\frac{1}{2}} \right) ds \\
&\leq \frac{c_0}{2} \int_0^t (t-s)^{-1+\epsilon} (s - \lfloor s \rfloor)^{1-\epsilon} ds \\
&\quad + \frac{c_0}{2} \int_0^t \left\| V_s^{N,K} - Y_s^{N,M,K} \right\|_H^2 \left\| V_s^{N,K} \right\|_V^2 \left\| AV_s^{N,K} \right\|_H^2 ds \\
&\quad + c_0 \int_0^t \left(\frac{3}{2} \left\| V_s^{N,K} - Y_s^{N,M,K} \right\|_H^2 + \frac{1}{2} \left\| Y_s^{N,M,K} - Y_{\lfloor s \rfloor}^{N,M,K} \right\|_H^2 \right) \\
&\qquad \times \left\| (-A) e^{A(t-s)} e^{A(t-\lfloor s \rfloor)} \right\|^{\frac{3}{4}}_{L(H)} \\
&\qquad \times \left(\left\| V_s^{N,K} \right\|_V^{\frac{1}{2}} \left\| V_s^{N,K} \right\|_H^{\frac{1}{2}} + \left\| Y_{\lfloor s \rfloor}^{N,M,K}) \right\|_V^{\frac{1}{2}} \left\| Y_{\lfloor s \rfloor}^{N,M,K} \right\|_H^{\frac{1}{2}} \right) ds.
\end{aligned}
$$

Young's inequality (Corollary 6.17) and $(s - \lfloor s \rfloor) \leq h$ yield

$$
\begin{aligned}
\frac{(B)}{2} &\leq h^{1-\epsilon} \frac{c_0}{2} \int_0^t (t-s)^{-1+\epsilon} ds \\
&\quad + \frac{c_0}{2} \int_0^t \left\| V_s^{N,K} - Y_s^{N,M,K} \right\|_H^2 \left\| V_s^{N,K} \right\|_V^2 \left\| AV_s^{N,K} \right\|_H^2 ds \\
&\quad + \frac{2}{3\eta} \int_0^t \left\| (-A) e^{A(t-s)} e^{A(t-\lfloor s \rfloor)} \right\|_{L(H)} \\
&\qquad \times \left(\frac{3}{2} \left\| V_s^{N,K} - Y_s^{N,M,K} \right\|_H^2 + \frac{1}{2} \left\| Y_s^{N,M,K} - Y_{\lfloor s \rfloor}^{N,M,K} \right\|_H^2 \right) ds \\
&\quad + 2c_0^4 \left(\frac{18\eta}{16} \right)^3 \int_0^t \left(\frac{3}{2} \left\| V_s^{N,K} - Y_s^{N,M,K} \right\|_H^2 + \frac{1}{2} \left\| Y_s^{N,M,K} - Y_{\lfloor s \rfloor}^{N,M,K} \right\|_H^2 \right) \\
&\qquad \times \left(\left\| V_s^{N,K} \right\|_V^2 \left\| AV_s^{N,K} \right\|_H^2 + \left\| Y_{\lfloor s \rfloor}^{N,M,K}) \right\|_V^2 \left\| AY_{\lfloor s \rfloor}^{N,M,K} \right\|_H^2 \right) ds
\end{aligned}
$$

We obtain for the bilinear term:

$$\begin{aligned}\frac{(B)}{2} \leq & \int_0^t \left\|V_s^{N,K} - Y_s^{N,M,K}\right\|_H^2 \left[\frac{1}{\eta}\left\|(-A)e^{A(t-s)}e^{A(t-\lfloor s\rfloor)}\right\|_{L(H)}\right. \\ & \left. + 5c_0^4\eta^3\left(\left\|V_s^{N,K}\right\|_V^2\left\|AV_s^{N,K}\right\|_H^2 + \left\|Y_{\lfloor s\rfloor}^{N,M,K})\right\|_V^2\left\|AY_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2\right)\right] ds \\ & + \int_0^t \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2 \left[\frac{1}{3\eta}\left\|(-A)e^{A(t-s)}e^{A(t-\lfloor s\rfloor)}\right\|_{L(H)}\right. \\ & \left. + 5c_0^4\eta^3\left(\left\|V_s^{N,K}\right\|_V^2\left\|AV_s^{N,K}\right\|_H^2 + \left\|Y_{\lfloor s\rfloor}^{N,M,K})\right\|_V^2\left\|AY_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2\right)\right] ds \\ & + h^{1-\epsilon}\frac{c_0}{2}\int_0^t (t-s)^{-1+\epsilon} ds. \end{aligned} \tag{7.4}$$

Since the stochastic integral (W) vanishes when taking the expectation value, we draw out attention to the (Itô)-term

$$\begin{aligned}(\text{Itô}) = & \sum_{j\in J_K,\ \mu_j\neq 0} \mu_j \int_0^t \left\|\left(e^{A(t-s)}\sigma^N(V_s^{N,K}) - e^{A(t-\lfloor s\rfloor)}\sigma^N(Y_{\lfloor s\rfloor}^{N,M,K})\right)g_j\right\|_H^2 ds \\ \leq & \int_0^t \left\|e^{A(t-s)}\sigma^N(V_s^{N,K}) - e^{A(t-\lfloor s\rfloor)}\sigma^N(Y_{\lfloor s\rfloor}^{N,M,K})\right\|_{HS(G_0,H)}^2 ds \\ \leq & 2\int_0^t \left\|\left(e^{A(t-s)} - e^{A(t-\lfloor s\rfloor)}\right)\sigma^N(V_s^{N,K})\right\|_{HS(G_0,H)}^2 ds \\ & + 2\int_0^t \left\|e^{A(t-\lfloor s\rfloor)}\left(\sigma^N(V_s^{N,K}) - \sigma^N(Y_{\lfloor s\rfloor}^{N,M,K})\right)\right\|_{HS(G_0,H)}^2 ds \\ \leq & 2\int_0^t \left\|(-A)^{-\frac{1}{2}}\left(e^{A(t-s)} - e^{A(t-\lfloor s\rfloor)}\right)\right\|_{L(H)}^2 \left\|(A^{\frac{1}{2}}\sigma^N(V_s^{N,K})\right\|_{HS(G_0,H)}^2 ds \\ & + 2\int_0^t \left\|(-A)^{\frac{\vartheta}{2}}e^{A(t-\lfloor s\rfloor)}\right\|_{L(H)}^2 \left\|(-A)^{-\frac{\vartheta}{2}}\left(\sigma^N(V_s^{N,K}) - \sigma^N(Y_{\lfloor s\rfloor}^{N,M,K})\right)\right\|_{HS(G_0,H)}^2 ds \\ \leq & 2\int_0^t \left\|(-A)^{-\frac{1}{2}}\left(1 - e^{A(s-\lfloor s\rfloor)}\right)\right\|_{L(H)}^2 \left\|\sigma^N(V_s^{N,K})\right\|_{HS(G_0,V)}^2 ds \\ & + 2\int_0^t (t-\lfloor s\rfloor)^{-\vartheta}\left\|V_s^{N,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2 ds \\ \leq & 2\int_0^t (s-\lfloor s\rfloor)c^2\left(1 + \left\|V_s^{N,K}\right\|_V^2\right) ds + 2\int_0^t (t-\lfloor s\rfloor)^{-\vartheta}\left\|V_s^{N,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2 ds \\ \leq & 2c^2h\int_0^t \left(1 + \left\|V_s^{N,K}\right\|_V^2\right) ds + 4\int_0^t (t-\lfloor s\rfloor)^{-\vartheta}\left\|V_s^{N,K} - Y_s^{N,M,K}\right\|_H^2 ds \\ & + 4\int_0^t (t-\lfloor s\rfloor)^{-\vartheta}\left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2 ds \end{aligned} \tag{7.5}$$

The initial equation (7.2) and the derived estimates (7.3),(7.4) and (7.5) give us the following, slightly unwieldy, error bound

$$
\begin{aligned}
&\left\|V_t^{N,K} - Y_t^{N,M,K}\right\|_H^2 - (W) \\
&\leq \int_0^t \left\|V_s^{N,K} - Y_s^{N,M,K}\right\|_H^2 \left[\frac{2}{\eta}\left\|(-A)e^{A(t-s)}e^{A(t-\lfloor s\rfloor)}\right\|_{L(H)}\right. \\
&\qquad + 10c_0^4\eta^3\left(\left\|V_s^{N,K}\right\|_V^2\left\|AV_s^{N,K}\right\|_H^2 + \left\|Y_{\lfloor s\rfloor}^{N,M,K})\right\|_V^2\left\|AY_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2\right) \\
&\qquad \left. + 2\left((t-s)^{-\frac{1}{2}} + \frac{3}{2}c(t-\lfloor s\rfloor)^{-\vartheta}\right) + 4(t-\lfloor s\rfloor)^{-\vartheta}\right]ds \\
&\quad + \int_0^t \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2 \left[\frac{2}{3\eta}\left\|(-A)e^{A(t-s)}e^{A(t-\lfloor s\rfloor)}\right\|_{L(H)}\right. \\
&\qquad + 10c_0^4\eta^3\left(\left\|V_s^{N,K}\right\|_V^2\left\|AV_s^{N,K}\right\|_H^2 + \left\|Y_{\lfloor s\rfloor}^{N,M,K})\right\|_V^2\left\|AY_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2\right) \\
&\qquad \left. + \frac{1}{2}c(t-\lfloor s\rfloor)^{-\vartheta} + 4(t-\lfloor s\rfloor)^{-\vartheta}\right]ds \\
&\quad + \int_0^t h^{1-\epsilon}\left[\frac{c_0}{2}(t-s)^{-1+\epsilon} + 2c^2\left(1+\left\|V_s^{N,K}\right\|_V\right)^2\left(1+(t-s)^{-\frac{1}{2}}\right)\right]ds.
\end{aligned}
$$

Introducing the following abbreviatory notations

$$
\begin{aligned}
\Psi_s &:= \left\|V_t^{N,K} - Y_t^{N,M,K}\right\|_H^2 \\
\Phi_s &:= 10c_0^4\eta^3\left(\left\|V_s^{N,K}\right\|_V^2\left\|AV_s^{N,K}\right\|_H^2 + \left\|Y_{\lfloor s\rfloor}^{N,M,K})\right\|_V^2\left\|AY_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2\right) \\
&\qquad + \frac{2}{\eta}\left\|(-A)e^{A(t-s)}e^{A(t-\lfloor s\rfloor)}\right\|_{L(H)} + 2(t-s)^{-\frac{1}{2}} + 7c(t-\lfloor s\rfloor)^{-\vartheta} \\
\chi_s &:= h^{1-\epsilon}\left[\frac{c_0}{2}(t-s)^{-1+\epsilon} + 2c^2\left(1+\left\|V_s^{N,K}\right\|_V\right)^2\left(1+(t-s)^{-\frac{1}{2}}\right)\right] \\
&\qquad + \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2\left[\frac{2}{3\eta}\left\|(-A)e^{A(t-s)}e^{A(t-\lfloor s\rfloor)}\right\|_{L(H)} + \frac{c+8}{2}(t-\lfloor s\rfloor)^{-\vartheta}\right. \\
&\qquad\qquad \left. + 10c_0^4\eta^3\left(\left\|V_s^{N,K}\right\|_V^2\left\|AV_s^{N,K}\right\|_H^2 + \left\|Y_{\lfloor s\rfloor}^{N,M,K})\right\|_V^2\left\|AY_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2\right)\right]
\end{aligned}
$$

$G_s :=$ stochastic integrand,

allows us to obtain the clearly arranged integral inequality

$$\Psi_t \leq \int_0^t \Phi_s\Psi_s ds + \int_0^t \chi_s ds + \int_0^t G_s dW_s.$$

Since $\Psi_0 = 0$ P-a.s., an application of the stochastic Gronwall Lemma 6.16 gives us

$$E\left[e^{-\int_0^t \Phi_s ds}\Psi_t\right] \leq E\left[\int_0^t \chi_s e^{-\int_0^s \Phi_u du} ds\right]. \tag{7.6}$$

The exponential term on the right hand side is bounded by one, due to $\Phi_t \geq 0$ P-a.s. for all $t \in [0,t]$. Therefore, in order to quantify the mean square discretization error $E[\Psi_t]$, we have to find both

- a P-a.s. upper bound for $\int_0^t \Phi_s ds$
- an upper bound on $E\left[\int_0^t \chi_s ds\right]$.

We start with the second task by looking at one of the most problematic terms:

$$
\begin{aligned}
&E\left[\int_0^t \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2 \left\|V_s^{N,K}\right\|_V^2 \left\|AV_s^{N,K}\right\|_H^2 ds\right] \qquad (7.7)\\
&= E\left[\int_0^t \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2 \left\|V_s^{N,K}\right\|_V^2 \left\|AV_s^{N,K}\right\|_H^{2-2\delta} \left\|AP_N V_s^{N,K}\right\|_H^{2\delta} ds\right]\\
&\le E\left[\int_0^t \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2 \left\|V_s^{N,K}\right\|_V^2 \left\|AV_s^{N,K}\right\|_H^{2-2\delta} \left\|A^{\frac12}V_s^{N,K}\right\|_H^{2\delta} \left\|A^{\frac12}P_N\right\|_{L(H)}^{2\delta} ds\right]\\
&\le |\lambda_N|^\delta E\left[\int_0^t \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2 \left\|V_s^{N,K}\right\|_V^{4\delta} \left\|V_s^{N,K}\right\|_V^{2-2\delta} \left\|AV_s^{N,K}\right\|_H^{2-2\delta} ds\right]\\
&\le |\lambda_N|^\delta E\left[\sup_{s\in[0,t]} \left\|V_s^{N,K}\right\|_V^{4\delta} \int_0^t \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2 \left\|V_s^{N,K}\right\|_V^{2-2\delta} \left\|AV_s^{N,K}\right\|_H^{2-2\delta} ds\right]
\end{aligned}
$$

for all $0 < \delta < 1$. Lemma 6.18 infers the existence of a constant k with

$$\sup_{s\in[0,t]} \left\|V_s^{N,K}\right\|_V^{4\delta} + \int_0^t \left\|V_s^{N,K}\right\|_V^2 \left\|AV_s^{N,K}\right\|_H^2 ds \le k$$

P-a.s. for all $t \in [0,T]$. Using Hölder's inequality with $p = \delta$ and $q = \frac{1}{1-\delta}$ leads to

$$
\begin{aligned}
&E\left[\int_0^t \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2 \left\|V_s^{N,K}\right\|_V^2 \left\|AV_s^{N,K}\right\|_H^2 ds\right]\\
&\le k|\lambda_N|^\delta\, E\Bigg[\left(\int_0^t \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^{2\delta} ds\right)^{\frac{1}{\delta}}\\
&\qquad\qquad \times \left(\int_0^t \left(\left\|V_s^{N,K}\right\|_V^{2-2\delta} \left\|AV_s^{N,K}\right\|_H^{2-2\delta}\right)^{\frac{1}{1-\delta}} ds\right)^{1-\delta}\Bigg]\\
&\le k|\lambda_N|^\delta\, E\left[\left(\int_0^t \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^{2\delta} ds\right)^{\frac{1}{\delta}} \left(\int_0^t \left\|V_s^{N,K}\right\|_V^2 \left\|AV_s^{N,K}\right\|_H^2 ds\right)^{1-\delta}\right]\\
&\le k^2|\lambda_N|^\delta\, E\left[\left(\int_0^t \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^{2\delta} ds\right)^{\frac{1}{\delta}}\right].
\end{aligned}
$$

The reverse Hölder inequality (Corollary 7.6) and the incremental estimate of Lemma 7.5 yield

$$
\begin{aligned}
&E\left[\int_0^t \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2 \left\|V_s^{N,K}\right\|_V^2 \left\|AV_s^{N,K}\right\|_H^2 ds\right]\\
&\le k^2|\lambda_N|^\delta\, E\left[t^{\frac{1}{\delta}-1} \int_0^t \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2 ds\right]\\
&= k^2|\lambda_N|^\delta t^{\frac{1}{\delta}-1} \int_0^t E\left[\left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2\right] ds\\
&\le k^2|\lambda_N|^\delta t^{\frac{1}{\delta}-1} \int_0^t ch\, ds\\
&\le ck^2 t^{\frac{1}{\delta}} |\lambda_N|^\delta h. \qquad (7.8)
\end{aligned}
$$

Remark 7.7. Coupling of Spatial and Temporal Discretization Error
It is precisely term (7.7) *that causes the coupling between temporal and spatial discretization error in* (7.8)*, and finally in Theorem 7.4. The underlying cause is the barely controlled bilinear advection term, which is responsible for the occurrence of* $\|AV^{N,K}\|_H^2$*. This term would be rather unproblematic on its own, since it is bounded in* L^1*. The multiplicatively linked interpolation error* $\|Y_s - Y_{\lfloor s\rfloor}\|_H^2$ *is unfortunately not bounded in* L^∞*. Therefore we have to take a mathematical detour, relocating a small share of the barely controlled* L^1 *term. This directly involves the maximal eigenvalue of the linear operator* AP_N*, which in turn depends on the spatial dimension* N*.*

Let us estimate the remaining terms that are necessary for an upper bound on $E[\int \chi_s ds]$. Since $\Phi_s \geq 0$ P-a.s. for all $s \in [0,T]$, we have $\exp(-\int_0^s \Phi_u du) \leq 1$ P-a.s. for all $s \in [0,T]$. This gives us

$$
\begin{aligned}
&E\left[\int_0^t \chi_s e^{-\int_0^s \Phi_u du} ds\right] \leq E\left[\int_0^t \chi_s ds\right] \\
&= \quad E\left[\int_0^t 10c_0^4\eta^3 \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2 \left\|V_s^{N,K}\right\|_V^2 \left\|AV_s^{N,K}\right\|_H^2 ds\right] \\
&\quad + E\left[\int_0^t 10c_0^4\eta^3 \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2 \left\|Y_{\lfloor s\rfloor}^{N,M,K})\right\|_V^2 \left\|AY_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2 ds\right] \\
&\quad + E\left[\int_0^t \left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2 \left(\frac{2}{3\eta}\left\|(-A)e^{A(t-s)}e^{A(t-\lfloor s\rfloor)}\right\|_{L(H)} + \frac{c+8}{2}(t-\lfloor s\rfloor)^{-\vartheta}\right) ds\right] \\
&\quad + E\left[\int_0^t h^{1-\epsilon}\left[\frac{c_0}{2}(t-s)^{-1+\epsilon} + 2c^2\left(1+\left\|V_s^{N,K}\right\|_V\right)^2\left(1+(t-s)^{-\frac{1}{2}}\right)\right] ds\right].
\end{aligned}
$$

While the first term is treated by (7.8), we use Assumption 7.3 for a P-a.s. upper bound on the stochastic parts of the second and fourth term. The remaining terms do not contain any products of stochastic processes and are therefore comparatively easy to control. We Combine constants independent of $N, M, K, \beta, \eta, \delta$ and ϵ into a constant c – which may change from line to line – and obtain

$$
\begin{aligned}
&E\left[\int_0^t \chi_s e^{-\int_0^s \Phi_u du} ds\right] \\
&\leq \quad c\eta^3 t^{\frac{1}{\delta}} |\lambda_N|^\delta h \\
&\quad + c\eta^3 \int_0^t E\left[\left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2\right] ds \\
&\quad + \frac{2}{3\eta}\int_0^t E\left[\left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2\right]\left\|(-A)e^{A(t-s)}e^{A(t-\lfloor s\rfloor)}\right\|_{L(H)} ds \\
&\quad + c\int_0^t E\left[\left\|Y_s^{N,M,K} - Y_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2\right](t-\lfloor s\rfloor)^{-\vartheta} ds \\
&\quad + ch^{1-\epsilon}\int_0^t 1 + (t-s)^{-1+\epsilon} + (t-s)^{-\frac{1}{2}} ds \\
&\leq \quad c\eta^3 t^{\frac{1}{\delta}} |\lambda_N|^\delta h \\
&\quad + c\eta^3 h \int_0^t 1\, ds \\
&\quad + \frac{c}{3\eta}h\int_0^t \left\|(-A)e^{A(t-s)}e^{A(t-\lfloor s\rfloor)}\right\|_{L(H)} ds \\
&\quad + c^2 h\int_0^t (t-\lfloor s\rfloor)^{-\vartheta} ds \\
&\quad + ch^{1-\epsilon}\int_0^t 1 + (t-s)^{-1+\epsilon} + (t-s)^{-\frac{1}{2}}\, ds \\
&\leq c\eta^3 t^{\frac{1}{\delta}} |\lambda_N|^\delta h + c\frac{t^\epsilon}{\epsilon}h^{1-\epsilon} + \frac{c}{3\eta}h\int_0^t \left\|(-A)e^{A(t-s)}e^{A(t-\lfloor s\rfloor)}\right\|_{L(H)} ds. \qquad (7.9)
\end{aligned}
$$

The last integral is almost identical to the one responsible for most of the complications during the proof of the Galerkin error estimate (Theorem 6.10). Recall that a direct estimate via Lemma 6.13 would lead to the divergent expression

$$
\int_0^t \left\|(-A)e^{A(t-s)}\right\|_{L(H)} ds \leq \int_0^t (t-s)^{-1} ds = \infty.
$$

However, the subtly differing exponent of the integral in (7.9), i.e., we consider the exponent $t - \lfloor s\rfloor$ instead of $t-s$, allows us to circumvent this issue. Indeed, using Lemma 6.13

the remaining integral is bounded by

$$\begin{aligned}
&\int_0^t \left\|(-A)e^{A(t-s)}e^{A(t-\lfloor s\rfloor)}\right\|_{L(H)} ds \\
&\quad\le \int_0^t \left\|(-A)e^{A(t-\lfloor s\rfloor)}\right\|_{L(H)} ds \\
&\quad\le \int_0^t (t-\lfloor s\rfloor)^{-1} ds \\
&\quad= \int_0^{\lfloor t\rfloor} (t-\lfloor s\rfloor)^{-1} ds + \int_{\lfloor t\rfloor}^t (t-\lfloor s\rfloor)^{-1} ds \\
&\quad= \int_{\lfloor t\rfloor}^t (t-\lfloor t\rfloor)^{-1} ds + \sum_{l=0}^{\lfloor t\rfloor/h-1} \int_{lh}^{(l+1)h} (t-lh)^{-1} ds
\end{aligned}$$

Since the first integral vanishes for $t = \lfloor t\rfloor$ and equals one for $t \neq \lfloor t\rfloor$, we obtain for all $t \in [0,T]$

$$\begin{aligned}
&\int_0^t \left\|(-A)e^{A(t-s)}e^{A(t-\lfloor s\rfloor)}\right\|_{L(H)} ds \\
&\quad\le 1 + \sum_{l=0}^{\lfloor t\rfloor/h-1} \frac{h}{t-lh} \\
&\quad\le 1 + \sum_{l=0}^{\lfloor t\rfloor/h-1} \frac{h}{\lfloor t\rfloor-lh} \\
&\quad= 1 + \sum_{l=0}^{\lfloor t\rfloor/h-1} \frac{1}{\lfloor t\rfloor/h-l} \\
&\quad= 1 + \sum_{l=1}^{\lfloor t\rfloor/h} \frac{1}{l} \\
&\quad\le 1 + \sum_{l=1}^{M} \frac{1}{l} \\
&\quad\le c + \ln M \\
&\quad\le c + \frac{M^\epsilon}{\epsilon}.
\end{aligned} \tag{7.10}$$

Therefore $h = T/M$ yields

$$\begin{aligned}
E\left[\int_0^t \chi_s e^{-\int_0^s \Phi_u du} ds\right] &\le c\eta^3 t^{\frac{1}{\delta}} |\lambda_N|^\delta M^{-1} + c\frac{t^\epsilon}{\epsilon} M^{-1+\epsilon} + \frac{c}{3\eta\epsilon} M^{-1+\epsilon} \\
&\le cM^{-1}\left(\eta^3 t^{\frac{1}{\delta}} |\lambda_N|^\delta + \frac{M^\epsilon}{\epsilon}\right).
\end{aligned} \tag{7.11}$$

This is the desired upper bound for the right hand side of

$$E\left[e^{-\int_0^t \Phi_s ds}\Psi_t\right] \le E\left[\int_0^t \chi_s e^{-\int_0^s \Phi_u du} ds\right]. \tag{7.6}$$

Since we are interested in an estimate for $E[\Psi_t]$, we are still left with the task to find a P-a.s. lower bound for the left hand side. To this end recall that

$$\begin{aligned}
\int_0^t \Phi_s ds = \int_0^t 10c_0^4\eta^3 &\left(\left\|V_s^{N,K}\right\|_V^2 \left\|AV_s^{N,K}\right\|_H^2 + \left\|Y_{\lfloor s\rfloor}^{N,M,K})\right\|_V^2 \left\|AY_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2\right) \\
&+ \frac{2}{\eta}\left\|(-A)e^{A(t-s)}e^{A(t-\lfloor s\rfloor)}\right\|_{L(H)} + 2(t-s)^{-\frac{1}{2}} + 7c(t-\lfloor s\rfloor)^{-\vartheta} ds.
\end{aligned} \tag{7.12}$$

As before we summarize constants independent of $N, M, K, \beta, \eta, \delta$ and ϵ into a single constant c. While Lemma 6.18 states

$$\int_0^t \left\|V_s^{N,K}\right\|_V^2 \left\|AV_s^{N,K}\right\|_H^2 ds \leq k$$

P-a.s. for all $t \in [0,T]$, Assumption 7.3 yields an analogous result for the process $Y_{\lfloor s \rfloor}^{N,M,K}$, which gives us control over the first term in (7.12). The third term is bound by (7.10), yielding

$$\begin{aligned}\int_0^t \Phi_s ds &\leq c\eta^3 \int_0^t \left\|V_s^{N,K}\right\|_V^2 \left\|AV_s^{N,K}\right\|_H^2 ds + c\eta^3 kt + \frac{2}{\eta}\int_0^t \left\|(-A)e^{A(t-s)}e^{A(t-\lfloor s \rfloor)}\right\|_{L(H)} ds \\ &\quad + \int_0^t 2(t-s)^{-\frac{1}{2}} + 7c(t-s)^{-\vartheta} ds \\ &\leq c\eta^3 + \frac{2}{\eta}\left(c + \ln(M)\right) \\ &\leq c\eta^3 + \frac{2}{\eta}\ln(M). \end{aligned} \tag{7.13}$$

Therefore a lower bound for the left hand side of (7.6) is given via (7.13), while the right hand side is controlled by the upper bound (7.11). We obtain

$$\begin{aligned} e^{-c\eta^3} M^{-\frac{2}{\eta}} E\left[\Psi_t\right] &\leq E\left[e^{-\int_0^t \Phi_s ds}\Psi_t\right] \\ &\leq E\left[\int_0^t \chi_s e^{-\int_0^s \Phi_u du} ds\right] \leq cM^{-1}\left(\eta^3 t^{\frac{1}{\delta}}|\lambda_N|^\delta + \frac{M^\epsilon}{\epsilon}\right). \end{aligned}$$

By definition of Ψ_t we have

$$E\left[\left\|V_t^{N,K} - Y_t^{N,M,K}\right\|_H^2\right] \leq e^{c\eta^3} M^{-1+\frac{2}{\eta}}\left(\eta^3 t^{\frac{1}{\delta}}|\lambda_N|^\delta + \frac{M^\epsilon}{\epsilon}\right),$$

for all $\eta, \epsilon > 0$ and $0 < \delta < 1$. The substitutions $\eta = 2\gamma$, $\epsilon = \gamma^{-1}$ and $\delta \mapsto \delta^{-1}$ yield

$$\begin{aligned} E\left[\left\|V_t^{N,K} - Y_t^{N,M,K}\right\|_H^2\right] &\leq e^{c\gamma^3}\left(M^{-1+\frac{2}{\gamma}} + t^\delta \lambda_N^{\frac{1}{\delta}} M^{-1+\frac{1}{\gamma}}\right) \\ &\leq e^{c\gamma^3}\left(M^{-1+\frac{2}{\gamma}} + e^{c\delta}\lambda_N^{\frac{1}{\delta}} M^{-1+\frac{1}{\gamma}}\right) \end{aligned}$$

for all $\gamma > 0$ and $\delta > 1$, completing the proof. □

7.1.1 Theorem 7.4 and the Primitive Equations

In order to gain a more intuitive understanding of the quite complex error bound in Theorem 7.4, we continue the discussion on the exemplary case of the PE from Section 6.4.1. Recall that the eigenvalues of the Laplace operator lead to

$$g(\beta^{-1}, N) \leq cN^{\frac{1}{\beta}},$$

for all $\beta > 0$. Let the eigenvalues of Q, which specify the stochastic processes, satisfy $\mu_j \leq cj^{-p}$, for a $p > 0$ and all $j \in \mathcal{J}$. Then Theorem 7.4 states

$$E\left[\left\|U_t - Y_t^{N,M,K}\right\|_H^2\right]^{1/2} \leq e^{c\beta^3} N^{\frac{1}{\beta}} \left(\lambda_{N+1}^{-\frac{1}{2}} + \mu_{K+1}^{\alpha}\right) + e^{c\gamma^3}\left(M^{-\frac{1}{2}+\frac{1}{\gamma}} + e^{c\delta}\lambda_N^{\frac{1}{2\delta}} M^{-\frac{1}{2}+\frac{1}{2\gamma}}\right).$$

Since $\lambda_{N+1} = (N+1)(N+2) \leq cN^2$ for all $N \in \mathbb{N}$, we have

$$E\left[\left\|U_t - Y_t^{N,M,K}\right\|_H^2\right]^{1/2} \leq e^{c\beta^3}\left(N^{-1+\frac{1}{\beta}} + N^{\frac{1}{\beta}} K^{-p\alpha}\right) + e^{c\gamma^3}\left(M^{-\frac{1}{2}+\frac{1}{\gamma}} + e^{c\delta} N^{\frac{1}{\delta}} M^{-\frac{1}{2}+\frac{1}{2\gamma}}\right).$$

To unify the exponential constants β, γ and δ, we define for an $0 < \epsilon < \frac{1}{2}$

$$\beta = \epsilon^{-1}, \quad \gamma = 2\epsilon^{-1}, \quad \delta = 2\epsilon^{-1},$$

yielding

$$E\left[\left\|U_t - Y_t^{N,M,K}\right\|_H^2\right]^{1/2} \leq e^{c\epsilon^{-3}}\left(N^{-1+\epsilon} + N^{\epsilon} K^{-p\alpha}\right) + e^{c\epsilon^{-3}}\left(M^{-\frac{1}{2}+\frac{1}{2}\epsilon} + e^{c\epsilon^{-1}} N^{\frac{1}{2}\epsilon} M^{-\frac{1}{2}+\frac{1}{4}\epsilon}\right).$$

In order to achieve the same order of convergence in spatial-, noise- and temporal-discretization, we set

$$M = N^2, \quad K = N^{\frac{1}{p\alpha}}, \tag{7.14}$$

leading to

$$E\left[\left\|U_t - Y_t^{N,N^2,N^{\frac{1}{p\alpha}}}\right\|_H^2\right]^{1/2} \leq e^{c\epsilon^{-3}} N^{-1+\epsilon},$$

for all $N \in \mathbb{N}$, $t \in [0,T]$ and all $0 < \epsilon < \frac{1}{2}$. The quadratic relation (7.14) between the dimensions of spatial and temporal discretization is a typical characteristic of stochastic Euler schemes, which is one topic of discussion in the following section.

7.2 Results

The Galerkin approximation for an abstract fluid model (6.1) discussed in Section 6 exhibits a rather complex discretization error, see Theorem 6.10. For many practical applications this reduces to clearly arranged root mean square convergence rates 1_- and α_- in spatial and stochastic dimensions, respectively, see Section 6.4.1 and 6.5. Recall that, due to Jensen's inequality, mean square convergence implies strong convergence (Definition 2.34). These results hold true for the exponential Euler scheme (Definition 7.1), allowing us to focus on the temporal discretization. Under the assumption of a priori bounds (Assumption 7.3), Theorem 7.4 shows that the numerical scheme converges with strong root mean square order $\frac{1}{2}-$ in time. It is a typical behavior of basic numerical schemes for both, ordinary as well as partial SDE, to exhibit only half the convergence order one would expect from their deterministic counterparts. An intuitive explanation is offered through the timescale of Brownian motions

$$\Delta W \sim \sqrt{\Delta t},$$

which directly transfers to the discretization of the corresponding integrals. This can only be compensated by more elaborate treatments of the stochastic integrals. In the context of low-dimensional SODE the Milstein scheme and stochastic Runge-Kutta methods [44] can be used for higher convergence rates. Unfortunately, they require global Lipschitz continuity and are therefore not applicable for the system at hand. Even if we could overcome this assumption, the classical approaches would still be impractical: the Milstein scheme, for instance, evaluates double stochastic integrals, i.e., for an m-dimensional SODE with d independent stochastic dimensions, we have to discretize the term

$$\int_{t_{n-1}}^{t_n} \int_{t_{n-1}}^{s} X_u^k \, dW_u^j dW_s^i,$$

with $1 \leq i, j \leq d$ and $1 \leq k \leq m$, for each time step. This leads to (at least) $d^2 \times m$ additional calculations in each time step. Even for rather low-dimensional systems, e.g., $d = m = 20$, this results in 8000 additional arithmetic operations required to go from X_{n-1} to X_n. Obviously, this approach does not work for SPDE, where we have to approximate infinite dimensions. Recently however, new techniques for global Lipschitz continuous SPDE were derived by Jentzen and Roeckner [166], and Wang [174]. They are based on the same mild interpretation of SPDE we used to formulate the exponential Euler scheme at hand. Unfortunately, like their low-dimensional counterparts they require global Lipschitz continuous drift operator. In a numerical experiment Jentzen and Roeckner [166, Ch. 5] apply their "*mild Milstein scheme*" to the SPDE

$$dX_t(x) = \left[\frac{1}{100} \frac{\partial^2}{\partial x} X_t(x) - X_t(x) \frac{\partial}{\partial x} X_t(x) \right] dt + X_t(x) dW_t,$$

known as "*stochastic Burgers equation*". This can be considered as a very basic, one-dimensional example for the fluid model at hand: the second order spatial derivative corresponds to A, while the bilinear drift term is the analog of B. This equation satisfies only a local but no global Lipschitz condition regarding its nonlinear drift. Although we can not use the rigorously obtained convergence rates for the mild Milstein scheme [166], the numerical experiment shows a *pathwise* convergence by a factor 10^3 faster than for the traditional linear implicit Euler scheme. We mention this fact not to suggest the uncritical application of this scheme on the system at hand, but as a motivation for further analysis. Note that Jentzen and Roeckner consider pathwise instead of mean square convergence, since "*it has been shown in* [175] *that Euler's method fails to converge in the root mean square sense to the exact solution of a SODE with a superlinearly growing drift coefficient*" [166, p. 40], whereas pathwise convergence often holds true [176]. On the other hand, Theorem 7.4 yields mean square convergence of the exponential Euler scheme, bearing the question whether or not the postulated a priori bounds (Assumption 7.3) are consistent with the physical setting and numerical implementation. An in-depth discussion on this subject is given in the following section.

7.2.1 Limitation of the Exponential Euler Scheme

A close look at the assumptions and results of this section reveals three major limitations of the explicit exponential Euler scheme for an abstract fluid model:

1. Galerkin approximations and the numerical scheme can only be applied up to a stopping time τ, for which the existence of solution of (6.1) is guaranteed.
2. The numerical scheme requires the assumption of a priori bounds, which may not be consistent with the physical setting.
3. The error bounds in Theorem 6.10 and 7.4 involve mixed terms, i.e., spatial, stochastic and temporal discretization, and cannot be treated independently.

Of these issues, the third one is certainly the least severe. Note that stochastic and temporal terms in Theorem 7.4 are independent from each other, and neither contributes to the spatial error. However, the spatial dimension affects both, stochastic and temporal error bound. Therefore, increasing the spatial dimension while keeping stochastic dimension and time-step size constant, leads to increasing error bounds in Theorem 7.4. On the other hand, changing stochastic dimension or time-step size for constant spatial dimensions, yields the expected decreasing error bound. For practical applications we are interested in the minimal overall computational effort. This is achieved balancing the arising error terms, which is discussed by example of the PE in Section 6.4.1 and 7.1.1. In this case, Theorem 7.4 yields strong mean square convergence, and the coupling between different error terms is not significant.

The coupling of spatial and stochastic dimensions has its origin in the estimate of the bilinear term via Assumption 6.5, i.e.

$$|\langle B(u,v),w\rangle| \leq c_0 \|u\|_H \|v\|_V^{\frac{1}{2}} \|Av\|_H^{\frac{1}{2}} \|w\|_V^{\frac{1}{2}} \|Aw\|_H^{\frac{1}{2}} \qquad \forall u \in H,\ v,w \in \mathcal{D}(-A).$$

During the proof of Theorem 6.10 this leads to the following term in equation (6.19), which is multiplicatively linked to the discretization error of stochastic dimensions:

$$\int_0^t \left\| (-A) e^{2A(t-s)} P_N \right\|_{L(H)}^{\frac{3}{4}} \left\| AV_s^{N,K} \right\|_H^{\frac{1}{2}} ds.$$

Due to the involvement of the linear operator A, the second factor is not bounded pointwise in time, but only in L^2. This in turn limits the possible exponents for an application of Hölder's or Young's inequality, finally leading to the complex function of spatial eigenvalues g, see Theorem 6.10. Analog, the aforementioned estimate on B causes the coupling of temporal and spatial discretization during the proof of Theorem 7.4, see (7.7), (7.8) and Remark 7.7. Conclusively, the underlying cause for the coupling of spatial dimensions with stochastic and temporal discretization lies in the ill-behaved bilinear term. However, recent results by Debussche et al. [177] hint at a possible way to gain better control over the bilinear term, which is set out in more detail in Section 7.2.4 below. Suppose we had an "*infinitesimal better*" estimate on B, leading, for instance, to the integral

$$\int_0^t \left\| (-A) e^{2A(t-s)} P_N \right\|_{L(H)}^{\frac{3}{4}} \left\| AV_s^{N,K} \right\|_H^{\frac{1}{2}-\epsilon} ds,$$

for an arbitrary small $\epsilon > 0$. Then, using Lemma 6.13 and Hölder's inequality, we could bound this term by a constant independent of spatial dimension N.

7.2.2 Discussion on A Priori Bounds

The need for a priori bounds is a common issue for explicit numerical schemes in the case of non-global Lipschitz continuous systems, see for instance [170, 172] and the references therein. Usually the assumptions involve moment bounds of the form:

$$E\left[\left\|Y_{\lfloor s\rfloor}^{N,M,K}\right\|_V^p\right] < \infty,$$

for some $p \geq 2$. However, for the system at hand we require P-a.s. bounds, since we have to find an P-a.s. upper bound for

$$\int_0^t 10c_0^4\eta^3\left(\left\|V_s^{N,K}\right\|_V^2\left\|AV_s^{N,K}\right\|_H^2 + \left\|Y_{\lfloor s\rfloor}^{N,M,K})\right\|_V^2\left\|AY_{\lfloor s\rfloor}^{N,M,K}\right\|_H^2\right)$$

during the proof of Theorem 7.4, see equation (7.12). In analogy to Lemma 6.18 we might use a stopping time argument to derive P-a.s. bounds from moment bounds, up to the maximal time of existence τ. An alternative approach is used by Debussche et al. during the proof for the existence of solutions for the abstract fluid model at hand [7]. They apply a different version of Gronwall's Lemma [178, Lemma 5.3], which circumvents the requirement of P-a.s. bounds by employing a partition of stopping times on moment bounds. However, for the present case, the resulting Gronwall-constant would depend exponentially on the bounds of the multiplicative process Φ in equation (6.17) and (7.6). As discussed above, the ill-behaved bilinear term leads to estimates depending on the spatial dimension N for these terms. This in turn would yield a Gronwall-constant exponentially growing in N. Therefore, this approach requires better control over the bilinear advection term to be implemented.

For an applicable numerical scheme a priori bounds must not just be postulated, but have to be derived based on the systems governing equations. This however is a nontrivial task, even for moment bounds, yet alone, P-a.s. estimates. In the context of one-dimensional SODE, Hutzenthaler et al. obtain such bounds using a "*tamed*" explicit Euler scheme for globally one-sided Lipschitz continuous drift functions with at most polynomial growing derivatives. Although the bilinear term at hand does not satisfy this assumptions, some of the employed ideas might be useful for the present system. Conclusively the underlying cause for both, the restricting structure of the a priori bounds in Assumption 7.3, as well as the difficulties obtaining more general moment bounds for the system at hand, lies in the ill-behaved advection term. We would like to emphasize that the postulated a priori bounds are only necessary for the time discretization in Theorem 7.4 and not for the Galerkin approximation in Theorem 6.10.

7.2.3 Discussion on the Maximal Time of Existence

The existence of solutions for the abstract fluid model (6.1) is only guaranteed up to a stopping time τ, defined in Corollary 6.9; see Section 5.4.2 for the PE, or [7] for the abstract case. Therefore the results of Theorem 6.10 and 7.4 only hold true up to this maximal time of existence. Since we have no estimate or lower bound on τ, this poses a serious obstacle for practical applications. A closer look at the proof of existence in [7] shows that the cause for this limitation is, once again, the barely controlled bilinear term. In order to control the growth of solutions, Debussche et al. introduced a "*tamed*" system (see proof of Lemma 6.18), which coincides with (6.1) as long as the solution stays close to the solution of the deterministic linear system (6.5). Analogous with the limitations of mixed discretization errors and postulated a priori bounds, we need a better control over the advection term in order to solve this issue.

7.2.4 Outlook

In order to derive a numerical scheme applicable to meteorological relevant systems, we have to address the aforementioned limitations of the exponential Euler scheme. Since the common cause of these issues can be found in the lacking control we have over the advection term, it is sensible to start further investigation at this point. Debussche et al. consider the special case of incompressible, stochastic PE in a confined, euclidean, cylindrical domain with vanishing boundary conditions [177]. They show existence of unique *global* pathwise solutions, employing the following advection estimate:

$$|\langle B(U,U^{\#}),U^{b}\rangle| \leq c\Big(|v|_{L^4}\left\|U^{\#}\right\|_V^{\frac{1}{4}}\left\|AU^{\#}\right\|_H^{\frac{3}{4}}\left\|U^{b}\right\|_H + \|v\|_V^{\frac{1}{2}}|v|_{H^2}^{\frac{1}{2}}\left\|\partial_z U^{\#}\right\|_H^{\frac{1}{2}}\left\|\partial_z U^{\#}\right\|_V^{\frac{1}{2}}\left\|U^{b}\right\|_H\Big),$$

for all $U=(v,T), U^{\#}\in\mathcal{D}(-A)$ and $U^{b}\in H$. Clearly, the horizontal and vertical directions are resolved more elaborately than in Assumption 6.5. Since this bound is derived along the lines of Lemma 5.20, using anisotrop Sobolev estimates, we recall that the central Sobolev inequality depends on the spatial dimension, i.e.

$$\|u\|_{L^q(\mathbb{R}^n)} \leq c\|\nabla u\|_{L^p(\mathbb{R}^n)}, \quad \text{with} \quad q=\frac{np}{n-p}.$$

Note that $\|\nabla u\|_{L^2}$ corresponds to $\|u\|_V$ by definition in Section 5.3.2. Therefore $\|u\|_V$ provides an upper bound for $\|u\|_{L^\infty}$ in space dimension two, but only for $\|u\|_{L^6}$ in space dimension three. As we stated above, an "*infinitesimal*" improvement on the bounds on B might be sufficient to overcome the limitations of the exponential Euler scheme. The elaborate calculations in [177] exploit the subtle but important fact, that the surface pressure does not depend on the vertical variable. Debussche et al. explicitly emphasize "*the crucial role played by the two-dimensional spatial dependence of the surface pressure p_s. This insight concerning the importance of the lower-dimensional pressure is key to the recent breakthroughs for global existence*" [177, Rem.4.2, p.14]. Although there are considerable differences to the global PE of the atmosphere – most notably compressibility and physical boundary conditions, see Section 5.2 – Debussche et al. state: "*We note however that equations of a quite similar structure may be given that describe the atmosphere and the coupled oceanic atmospheric system. See e.g., [136]. The methods developed here could thus be extended to treat these systems*" [177, p.4]. Following this line of thought may allow us to utilize the alternative Gronwall approach discussed in Section 7.2.2, which in turn would yield a stronger convergence result of the form

$$E\left[\sup_{t\in[0,T]}\left\|U_t-Y_t^{N,M,K}\right\|_H^2\right]\to 0, \quad \text{for} \quad N,M,K\to\infty.$$

Furthermore, it is worth to consider the aforementioned infinite-dimensional Milstein scheme [166], as well as the Runge-Kutta scheme for SPDE [174], in order to improve the temporal convergence order.

It is only natural for practical meteorological applications to expect fluctuations that are correlated in time and space. Therefore a potential extension of the system at hand is given by the description of spatial correlation via so called "*Brownian Sheets*". They can be considered as the continuous extensions of discrete Gaussian random fields [179], in analogy to the relation between Brownian motions and random walks. Therefore, Brownian sheets broaden the concept of Brownian motions, or more generally of Itô processes, using multidimensional "*time*" parameters, i.e., $W\equiv W(t,x_1,\ldots,x_n)$. For a theoretical treatment we refer to [180] while numerical schemes for SPDE driven by Brownian sheets are, for instance, discussed in [181, 162, 163].

8 Conclusion

This final section is a brief summary of the central results of this work. For more elaborate discussions we refer to the conclusions at the end of each section. The central statements of this thesis are as follows:

- Stochastic formulations in numerical weather prediction and climate modeling must not be arbitrarily chosen but have to be physically based.
- Stochastic models require a specific numerical treatment which fundamentally differs from deterministic schemes.

These objectives are motivated by the exemplary discussion of a linear system based on climate sensitivity in Section 3. We observe fundamentally different systems for different stochastic formulations, including random variables (Section 3.2), white noise (Section 3.3.1), red noise (Section 3.3.2) and a coupled noise approach (Section 3.4). Lemma 3.2 shows that state dependent fluctuations induce a stochastic drift, which has direct consequences for parameter estimations via time series data (Remark 3.3). Systems that do not exclude (short term) negative values for the climate sensitivity, exhibit regimes of instability, depending on the ratio between mean value and variance of the driving stochastic process (Prop. 3.5 and 3.13). The analysis of these regimes points out a significant discrepancy between the expectations on the ability of stochastic meteorological models to quantify the probability of extreme events on the one hand, and their low number of sample paths on the other hand. The importance of a careful numerical implementation is demonstrated in Section 3.5, where non-Lipschitz continuous coefficients lead to the failure of an explicit Milstein scheme for the coupled noise system.

Using the example of an energy budget model (EBM), we demonstrate a method to implement stochastic terms into a formerly deterministic model in Section 4. The characteristics of the stochastic terms are derived from ice core data using a Yule-Walker technique (Section 4.3.3). The three EBM models based on Vostok, Byrd and Taylor ice core data, are all driven by the same kind of noise, i.e. an Ornstein Uhlenbeck process, where only the parameters differ. Nevertheless, the resulting models differ fundamentally, particularly with respect to the interactions between temperature and insolation process (Section 4.4). This emphasizes the importance to choose a stochastic formulation very carefully. Equally important, this section shows that we have the necessary mathematical tools to derive physically based stochastics from data.

Regarding the numerical aspects, there exist rigorous schemes for Lipschitz continuous SODE [44] and SPDE [8]. Unfortunately many meteorological relevant applications, including the primitive equations and Navier Stokes equations, do not satisfy these assumptions. A numerical treatment of SPDE comprises spatial, stochastic and temporal discretizations. Based on recent theoretical progress [7], we propose a Galerkin scheme for the atmospheric three-dimensional PE on the sphere (Sections 5 and 6). Theorem 6.10 guarantees strong mean square convergence and quantitative error bounds up to a maximal time of existence. Furthermore, the results for the discretization of spatial and stochastic dimensions hold true for abstract fluid models (6.1) with a wide range of stochastic terms, including white, red, additive, multiplicative, linear and Lipschitz continuous nonlinear noise. Therefore, Theorem 6.10 provides rigorous quantitative error bounds for Galerkin approximations for a large class of meteorological relevant systems.

Under the additional assumption of a priori bounds (Assumption 7.3), Theorem 7.4 shows the temporal convergence of an exponential Euler scheme (Section 7). However, for an applicable numerical scheme, the a priori bounds have to be derived from the system's governing equations instead of being postulated. This issue is directly related to the behavior of the bilinear advection term. In the context of three-dimensional PE, recent results [177] indicate that it should be possible to gain a better control over this term, promising the existence and uniqueness of global pathwise solutions (Section 7.2.4). Simultaneously, this approach bears the potential to derive a (possibly implicit) numerical scheme without the need for postulated a priori bounds.

References

[1] H. von Storch, “Conditional Statistical Models: A Discourse About the Local Scale in Climate Modelling.,” in *Monte Carlo Simulations in Oceanography: Proceedings of the 9th ’Aha Huliko’a Winter Workshop* (P. Müller and D. Henderson, eds.), pp. 49–58, University of Hawaii at Manoa, January 14-17, 1997, 49-58 (Reprint: GKSS 97/E/59), 1997.

[2] J. Honerkamp, *Statistical Physics, An Advanced Approach With Application*. Springer, New York, Berlin, 2002. ISBN 978-3642077036.

[3] E. N. Lorenz, “Climatic Predictability, The Physical Basis of Climate and Climate Modeling,” *GARP Publ. Ser.*, vol. 16, pp. 132–136, 1976.

[4] K. Hasselmann, “Stochastic Climate Models. Part I: Theory,” *Tellus*, vol. 28, pp. 473–485, 1976.

[5] G. Yule, “On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer’s Sunspot Numbers,” *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 226, pp. 267–298, 1927.

[6] G. Walker, “On Periodicity in Series of Related Terms,” *Proceedings of the Royal Society of London. Series A*, vol. 131, no. 818, pp. 518–532, 1931.

[7] A. Debussche, N. Glatt-Holtz, and R. Temam, “Local Martingale and Pathwise Solutions for an Abstract Fluid Model,” *Physica D: Nonlinear Phenomena*, 2011.

[8] A. Jentzen and P. Kloeden, *Taylor Approximations for Stochastic Partial Differential Equations*. SIAM, 2011.

[9] M. Leutbecher and T. Palmer, “Ensemble Forecasting,” *Journal of Computational Physics*, vol. 227, no. 7, pp. 3515–3539, 2008.

[10] R. Buizza, M. Miller, and T. N. Palmer, “Stochastic simulation of model uncertainties,” *Quart. J. Roy. Meteor. Soc.*, vol. 125, pp. 2887–2908, 1999.

[11] P. Mason and D. Thomson, “Stochastic Backscatter in Large-Eddy Simulations of Boundary Layers,” *Journal of Fluid Mechanics*, vol. 242, no. 1, pp. 51–78, 1992.

[12] G. Shutts, “A Kinetic Energy Backscatter Algorithm for Use in Ensemble Prediction Systems,” *Q. J. R. Meteor. Soc.*, vol. 131, pp. 3079–3102, 2005.

[13] T. Palmer, R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. Shutts, M. Steinheimer, and A. Weisheimer, “Stochastic Parametrization and Model Uncertainty,” *ECMWF, Technical Memorandum*, vol. 598, pp. 1–44, 2009.

[14] D. S. Wilk, “Effects of Stochastic Parameterization on Conceptual Climate Models,” *Phil. Trans. R. Soc.*, vol. 366, pp. 2477–2490, 2008.

[15] T. Palmer, G. Shutts, R. Hagedorn, F. Doblas-Reyes, T. Jung, and M. Leutbecher, “Representing Model Uncertainty in Weather and Climate Prediction,” *Ann.Rev. Earth Planet*, vol. 33, pp. 163–193, 2005.

[16] C. Penland and B. Ewald, “On Modelling Physical Systems with Stochastic Models: Diffusion versus Lévy Processes,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 366, no. 1875, pp. 2455–2474, 2008.

[17] A. Majda, C. Franzke, and B. Khouider, “An Applied Mathematics Perspective on Stochastic Modelling for Climate,” *Phil. Trans. R. Soc.,* A, vol. 366, pp. 2429–2455, 2008.

[18] A. J. Majda, I. Timofeyev, and Vanden-Eijnden, “Models for Stochastic Climate Prediction,” *Proc. Natl Acad. Sci. USA*, vol. 96, pp. 14687–14691, 1999.

[19] A. J. Majda, I. Timofeyev, and Vanden-Eijnden, “A Mathematical Framework for Stochastic Climate Models,” *Commun. Pure Appl. Math.*, vol. 54, pp. 891–974, 2001.

[20] A. J. Majda, I. Timofeyev, and Vanden-Eijnden, “A Priori Tests of a Statistic Mode Reduction Strategy,” *Physica D*, vol. 170, pp. 206–252, 2002.

[21] A. J. Majda, I. Timofeyev, and Vanden-Eijnden, “Systematic Strategies for Stochastic Mode Reduction in Climate,” *J. Atmos. Sci.*, vol. 60, pp. 1705–1722, 2003.

[22] A. J. Majda, I. Timofeyev, and Vanden-Eijnden, “Stochastic Models for Selected Slow Variables in Large Deterministic Systems,” *Nonlinearity*, vol. 19, pp. 769–794, 2006.

[23] C. Franzke, A. J. Majda, and Vanden-Eijnden, “Low-Order Statistic Mode Reduction for a Realistic Barotropic Model Climate,” *J. Atmos. Sci.*, vol. 62, pp. 1722–1745, 2005.

[24] C. Franzke and A. J. Majda, “Low-Order Statistic Mode Reduction for a Prototype Atmopsheric GCM,” *J. Atmos. Sci.*, vol. 63, pp. 457–479, 2006.

[25] P. Sura and P. D. Sardeshmukh, “A Global View of Non-Gaussian SST Variability,” *J. Phys. Oceanogr.*, vol. 38, pp. 639–647, 2008.

[26] P. D. Sardeshmukh, C. Penland, and M. Newman, “Rossby Waves in a Stochastically Fluctuating Medium,” *Progress in Probability*, vol. 49, pp. 369–384, 2001.

[27] P. D. Sardeshmukh, C. Penland, and M. Newman, “Drifts Induced by Multiplicative Red Noise with Application to Climate,” *Europhysics Letters*, vol. 63, pp. 498–504, 2003.

[28] Monahan, Pandolfo, and Imkeller, “Stochastic Confinement of Rossby Waves by Fluctuating Eastward Flows,” *Progress in Probability*, vol. 49, pp. 325–342, 2001.

[29] A. Jentzen and P. E. Kloeden, “The Numerical Approximation of Stochastic Partial Differential Equations,” *Milan J. Math.*, vol. vol. 77 no. 1, pp. 205–244, 2009.

[30] B. Oksendal, *Stochastic Differential Equations: An Introduction with Applications*. Springer Verlag, 2003.

[31] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, vol. 293. Springer Verlag, 1999.

[32] P. Lee, *Bayesian Statistics*. Arnold London, UK:, 1997.

[33] O. Nikodym, “Sur une Généralisation des Intégrales de MJ Radon,” *Fund. Math*, vol. 15, pp. 131–179, 1930.

[34] R. Mansuy, “The Origins of the Word "Martingale"," *Electronic Journal for History of Probability and Statistics*, vol. 5, no. 1, pp. 1–10, 2009.

[35] R. Cotgrave, *A Dictionarie of the French and English Tongues*. Islip, 1611.

[36] J. Ville, “Etude Critique de la Notion de Collectif,” *Bull. Amer. Math. Soc. 45 (1939), 824. DOI: 10.1090/S0002-9904-1939-07089-4 PII: S*, vol. 2, no. 9904, pp. 07089–4, 1939.

[37] J. Ville, “A Letter to P. Crépel 5th February 1985,” 1985.

[38] *Académie Française: Dictionnaire de l’Académie Françoise*. Paris: Veuve Brunet, 5th ed., 1762.

[39] D. Williams, *Probability with Martingales*. Cambridge Univ Pr, 1991.

[40] R. Aumann, “The St. Petersburg Paradox: A Discussion of Some Recent Comments,” *Journal of Economic Theory*, vol. 14, no. 2, pp. 443–445, 1977.

[41] P. Protter, *Stochastic Integration and Differential Equations*, vol. 21. Springer Verlag, 2004.

[42] R. Manella and V. Peter, "Itô versus Stratonovich: 30 Years Later," *Fluctuation and Noise Letters*, vol. 11, no. 01, 2012.

[43] R. Yuan and P. Ao, "Beyond Itô vs. Stratonovich," *Arxiv preprint arXiv:1203.6600*, 2012.

[44] P. E. Kloeden and E. Platen, *Numerical Solutions of Stochastic Differential Equation*. Springer Verlag, Berlin, 1992.

[45] R. Miller, "Topics in Data Assimilation: Stochastic Processes," *Physica D: Nonlinear Phenomena*, vol. 230, no. 1-2, pp. 17–26, 2007.

[46] B. West, A. Bulsara, K. Lindenberg, V. Seshadri, and K. Shuler, "Stochastic Processes with Non-Additive Fluctuations: I. Itô and Stratonovich Calculus and the Effects of Correlations," *Physica A: Statistical and Theoretical Physics*, vol. 97, no. 2, pp. 211–233, 1979.

[47] A. Bulsara, K. Lindenberg, V. Seshadri, K. Shuler, and B. West, "Stochastic Processes with Non-Additive Fluctuations: II. Some Applications of Itô and Stratonovich Calculus," *Physica A: Statistical Mechanics and its Applications*, vol. 97, no. 2, pp. 234–243, 1979.

[48] J. Smythe, F. Moss, P. McClintock, and D. Clarkson, "Itô versus Stratonovich Revisited," *Physics Letters A*, vol. 97, no. 3, pp. 95–98, 1983.

[49] I. Sokolov, "Itô, Stratonovich, Hänggi and all the Rest: The Thermodynamics of Interpretation," *Chemical Physics*, vol. 375, no. 2-3, pp. 359–363, 2010.

[50] S. Sethi and J. Lehoczky, "A Comparison of the Itô and Stratonovich Formulations of Problems in Finance," *Journal of Economic Dynamics and Control*, vol. 3, pp. 343–356, 1981.

[51] C. Braumann, "Harvesting in a Random Environment: Itô or Stratonovich Calculus?," *Journal of Theoretical Biology*, vol. 244, no. 3, pp. 424–432, 2007.

[52] C. Braumann, "Population Growth in Random Environments: Which Stochastic Calculus?," *Bulletin of the International Statistical Institute, LXII*, vol. 56^{th} Session, 2008.

[53] P. Ao, "Potential in Stochastic Differential Equations: Novel Construction," *Journal of physics A: mathematical and general*, vol. 37, p. L25, 2004.

[54] L. Yin and P. Ao, "Existence and Construction of Dynamical Potential in Nonequilibrium Processes without Detailed Balance," *Journal of Physics A: Mathematical and General*, vol. 39, p. 8593, 2006.

[55] P. E. Kloeden and A. Jentzen, "Pathwise Convergent Higher Order Numerical Schemes for Random Ordinary Differential Equations," *Proc. Roy. Soc. A*, vol. vol. 463 no. 2087, pp. 2929–2944, 2007.

[56] "G8 Summit of Deauville, Declaration: Renewed Commitment for Freedom and Democracy." http://www.g20-g8.com/g8-g20/g8/english/live/news/renewed-commitment-for-freedom-and-democracy.1314.html, May 26-27 2011.

[57] U. Marconi, A. Puglisi, L. Rondoni, and A. Vulpiani, "Fluctuation–Dissipation: Response Theory in Statistical Physics," *Physics Reports*, vol. 461, no. 4, pp. 111–195, 2008.

[58] E. Mach, *The Science of Mechanics*. Chicago: Open Court, 1883.

[59] J. Gibbs, "Elementary Principles in Statistical Physics," *Yale University Press*, 1902.

[60] P. van der Pas, “The Discovery of the Brownian Motion,” *Scien. Historiae*, vol. 13, no. 17, 1971.

[61] A. Einstein, “Motion of Suspended Particles on the Kinetic Theory,” *Annalen der Physik*, vol. 17, no. 3, pp. 549–560, 1905.

[62] A. Einstein, “On the Theory of the Brownian Movement,” *Annalen der Physik*, vol. 4, pp. 371–381, 1906.

[63] M. von Smoluchowski, “Zur Kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen,” *Annalen der Physik*, vol. 326, no. 14, pp. 756–780, 1906.

[64] P. Langevin, “Sur la Théorie du Mouvement Brownien,” *CR Acad. Sci. Paris*, vol. 146, no. 530-533, 1908.

[65] J. Perrin, *Les Atoms*. Paris: Alcan, 1913.

[66] H. Nyquist, “Thermal Agitation of Electric Charge in Conductors,” *Physical Review*, vol. 32, no. 1, pp. 110–113, 1928.

[67] L. Onsager, “Reciprocal Relations in Irreversible Processes. I.,” *Physical Review*, vol. 37, pp. 405–426, 1931.

[68] L. Onsager, “Reciprocal Relations in Irreversible Processes. II.,” *Physical Review*, vol. 38, pp. 2265–2279, 1931.

[69] H. Callen and T. Welton, “Irreversibility and Generalized Noise,” *Physical Review*, vol. 83, no. 1, pp. 34–40, 1951.

[70] R. Kubo, “The Fluctuation-Dissipation Theorem,” *Reports on Progress in Physics*, vol. 29, p. 255, 1966.

[71] M. Green, “Markoff Random Processes and the Statistical Mechanics of Time-Dependent Phenomena. II. Irreversible Processes in Fluids,” *The Journal of Chemical Physics*, vol. 22, p. 398, 1954.

[72] R. Kubo, “Statistical-Mechanical Theory of Irreversible Processes. I. General Theory and Simple Applications to Magnetic and Conduction Problems,” *Journal of the Physical Society of Japan*, vol. 12, no. 6, pp. 570–586, 1957.

[73] N. Van Kampen, “The Case Against Linear Response Theory,” *Physica Norvegica*, vol. 5, no. 3–4, pp. 279–284, 1971.

[74] R. Kubo, “Brownian Motion and Nonequilibrium Statistical Mechanics,” *Science*, vol. 233, no. 4761, pp. 330–334, 1986.

[75] D. Evans, E. Cohen, and G. Morriss, “Probability of Second Law Violations in Shearing Steady States,” *Physical Review Letters*, vol. 71, no. 15, pp. 2401–2404, 1993.

[76] D. Evans and D. Searles, “Steady States, Invariant Measures, and Response Theory,” *Physical Review E*, vol. 52, no. 6, p. 5839, 1995.

[77] D. Evans and D. Searles, “Equilibrium Microstates which Generate Second Law Violating Steady States,” *Physical Review E*, vol. 50, pp. 1645–1648, 1994.

[78] T. Tatsumi, “Non-equilibrium Statistical Mechanics of Fluid Turbulence,” *Progress in Turbulence and Wind Energy IV*, pp. 49–55, 2012.

[79] L. Leuzzi and T. Nieuwenhuizen, *Thermodynamics of the Glassy State*. Taylor & Francis Group, 2008.

[80] B. Dybiec, J. Parrondo, and E. Gudowska-Nowak, “Fluctuation-Dissipation Relations under Levy Noises,” *Arxiv preprint, arXiv:1201.1752*, 2012.

[81] P. Ditlevsen, “The Role of Stochastic Forcing in Rapid Climate Transitions,” in *EGU General Assembly Conference Abstracts*, vol. 14, p. 7873, 2012.

[82] C. Leith, “Climate Response and Fluctuation Dissipation,” *Journal of the Atmospheric Sciences*, vol. 32, no. 10, pp. 2022–2026, 1975.

[83] C. Leith, “Predictability of Climate,” 1978.

[84] A. Majda, R. Abramov, and M. Grote, *Information Theory and Stochastics for Multiscale Nonlinear Systems*, vol. 25. Amer Mathematical Society, 2005.

[85] T. Bell, “Climate Sensitivity from Fluctuation Dissipation: Some Simple Model Tests.,” *Journal of Atmospheric Sciences*, vol. 37, pp. 1700–1707, 1980.

[86] G. Carnevale, M. Falcioni, S. Isola, R. Purini, and A. Vulpiani, “Fluctuation-Response Relations in Systems with Chaotic Behavior,” *Physics of Fluids A: Fluid Dynamics*, vol. 3, p. 2247, 1991.

[87] A. Gritsoun and V. Dymnikov, “Barotropic Atmosphere Response to Small External Actions: Theory and Numerical Experiments,” *Izvestiia-Russian Acedemy of Sciences Atmospheric and Oceanic Physics*, vol. 35, pp. 511–525, 1999.

[88] A. Gritsoun, “Fluctuation-Dissipation Theorem on Attractors of Atmospheric Models,” *Russian Journal of Numerical Analysis and Mathematical Modelling*, vol. 16, no. 2, pp. 115–134, 2001.

[89] A. Gritsoun, G. Branstator, and V. Dymnikov, “Construction of the Linear Response Operator of an Atmospheric General Circulation Model to Small External Forcing,” *Russian Journal of Numerical Analysis and Mathematical Modelling*, vol. 17, no. 5, pp. 399–416, 2002.

[90] D. Ruelle, “General Linear Response Formula in Statistical Mechanics, and the Fluctuation-Dissipation Theorem far from Equilibrium,” *Physics Letters A*, vol. 245, no. 3-4, pp. 220–224, 1998.

[91] D. Ruelle, “A Review of Linear Response Theory for General Differentiable Dynamical Systems,” *Nonlinearity*, vol. 22, p. 855, 2009.

[92] R. Abramov and A. Majda, “New Approximations and Tests of Linear Fluctuation-Response for Chaotic Nonlinear Forced-Dissipative Dynamical Systems,” *Journal of Nonlinear Science*, vol. 18, no. 3, pp. 303–341, 2008.

[93] A. Majda, R. Abramov, and B. Gershgorin, “High Skill in Low-Frequency Climate Response Through Fluctuation Dissipation Theorems Despite Structural Instability,” *Proceedings of the National Academy of Sciences*, vol. 107, no. 2, pp. 581–586, 2010.

[94] A. Majda and X. Wang, “Linear Response Theory for Statistical Ensembles in Complex Systems with Time-Periodic Forcing,” *Communications in Mathematical Sciences*, vol. 8, no. 1, pp. 145–172, 2010.

[95] G. North, R. Bell, and J. Hardin, “Fluctuation Dissipation in a General Circulation Model,” *Climate Dynamics*, vol. 8, no. 6, pp. 259–264, 1993.

[96] I. Cionni, G. Visconti, and F. Sassi, “Fluctuation Dissipation Theorem in a General Circulation Model,” *Geophysical Research Letters*, vol. 31, p. L09206, 2004.

[97] R. Haskins, R. Goody, and L. Chen, “Radiance Covariance and Climate Models,” *Journal of Climate*, vol. 12, no. 5, pp. 1409–1422, 1999.

[98] J. Von Storch, “On Statistical Dissipation in GCM-Climate,” *Climate dynamics*, vol. 23, no. 1, pp. 1–15, 2004.

[99] J. Houghton, Y. Ding, D. Griggs, M. Noguer, P. van der Linden, X. Dai, K. Maskell, and C. Johnson, *IPCC TAR WG1 Climate Change 2001: The Scientific Basis, Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2001.

[100] R. Pachauri and A. Reisinger, eds., *IPCC AR4 SYR Climate Change 2007: Synthesis Report, Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC, 2007.

[101] F. Hollander, "Large Deviations (Fields Institute Monographs, 14)," *American Mathematical Society*, 2000.

[102] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, vol. 38. Springer Verlag, 2009.

[103] P. Häunggi and P. Jung, "Colored Noise in Dynamical Systems," *Advances in Chemical Physics*, pp. 239–326, 1995.

[104] L. Dubins and G. Schwarz, "On Continuous Martingales," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 53, no. 5, p. 913, 1965.

[105] K. Itô, "0n Stochastic Differential Equations," *American Mathematical Society, New York*, vol. 4, pp. 1–51, 1951.

[106] M. Milankovic, "Kanon der Erdbestrahlung," *Acad. Roy. Serbe, Editions spec, Tome CXXXIII.-1941*, 1941.

[107] Budyko, "The Effect of Solar Radiation Variations on the Climate of the Earth," *Tellus*, vol. 21, pp. 611–619, 1969.

[108] W. Sellers, "A Global Climate Model Based on the Energy Balance of the Earth-Atmosphere System," *J. Appl. Meteor.*, vol. 8, pp. 396–400, 1969.

[109] K. Fraedrich, "Structural and Stochastic Analysis of a Zero-Dimensional Climate System," *Q.J.R. Meteorol. Soc.*, vol. 104, pp. 461–474, 1978.

[110] G. North, "Energy Balance Models," *Rev. Geophys.*, vol. 19, pp. 91–121, 1981.

[111] M. Ghil, *Climate Variations and Variability: Facts and Theory*, pp. 461–480. D. Reidel Publishing Company, 1981.

[112] R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani, "Sochastic Resonance in Climate Changes," *Tellus*, vol. 34, pp. 10–16, 1982.

[113] M. Wahlen *et al.*, "Vostok Ice Core CO2 Data, 1105-2856m." IGBP PAGES/World Data Center-A for Paleoclimatology Data Contribution Series No. 2000-003. NOAA/NGDC Paleoclimatology Program, Boulder CO, USA, 2000.

[114] K. Kawamura *et al.*, "Dome Fuji Ice Core 338KYr Wet Extraction CO2 Data." IGBP PAGES/World Data Center for Paleoclimatology Data Contribution Series No. 2007-074. NOAA/NCDC Paleoclimatology Program, Boulder CO, USA., 2007.

[115] J. Ahn and E. Brook, "Byrd Ice Core CO2 Data for 20-90 KYrBP." IGBP PAGES/World Data Center for Paleoclimatology Data Contribution Series No.2008-095 NOAA/NCDC Paleoclimatology Program, Boulder CO, USA., 2008.

[116] E. Monnin *et al.*, "EPICA Dome C Ice Core High Resolution Holocene and Transition CO2 Data." IGBP PAGES/World Data Center for Paleoclimatology Data Contribution Series No. 2004-055. NOAA/NGDC Paleoclimatology Program, Boulder CO, USA, 2004.

[117] A. Indermuehle, B. Stauffer, T. Stocker, and M. Wahlen, "Taylor Dome Ice Core CO2 Holocene Data." IGBP PAGES/World Data Center-A for Paleoclimatology Data Contribution Series No.1999-021 NOAA/NGDC Paleoclimatology Program, Boulder CO, USA NOAA/NGDC Paleoclimatology Program, Boulder CO, USA, 1999.

[118] A. Indermuehle, B. Stauffer, T. Stocker, and M. Wahlen, “Taylor Dome 60-20 KYrBP Ice Core CO2 Data.” IGBP PAGES/World Data Center-A for Paleoclimatology Data Contribution Series No. 2000-016. NOAA/NGDC Paleoclimatology Program, Boulder CO, USA, 2000.

[119] J. Doob, “The Brownian Movement and Stochastic Equations,” *The annals of Mathematics*, vol. 43, no. 2, pp. 351–369, 1942.

[120] H. Von Storch and F. Zwiers, *Statistical analysis in climate research*. Cambridge Univ Pr, 2002.

[121] J. Hamilton, *Time Series Analysis*, vol. 2. Cambridge Univ Press, 1994.

[122] T. Mills, *Time Series Techniques for Economists*. Cambridge Univ Pr, 1991.

[123] M. Priestley, “Spectral Analysis and Time Series,” 1981.

[124] M. De Hoon, T. Van der Hagen, H. Schoonewelle, and H. Van Dam, “Why yule-walker should not be used for autoregressive modelling,” *Annals of Nuclear Energy*, vol. 23, no. 15, pp. 1219–1228, 1996.

[125] J. Burg, “A New Analysis Technique for Time Series Data,” *NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics*, pp. 12–23, 1968.

[126] J. Burg, “Maximum Entropy Spectral Analysis,” *SEP6*, vol. 6, 1975.

[127] J. Laskar, P. Robutel, F. Joutel, M. Gastineau, A. Correia, and B. Levrard, “A Long-Term Numerical Solution for the Insolation Quantities of the Earth,” *Astronomy and Astrophysics*, vol. 428, no. 1, pp. 261–285, 2004.

[128] C. Navier, “Memoir on the Laws of Fluid Motion,” *Mem., acad. sci.(Paris)*, vol. 6, p. 389, 1827.

[129] G. Stokes, *On the Friction of Fluids in Motion, and the Equilibrium and Motion of Elastic Solids*. 1846.

[130] L. Euler, “Principes Généraux du Mouvement des Fluides,” *Mémoires de l’Académie des Sciences de Berlin*, vol. 11, pp. 274–315, 1757.

[131] C. Fefferman, “Official Clay Prize Problem Description: Existence and Smoothness of the Navier-Stokes Equation,” 2000.

[132] O. Ladyzhenskaia, R. Silverman, and J. Chu, *The Mathematical Theory of Viscous Incompressible Flow*. Gordon & Breach, 1969.

[133] J. Holton, *An Introduction to Dynamic Meteorology*, vol. 1. Academic press, 2004.

[134] L. Richardson, “Weather Prediction by Numerical Methods.” Cambridge University Press, London, 1922.

[135] J. Lions, R. Temam, and S. Wang, “New Formulations of the Primitive Equations of Atmosphere and Applications,” *Nonlinearity*, vol. 5, pp. 237–288, 1992.

[136] R. Temam and M. Ziane, “Some Mathematical Problems in Geophysical Fluid Dynamics,” *Handbook of Mathematical Fluid Dynamics*, vol. 3, pp. 535–658, 2005.

[137] M. Petcu, R. Temam, and D. Wirosoetisno, “Existence and Regularity Results for the Primitive Equations in Two Space Dimensions,” *Communications on Pure and Applied Analysis*, vol. 3, no. 1, pp. 115–132, 2004.

[138] C. Cao and T. C.S., “Global Well-posedness of the Three-dimensional Viscous Primitive Equations of Large Scale Ocean and Atmosphere Dynamics,” *Annals of Mathematics*, vol. 165, pp. 245–267, 2007.

[139] B. Ewald and C. Penland, “Numerical Generation of Stochastic Differential Equations in Climate Models,” *Handbook of Numerical Analysis*, vol. 14, pp. 279–306, 2009.

[140] C. Penland and P. Sardeshmukh, “The Optimal Growth of Tropical Sea Surface Temperature Anomalies,” *Journal of climate*, vol. 8, no. 8, pp. 1999–2024, 1995.

[141] H. Rose, “Eddy Diffusivity, Eddy Noise and Subgrid-Scale Modelling,” *Journal of Fluid Mechanics*, vol. 81, pp. 719–734, 1977.

[142] D. Leslie and G. Quarini, “The Application of Turbulence Theory to the Formulation of Subgrid Modelling Procedures,” *Journal of fluid mechanics*, vol. 91, no. 01, pp. 65–91, 1979.

[143] J. Berner, G. Shutts, M. Leutbecher, and T. Palmer, “A Spectral Stochastic Kinetic Energy Backscatter Scheme and its Impact on Flow-Dependent Predictability in the ECMWF Ensemble Prediction System,” *Journal of the Atmospheric Sciences*, vol. 66, no. 3, pp. 603–626, 2009.

[144] M. Zidikheri and J. Frederiksen, “Stochastic Subgrid-Scale Modelling for Non-Equilibrium Geophysical Flows,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 368, no. 1910, pp. 145–160, 2010.

[145] N. Glatt-Holtz and M. Ziane, “The Stochastic Primitive Equations in Two Space Dimensions with Multiplicative Noise,” *Discrete Contin. Dyn. Syst. Ser. B*, vol. 10, no. 4, pp. 801–822, 2008.

[146] B. Ewald, M. Petcu, and R. Temam, “Stochastic Solutions of the Two-Dimensional Primitive Equations of the Ocean and Atmosphere with an Additive Noise,” *Analysis and Applications*, vol. 5, no. 2, p. 183, 2007.

[147] N. Glatt-Holtz and R. Temam, “Pathwise Solutions of the 2-D Stochastic Primitive Equations,” *Applied Mathematics & Optimization*, vol. 63, no. 3, pp. 401–433, 2011.

[148] B. Guo and D. Huang, “3D Stochastic Primitive Equations of the Large-Scale Ocean: Global Well-Posedness and Attractors,” *Communications in Mathematical Physics*, vol. 286, no. 2, pp. 697–723, 2009.

[149] P. Azérad and F. Guillén, “Mathematical Justification of the Hydrostatic Approximation in the Primitive Equations of Geophysical Fluid Dynamics,” *SIAM journal on mathematical analysis*, vol. 33, no. 4, pp. 847–859, 2001.

[150] J. Lions, R. Temam, and S. Wang, “Mathematical Study of the Coupled Models of Atmosphere and Ocean (CAO III),” *J. Math. Pures Appl*, vol. 74, pp. 105–163, 1995.

[151] Q. Zeng, “Mathematical and Physical Foundations of Numerical Weather Prediction,” 1979.

[152] B. Ewald and R. Temam, “Maximum Principles for the Primitive Equations of the Atmosphere,” *Discrete and Continuous Dynamical Systems*, vol. 7, no. 2, pp. 343–362, 2001.

[153] C. Jifan, “Long-Term Numerical Weather Prediction,” *China Meteorological Press, Beijing*, vol. 216, p. 231, 1986.

[154] S. Wang, J. Huang, and J. Chou, “Some Properties of the Equations of Large Scale Atmospheric Motion,” *Science in China*, pp. 328–36, 1989.

[155] S. Wang, *On Solvability for the Equations of the Large-Scale Atmospheric Motion*. PhD thesis, Ph. D. thesis, Lanzhou University, China, 1988.

[156] R. Temam, *Navier-Stokes Equations: Theory and Numerical Analysis*, vol. 2. Amer Mathematical Society, 2001.

[157] L. Brouwer, “Über Abbildung von Mannigfaltigkeiten,” *Mathematische Annalen*, vol. 71, no. 4, pp. 598–598, 1912.

[158] J. Schauder, “Der Fixpunktsatz in Funktionalräumen,” *Studia math*, vol. 2, pp. 171–180, 1930.

[159] G. Da Prato and J. Zabczyk, “Stochastic Equations in Infinite Dimensions,” *Encyclopedia of Mathematics and its Applications*, vol. 44, 1992.

[160] C. Prévôt and M. Röckner, *A Concise Course on Stochastic Partial Differential Equations*. No. 1905, Springer Verlag, 2007.

[161] G. Da Prato and J. Zabczyk, *Stochastic Equations in Infinite Dimensions*. Cambridge Univ Press, 2008.

[162] I. Gyöngy, “Lattice Approximations for Stochastic Quasi-Linear Parabolic Partial Differential Equations Driven by Space-Time White Noise I,” *Potential Analysis*, vol. 9, no. 1, pp. 1–25, 1998.

[163] I. Gyöngy, “Lattice Approximations for Stochastic Quasi-Linear Parabolic Partial Differential Equations Driven by Space-Time White Noise II,” *Potential Analysis*, vol. 11, no. 1, pp. 1–37, 1999.

[164] A. Davie and J. Gaines, “Convergence of Numerical Schemes for the Solution of Parabolic Stochastic Partial Differential Equations,” *Mathematics of Computation*, vol. 70, no. 233, pp. 121–134, 2001.

[165] A. Jentzen and P. Kloeden, “Overcoming the Order Barrier in the Numerical Approximation of Stochastic Partial Differential Equations with Additive Space–Time Noise,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, vol. 465, no. 2102, pp. 649–667, 2009.

[166] A. Jentzen and M. Röckner, “A Break of the Complexity of the Numerical Approximation of Nonlinear SPDEs with Multiplicative Noise,” *Arxiv preprint arXiv:1001.2751*, 2010.

[167] A. Jentzen and M. Röckner, “Regularity Analysis for Stochastic Partial Differential Equations with Nonlinear Multiplicative Trace Class Noise,” *Journal of Differential Equations*, 2011.

[168] T. Tachim Medjo, “Optimal Control of the Primitive Equations of the Ocean with State Constraints,” *Nonlinear Analysis: Theory, Methods & Applications*, vol. 73, no. 3, pp. 634–649, 2010.

[169] G. Da Prato, A. Jentzen, and M. Roeckner, “A Mild Itô Formula for SPDEs,” *Arxiv preprint, arXiv:1009.3526*, 2010.

[170] Y. Hu, “Semi-Implicit Euler-Maruyama Scheme for Stiff Stochastic Equations,” *Progress in Probability*, pp. 183–202, 1996.

[171] D. Higham, X. Mao, and A. Stuart, “Strong Convergence of Euler-Type Methods for Nonlinear Stochastic Differential Equations,” *SIAM Journal on Numerical Analysis*, pp. 1041–1063, 2003.

[172] M. Hutzenthaler, A. Jentzen, and P. Kloeden, “Strong Convergence of an Explicit Numerical Method for SDEs with Non-Globally Lipschitz Continuous Coefficients,” *Arxiv preprint arXiv:1010.3756*, 2010.

[173] G. Hardy, J. Littlewood, and G. Pólya, *Inequalities*. Cambridge Univ Press, 1952.

[174] X. Wang, “A Runge-Kutta Type Scheme for Nonlinear Stochastic Partial Differential Equations with Multiplicative Trace Class Noise,” *Arxiv preprint arXiv:1106.1889*, 2011.

[175] M. Hutzenthaler and A. Jentzen, “Non-Globally Lipschitz Counterexamples for the Stochastic Euler Scheme,” *Arxiv preprint ArXiv:0905.0273*, 2009.

[176] A. Jentzen, P. Kloeden, and A. Neuenkirch, “Pathwise Approximation of Stochastic Differential Equations on Domains: Higher Order Convergence Rates without Global Lipschitz Coefficients,” *Numerische Mathematik*, vol. 112, no. 1, pp. 41–64, 2009.

[177] A. Debussche, N. Glatt-Holtz, R. Temam, and M. Ziane, “Global Existence and Regularity for the 3D Stochastic Primitive Equations of the Ocean and Atmosphere with Multiplicative White Noise,” *Arxiv preprint, arXiv:1104.4754*, 2011.

[178] N. Glatt-Holtz and M. Ziane, “Strong Pathwise Solutions of the Stochastic Navier-Stokes System,” *Advances in Differential Equations*, vol. 14, no. 5-6, pp. 567–600, 2009.

[179] P. Abrahamsen, *A Review of Gaussian Random Fields and Correlation Functions*. Norsk Regnesentral/Norwegian Computing Center, 1997.

[180] D. Khoshnevisan, *Multiparameter Processes: An Introduction to Random Fields*. Springer Verlag, 2002.

[181] I. Gyöngy and D. Nualart, “Implicit Scheme for Stochastic Parabolic Partial Diferential Equations Driven by Space-Time White Noise,” *Potential Analysis*, vol. 7, no. 4, pp. 725–757, 1997.

[182] K. Triantafyllopoulos, “On the Central Moments of the Multidimensional Gaussian Distribution,” *The Mathematical Scientist*, vol. 28, pp. 125–128, 2003.

[183] H. Risken, *The Fokker-Planck Equation*. Springer, Berlin, 1996.

List of Figures

List of Tables

A Pending Proofs

For the technical calculations in the remaining proofs we need some results on the moments of Brownian motions.

Lemma A.1. Mixed Moments of Gaussian Random Variables
Let $\boldsymbol{\xi} = (\xi_i)_{i=1}^r$, $r \in \mathbb{N}$ be a real-valued multivariate Gaussian distributed random variable with zero mean. Denote by

$$\sigma_{i,j} = \mathit{cov}\left(\xi_i, \xi_j\right), \quad i,j \in \{1,2,...,r\}$$

the pairwise covariances. Then the mixed moments satisfy

$$\begin{aligned} E\left[\xi_1 ... \xi_{2n+1}\right] &= 0, && (2n+1) \in \{1,2,...,r\} \\ E\left[\xi_1 ... \xi_{2n}\right] &= \sum_{P_d} \sigma_{i_1,i_2} ... \sigma_{i_{2n-1},i_{2n}}, && 2n \in \{1,2,...,r\}, \end{aligned}$$

where P_d is the set only of those index permutations of $(1,...,2n)$ which yield different products $\sigma_{i_1,i_2} ... \sigma_{i_{2n-1},i_{2n}}$. Regarding the cardinal number of P_d, one gets

$$|P_d| = \frac{(2n)!}{2^n n!}.$$

Proof. *Lemma A.1*
Lemma A.1 is a well known result in the field of multivariate Gaussian statistics and is for instance derived in [182] or [183]. □

Corollary A.2.
For central moments of a Gaussian distributed random variable $X \sim \mathcal{N}(0,\sigma^2)$, it holds true that

$$E\left[X^n\right] = \begin{cases} \Psi_k \sigma^{2k} & , \quad n = 2k \\ 0 & , \quad n = 2k+1 \end{cases}, \qquad \Psi_k := \frac{(2k)!}{2^k k!}. \tag{A.1}$$

Proof. *Corollary A.2*
Corollary A.2 is a direct consequence of Lemma A.1. □

Lemma A.3.
For $i + j = 2m$ we have

$$E\left[W_{s\wedge t}^i W_{s\vee t}^j\right] = \sum_{0 \le k \le \frac{j}{2}} \binom{j}{2k} \Psi_{m-k} \Psi_k (s \wedge t)^{m-k} (s \vee t - s \wedge t)^k,$$

and

$$E\left[W_{s\wedge t}^i W_{s\vee t}^j\right] = 0,$$

for odd $i + j$.

Proof. *Lemma A.3*

$$\begin{aligned} E\left[W_{s\wedge t}^i W_{s\vee t}^j\right] &= E\left[W_{s\wedge t}^i \Big(W_{s\wedge t} + \big(W_{s\vee t} - W_{s\wedge t}\big)\Big)^j\right] \\ &= E\left[W_{s\wedge t}^i \sum_{0 \le k \le j} \binom{j}{k} W_{s\wedge t}^{j-k} \big(W_{s\vee t} - W_{s\wedge t}\big)^k\right] \\ &= \sum_{0 \le k \le j} \binom{j}{k} E\left[W_{s\wedge t}^{i+j-k}\right] E\left[\big(W_{s\vee t} - W_{s\wedge t}\big)^k\right] \end{aligned}$$

Where the last equation holds true due to the independence of increments of Brownian motions. Considering (A.1), the first expectation value vanishes for odd $i+j$, and the second one vanishes for odd k. It follows for $i+j=2m$

$$E\left[W^i_{s\wedge t}W^j_{s\vee t}\right] = \sum_{0\le k\le \frac{j}{2}} \binom{j}{2k} E\left[W^{2m-2k}_{s\wedge t}\right] E\left[\left(W_{s\vee t}-W_{s\wedge t}\right)^{2k}\right].$$

Finally, note that $W_{t_i} - W_{t_j} \sim \mathcal{N}(0, t_i - t_j)$. □

Corollary A.4.

i) $E[W_t] = E\left[W_s^2 W_t\right] = 0$

ii) $E\left[W_t^2\right] = t$

iii) $E[W_s W_t] = s \wedge t$

iv) $E\left[W_s^2 W_t^2\right] = st + 2(s\wedge t)^2$

v) $E\left[W_t^4\right] = 3t^2$

Proof. *Corollary A.4*
Corollary A.4 is derived from basic properties of Brownian motions and Lemma A.3. □

Proposition A.5.
For all $0 < s \le t$ *we have*

$$\int_0^s \int_0^t e^{-\Theta|u-r|} du\, dr = \frac{1}{\Theta}\Big(2s - \big(1+e^{-\Theta(t-s)}\big)\varphi_s\Big).$$

Proof. *Proposition A.5*

$$\begin{aligned}
\int_0^s \int_0^t e^{-\Theta|u-r|} \mathrm{du}\, \mathrm{dr} &= \int_0^s \left(\int_0^r e^{\Theta(u-r)} \mathrm{du} + \int_r^t e^{-\Theta(u-r)} \mathrm{du}\right) \mathrm{dr} \\
&= \int_0^s \left(\frac{1}{\Theta}\Big[e^{\Theta(u-r)}\Big]_0^r - \frac{1}{\Theta}\Big[e^{-\Theta(u-r)}\Big]_r^t\right) \mathrm{dr} \\
&= \frac{1}{\Theta}\int_0^s \left(2 - e^{-\Theta r} - e^{-\Theta(t-r)}\right) \mathrm{dr} \\
&= \frac{1}{\Theta}\left(2s + \frac{1}{\Theta}\Big[e^{-\Theta r} - e^{-\Theta(t-r)}\Big]_0^s\right) \\
&= \frac{1}{\Theta}\Big(2s + \frac{1}{\Theta}\Big(e^{-\Theta s} - 1 - e^{\Theta(s-t)} + e^{-\Theta t}\Big)\Big) \\
&= \frac{1}{\Theta}\Big(2s + \frac{1}{\Theta}\Big(1+e^{\Theta(s-t)}\Big)\Big(e^{-\Theta s} - 1\Big)\Big) \\
&= \frac{1}{\Theta}\Big(2s - \big(1+e^{-\Theta(t-s)}\big)\varphi_s\Big)
\end{aligned}$$

□

A.1 Proof: Lemma 3.12

For the convenience of the reader we recall

Lemma 3.12. Statistical Properties of T
Consider system (3.8) *driven by a stationary OUP* ϵ *with parameters* μ, Θ, D. *Then*

$$E[T_t] = T_0 \exp\Big(-\mu t + \frac{D}{2\Theta^2}\big(t - \varphi_t\big)\Big) \tag{3.12}$$

$$\mathit{var}(T_t) = T_0^2 \exp\Big(-2\mu t + \frac{D}{\Theta^2}\big(t-\varphi_t\big)\Big)\Big[\exp\Big(\frac{D}{\Theta^2}(t-\varphi_t)\Big) - 1\Big] \tag{3.13}$$

$$\begin{aligned}\mathit{cov}(T_s, T_t) = T_0^2 \exp\Big(&-\mu(s+t) + \frac{D}{2\Theta^2}\big(s+t-(\varphi_s+\varphi_t)\big)\Big) \\ &\times\Big[\exp\Big(\frac{D}{2\Theta^2}\big(2(s\wedge t) - (1+e^{-\Theta|t-s|})\varphi_{s\wedge t}\big)\Big) - 1\Big]\end{aligned} \tag{3.14}$$

Proof. *Lemma 3.12*
We aim to derive an expression for the mixed moment $E[T_sT_t]$, which yields (3.12) for $s = 0$. Using these two expressions, the covariance can be calculated, which shows (3.14) and gives us (3.13) for $s = t$. We denote the characteristic function of an interval $[a, b]$ by

$$I_{[a,b]}(x) = \begin{cases} 1, \text{ for } x \in [a,b] \\ 0, \text{ else} \end{cases} .$$

By means of Taylor expansions and due to the linearity of the Lebesgue integral, we obtain

$$\begin{aligned}
&E[T_sT_t]\, T_0^{-2} e^{\mu(s+t)} \\
&= T_0^{-2} e^{\mu(s+t)} E\Big[T_0^2 e^{-\mu(s+t)} \exp\Big(\int_0^s \epsilon_u du + \int_0^t \epsilon_u du\Big)\Big] \\
&= E\Big[\exp\Big(\int_0^s \epsilon_u du + \int_0^t \epsilon_u du\Big)\Big] \\
&= E\Big[\exp\Big(\int_0^{s\vee t} \big(I_{[0,s]}(u) + I_{[0,t]}(u)\big)\epsilon_u du\Big)\Big] \\
&= \sum_{m=0}^{\infty} \frac{1}{m!} E\Big[\Big(\int_0^{s\vee t} \big(I_{[0,s]}(u) + I_{[0,t]}(u)\big)\epsilon_u du\Big)^m\Big] \\
&= \sum_{m=0}^{\infty} \frac{1}{m!} E\Big[\int_0^{s\vee t} \dots \int_0^{s\vee t} \big(I_{[0,s]}(u_1) + I_{[0,t]}(u_1)\big)\epsilon_{u_1} \times \dots \\
&\qquad\qquad \dots \times \big(I_{[0,s]}(u_m) + I_{[0,t]}(u_m)\big)\epsilon_{u_m}\, du_1 \dots du_m\Big] \\
&= \sum_{m=0}^{\infty} \frac{1}{m!} \int_0^{s\vee t} \dots \int_0^{s\vee t} \big(I_{[0,s]}(u_1) + I_{[0,t]}(u_1)\big) \times \dots \\
&\qquad\qquad \dots \times \big(I_{[0,s]}(u_m) + I_{[0,t]}(u_m)\big) E[\epsilon_{u_1} \dots \epsilon_{u_m}]\, du_1 \dots du_m.
\end{aligned}$$

Since ϵ_t is a Gaussian process, the m-dimensional random variable $(\epsilon_{u_i})_{i=1}^m$ obeys a multivariate Gaussian distribution with zero mean and

$$\mathrm{cov}\left(\epsilon_{u_i}, \epsilon_{u_j}\right) = \frac{D}{2\Theta} e^{-\Theta|u_i - u_j|} =: \sigma_{u_i,u_j}.$$

Lemma A.1 yields

$$\begin{aligned}
E\left[\epsilon_{u_1} \dots \epsilon_{u_{2n+1}}\right] &= 0 \\
E[\epsilon_{u_1} \dots \epsilon_{u_{2n}}] &= \sum_{P_d} \sigma_{u_{i_1},u_{i_2}} \dots \sigma_{u_{i_{2n-1}},u_{i_{2n}}},
\end{aligned}$$

which leads to

$$
\begin{aligned}
&E\left[T_s T_t\right] T_0^{-2} e^{\mu(s+t)} \\
&= \sum_{n=0}^{\infty} \frac{1}{(2n)!} \int_0^{s\vee t} \dots \int_0^{s\vee t} \sum_{P_d} \sigma_{u_{i_1},u_{i_1}} \dots \sigma_{u_{i_{2n}},u_{i_{2n}}} \\
&\qquad \times \big(I_{[0,s]}(u_1) + I_{[0,t]}(u_1)\big) \times \dots \times \big(I_{[0,s]}(u_{2n}) + I_{[0,t]}(u_{2n})\big)\, du_1 \dots du_{2n} \\
&= \sum_{n=0}^{\infty} \sum_{P_d} \frac{1}{(2n)!} \int_0^{s\vee t} \dots \int_0^{s\vee t} \sigma_{u_{i_1},u_{i_1}} \times \dots \times \sigma_{u_{i_{2n}},u_{i_{2n}}} \\
&\qquad \times \big(I_{[0,s]}(u_1) + I_{[0,t]}(u_1)\big) \times \dots \times \big(I_{[0,s]}(u_{2n}) + I_{[0,t]}(u_{2n})\big)\, du_1 \dots du_{2n} \\
&= \sum_{n=0}^{\infty} \sum_{P_d} \frac{1}{(2n)!} \int_0^{s\vee t} \int_0^{s\vee t} \big(I_{[0,s]}(u_{i_{2n}}) + I_{[0,t]}(u_{i_{2n}})\big) \\
&\qquad\qquad \times \big(I_{[0,s]}(u_{i_{2n-1}}) + I_{[0,t]}(u_{i_{2n-1}})\big) \sigma_{u_{i_{2n-1}},u_{i_{2n}}} \\
&\qquad \dots \int_0^{s\vee t} \int_0^{s\vee t} \big(I_{[0,s]}(u_{i_2}) + I_{[0,t]}(u_{i_2})\big) \\
&\qquad\qquad \times \big(I_{[0,s]}(u_{i_1}) + I_{[0,t]}(u_{i_1})\big) \sigma_{u_{i_1},u_{i_2}} du_1 du_2 \dots du_{2n-1} du_{2n}.
\end{aligned}
$$

These n double integrals are pairwise independent and can therefore be written as a product. After renaming the integration variables, we obtain

$$
\begin{aligned}
&E\left[T_s T_t\right] T_0^{-2} e^{\mu(s+t)} \\
&= \sum_{n=0}^{\infty} \sum_{P_d} \frac{1}{(2n)!} \Big(\int_0^{s\vee t} \int_0^{s\vee t} \big(I_{[0,s]}(u) + I_{[0,t]}(u)\big)\big(I_{[0,s]}(r) + I_{[0,t]}(r)\big)\, \sigma_{u,r}\, dudr \Big)^n \\
&= \sum_{n=0}^{\infty} \frac{1}{(2n)!} \frac{(2n)!}{2^n n!} \Big(\int_0^{s\vee t} \int_0^{s\vee t} \big(I_{[0,s]}(u) + I_{[0,t]}(u)\big)\big(I_{[0,s]}(r) + I_{[0,t]}(r)\big)\, \sigma_{u,r}\, dudr \Big)^n \\
&= \sum_{n=0}^{\infty} \frac{1}{n!} \Big(\frac{1}{2} \int_0^{s\vee t} \int_0^{s\vee t} \big(I_{[0,s]}(u) + I_{[0,t]}(u)\big)\big(I_{[0,s]}(r) + I_{[0,t]}(r)\big)\, \sigma_{u,r}\, dudr \Big)^n \\
&= \exp\Big(\frac{1}{2} \int_0^{s\vee t} \int_0^{s\vee t} \big(I_{[0,s]}(u) + I_{[0,t]}(u)\big)\big(I_{[0,s]}(r) + I_{[0,t]}(r)\big)\, \sigma_{u,r}\, dudr \Big)
\end{aligned}
$$

To calculate the double integral, we use Proposition A.5 and the fact that the double integral is symmetric in s and t since the integrand is symmetric in u and r. Furthermore, note that for every symmetric function f it holds true that $f(s,t) = f(t,s) = f(s \wedge t, s \vee t)$.

$$
\begin{aligned}
&\int_0^{s\vee t} \int_0^{s\vee t} \big(I_{[0,s]}(u) + I_{[0,t]}(u)\big)\big(I_{[0,s]}(r) + I_{[0,t]}(r)\big)\, \sigma_{u,r}\, dudr \\
&\qquad = \int_0^s \int_0^s \sigma_{u,r}\, dudr + \int_0^s \int_0^t \sigma_{u,r}\, dudr \\
&\qquad\qquad + \int_0^t \int_0^s \sigma_{u,r}\, dudr + \int_0^t \int_0^t \sigma_{u,r}\, dudr \\
&\qquad = \int_0^s \int_0^s \sigma_{u,r}\, dudr + 2 \int_0^{s\wedge t} \int_0^{s\vee t} \sigma_{u,r}\, dudr + \int_0^t \int_0^t \sigma_{u,r}\, dudr \\
&\qquad = \frac{D}{2\Theta^2} \Big(2s - 2\varphi_s + 4(s \wedge t) - 2\big(1 + e^{-\Theta|t-s|}\big)\varphi_{s\wedge t} + 2t - 2\varphi_t \Big).
\end{aligned}
$$

This finally yields

$$
E\left[T_s T_t\right] = T_0^2 e^{-\mu(s+t)} \exp\left(\frac{D}{2\Theta^2} \Big(s + t + 2(s \wedge t) - \big[\varphi_s + \varphi_t + \big(1 + e^{-\Theta|t-s|}\big)\varphi_{s\wedge t}\big] \Big) \right).
$$

We can now calculate (3.12) by using $E\left[T_t\right] = \frac{1}{T_0} E\left[T_0 T_t\right]$. (3.14) follows since $\operatorname{cov}\left(T_s, T_t\right) = E\left[T_s T_t\right] - E\left[T_s\right] E\left[T_t\right]$, which yields (3.13) for $s = t$.

□

A.2 Proof: Lemma 3.15

For the convenience of the reader we recall

Lemma 3.15. Statistical Properties of an OUP-Square Process
Let ϵ be a stationary OUP with parameters μ, Θ and D. Then the square process $(\gamma_t)_{t\geq 0} = (\epsilon_t^2)_{t\geq 0}$ satisfies

i) γ_t *is a stationary process*

ii) $E[\gamma_t] = \mu^2 + \frac{D}{2\Theta}$

iii) $var(\gamma_t) = \frac{2D}{\Theta}\mu^2 + \frac{D^2}{2\Theta^2}$

iv) $cov(\gamma_s, \gamma_t) = \frac{2D}{\Theta}\mu^2 e^{-\Theta|t-s|} + \frac{D^2}{2\Theta^2}e^{-2\Theta|t-s|}$

v) $\tau = -\frac{1}{\Theta}\ln\left(\sqrt{\frac{4\Theta^2}{D^2}\mu^4 + \frac{2\Theta}{D}\mu^2 + \frac{1}{2}} - \frac{2\Theta}{D}\mu^2\right)$.

Proof. *Lemma 3.15*

i) $\gamma_t = \epsilon_t^2$ is a stationary process because ϵ_t is stationary.

ii) Basic properties of Brownian motions directly yield

$$\begin{aligned} E[\gamma_t] &= E\left[\epsilon_t^2\right] \\ &= E\left[\mu^2 + \mu\sqrt{\frac{2D}{\Theta}}e^{-\Theta t}W(e^{2\Theta t}) + \frac{D}{2\Theta}e^{-2\Theta t}W^2(e^{2\Theta t})\right] \\ &= \mu^2 + \mu\sqrt{\frac{2D}{\Theta}}e^{-\Theta t}E\left[W(e^{2\Theta t})\right] + \frac{D}{2\Theta}e^{-2\Theta t}E\left[W^2(e^{2\Theta t})\right] \\ &= \mu^2 + \frac{D}{2\Theta}. \end{aligned}$$

iii) Follows from *iv)* for $s = t$.

iv) For a convenient notation, define $\tilde{s}, \tilde{t} = \frac{1}{2\Theta}\ln(s,t)$ for $s, t > 0$. Then we have

$$\begin{aligned} E[\gamma_{\tilde{s}}\gamma_{\tilde{t}}] &= E\left[\epsilon_{\tilde{s}}^2\epsilon_{\tilde{t}}^2\right] \\ &= E\left[\left(\mu^2 + \mu\sqrt{\frac{2D}{\Theta}}\frac{1}{\sqrt{s}}W_s + \frac{D}{2\Theta}\frac{1}{s}W_s^2\right)\left(\mu^2 + \mu\sqrt{\frac{2D}{\Theta}}\frac{1}{\sqrt{t}}W_t + \frac{D}{2\Theta}\frac{1}{t}W_t^2\right)\right] \\ &= \mu^4 + \mu^2\frac{D}{2\Theta}\left(\frac{1}{s}E\left[W_s^2\right] + \frac{1}{t}E\left[W_t^2\right]\right) \\ &\qquad + \mu^2\frac{2D}{\Theta}\frac{1}{\sqrt{st}}E[W_sW_t] + \frac{D^2}{4\Theta^2}\frac{1}{st}E\left[W_s^2W_t^2\right] \\ &= \mu^4 + \mu^2\frac{D}{\Theta} + \frac{D^2}{4\Theta^2} + \mu^2\frac{2D}{\Theta}\frac{s\wedge t}{\sqrt{st}} + \frac{D^2}{2\Theta^2}\frac{(s\wedge t)^2}{st} \end{aligned}$$

Returning to the original time scale via the monotone mappings $s \mapsto e^{2\Theta s}$ and $t \mapsto e^{2\Theta t}$, which imply $s \wedge t \mapsto e^{2\Theta(s\wedge t)}$, we get

$$E[\gamma_s\gamma_t] = \left(\mu^2 + \frac{D}{2\Theta}\right)^2 + \mu^2\frac{2D}{\Theta}e^{\Theta(2s\wedge t-(s+t))} + \frac{D^2}{2\Theta^2}e^{2\Theta(2s\wedge t-(s+t))}.$$

Since $\text{cov}(\gamma_s, \gamma_t) = E[\gamma_s\gamma_t] - E[\gamma_s]E[\gamma_t]$ and $2(s\wedge t) = s + t - |s-t|$, this proves *iv)*.

v) By definition of the decorrelation time τ, we have

$$\begin{aligned}
& \mathsf{cov}\left(\gamma_t, \gamma_{t+\tau}\right) = \frac{1}{2}\mathsf{var}\left(\gamma_t\right) \\
\Leftrightarrow \quad & \frac{2D}{\Theta}\mu^2 e^{-\Theta\tau} + \frac{D^2}{2\Theta^2}e^{-2\Theta\tau} = \frac{D}{\Theta}\mu^2 + \frac{D^2}{4\Theta^2} \\
\Leftrightarrow \quad & \left(e^{-\Theta\tau} + \mu^2\frac{2\Theta}{D}\right)^2 = \mu^4\frac{4\Theta^2}{D^2} + \mu^2\frac{2\Theta}{D} + \frac{1}{2} \\
\Leftrightarrow \quad & e^{-\Theta\tau} = \sqrt{\mu^4\frac{4\Theta^2}{D^2} + \mu^2\frac{2\Theta}{D} + \frac{1}{2}} - \mu^2\frac{2\Theta}{D} \\
\Leftrightarrow \quad & \tau = -\frac{1}{\Theta}\ln\left(\sqrt{\frac{4\Theta^2}{D^2}\mu^4 + \frac{2\Theta}{D}\mu^2 + \frac{1}{2}} - \frac{2\Theta}{D}\mu^2\right).
\end{aligned}$$

□

BONNER METEOROLOGISCHE ABHANDLUNGEN

Herausgegeben vom Meteorologischen Institut der Universität Bonn durch Prof. Dr. H. FLOHN (Hefte 1-25), Prof. Dr. M. HANTEL (Hefte 26-35), Prof. Dr. H.-D. SCHILLING (Hefte 36-39), Prof. Dr. H. KRAUS (Hefte 40-49), ab Heft 50 durch Prof. Dr. A. HENSE.

Heft 1-39: siehe `http://www2.meteo.uni-bonn.de/bibliothek/bma.html`

Heft 40: ***Hermann Flohn***: Meteorologie im Übergang Erfahrungen und Erinnerungen (1931-1991). 1992, 81 S. + XII. € 23

Heft 41: ***Adnan Alkhalaf and Helmut Kraus***: Energy Balance Equivalents to the Köppen-Geiger Climatic Regions. 1993, 69 S. + IX. € 19

Heft 42: ***Axel Gabriel***: Analyse stark nichtlinearer Dynamik am Beispiel einer rei- bungsfreien 2D-Bodenkaltfront. 1993, 127 S. + XIV. € 30

Heft 43: ***Annette Münzenberg-St.Denis***: Quasilineare Instabilitätsanalyse und ihre Anwendung auf die Strukturaufklärung von Mesozyklonen im östlichen Weddellmeergebiet. 1994, 131 S. + XIII. € 33

Heft 44: ***Hermann Mächel***: Variabilität der Aktionszentren der bodennahen Zirkulation über dem Atlantik im Zeitraum 1881-1989. 1995, 188 S. + XX. € 48

Heft 45: ***Günther Heinemann***: Polare Mesozyklonen. 1995, 157 S. + XVI. € 46

Heft 46: ***Joachim Klaßen:*** Wechselwirkung der Klima-Subsysteme Atmosphäre, Meereis und Ozean im Bereich einer Weddellmeer-Polynia. 1996, 146 S. + XVI. € 43

Heft 47: ***Kai Born***: Seewindzirkulationen: Numerische Simulationen der Seewind- front. 1996, 170 S. + XVI. € 48

Heft 48: ***Michael Lambrecht***: Numerische Untersuchungen zur tropischen 30-60-tägigen Oszillation mit einem konzeptionellen Modell. 1996, 48 S. + XII. € 15

Heft 49: ***Cäcilia Ewenz***: Seewindfronten in Australien: flugzeuggestützte Messungen und Modellergebnisse. 1999, 93 S. + X. € 30

Heft 50: ***Petra Friederichs***: Interannuelle und dekadische Variabilität der atmosphärischen Zirkulation in gekoppelten und SST-getriebenen GCM-Experimenten. 2000, 133 S. + VIII. € 25

Heft 51: ***Heiko Paeth***: Anthropogene Klimaänderungen auf der Nordhemisphäre und die Rolle der Nordatlantik-Oszillation. 2000, 168 S.+ XVIII. € 28

Heft 52: ***Hildegard Steinhorst***: Statistisch-dynamische Verbundsanalyse von zeitlich und räumlich hoch aufgelösten Niederschlagsmustern: eine Untersuchung am Beispiel der Gebiete von Köln und Bonn. 2000, 146 S. + XIV. € 25

Heft 53: ***Thomas Klein***: Katabatic winds over Greenland and Antartica and their interaction with mesoscale and synoptic-scale weather systems: three-dimensional numerical models. 2000, 146 S. + XIV. € 25

Heft 54: ***Clemens Drüe***: Experimentelle Untersuchung arktischer Grenzschichtfronten an der Meereisgrenze in der Davis-Straße. 2001, 165 S. + VIII. €

Heft 55: ***Gisela Seuffert***: Two approaches to improve the simulation of near surface processes in numerical weather prediction models. 2001, 128 S. + VI. € 25

Heft 56: ***Jochen Stuck***: Die simulierte axiale atmosphärische Drehimpulsbilanz des ECHAM3-T21 GCM. 2002, 202 S. + VII. € 30

Heft 57: ***Günther Haase***: A physical initialization algorithm for non-hydrostatic weather prediction models using radar derived rain rates. 2002, 106S. + IV. € 25

Heft 58: ***Judith Berner***: Detection and Stochastic Modeling of Nonlinear Signatures in the Geopotential Height Field of an Atmospheric General Circulation Model. 2003, 157 S. + VIII. € 28

Heft 59: ***Bernd Maurer***: Messungen in der atmosphärischen Grenzschicht und Validation eines mesoskaligen Atmosphärenmodells über heterogenen Landoberflächen. 2003, 182 S. + IX. € 30

Heft 60: ***Christoph Gebhardt***: Variational reconstruction of Quaternary temperature fields using mixture models as botanical – climatological transfer functions. 2003, 204 S. + VIII. € 30

Heft 61: ***Heiko Paeth***: The climate of tropical and northern Africa – A statistical-dynamical analysis of the key factors in climate variability and the role of human activity in future climate change. 2005, 316 S. + XVI. € 15

Heft 62: ***Christian Schölzel***: Palaeoenvironmental transfer functions in a Bayesian framework with application to Holocene climate variability in the Near East. 2006, 104 S. + VI. € 15

Heft 63: ***Susanne Bachner***: Daily precipitation characteristics simulated by a regional climate model, including their sensitivity to model physics, 2008, 161 S. € 15

Heft 64: ***Michael Weniger***: Stochastic parameterization: a rigorous approach to stochastic three-dimensional primitive equations, 2014, 148 S. + XV. open access[1]

[1] Available at `http://hss.ulb.uni-bonn.de/fakultaet/math-nat/`

Meteorologisches Institut
Mathematisch Naturwissenschaftliche Fakultät
Universität Bonn

Meteorologisches Institut
Universität Bonn