Intermittency and Irreversibility in the

Soil-Plant-Atmosphere System

by

J.R. Rigby

Department of Civil and Environmental Engineering
Duke University

Date: ____________________________

Approved:

______________________________
Amilcare Porporato, Advisor

______________________________
John Albertson

______________________________
Ana Barros

______________________________
Gabriel Katul

______________________________
Jonathan Mattingly

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Civil and Environmental Engineering in the Graduate School of Duke University
2009
Abstract

(Civil and Environmental Engineering)

Intermittency and Irreversibility in the Soil-Plant-Atmosphere System

by

J.R. Rigby

Department of Civil and Environmental Engineering
Duke University

Date: __________________________

Approved:

_______________________________
Amilcare Porporato, Advisor

_______________________________
John Albertson

_______________________________
Ana Barros

_______________________________
Gabriel Katul

_______________________________
Jonathan Mattingly

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Civil and Environmental Engineering in the Graduate School of Duke University

2009
Abstract

The hydrologic cycle may be described in essence as the process of water rising and falling in its various phases between land and atmosphere. In this minimal description of the hydrologic cycle two features come into focus: intermittency and irreversibility. In this dissertation intermittency and irreversibility are investigated broadly in the soil-plant-atmosphere system. The framework of intermittency and irreversibility is used to highlight a method of approaching the ecohydrological dynamics which uses simplified deterministic models coupled with appropriate stochastic terms. The theory of intermittency and irreversibility is addressed here in three ways: (1) through its effect on components of the soil-plant-atmosphere system, (2) through development of a measure of the degree of irreversibility in time-series, and (3) by the investigation of the dynamical sources of this intermittency. First, soil infiltration and spring frost risk are treated as two examples of hydrologic intermittency with very different characters and implications for the soil plant system. An investigation of the water budget in simplified soil moisture models reveals that simple bucket models of infiltration perform well against more accurate representation of intra-storm infiltration dynamics in determining the surface water partitioning. Damaging spring frost is presented as a “biologically-defined extreme event” and thus as a more subtle form of hydrologic intermittency. This work represents the first theoretical development of a biologically-defined extreme and highlights the importance of the interplay between daily temperature mean and variance in determining the changes in damaging frost
risk in a warming climate. Second, a statistical measure of directionality/asymmetry is developed for stationary time-series based on analogies with the theory of nonequilibrium thermodynamics. This measure is then applied to a set of DNA sequences as an example of a discrete sequence with limited state-space. The DNA sequences are found to be statistically asymmetric and further that the local degree of asymmetry is a reliable indicator of the coding/noncoding status of the DNA segment. Third, the phenomenology of rainfall occurrence is compared with canonical examples of dynamical intermittency to determine whether these simple dynamical features may display a dominant signature in rainfall processes. Summer convective rainfall is found to be broadly consistent with Type-III intermittency. Following on this result we studied daytime atmospheric boundary layer dynamics with a view toward developing simplified models that may further elucidate the interaction between land surface conditions and convective rainfall triggering.
To Hannah Rigby
Contents

Abstract iv
List of Tables xi
List of Figures xii
Acknowledgements xix

1 Introduction 1

2 Simplified Stochastic Soil Moisture Models: a look at infiltration 7
  2.1 Introduction ................................................................. 7
  2.2 Description of Models .................................................. 10
     2.2.1 Instantaneous Event Model (IEM) ............................. 11
     2.2.2 Finite Duration Event Model (FDEM) ........................ 14
  2.3 Model Comparisons ...................................................... 20
  2.4 Conclusions ............................................................... 28

3 Spring Frost Risk in a Changing Climate 30
  3.1 Introduction ............................................................... 30
  3.2 Theory ........................................................................... 32
     3.2.1 Timing of Budbreak ................................................ 32
     3.2.2 Frost Risk ............................................................... 33
  3.3 Data and Model ............................................................ 35
     3.3.1 Temperature Data from Durham, NC ........................ 35
8.2.1 The Free Atmospheric Temperature Profile ................. 110
8.2.2 Conserved Quantities in the Cloud-Topped Atmospheric Boundary Layer ........................................ 112
8.3 Thresholds of Moist Convection .............................. 112
  8.3.1 Lifting Condensation Level ................................. 113
  8.3.2 Level of Free Convection .................................. 114
  8.3.3 Level of Neutral Buoyancy and CAPE .................... 115
8.4 Cloud-Topped Boundary Layer Dynamical Model ............ 116
  8.4.1 Energy Budget .............................................. 116
  8.4.2 ABL Growth ............................................... 117
  8.4.3 Entrainment Condition .................................... 118
  8.4.4 Relationship Between ABL Height and Inversion Strength .. 122
  8.4.5 Model Summary ............................................ 122
8.5 Simulation and Preliminary Results ............................. 123
8.6 Conclusions ................................................... 126

9 Conclusion .......................................................... 127

Bibliography ......................................................... 130

Biography ............................................................ 142
## List of Tables

2.1 Table of parameter values used in simulation of soil-moisture and rainfall processes. ........................................... 21

3.1 Summary of Parameter Values ........................................... 35

6.1 Description of data series used. Temperatures are the maximum of monthly mean high temperatures in June-August obtained from www.worldclimate.com........................................... 76

7.1 Typical values of parameters ........................................... 96
List of Figures

1.1 A schematic of the hydrologic cycle highlighting the intermittency of precipitation as well as the directionality in the cycle.

2.1 Summary of the stochastic rainfall model used by [34]. The frequency, $\lambda$, for the corresponding marked Poisson process, used in the IEM, is also shown. The mean rainfall depth $\alpha$ represents the mean area of the rectangular pulses.

2.2 Plot showing the domain in which the [72] model by ignoring percolation during storm events produces the unphysical result that the time to ponding is greater than the time to saturation. Parameter values used for simulation may be found in Table 2.1.

2.3 Comparison of FDEM (solid line) and IEM (dotted line) soil-moisture models over one hundred days. The stochastic rainfall series of rectangular pulses is shown above. To the right is shown the simulated p.d.f. of the FDEM model (bars) with the analytic p.d.f. of the IEM model.

2.4 Comparison of simulated FDEM (lines) and analytic IEM (shaded area) probability distributions for soil moisture. The four plots show varying soil depth and rainfall arrival rates. The two lines on each plot are for mean rainfall durations of 4 (solid) and 6 (dotted) hours. The plots on the left correspond to an index of dryness $(\lambda\alpha/E_{max})$ of 1.2 while those on the right have an index of 0.6.

2.5 Comparison of simulated FDEM (lines) and analytic IEM (shaded area) probability distributions for soil moisture. The two plots show the effect of soil saturated hydraulic conductivity on the correspondence between pdf’s of soil moisture according to the IEM and FDEM. The two lines on each plot are for mean rainfall durations of 4 (solid) and 6 (dotted) hours.
2.6 Rainfall partitioning during a storm event for the IEM (left) and the FDEM (Right). The vertical axis represents the rainfall depth transformed by processes of infiltration and percolation. The normalized curves in the FDEM plot are, from highest to lowest: depth of rainfall event (slope = unity), infiltrated depth according to [72], infiltrated depth minus percolation according to FDEM without bound at $s = 1$, and the bold line represents the actual change in soil-moisture state as a function of rainfall depth according to the FDEM with the bound at $s = 1$.

2.7 Fraction of losses attributed to runoff as a function of total rainfall depth for both (a) very large depths and (b) typical rainfall depths. The curves represent different precipitation intensities increasing from the right: 3.5, 4, 4.5, 5, 6, 8, and 10, respectively.

2.8 Change in soil moisture, $y = \Delta s$, representing normalized net infiltration, for different values of $\tilde{P}$ (left) and $s_0$ (right) for both the IEM and FDEM models. The dotted lines represent the IEM model.

2.9 Derived distributions of the normalized net infiltration for both the IEM and FDEM. Note that the IEM distribution (bold) is a truncated exponential with an atom of probability at $\tilde{D} = 1 - s_0$ represented by the corresponding Dirac delta functions.

3.1 (a) NDVI observed at Walker’s Branch Watershed, Oak Ridge, TN for all of 2006 and January-May 2007 (reprinted with permission of Tilden Meyers). Both the warm March and the April freeze show clear signatures in the 2007 trace in comparison with 2006. (b) Minimum temperature time series for Durham, NC for the spring of 2006 and 2007.

3.2 Sensitivity of the probability of survival, $F(\infty)$, (circles) and the most probable date of budbreak (squares) to parameters of the temperature and phenology models. Each plot was generated by varying a single parameter while holding the others at their mean value (see Table 1). The vertical line on each plot represents the mean value of the varied parameter as estimated from the Durham, NC data set.

3.3 Survival probability as a function of the percent change to the current values of variance ($k_3$) shown for multiple increases in mean daily temperature ($k_0$). The figure shows that an increase in mean temperature of $1^\circ$C accompanied by an 6% increase in variance would result in the same value of $F(\infty)$. 

xiii
4.1 Asymmetry statistic, \( \langle Z_n \rangle / n \), as a function of \( \alpha \) (i.e., for increasingly Gaussian signals) for different block lengths \( n = 2 \), circles; \( n = 3 \), squares; \( n = 4 \), diamonds). The insets show portions of time series, \( x_t \), for three values of \( \alpha \). Values used for simulations: \( k = 0.07, \gamma = 0.04, \lambda = 0.28 \).

4.2 Illustration of the FT applied to Po River (Italy) discharge series. a) Subset of the Po River discharge series, b) probability distribution of time derivative given \( x \) chosen as the mode of the distribution, c) the functional form of \( Z_2^*(\dot{x}|x) \), d) the probability distribution of \( Z_2^*(\dot{x}|x) \) derived from b) and c), and e) the resulting FT for the pdf of \( Z_2^*|x \) where \( x \) is chosen respectively as the mode (circles), mean (squares), and the median (crosses) of the discharge series. Note that b) and d) are shown with suitably coarse binning for the purposes of clear illustration while do not match, computationally, the values reported in c) and e) which are computed with finer resolution.

5.1 The asymmetry rate as a function of word length (panel a) in coding and non-coding regions for both \( E. coli \) and \( A. thaliana \). Note that in both cases there is a significant difference between the coding and non-coding regions as well as dramatic differences in slope between species. Panel b then shows the performance of using \( Z_2 \) as a statistic for locating coding regions in \( A. thaliana \) (see Methods). The curves show the proportion of cases in which \( Z_2 > t \) correctly indicated a coding region (circles) and the proportion of coding regions for which the corresponding \( Z_2 \) value was above the threshold (x’s). The histogram shows the probability distribution of all local \( Z_2 \) estimates from the sequence (magnified by a factor of 7 for comparison). The mean standard error across the estimates, \( \sigma_z \), is also shown, as well as the region (shaded) in which the proportion for both indicators is between 0.8 and 0.9. Notice that the statistic performs well for thresholds above the standard error. Thus any region that has an asymmetry significantly different from zero is highly likely to be coding.

5.2 The asymmetry measured for sequences of codons and amino acids for each of the four species. Of particular interest is the apparent negative trend with organismal complexity. Note also that the asymmetry is much greater in codon sequences suggesting that significant information is coded in ordered pairs of synonymous codons for the same amino acid pair.
6.1 Plot showing time series from (a) the Manneville map, (b) the stochastic model of type III intermittency, and (c) a static transformation of the series in (b) according to $z_t = \frac{1}{1-x_t} - 1$ which produces a strikingly rainfall-like appearance. ........................................ 70

6.2 Plot showing On-Off time series from (a) the random logistic map, and (b) the stochastic continuous version. ............................. 72

6.3 Log-log plot of the probability density of laminar (i.e., no rainfall) periods from four data sites. The solid black line with slope $-3/2$ is included as a visual aid. ........................................ 77

6.4 Log-log plots of the power spectral density for each of the four data sites. The straight line in each case indicates the $-1/2$ slope expected from Type-III and On-Off intermittency for low frequencies. .... 78

6.5 Log-log plots of the power spectral density at high frequencies for summer (left) and winter (right). Straight lines indicate slopes of $-1/2$ for summer and $-1$ for winter. While the $-1/2$ scaling is suggested by Type-III intermittency, the consistent $-1$ scaling for winter is an unanticipated result. ............................. 79

6.6 Log-log plot of the power spectral density for the 1min resolution series from Chiavari. For periods between 1hr and 1day the spectra show scaling near $-1/2$, while for higher frequencies an apparent $-3/2$ scaling appears. The $-1$ scaling in winter is again apparent. .... 80

6.7 Log-log plots of the probability density of rainfall intensities for each of the four data sites. The straight line in each panel offers a visual aid for comparison with expected scaling for On-Off and Type-III intermittency. No consistent scaling appears across sites. .... 81

6.8 Log-log plots of the pdf’s of summer (left) and winter (right) precipitation intensities. Lines are offered as visual aid for slope of $-1$. Note that summer intensity slopes are very close to $-2$ for three sites (except Valentia), while winter series shows no discernible pattern across sites. Insert shows consistency between summer and winter for Valentia. 82
6.9 Plot of the block entropy, $H(n)$, (left) and the slope of the logarithm of the block entropy, $\alpha = \log\left( \frac{H(n+1)}{H(n)} \right) / \log\left( \frac{n+1}{n} \right)$, (right) as a function of the block length. For sporadic signals $H(n) \sim n^{-\alpha}$, thus the slope on log-log scale should converge to a constant value, $\alpha$. For each site the maximum $n$ for which the entropy was estimated is based on $n_{\text{max}} = \log_2[N/100]$, where $N$ is the length of the series, to avoid finite-size effects. The log-entropy rate for each of the sites appears to reach a plateau consistent with sporadicity.

7.1 Schematic representation of the vertical profiles of specific humidity $q$ and potential temperature $\theta$.

7.2 Exact numerical solution of the coupled ABL equations for (a) the height of the mixed-layer, (b) the potential temperature, (c) the specific humidity and (d) the Bowen ratio.

7.3 ... saturation specific humidity curve with linear approximation over the range $T=288-295$K.

7.4 The term $\xi = \tilde{Q}/(q^* - q)$ with the saturation deficit computed exactly (solid line) compared with the approximate representation $\xi \approx \alpha Q$ for $\alpha \approx 0.015$ (dashed line).

7.5 A comparison of $h(t)$ solutions. The solid curve represents the exact solution to the coupled ABL equations. Increasing degrees of approximation are then represented by the dotted/dashed lines. Dashed curve: numerical solution to Eqs. (8.14) and (7.26). Dot-Dashed curve: analytic solution given in Eq. (7.32). Dotted curve: analytic solution for the case of constant Bowen ratio (Porporato 2009 Eq. 37) with $Bo = Bo_{\text{max}}$.

7.6 A comparison of $q(t)$ solutions of varying degrees of approximation. The solid curve represents the exact solution to the coupled ABL equations. The dashed curve shows the numerical solution to Eq. 8.13 using the analytic solution of $h(t)$ given in Eq. 7.32 and the Bowen ratio given by the Penman-Monteith approximation (Eq. 7.23). The dash-dot curve shows the solution using the same analytic $h(t)$ but with the approximate Bowen ratio given in Eq. 7.29. Finally, the two dotted curves show the solution for constant Bowen ratio (taken as 1/2 of the maximum). The lower dotted curve shows the analytic relation for strictly constant Bowen ratio given in [112]. The upper dotted line shows the numerical solution to Eq. 8.13 with the analytic $h(t)$ but with constant Bowen ratio.
7.7 A comparison of Bowen ratios computed using different approximations. Curves represent the Bowen ratio computed exactly (solid) compared with the Penman-Monteith approximate Bowen ratio (dashed) corresponding to Eq. (7.23), as well as that of Eq. (7.26) (dot-dashed) and that of Eq. (7.29) which corresponds to the analytical solution for $h$ given in (7.32) (dotted).

8.1 Illustration of the profile of $\theta$ assuming hydrostatic conditions (dashed line) and the linear approximation (solid line) through the troposphere.

8.2 Schematic representation of the lifting condensation level. In the unsaturated ABL the lapse rate is approximately adiabatic while in the free atmosphere the lapse rate is sub-adiabatic (i.e., stable). The lifting condensation level is the height at which a parcel of air of given $q$ would saturate if lifted adiabatically.

8.3 Schematic representation of the lifting condensation level and the level of free convection. In the unsaturated ABL the lapse rate is approximately adiabatic while in the free atmosphere the lapse rate is sub-adiabatic (i.e., stable). Above the LCL the lapse rate follows the saturated adiabat (dash-dot line). The level of free convection is the height at which a parcel lifted adiabatically first becomes warmer than the surrounding atmosphere.

8.4 Numerical simulation of the daytime boundary layer evolution: (a) shows the height of boundary layer (solid line), the solution for the dry ABL (dashed line), and the cloud base at the LCL (dotted line); (b) shows the equivalent potential temperature of the mixed-layer; (c) shows the total water mixing ratio in the mixed-layer (solid line) and the saturation specific humidity curve (dashed line); (d) shows the evolution of the LNB (solid line), LFC (dashed line) and the LCL (dotted line) during the day.

8.5 The temperature for a parcel lifted along an adiabatic path from the surface (solid line) and the free atmospheric temperature profile (dashed line).

8.6 Instantaneous profile of the temperature profile with the jump at the top of the boundary layer (solid line) and the extrapolated free atmospheric profile for comparison (dashed line): (a) before reaching the LCL and (b) at the LFC.

8.7 The partitioning of water into liquid and vapor phases at the top of the boundary layer.
8.8 The evolution of CAPE during daytime hours up to the initiation of free convection.
Acknowledgements

This research was supported by a National Science Foundation Graduate Fellowship, a James B. Duke Fellowship, and a Pratt-Gardner Graduate Fellowship. Figure 2.1a was provided by Tilden Meyers of Oak Ridge National Laboratory. Rainfall data for Valencia and Florence was provided by G. Kiely and F. Castelli. The Meteorological Observatory “Andrea Bianchi” provided the Chiavari time series and the Department of Construction, Environmental and Territorial Engineering of the University of Genova provided the high resolution time series of Villa Cambiaso. DNA sequences were provided by the National Center for Biotechnology Information.
The hydrologic cycle may be described in essence as the process of water rising and falling in its various phases between land and atmosphere. Figure 1.1 illustrates this simple view of the hydrologic cycle. In this minimal description of the hydrologic cycle two features come into focus. The first, and perhaps obvious, aspect is that precipitation introduces an intermittent character (at least locally) into what is otherwise basically a continuous flow of water through the cycle. Stream and groundwater flows, evapotranspirative flux, and atmospheric advection of vapor are all continuous fluxes of water. However, the return of water to the Earth’s surface is only accomplished intermittently by precipitation. It is this intermittency that introduces much of the rich behavior, as well as uncertainty, into hydrology and those systems coupled with the hydrologic cycle.

In addition to intermittency the second feature that comes into focus in this simple picture of the hydrologic cycle is its directionality. During precipitation and terrestrial redistribution water moves primarily by the force of gravity. It requires the expenditure of energy (supplied of course by the sun) to evaporate surface water and then transport it aloft by turbulent mixing against the force of gravity. On the other
hand, water is not typically transported aloft in its condensed phase. Thus a preferred direction in the cycle of water lifting and falling, evaporating and condensing, is manifest. In this sense the hydrologic cycle functions as a natural heat engine driven by solar radiation [74].

In hydrologic research intermittency has been ubiquitous if also almost too obvious to make explicit. This is especially true of some recent applications of stochastic methods to hydrological and ecohydrological problems [119]. While intermittency has been understood as persistent and fundamental property of hydrologic systems, it may also be treated as a research problem in itself. That is, we may look for the sources of that intermittency. This is the rationale behind physical modeling of precipitation at a variety of spatial and temporal scales (e.g., [7, 69]). These models may be generally divided into phenomenological and physical categories. The former seek to capture the phenomenological structure of, say, rainfall intermittency using purely mathematical techniques [147, 55, 144]. The latter, physical, models attempt to model dominant physical processes at some spatial and temporal scale in order to
understand the physics of rainfall intermittency [7, 78].

Irreversibility has by comparison received very little attention. The hydrologic cycle has been commonly described as a heat engine [74, 103] with concomitant discussion of the second law of thermodynamics and reversibility. However, a more general investigation of temporal asymmetry has been absent from the literature. There exist no measures of temporal asymmetry in environmental series nor a general understanding of how temporal asymmetry is related to more commonly understood features of hydrologic systems. In particular the possible link between intermittency and irreversibility in the hydrologic cycle has not been investigated.

The work presented in the following chapters investigates intermittency and irreversibility as they are manifest broadly in the soil-plant-atmosphere system. While the chapters cover a diverse set of topics, the central theme is the various manifestations of intermittency and irreversibility in the soil-plant-atmosphere system with increasing emphasis on laying the groundwork for understanding how these two features are related.

The chapters may be divided into three thematic sections. The first section, comprising the first two chapters, is focused on investigating two manifestations of intermittency in hydrology and ecohydrology. Chapter one presents a model investigation of infiltration, a classic hydrologic topic. A simplified, vertically-averaged model of soil moisture interpreted at the daily time scale and forced by a stochastic process of instantaneous rainfall events is compared with a vertically-averaged model which uses a non-overlapping rectangular pulse rainfall model and a more physically based description of infiltration. The models are compared with respect to the importance of short time-scale (intra-storm) variable infiltration in determining the probabilistic structure of soil-moisture dynamics at the daily time-scale. This chapter illustrates the fundamental role of intermittency in partitioning water at the Earth’s surface.
Chapter two introduces a more subtle form of intermittency at work in the environment. In this chapter a probabilistic framework for understanding spring frost risk to vegetation is developed in order to determine whether risk of frost damage is likely to increase or decrease in a warming climate based on several statistical parameters of the daily temperature process. The unique aspects of this study are two-fold. First, it highlights frost as a subtle intermittent hydrologic phenomenon with important consequences for plant communities. In this case frost risk is dependent on the correlation between integrated daily temperature during early spring and the probability of low temperature excursions. Second, it introduces the notion of defining extreme events from a biological perspective. Frost is not per se an extreme event. It is only extreme, even to a plant, if it occurs during a particularly sensitive period in the life cycle of the plant. This sort of dependence is a widespread phenomenon in biological systems. The analysis of spring frost should open several avenues of research investigating the intermittency of biologically meaningful events.

The second broad section consists of chapters three and four. This section deals explicitly with the definition and measurement of irreversibility. In chapter three a statistical measure of temporal asymmetry (irreversibility) is developed for stationary time series. This chapter is based on an article for which the author (as second author on the article) was an integral part of the theoretical development of the paper, performed the modeling and data analysis, generated the figures, and contributed in part to the writing of the paper. This chapter presents a measure of asymmetry that can be used to detect directionality in time series that is invisible to the naked eye. It is therefore a basis for future exploratory studies into temporal asymmetry in environmental time-series.

Chapter four applies the measure developed in chapter three to a, perhaps surprising, data set. To avoid initially the difficulties of continuous variables we sought a data set that was truly discrete in order to apply the measure of irreversibility
unambiguously. DNA is an ideal candidate as it has a very limited state space (4 nucleotides, 64 codons) and is inherently discrete. While exploring whether DNA exhibits directional asymmetry we found some surprising results. First, we found that the local asymmetry within a DNA sequence is a good predictor of whether the sequence is likely to be transcribed by RNA for the coding of proteins or not. Second, we found an apparent trend between the degree of asymmetry in DNA sequences and the phylogenetic complexity of organisms (i.e., bacteria, fungi, plant).

The final three chapters form the third and final section of the dissertation. In this section we take up rainfall specifically. In chapter five we approach the source of rainfall intermittency from a novel perspective. Rather than starting with a physical model or trying to capture the phenomenology with a stochastic model, we use a hybrid approach to investigate the phenomenology for clues to the physics. That is, while rainfall is an ultra-high dimensional system, there may be dominant dynamical features (e.g., specific types of instabilities) that predominate in the phenomenology of the system. Thus, we begin by outlining a typology of dynamical intermittencies present in the physics and mathematics literature. We then compare the properties of these canonical types to the phenomenology of rainfall. In this way we hope to gain insight into the types of dynamical features that might be responsible for observed properties of rainfall at particular scales. The results in this vein are not conclusive but suggest that convective precipitation at the temporal scale of a single storm is consistent with Type-III intermittency.

From the phenomenological comparison in chapter five, we move to the final two chapters which investigate the deterministic modeling of the atmospheric boundary layer. These chapters are early drafts of work that is ongoing. The motivation for these chapters is to investigate the triggering of convective rainfall in hopes of tying this back in to the phenomenological investigation of rainfall intermittency. We limit the time scale for the deterministic model to the daytime evolution of
the atmospheric boundary layer. In chapter six we extend the work of [112] on simplified solutions to the dry boundary layer dynamics. We present a variety of numerical and analytical simplifications associated with approximate forms of the Bowen ratio that may allow for a fully analytic approximate solution to the land-atmosphere interaction. In chapter seven we extend the boundary layer model to include a strato-cumulus layer with a highly simplified entrainment assumption in order to study the land-atmosphere interaction in a cloud-topped boundary layer and specifically how the surface conditions affect the initiation and strength of free convection.

2.1 Introduction

As both a reservoir and a regulator of water movement in the soil-plant-atmosphere continuum, the soil is an enormously rich and complicated domain for hydrologic enquiry. In ecosystems where water is the limiting resource, understanding the dynamics and variability of soil water is essential not only for understanding the cycling of water, but also for understanding ecosystem dynamics, such as patterns of vegetation form, adaptation, and distribution (both spatially and temporally) [119]. However, these are complex, nonlinear systems making mathematical analysis of the dynamics difficult. Development of simplified soil-moisture models (e.g., [34, 93, 72, 120, 80, 113, 119, 25]) is therefore an important step in assembling the analytical tools necessary to unravel the intertwined dynamics of ecosystems and
the hydrologic cycle. The aim of developing such models is to balance the faithful representation of physical dynamics (e.g., nonlinearities of infiltration and plant dynamics) against the mathematical simplicity that may allow analytical solutions. These solutions in turn provide insight into the relationships between component processes in determining the character of soil water dynamics.

One of the many tasks in developing simplified models of soil moisture is determining how to represent the partitioning of rainfall into runoff and infiltration. Two mechanisms are commonly associated with runoff: that of subsurface control or saturation deficit, and surface control (often associated with Horton). While the distinction is somewhat artificial, it is useful for describing approximate models of infiltration which require an imposed discontinuity in the infiltration curve at saturation \((s = 1)\) to avoid supersaturating the soil. By surface-controlled runoff here we mean runoff generated due to an explicitly time dependent soil infiltrability. By subsurface control we simply mean that, for time resolved events, at saturation the infiltrability is instantaneously reduced to the saturated conductivity or that, for instantaneous events, the infiltrated depth cannot exceed the saturation deficit. In simplified models it is often convenient to ignore surface-controlled runoff in favor of the saturation deficit approach given its simple implementation [120, 119]. In this paper we examine the relationship between models treating runoff solely from the saturation deficit approach in favor of analytical (probabilistic) solutions and models which take into account surface-controlled runoff at some analytical cost.

To make such a comparison we have selected two models (each with some modifications for the purposes of this investigation) of soil moisture at a point which broadly illustrate the differing treatments of infiltration while otherwise remaining similar in structure. The first model is that of [120] (see also [93], [80], and [113]) which models soil moisture at the daily time-scale using instantaneous rainfall events and ignoring surface-controlled runoff. We will hereafter refer to this model as the Instantaneous
Event Model (IEM). The second model is derived from those of [33, 34] and [72] which take into account rainfall duration/intensity and the associated possibility of surface-controlled runoff. This model will be referred to as the Finite Duration Event Model (FDEM). Both the IEM and the FDEM treat soil-moisture content averaged vertically over the root zone (i.e., instantaneous vertical redistribution). For a comparison of vertically lumped versus distributed models see [57].

The fundamental differences between the two models are in the representation of rainfall and infiltration. For models using the saturation deficit approach it is not necessary (at the daily time scale) to resolve the dynamics of soil moisture during the rainfall event (since only the initial soil saturation deficit and the rainfall depth determine the infiltration response). In such models an instantaneous pulse of rainfall containing a finite depth may then be used as a model for rain events. Alternatively, in order to resolve surface-controlled runoff the model must also ascribe an intensity to the rainfall event in order to determine the infiltration. This amounts to assigning a (stochastic) duration to each rainfall event and then defining a function which transforms a given rainfall depth and duration into an infiltrated depth. In the absence of an analytical solution for this transformation, storm events must be resolved numerically. Otherwise, at the daily time scale a new probability density of jumps in soil moisture could be derived from the known distribution of rainfall depths. For the second model in this paper we follow the approach of [34] and [72] in using Philip’s (1957) infiltration solution modified by the time compression approximation (e.g., [129]) as the basis for this function. The two models for comparison differ then only in accounting losses during storm events. As the stochastic forcing is generally the factor determining analytical tractability of the problem, it is of particular interest to understand what is gained from the added complexity of resolving storm duration and whether modifications of the instantaneous storm models are available which might retain the possibility of analytical solutions while improving the accuracy of
the model.

**Figure 2.1:** Summary of the stochastic rainfall model used by [34]. The frequency, $\lambda$, for the corresponding marked Poisson process, used in the IEM, is also shown. The mean rainfall depth $\alpha$ represents the mean area of the rectangular pulses.

### 2.2 Description of Models

The basic structure of vertically-averaged models of soil moisture at the daily time-scale is that of a stochastic differential equation describing the rate of change in soil moisture as the sum of inputs and losses associated with the active soil layer. The balance equation is then given by

$$nZ_r \frac{ds}{dt} = \phi(R_t) - ET - L, \quad (2.1)$$

where $n$ is the soil porosity, $Z_r$ is the soil rooting depth (active layer), $s$ is the vertically averaged relative soil-moisture content, $\phi$ is an infiltration function, $R_t$ represents a stochastic rainfall process, $ET$ is the rate of evapotranspiration, and $L$ represents the losses to deep percolation. Runoff (and infiltration) mechanisms are contained in $\phi$ which may be a nonlinear function including thresholds (e.g., at $s = 1$).

In this section we describe two models that may be expressed in the manner of Eq. (2.1): the IEM, which models rainfall as a marked Poisson process, and the FDEM, which models rainfall using random rectangular pulses. As the models
differ primarily in the processes at work during a rainfall event, we will divide the description of the models into “during storm” and “between storm” components.

Between storm events both models evolve according to the same equation representing losses due to evapotranspiration and percolation, following [72],

\[ nZ_r \frac{ds}{dt} = -(k_s s^{c+1} + E_{max} s), \]  

(2.2)

where \( k_s \) is the saturated hydraulic conductivity, \( c = 2(1 + m)/m \) where \( m \) is the exponent in the [11] water retention relation, and \( E_{max} \) is the potential evapotranspiration. Here percolation is modelled after the [11] relation for unsaturated conductivity. Evapotranspiration is assumed to decrease linearly with soil moisture from a maximum at saturation, \( E_{max} \), following Kim et al 1996. While in general the evapotranspiration tends to be a nonlinear function of soil moisture [120], the linear evapotranspiration losses have been found to be reasonable for a broad range of soil-moisture values (see [113] and references therein). For the purposes of this paper, the loss function given in Eq. (2.2) is adopted for both the instantaneous and finite duration models.

While the models are identical in their representation of soil moisture between storms, the models differ significantly in their treatments during a rainfall event. In the following sections we describe the particulars of the stochastic rainfall process and soil-moisture accounting in each model.

2.2.1 Instantaneous Event Model (IEM)

Rainfall

Since both the occurrence and amount of rainfall can be considered to be stochastic, the occurrence of rainfall is here idealized as a series of point events in continuous time, arising according to a Poisson process of rate \( \lambda \), each carrying a random amount
of rainfall extracted from a given distribution. The temporal structure within each rain event is ignored and the marked Poisson process representing precipitation is physically interpreted at a daily time-scale, where the pulses of rainfall corresponding to daily precipitation are assumed to be concentrated at an instant in time.

With these assumptions, the distribution of the times between precipitation events is exponential with mean $1/\lambda$ (e.g., [22]). Furthermore, the depth of rainfall events is assumed to be an independent random variable $D$, described by an exponential probability distribution where $\alpha$ is the mean depth of rainfall events.

Both the Poisson process and the exponential distribution are of common use in simplified models of rainfall at the daily time scale. The exponential distribution fits well daily rainfall data and, at the same time, allows analytical tractability [8, 32, 34]. The values of $\alpha$ and $\lambda$ are assumed to be time-invariant quantities, representative of a typical growing season.

**Infiltration**

In the IEM the Poisson rainfall process creates an instantaneous jump in soil moisture such that the infiltration depth, $I_D$, is assumed equal to the minimum value between the soil saturation deficit and the depth of the rainfall event, i.e.,

$$I_D = \min[nZ_r(1 - s_0), D], \quad (2.3)$$

where $s_0$ is the relative soil moisture at the beginning of the event and $D$ represents the total depth of the rainfall event [120, 119]. For later comparison with the FDEM a normalized infiltration function, $y(\bar{D}, s_0) = I_D/nZ_r$, representing the net increase in relative soil moisture due to a rainfall event of dimensionless depth, $\bar{D} = D/nZ_r$, can be defined as
\[ y(\tilde{D}, s_0) = \begin{cases} 
\tilde{D}, & 0 \leq \tilde{D} \leq (1 - s_0) \\
1 - s_0, & \tilde{D} > (1 - s_0). 
\end{cases} \] (2.4)

Any rainfall in excess of \(1 - s_0\) is attributed to cumulative losses (i.e., the combined effect of runoff and percolation).

**Model Summary**

The IEM is a vertically averaged model of soil moisture interpreted at the daily time-scale, driven by a marked Poisson rainfall process of rate \(\lambda\) with exponentially distributed depths of mean \(\alpha\). The instantaneous jump in soil-moisture state for a particular event is determined completely by the subsurface state, or saturation deficit, and the depth of the rainfall event. Losses between storms are assumed due only to evapotranspiration and percolation. This may be expressed by the stochastic differential equation,

\[ nZ_r \frac{ds}{dt} = I_D(R_t, s_0) - (k_s s^{c+1} + E_{max} s). \] (2.5)

The stochastic soil-moisture process described by Eq. (2.5) may be solved analytically under steady state conditions [119]. The resulting probability distribution is, in this case,

\[ p(s) = C s^{\frac{\lambda}{\eta} - 1} e^{-\gamma s} (E_{max} + k_s s^{c})^{-\frac{\lambda}{\eta} - 1}, \] (2.6)

where \(\eta = E_{max}/nZ_r\) and \(C\) is a normalization constant that must be evaluated numerically.
2.2.2 Finite Duration Event Model (FDEM)

Rainfall

[34] offered an alternative to the Poisson rainfall process to allow for surface-controlled runoff by modelling rainfall with non-zero storm durations. In contrast to the marked Poisson process, each rainfall event is a rectangular pulse occupying a finite time, with the time between storms distributed exponentially with mean \( \tau \). A probability distribution is also assigned to the storm durations as well as to either the intensity or the total depth of rainfall. The remaining distribution may then be derived from the other two. Drawing on data from Massachusetts and California, [33] found that the durations were fit reasonably well by the exponential distribution (with mean \( \delta \), see Fig. 2.1) and that the event depths fit a two parameter gamma distribution. [33] then employed a model based on assumed distributions for the depth and duration of rainfall events. Given that the exponential distribution is a special case of the two parameter gamma distribution, we will use the simpler exponential form in this paper so that the two rainfall models (IEM and FDEM) agree with respect to the distribution of depths. Thus, for our finite duration model, each rainfall event is determined by three random variables (depth, duration, and inter-arrival time), each of which is drawn from an exponential distribution.

Assuming statistical independence between rainfall depth and duration, one may now derive the distribution of rainfall intensities dictated by fixing the distribution of depths and durations as an exponential. The resulting probability density function is

\[
f_P(P) = \frac{\alpha \delta}{(\alpha + \delta P)^2}
\]

which is the positive part \( (P > 0) \) of a Cauchy distribution. As [33] found that measured rainfall intensities were modelled well as an exponential distribution, the
Cauchy distribution, with power law tails, should overestimate the frequency of intense rain events and the corresponding runoff. Furthermore, while depth and duration are assumed independent variables, sampling from these two distributions for each rainfall event leads to a statistically dependent intensity. The conditional distribution of intensity, \( P \), given the event duration, \( w \), is then, 

\[
f_{P|w}(P|w) = \frac{w}{\alpha} e^{-\frac{w}{\alpha} P},
\]

which shows the negative correlation between intensity and duration. Such negative correlation is consistent with observed rainfall frequency-duration patterns, though the negative correlation here is probably exaggerated due to the simple rainfall model.

**Infiltration**

To treat infiltration, the FDEM follows [34] and the improvements of [72] by employing Philip’s (1957) approximate solution (hereafter Philip solution) to Richards’ equation combined with the time compression approximation.

Assuming a constant hydraulic head at the soil surface with an initially uniform (semi-infinite) vertical soil-moisture profile, [107] obtained a series solution to Richards’ equation. In its truncated form, the approximate solution states that the infiltration rate, \( i(t) \), decreases in time as

\[
i(t) = \frac{1}{2} S(s_0) t^{-1/2} + ak_s
\]

where \( t \) is the time since the inception of the rainfall event, and \( S(s_0) \) represents the soil sorptivity and may be expressed as

\[
S(s_0) = \left( \frac{2n(1-s_0)\psi_s(s_0^{(1+3m)/m}}{1+3m} - 1 \right)^{1/2} k_s^{1/2}
\]

where \( \psi_s \) is the [11] air entry pressure [130]. The constant \( a \) in Eq. (2.8) which depends on unsaturated hydraulic conductivity near saturation (see [102]) is here
taken to be unity for consistency with percolation losses at very long event durations (see Sect. 2.2.2).

For small $t$, according to the Philip solution, the potential rate of infiltration of the soil will exceed the precipitation rate. With these assumptions the infiltration rate curve would then be equal to the precipitation rate, $P$, up to time $t_e$ when the Philip potential infiltration rate equals the precipitation rate, after which ponding should begin. Thus,

$$i(t) = \begin{cases} P, & 0 \leq t \leq t_e \\ \frac{4}{1}S(s_0)t^{-1/2} + k_s, & t > t_e \end{cases}$$  \hspace{1cm} (2.10)$$

(see Fig. 2.1). Setting Eq. (2.8) equal to $P$ and solving for time yields,

$$t_e = \frac{S(s_0)^2}{4(P - k_s)^2}, \hspace{1cm} P > k_s.$$ \hspace{1cm} (2.11)

However, initially, the boundary condition is that of constant flux (equal to $P$) rather than the constant head assumed in the Philip solution. The result is that the time to ponding, $t_p$, is not generally equal to $t_e$ and is found to be somewhat larger. [87] provide a nice description of the exact solution for one dimensional linearized infiltration. As an approximate correction for the difference between the exact infiltration solution and the Philip solution, according to the time-compression approximation (TCA) (also termed the Infiltrability-Depth Approximation, see [129] for detailed discussion), cumulative infiltration may be used as a surrogate for time [127, 87]. Accordingly, one assumes that at time $t_p$ the cumulative infiltration under the constant flux is equal to the cumulative infiltration under the Philip curve up to time $t_e$. The time to ponding, $t_p = t_e + t_c$ where $t_c$ is the time of compression representing the difference between ponding in the Philip solution and the actual time to ponding, is then defined by
\[
\int_0^{t_p} P \, dt = \int_0^{t_c} i(t) \, dt
\]  
(2.12)

where \(i(t)\) is the Philip solution from Eq. (2.8). From this definition it follows that

\[
t_p = \begin{cases} 
\frac{S(s_0)^2(2P - k_s)}{4P(P - k_s)^2}, & P > k_s \\
\infty, & P \leq k_s
\end{cases}
\]  
(2.13)

Making the added assumption that for \(t \geq t_p\) the infiltration rate follows the Philip curve, the infiltration rate from Eq. (2.10) becomes

\[
i(t) = \begin{cases} 
P, & 0 \leq t \leq t_p \\
\frac{1}{2}S(s_0)(t - t_c)^{-1/2} + k_s, & t > t_p.
\end{cases}
\]  
(2.14)

Furthermore, we can express the cumulative infiltration depth (i.e., the cumulative depth of infiltrated rainfall) analytically by integrating Eq. (2.14)

\[
I_D(t) = \begin{cases} 
Pt, & 0 \leq t \leq t_p \\
Pt_p + S(s_0)\left( (t - t_c)^{1/2} - t_e^{1/2} \right) + k_s(t - t_p), & t > t_p
\end{cases}
\]  
(2.15)

From Eq. (2.15) one may now derive the normalized cumulative infiltration \(y(\tilde{D}, s_0)\) in analogy with that for the IEM, Eq. (2.4), as a function of the non-dimensional rainfall depth by dividing Eq. (2.15) by \(nZ_r\), substituting \(D/P\) for \(t\), and then non-dimensionalizing the precipitation rate by \(\tilde{P} = P/k_s\). The result is the somewhat complicated expression,

\[
y(\tilde{D}, \tilde{P}, s_0) = \begin{cases} 
\frac{\tilde{D}}{\tilde{P}}, & 0 \leq \tilde{D} \leq \frac{k_p}{nZ_r} \tilde{P}t_p \\
\frac{S(s_0)}{nZ_r} \left( \frac{nZ_r \tilde{D}}{k_s \tilde{P}} - t_c \right)^{1/2} - t_e^{1/2} + k_s \left( \frac{nZ_r \tilde{D}}{k_s \tilde{P}} - t_p \right), & \tilde{D} > \frac{k_p}{nZ_r} \tilde{P}t_p
\end{cases}
\]  
(2.16)
where $t_p$, $t_e$, and $t_c$ are all functions of both $s_0$ and $\tilde{P}$. Notice that since Philip’s solution assumes a semi-infinite domain, the cumulative infiltration is potentially infinite.

**Losses During Rainfall**

The model of infiltration described in the previous section only accounts for the cumulative infiltration across the soil surface and does not provide explicitly a method for determining the soil-moisture content of an active layer of soil. In order to model the change in mean soil moisture content in the upper soil layer (of depth $Z_r$) it is necessary to keep an account of the flux of water across the lower bound of this layer (i.e., percolation) during the rainfall event.

In the [72] model, however, losses were only included during the inter-storm periods. One consequence is shown clearly by comparing the time to soil saturation (given the linear increase in relative soil moisture during the period prior to ponding) with the calculated time to ponding derived from the time compression approximation. Combining Eqs. (2.9) and (2.13),

$$t_p = \left(\frac{nZ_r(1-s_0)}{P}\right) \left(\frac{\psi_s k_s (s_0^{(1+3m)/m} - 1)(2P - k_s)}{2Z_r(1 + 3m)(P-k_s)^2}\right),$$

from which it is clear that the first bracketed term represents the time to saturation if ponding does not occur, and thus that the second bracketed term must be less than or equal to unity in order for ponding to occur before the soil is saturated (a reasonable physical requirement). This condition is, in fact, not met identically.

Figure 2.2 illustrates the domain in which ponding occurs before saturation. [72] account for this possibility by including it as part of “infiltration excess.” So, while the time to ponding may in some cases violate physical sense, it presents no problem for simulation due to the bound imposed at $s = 1$.  

18
In order to avoid this unphysical result, the FDEM first incorporates percolation during storm events in the same form as Eq. (2.2). Following [72] we assume that evapotranspiration is negligible during storm events. One then has $-k_s s^{c+1}$, which is simply the loss for periods between storms, Eq. (2.2), without the evapotranspiration term. However, taking into account leakage during the storm in this manner does not solve the problem satisfactorily. While the percolation losses will strictly balance with Eq. (2.14) for very long durations, for finite times the infiltration term involving the sorptivity will still produce a supersaturation. This is due to the approximate nature of both the infiltration and the loss equations. In reality, soil moisture should asymptotically, and monotonically, approach saturation during a precipitation event of constant intensity. Thus, to prevent the supersaturation we further impose a bound at $s = 1$ which, however, produces a discontinuity in the $s(t)$ curve which is strictly an artifact of the model. The difference between the soil-moisture curve without the bound at $s = 1$ and that with the bound represents the model’s error in properly allocating moisture to runoff, storage, or leakage. Under our distinctions between runoff mechanisms it is therefore attributed to subsurface-controlled runoff.
In the FDEM this term is typically small compared with the other losses.

Model Summary

Following [33, 34] and [72] the FDEM is a physically-based model of vertically averaged soil moisture at the daily time scale which incorporates Philip’s (1957) infiltration solution coupled with the time compression approximation and the [11] model for percolation. The FDEM uses a non-overlapping, rectangular pulse model for rainfall for which the depths and durations are drawn from corresponding exponential distributions with means $\alpha$ and $\delta$. The mean inter-arrival time, $\tau$, is then chosen to be consistent with that of the IEM, $\lambda = (\tau + \delta)^{-1}$.

The evolution of soil moisture during storm events is described by the equation

$$n Z_r \frac{ds}{dt} = \begin{cases} 
  i(P, s_0, t) - k_s s^{c+1}, & s < 1 \\
  0, & s = 1 
\end{cases} \quad (2.18)$$

where $i(P, s_0, t)$ is the time dependent infiltration rate (given by Eq. (2.14) for a single rainfall event) and $P$ is the rainfall intensity. In the FDEM, as with the IEM, a bound is imposed at $s = 1$. Between storm events the model evolves according to Eq. (2.2). The steady-state probabilistic structure of this process is not known analytically and is thus determined by numerical simulation.

2.3 Model Comparisons

A combination of numerical simulations and analytic solutions were used to compare the two models. Analytic solutions exist for Philip’s infiltration with time compression approximation, as well as for the probability density of the full soil-moisture process defined in the IEM, Eq. (2.6).

Figure 2.3 illustrates the correspondence between the IEM and the FDEM for a simulation period of 100 days. The traces are almost identical with a notable excep-
Figure 2.3: Comparison of FDEM (solid line) and IEM (dotted line) soil-moisture models over one hundred days. The stochastic rainfall series of rectangular pulses is shown above. To the right is shown the simulated p.d.f. of the FDEM model (bars) with the analytic p.d.f. of the IEM model.

The net effect of the differences in infiltration modelling between the IEM and FDEM is illustrated in Figs. 2.4 and 2.5 which show the probability distributions of relative soil moisture generated with the FDEM agrees well with the analytical solution to the IEM.

Table 2.1: Table of parameter values used in simulation of soil-moisture and rainfall processes.

<table>
<thead>
<tr>
<th>Name</th>
<th>Units</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ψₕ</td>
<td>mm</td>
<td>-500</td>
</tr>
<tr>
<td>kₛ</td>
<td>mm day⁻¹</td>
<td>50, 200, 2000</td>
</tr>
<tr>
<td>m</td>
<td>[−]</td>
<td>0.5</td>
</tr>
<tr>
<td>Zᵣ</td>
<td>mm</td>
<td>300, 600</td>
</tr>
<tr>
<td>τ</td>
<td>hr</td>
<td>74</td>
</tr>
<tr>
<td>δ</td>
<td>hr</td>
<td>4, 6</td>
</tr>
<tr>
<td>λ</td>
<td>day⁻¹</td>
<td>0.15, 0.3</td>
</tr>
<tr>
<td>α</td>
<td>mm</td>
<td>12</td>
</tr>
<tr>
<td>Eₘₐₓ</td>
<td>mm day⁻¹</td>
<td>3</td>
</tr>
</tbody>
</table>
soil moisture for the two models. The four plots in Fig. 2.4 represent independent simulations between which the soil depth, mean rainfall frequency, and mean rainfall duration were varied. As one would expect, the FDEM simulation shows the greatest departure from the IEM when the soil is deep and rainfall is infrequent. Under these conditions the mean soil-moisture state is relatively dry leading to a high mean saturation deficit, while rainfall intensities are also high, leading to significant losses to runoff for the FDEM. However, even in these cases the correspondence between the two is very good. Figure 2.5 shows similar results for two different saturated conductivity values representing two orders of magnitude difference. Unsurprisingly, the deep clay soil shows the greatest discrepancy. Still the difference appears to be primarily in the position and less in the shape of the distribution.

Figure 2.4: Comparison of simulated FDEM (lines) and analytic IEM (shaded area) probability distributions for soil moisture. The four plots show varying soil depth and rainfall arrival rates. The two lines on each plot are for mean rainfall durations of 4 (solid) and 6 (dotted) hours. The plots on the left correspond to an index of dryness ($\lambda \alpha / E_{\text{max}}$) of 1.2 while those on the right have an index of 0.6.
Figure 2.5: Comparison of simulated FDEM (lines) and analytic IEM (shaded area) probability distributions for soil moisture. The two plots show the effect of soil saturated hydraulic conductivity on the correspondence between pdf’s of soil moisture according to the IEM and FDEM. The two lines on each plot are for mean rainfall durations of 4 (solid) and 6 (dotted) hours.

Given the correspondence between the two models evident from Figs. 2.3 and 2.4, it is worth taking a closer look at the relative importance of runoff and percolation in determining the change in soil moisture state due to a single event. Figure 2.6 illustrates the relationships between the models as they account for the partitioning of a rainfall event into constituent depths. The plot on the left shows the simple partitioning of the IEM into the depth contributing to a change in soil moisture and cumulative losses for a storm event as a function of rainfall depth for a given rainfall intensity and initial soil moisture state. The plot on the right of Fig. 2.6 gives a detailed account of the partitioning in the FDEM: The diagonal line of unit slope represents the dimensionless depth of water input to the system (equal to the event depth normalized by \( nZ_r \)). The curve just below this represents the infiltration model of [72] comprising the [107] infiltration solution and the time compression approximation. The difference between the two upper curves is that portion of the total depth which is lost to surface-controlled runoff. The next lowest curve in the diagram is that of the FDEM without the bound at \( s = 1 \). The difference between the [72] and FDEM curves is the effective portion of rainfall contributing to percolation. The bold curve represents the FDEM taking into account the bound at \( s = 1 \) and
Figure 2.6: Rainfall partitioning during a storm event for the IEM (left) and the FDEM (Right). The vertical axis represents the rainfall depth transformed by processes of infiltration and percolation. The normalized curves in the FDEM plot are, from highest to lowest: depth of rainfall event (slope = unity), infiltrated depth according to [72], infiltrated depth minus percolation according to FDEM without bound at $s = 1$, and the bold line represents the actual change in soil-moisture state as a function of rainfall depth according to the FDEM with the bound at $s = 1$.

represents the portion of a rainfall event that is stored in the rooting zone (i.e., the change in soil-moisture state). The difference between the FDEM curve without the bound at $s = 1$ and this bold curve is then a loss associated with the model error due to the approximation of infiltration and percolation functions and is here termed subsurface-controlled runoff.

From the point of view of simplified soil-moisture models one should notice that for all event depths the dominant loss during rainfall events is percolation (Fig. 2.6, shown for $\bar{P} = 2$). The character changes significantly for $\bar{P} > 2.5$ (not shown) as runoff plays a strongly increasing role. Secondly, the diagram in Fig. 2.6 may be somewhat misleading with respect to the values of $\bar{D}$ one may expect to encounter. A typical mean event depth, $\alpha = 12$ mm (used for the simulations in this paper), yields a mean value of $\bar{D}$ between 0.1 and 0.2 (depending on $Z_r$). In fact, $\bar{D} < 0.3$ for 95% of the rainfall events drawn from this exponential distribution.
Figure 2.7: Fraction of losses attributed to runoff as a function of total rainfall depth for both (a) very large depths and (b) typical rainfall depths. The curves represent different precipitation intensities increasing from the right: 3.5, 4, 4.5, 5, 6, 8, and 10, respectively.

From the diagram, at $\tilde{D} = 0.3$ the losses are almost entirely due to percolation. For larger rainfall intensities the proportion of losses due to runoff will increase, though for sites with moderate clay content the average rainfall intensity is unlikely to be much greater than that shown, particularly for longer durations. Figure 2.7 shows the fraction of total losses (runoff and percolation) due to surface-controlled runoff in the FDEM. The multiple curves can be interpreted as either increasing clay content for a given intensity or increasing precipitation intensity for a given soil.

For large depth events (long duration) the fraction of losses due to runoff approach (approximately) $L_R / L_{tot} = 1 - 1/\tilde{P}$. The relation is only approximate because of the model error which produces the small subsurface-controlled runoff term. Notice that for ponding to occur at such small event depths ($D \approx 25$ mm) for $s_0$ near the mode of the distribution, the intensity must be at least 3.5 times the saturated hydraulic conductivity.

While Fig. 2.6 illustrates the deterministic partitioning of a rainfall event into infiltration, runoff and percolation, this reveals little of the behavior of the two models as the parameters $s_0$ and $\tilde{P}$ vary (stochastically) during a growing season.
Figure 2.8: Change in soil moisture, \( y = \Delta s \), representing normalized net infiltration, for different values of \( \tilde{P} \) (left) and \( s_0 \) (right) for both the IEM and FDEM models. The dotted lines represent the IEM model.

Figure 2.8 shows how the change in relative soil-moisture state, \( y \), due to a single rainfall event varies with rainfall intensity and initial soil-moisture state in the two models. Notice that the change in soil moisture, especially for small values of \( \tilde{D} \), is strongly controlled by \( s_0 \).

Given the one-to-one relation between event depth and change in soil-moisture state (for given values of \( s_0 \) and \( P \)) represented by these curves along with the distribution of event depths, we may derive the probability distribution of change in soil-moisture state simply by transformation of variables. The result of the transformation, performed numerically, is shown in Fig. 2.9. Comparison of the two plots in Fig. 2.9 again supports the observation that the change in soil moisture due to a storm event is significantly more sensitive to initial soil-moisture state than to rainfall intensity. For \( s_0 = 0.8 \) the IEM significantly overestimates the probability of saturation (represented by the Dirac delta function at \( \tilde{D} = 1 - s_0 \)). The shape of the distributions from the FDEM as \( s_0 \) increases may be somewhat counterintuitive. Taking the \( s_0 = 0.8 \) case as an example, the shape can be understood in the following way. For \( t < t_p \) the change of variables is just a re-scaling of the exponential
Figure 2.9: Derived distributions of the normalized net infiltration for both the IEM and FDEM. Note that the IEM distribution (bold) is a truncated exponential with an atom of probability at $\tilde{D} = 1 - s_0$ represented by the corresponding Dirac delta functions.

curve. For durations (where storm duration and time are used here interchangeably) longer than $t_p$ the duration necessary to saturate the soil is significantly longer for the FDEM. In effect, a larger domain of event depths contributes to a smaller range of changes in soil moisture, which results in a redistribution of probability from the atom at saturation for the IEM to values of $y < 1 - s_0$. For $s_0 = 0.8$ the IEM has an atom of probability (exceedence probability for $\tilde{D} = 1 - s_0$) of approximately 0.14, while that for the FDEM model has an atom of only about 0.02.

Examination of the distribution of net infiltration, $y$, as $s_0$ and $\tilde{P}$ vary suggests no particularly straightforward method to improve the IEM with respect to losses during rainfall events. One possible correction is introduce another element of state dependence into the jump distribution. Whereas the IEM currently uses a jump distribution that is an exponential truncated at $y = 1 - s_0$ with mean $\gamma = \alpha/nZ_r$, one might define a state dependent mean which maps an exponential probability distribution with the same atom of probability at $1 - s_0$ as the FDEM distribution onto each value of $s_0$. Such an approach is the subject of future research and may still yield to analytical solution. This sort of correction is most likely to be of use
in wetter climates where the probability of high soil-moisture values is significant. Otherwise, as can be seen in Fig. 2.9, the effect of corrections will probably be of little value.

2.4 Conclusions

We have presented two models to compare the importance of resolving variable infiltration during storm events in capturing the dominant characteristics of soil-moisture dynamics. The first is a model of vertically averaged soil moisture forced by a marked Poisson arrival process. The second model is rooted in the treatment by [34] and [72] with a physically based description of infiltration which was further modified in this paper to include percolation losses.

In resolving both runoff and percolation, we have shown evidence that accounting for fractional loss to leakage during a storm event is probably of equal or more concern for improving the accuracy of simplified models than is runoff, particularly for events of lower intensity and longer duration. It is worth noting once more the significant difference between the IEM and the model of [72] in which losses during the storm event were neglected. The latter model is similar to the IEM except that it accounts for variable infiltration during the rainfall event. However, neglecting the losses to percolation (particularly for long durations) is a significant weakness for the [72] model. Since in the IEM events are instantaneous, percolation continues essentially uninterrupted. The IEM error is thus concentrated at an instant in time and is then damped quickly by the strongly nonlinear character of percolation, while the [72] model spreads the error over the duration of the event. For longer rainfall durations, therefore, the [72] model may be expected to overestimate infiltration to a greater extent than the IEM. In such cases the gains of representing temporally extended rainfall events with variable infiltration are outweighed by the error of neglecting percolation.
The highly simplified IEM performs well against more complex, physically-based models such as the FDEM (Fig. 2.3) in reproducing the probabilistic structure of soil-moisture dynamics (Fig. 2.4). As expected, the most significant difference between the models occurs under conditions of intense rainfall over short duration, in which case the IEM will consistently overestimate infiltration. However, Fig. 2.5 suggests that the primary difference in the probability density is one of location and not shape. Our analysis has been conservative with respect to the frequency of intense rainfall, as the use of Eq. (2.7) likely overestimates its frequency, thus likely exaggerating the importance of runoff in simulations. Also, while the IEM used here incorporates a very simple mechanism for losses during storm events, the model described by Eq. (2.5) retains a significant amount of flexibility through the definition of the $I_D$ function. We find, however, that even in this conservative analysis the IEM reproduces well the probabilistic structure of soil-moisture dynamics.
3

Spring Frost Risk in a Changing Climate


3.1 Introduction

Spring frost damage to vegetation is determined by the relative timing of the opening of vegetation buds (hereafter “budbreak”) and freezing temperatures. While frost risk to trees has received significant attention in the literature [60, 76, 86], the emphasis has been on improving the phenology component of the model for more accurate prediction of budbreak. Under current climatic conditions the emphasis on developing the phenological model rightly reflects the goal of understanding the processes involved in phenological development. However, inquiry into the likely effects of changing climate (i.e., changes in the temperature process) on frost risk in these studies has relied predominantly on simple increases in monthly mean temperatures to represent future temperature scenarios [60, 61, 76, 77, 86, 62, 128]. The drawback of this approach is that climatic change probably will not result in a simple translation of mean daily temperature and that vegetation development as a dynamical
phenomenon will be sensitive to other statistical characteristics of daily temperature besides the mean. We may assume that vegetation is adapted to a particular risk environment (i.e., to the particular character of spring temperature fluctuations) and that subtle changes in the statistical character of daily temperature might still produce noticeable effects in vegetation.

Heuristically, frost risk may be thought of as a two-stage problem. There is first the early spring, temperature-dependent development of vegetation up to the point of budbreak. Following budbreak there is then the risk of temperature dropping below a damaging threshold. By responding to warmer temperatures in early spring, plants face a risk of damaging fluctuations after budbreak. As an example of such events Figure 3.1 shows NDVI data for Walker Branch, TN for 2006 and early 2007 along with minimum temperature data from Durham, NC. The damaging frost indicated by the NDVI was a widespread, abnormally warm period in March followed by almost a week of temperatures well below freezing in early April which affected much of the southeastern and central United States with devastating effects for several agricultural crops [see [53] for detailed description]. Notice from the figure that freezes of similar character occurred almost symmetrically around the mean date of last spring frost [29] in 2007 with only the later one causing damage, thereby highlighting the threshold-like dependence on temperature characterized by budbreak.

Our approach begins with a theoretical analysis using a stochastic temperature process. Assuming that the rate of bud development prior to budbreak is a function of temperature, we treat the time of budbreak as the stochastic-crossing of a threshold in the development variable. The risk of damaging frost then becomes a second crossing problem, this time concerning the crossing of a low-temperature threshold after budbreak. We then illustrate the general theory with a particular model of temperature and phenology, using temperature data from Durham, NC to calibrate the temperature process, in order to quantify potential changes in risk due to these
3.2 Theory

3.2.1 Timing of Budbreak

The development of vegetation buds is generally a function of many environmental variables with the strongest contribution typically coming from temperature [15]. Following the approach of so called “Thermal Time” models of plant phenology, budbreak may be modeled as a threshold, \( D_* \), in the development process \( D(t) \), such that the rate of development is a function of some representative temperature [60, 86]. The corresponding time of budbreak, \( t_b \), at which \( D(t_b) = D_* \) will have probability density function (pdf) \( p_{t_b}(t_b) \) which is determined by the particular species and its adaptation to the local climate. Thus \( p_{t_b}(t_b) \) is the solution to the first-passage time of the biological development driven by the stochastic temperature process.

For the purposes of determining subsequent frost risk the variable of interest is
clearly daily minimum temperature. The representative temperature used for model-
ing the rate of bud development is, however, generally the daily average temperature. We will assume that driving phenology with daily minimum temperature (as opposed to daily average temperature, for example) results only in a change in the parameter values of the plant phenology model and does not qualitatively change the statistical characteristics of development. The daily minimum temperature will be modeled as a stochastic process with parameters $k = \{k_0...k_n\}$ (corresponding to mean, variance, etc.) and probability distribution, $p_T(T; t)$, which represents the probability density at time $t$ (interpreted at the daily scale) for the stochastic temperature process.

3.2.2 Frost Risk

To characterize the risk of frost we consider the problem of temperature dropping below some threshold $T_*$ (such as freezing) after budbreak. If we assume that any damaging event is a killing event (i.e., there is no variation in the degree of damage with temperature), such that no open vegetation buds survive the freeze, then we are most interested in the pdf of the time between budbreak and the first subsequent freeze, $\tau = t_f - t_b$, where $t_f$ is the time of first freeze after budbreak. This distribution may be expressed as $p_{\tau|t_b}(\tau|t_b)$. Note that $\tau$, $t_b$ and $t_f$ are clearly dependent variables. Note also that since it is not certain that a damaging frost will occur in spring, the pdf’s of $\tau$ and $t_f$ will generally have an atom of probability at infinity corresponding to the probability that no damaging frost occurs. As the probability of an infinite $\tau$ is not of practical use, we will use $f_{\tau|t_b}(\tau|t_b)$ and $f_{t_f|t_b}(t_f|t_b)$ to denote these densities without the atom at infinity. In this case, rather than integrating to unity, the distribution functions integrate to the probability of a freezing event occurring after $t_b$ in spring.

The distribution of times of first freeze after budbreak is given by,
\[
f_\tau(\tau) = \int_0^\infty f_{\tau|t_b}(\tau|t_b) p_{t_b}(t_b) dt_b \tag{3.1}
\]
and reflects the distribution of times to first frost after budbreak for an ensemble of years.

From the distribution of \( \tau \) we may derive the distribution of the date of first frost as the distribution of the sum of two random variables \( t_f = \tau + t_b \) where the joint probability of \( \tau \) and \( t_b \) is given by \( f_{\tau,t_b}(\tau,t_b) = f_{\tau|t_b}(\tau|t_b) p_{t_b}(t_b) \).

Then,

\[
f_{t_f}(t_f) = \int_0^\infty f_{\tau|t_b}(\tau|t_b) p_{t_b}(t_f - \tau) d\tau. \tag{3.2}
\]

From Eq. (3.2) we can define a “survivor function” of time \([19]\),

\[
F(t) = 1 - \int_0^t f_{t_f}(u) du \tag{3.3}
\]

which represents the probability that a plant survives at least to time \( t \). From the survivor function the ensemble “age-specific failure rate” \([19]\), or “hazard rate function” \([5]\) may be defined as,

\[
\lambda(t) = -\frac{F'(t)}{F(t)} \tag{3.4}
\]

which gives the conditional rate of damaging events occurring at time \( t \) given no previous freezing event. It can be shown \([19]\) that the age-specific failure rate completely determines \( f_{t_f}(t) \) with the relationship,

\[
f_{t_f}(t) = \lambda(t) e^{-\int_0^t \lambda(u) du}, \tag{3.5}
\]

which formally corresponds to the distribution of time between hypothetical Poisson events with an inhomogeneous rate, \( \lambda(t) \). From the formal analogy there is a direct
relationship between the integral of $\lambda(t)$ and the survivor function,

$$F(t) = e^{-\int_0^t \lambda(u)du},$$

so that clearly there is a one-to-one relationship between the probability of survival and the area under $\lambda(t)$. Thus, $\lambda(t)$ indicates the temporal distribution of the risk of mortality due to low temperatures, thus the name “hazard rate”.

3.3 Data and Model

3.3.1 Temperature Data from Durham, NC

As a guide for developing a temperature model, we used daily minimum temperature data for Durham, NC, from 1919 to 2007. The data set, available from the National Climate Data Center (NCDC), station number 312515.

The average spring trend (March-May) was found to be linear. For each year we then performed a linear least-squares fit to the spring temperatures, removed the trend and calculated the autocorrelation of detrended process. The resulting (average) autocorrelation was found to be approximately exponential (rate given in Table 1). The residuals from the linear fits were also approximately Gaussian.

3.3.2 Temperature Model

The logical choice of model for the Durham data is an Ornstein-Uhlenbeck process with an added linearly increasing trend (see e.g.,[73]). The Langevin equation for the standard Ornstein-Uhlenbeck process is

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$k_0$</th>
<th>$k_1$</th>
<th>$k_2$</th>
<th>$k_3$</th>
<th>$\alpha$</th>
<th>$T_{th}$</th>
<th>$t_0$</th>
<th>$D_*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.78</td>
<td>0.81</td>
<td>0.16</td>
<td>32.6</td>
<td>1.62</td>
<td>2.8</td>
<td>March 1</td>
<td>100</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>4.7</td>
<td>0.38</td>
<td>0.05</td>
<td>17</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Table 3.1: Summary of Parameter Values
\[ \frac{dx}{dt} = -k_1x + \sqrt{k_3}\eta_t, \] (3.7)

where \( x \) is the state variable and \( k_1 \) and \( k_3 \) are the drift and diffusion parameters, respectively, and \( \eta_t \) is a delta correlated Gaussian noise with unit variance. For our purposes we consider \( x \) to be the “detrended” temperature, such that \( x = T - T_\mu(t) \).

For the specifics of our case we treat the spring daily mean temperature as increasing linearly during the months of March through May with trend \( T_\mu(t) = k_0 + k_2t \).

Making the substitution gives us the final equation,

\[ \frac{dT}{dt} = k_2 - k_1(T - k_0 - k_2t) + \sqrt{k_3}\eta_t \] (3.8)

where \( k = \{k_0, k_1, k_2, k_3\} \) are the parameters of the temperature process. The expected value of the temperature process can be checked easily by integration to arrive again at \( T_\mu \). Our temperature process is then Gaussian with mean \( T_\mu = k_0 + k_2t \), variance \( \sigma^2 = k_3/2k_1 \), and autocorrelation function (of the detrended process) \( \rho(\tau) = e^{-k_1\tau} \).

3.3.3 Phenology Model

We used a simple Thermal Time (or Degree Day) model for the phenology of bud-break (see e.g., [14, 15]). Let the rate of daily ontogenetic development [61] be a function of temperature given by,

\[ \frac{dD_o}{dt} = 0, \quad T < T_{th}, \] (3.9)

\[ \frac{dD_o}{dt} = \alpha(T(t) - T_{th}), \quad T \geq T_{th}, \] (3.10)
where $D_0$ is the level of ontogenetic development, $\alpha$ is a rate constant (which may more generally be a function of other environmental variables) and $T_{th}$ is a threshold temperature above which ontogenetic development proceeds. It is further assumed that degree days begin to accumulate on an arbitrary date, $t_0$ (here March 1), and that the time of budbreak, $t_b$, is defined arbitrarily as $D_* = 100$. Parameters of the phenology model were chosen to give a date of budbreak in early April for the mean temperature process (see Table 1 for values).

### 3.3.4 Simulation

For each parameter set, $10^5$ realizations of temperature and phenology were generated. Each realization represents a simulation of a single spring season from March 1 for 120 days or until a damaging frost, whichever occurred first. From these iterations we constructed the pdf of the day of budbreak and the distribution of times of damaging frost, as well as the survivor function and the frost hazard rate.

### 3.4 Results

We simulated temperature and time of budbreak with a simple, coupled stochastic model containing 8 parameters (see Table 1), four each pertaining to the temperature and phenology components. The four temperature parameters correspond to the mean temperature on March 1 ($k_0$), the decay rate of the autocorrelation function for the detrended temperature process ($k_1$), the mean linear temperature trend ($k_2$), and the variance ($k_3 = 2\sigma^2k_1$). The four phenology parameters are the date on which ontogenetic development begins to accumulate ($t_0$ = March 1), the arbitrary level of development at which budbreak occurs ($D_* = 100$), the rate of development ($\alpha$), and the temperature threshold above which development accumulates ($T_{th}$). Of the phenology parameters, $t_0$ is of little importance as it is largely controlled by
photoperiod and will therefore not vary with climate. Thus, as long as we choose $t_0$ reasonably, it is of little consequence what the exact value is [86]. Similarly, $D_*$ is an arbitrary level the effect of which may be absorbed into the rate and threshold parameters. Therefore, our analysis will focus on the effect of six parameters: the four temperature parameters, $\{k_0, k_1, k_2, k_3\}$, and the two phenology parameters, $\{\alpha, T_{th}\}$.

Figure 3.2 shows the effect of parameter variability on the temperature and phenology models. Each plot shows the effect on the survivor function’s value at the end of spring (representing the probability of a plant surviving spring temperatures), $F(\infty)$, and the most probable date of budbreak. Each plot was generated by varying a single parameter while the remaining parameters were held fixed at their mean values (see Table 1). The vertical line in each plot shows the sample mean from the Durham data. Notice that the greatest changes in both survival probability and date of budbreak are induced by changing the mean daily temperature, $k_0$. An increase in mean daily temperature ($k_0$) of a single degree, with all other parameters held constant, would result in roughly a 0.05 increase in probability of survival. Changes in both mean slope ($k_2$) and variance ($k_3$) have significant effects on the probability of survival but significantly less influence on the date of budbreak, such that a 10% change in either parameter results in a approximate change in probability of survival of 0.05. Finally, changes in autocorrelation ($k_1$) tend to have almost no effect on the date of budbreak and only a change in survival probability of 0.1 for the range $\pm 35\%$ of the estimated mean value of $k_1$ at Durham, NC. Note that the low probability of survival predicted under current climate ($F(\infty) \approx 0.4$) is a result both of model assumptions (i.e., all buds open simultaneously and any damaging event results in complete mortality) and the fact that the phenology parameters have not been estimated here against data as they would be in practice.

With regard to the phenology parameters ($\alpha$ and $T_{th}$), as one might expect,
Figure 3.2: Sensitivity of the probability of survival, $F(\infty)$, (circles) and the most probable date of budbreak (squares) to parameters of the temperature and phenology models. Each plot was generated by varying a single parameter while holding the others at their mean value (see Table 1). The vertical line on each plot represents the mean value of the varied parameter as estimated from the Durham, NC data set.
increasing the rate of development advances the date of budbreak and increases the risk of frost damage. Conversely, increasing the threshold for development has the opposite effect. Thus the phenology parameters have a straightforward effect on the probability of survival by adjusting the date of budbreak.

3.5 Discussion and Conclusions

In this paper we have presented a theory for characterizing the probabilistic risk of spring frost to vegetation based on the date of budbreak. Spring bud phenology is characterized by two stochastic crossing problems, one of ontogenetic development up to budbreak which is driven by temperature and the other concerning the probability of temperature down-crossing a minima threshold representing frost damage. Using a stochastic process to simulate temperature that in turn forced a simple Degree Day model of plant phenology we have shown that noticeable changes to the risk of frost may be effected by changes in any of the parameters controlling the temperature process (Fig. 3.2). Furthermore, the sensitivity of spring frost risk (represented
by $F(\infty)$ to a changing temperature regime is not necessarily concentrated in the
parameters defining mean temperature behavior ($k_0$ and $k_2$). That is, judging by
the change in $F(\infty)$ induced by a given percentage change in each parameter, the
variance (not the mean) appears to be the strongest determinant. Specifically, an
increase in $k_0$ of a few degrees would increase the probability of survival by about 0.1,
a gain which could be lost through a concomitant 10% increase in the variance. Figure
3.3 shows the relationship between survival probability and temperature variance for
multiple increases in $k_0$ (corresponding to daily mean temperature increases).

In our analysis we have shown that changes in daily temperature correlation and
variance must be taken into account in any analysis of projected effects of climate
change on frost risk. More subtly, the results shown in Fig. 3.2 suggest that a diag-
nostic phenological study (e.g., one focused on detecting evidence of changing climate
via changes in date of budbreak) may overlook linkages between phenology dynamics
and changing climate driven by other statistical characteristics of temperature. For
example, changes in species composition driven by changes in risk of spring frost
damage (and thus survival probability) may not necessarily be accompanied by, nor
be caused by, significant changes in date of budbreak. Such might be the case if the
daily temperature displayed a subtle “whitening” (that is, faster decay in temporal
autocorrelation, see Fig. 3.2).

Gu et al. [2008] have recently shown through a detailed characterization of the
April 2007 freeze event that spring frost events can have extreme short-term bio-
logical effects, and may yet have important long-term effects, for both natural and
agricultural systems over a wide areal extent. Our study highlights the delicate
balance of risk for vegetation in spring that which vegetation seeks to avoid frost
damage. It is thus important for climate change research to address the sensitivity
of biological systems to subtle changes in spring climate as this linkage may lead to
unexpectedly strong responses in ecosystems. We have approached the problem of
“extreme spring temperatures” from the plant’s point of view by setting the problem in terms of stochastic environmental variables and associated biological response. This is in line with [59] who outlined a need for understanding “how extreme events shape individuals, communities, and ecosystems” by, in part, seeking to define environmental extremes relative to the biological adaptation of a community. We believe that in such a way the framework of coupled stochastic crossing properties may be of use in quantitative analysis of “biologically-defined” extremes in a variety of biological settings. This research represents only a first step in providing a quantitative approach to understanding “biologically-defined” extreme climatic events.
4

Irreversibility and Fluctuation Theorem in Stationary Time-Series


4.1 Introduction

Detecting time asymmetry or irreversibility in stationary time series is important not only because time series may be easier to predict and model in one direction, but especially because irreversibility is a symptom of non-Gaussian forcing and dynamic nonlinearities [21, 82]. Moreover, when time series represent the evolution of either stochastic or deterministic dynamical systems, irreversibility takes on a special meaning linked to the lack of equilibrium and detailed balance of the probability fluxes among the system states. Despite its importance, however, the issue of time irreversibility has received relatively less attention compared to other aspects of nonlinear time series analysis, with a few notable exceptions [20, 82, 109, 21, 136, 26,
In thermodynamics and statistical mechanics, reversibility is synonymous with equilibrium. Steady state systems that are in equilibrium obey detailed balance, while non-equilibrium steady-state (NESS) systems are time irreversible and have a positive internal entropy production rate \[75\]. Recently, a renewed interest has been sparked by the discovery of general relationships valid also far from equilibrium \[13, 38\]. The fluctuation theorem (FT), in particular, links the probability of realizations that consume entropy to those that produce it as a function of the system size. The theorem was originally proposed for many particle systems \[37, 44\], and was then verified experimentally \[16\] and derived theoretically for Markovian processes \[83, 48, 68, 79, 23\].

In this paper we borrow from the thermodynamics of NESS and information theory to propose a consistent framework to quantify the degree of asymmetry in stationary time series and show that FT-type relationships derive from simple symmetries between forward and backward sequences which are valid with great generality for any discrete and continuous stationary time series.

4.2 Measure of Asymmetry

Let us begin by considering a stationary time series, \(x\), assumed for now to be discrete in time and with finite state space. The series may be naturally discrete or have been discretized from a continuous series after suitable coarse graining (i.e., a partitioning as in \[50\]). Let the joint probability distribution of a sequence of \(n\) consecutive values of the time series be \(p(x_1, x_2, ..., x_n)\), and the corresponding distribution of the reverse sequence \(\hat{p}(x_1, x_2, ..., x_n)\). A fundamental symmetry links the forward and the backward distributions,

\[
\hat{p}(x_1, x_2, ..., x_n) = p(x_n, x_{n-1}, ..., x_1). \tag{4.1}
\]

44
This property is very general and independent of any stationarity assumption.

With these definitions, a stationary time series is said to be reversible (or time symmetric) if and only if \( p(x_1, x_2, \ldots, x_n) = \hat{p}(x_1, x_2, \ldots, x_n) \) for any \( n \). Thus, clearly, any time series generated as a Bernoulli sequence is reversible. In the case of Markov chains with transition matrix \( P_{ij} \) and steady state distribution \( \pi_i \), time symmetry (or reversal) holds if \( \pi_i P_{ij} = \pi_j P_{ji} \), and the chain is said to be in detailed balance. Although the backward sequence is still generated by a Markov chain [98], this has a different transition matrix, \( \hat{P}_{ij} = \frac{\pi_j}{\pi_i} P_{ji} \). For general Markov processes this property is generalized as \( \hat{p}(x_1|x_2)p(x_2) = p(x_2|x_1)p(x_1) \) (see [70], p. 83).

For general stationary time series the degree of time irreversibility and non-equilibrium can be determined by how different the backward and forward joint probability distributions are. A natural statistic to quantify the difference between \( p \) and \( \hat{p} \) is given by the so-called relative entropy or Kullback-Leibler distance [18]

\[
\langle Z_n \rangle = \sum p(x_1, x_2, \ldots, x_n) \log \frac{p(x_1, x_2, \ldots, x_n)}{\hat{p}(x_1, x_2, \ldots, x_n)},
\]

where the sum is intended over all the possible states \( x_1, x_2, \ldots, x_n \). Eq. (2) can be interpreted as the mean of the difference between the ‘surprise’ of finding a given sequence in forward time, i.e., \( \log p \), and in reverse time, i.e., \( \log \hat{p} \), or equivalently

\[
Z_n = \log \frac{p(x_1, x_2, \ldots, x_n)}{\hat{p}(x_1, x_2, \ldots, x_n)}.
\]

\( \langle Z_n \rangle \) is always positive and it is zero only if the two distributions are equal [18]. It can be shown to be symmetric, \( \langle Z_n \rangle = \langle \hat{Z}_n \rangle \), because of Eq. (4.1), and that \( \langle Z_n \rangle \geq \langle Z_{n-1} \rangle \). The same measure is used to quantify the lack of equilibrium and detailed balance in NESS systems [83, 48].

\( \langle Z_n \rangle \) can also be expressed as a difference between the so-called block entropy...
[24], $H_n$, and another form of entropy, $H_n^R$, introduced in the context of NESS thermodynamics [48]. Division by $n$ transforms these quantities into entropy rates, the limits of which converge to

$$\lim_{n \to \infty} \frac{\langle Z_n \rangle}{n} = \frac{H_n^R}{n} - \frac{H_n}{n} = h^R - h = \sigma,$$  \hspace{1cm} (4.4)

where $h$ is the Kolmogorov-Sinai entropy, and $\sigma$ is related to the internal entropy generation rate (it is always positive and zero only for time-reversible stationary time series). It is important to note that $\langle Z_n \rangle$ is infinite if at least one sequence is not found in reverse, i.e., if $\hat{p}$ is zero for at least one sequence. This is the case of irreversible periodic signals and of Markov chains that have asymmetric zeros in their transition matrices. The latter fact was already noted in [68].

The behavior of $\langle Z_n \rangle/n$ as a function of $n$ is of particular interest in time series as it describes the degree of temporal asymmetry at different scales. In the trivial case of Bernoulli sequences, as well as in the case of two state Markov processes [109], $\langle Z_n \rangle$ is always zero, while for Markov chains with three or more states one gets [48, 68]

$$\langle Z_n \rangle = n \sum \pi_i P_{ij} \log \frac{\pi_i P_{ij}}{\pi_i P_{ji}},$$  \hspace{1cm} (4.5)

for $n \geq 2$, which shows, as expected from the behavior of $H_n/n$ [24], that the corresponding rates are constant and equal to $\sigma$ for $n \geq 2$.

### 4.3 Fluctuation Theorem

An important property for $Z_n$ can be derived solely from the symmetry (1) and using the moment generating function (MGF) of $Z_n$, $G_n(k) = \langle e^{-kZ_n} \rangle$. In fact, writing
\[ G_n(k) = \sum p(x_1, x_2, ..., x_n) \left[ \frac{p(x_1, x_2, ..., x_n)}{\hat{p}(x_1, x_2, ..., x_n)} \right]^{-k} \]
\[ = \sum p(x_1, x_2, ..., x_n) \left[ \frac{p(x_1, x_2, ..., x_n)}{\hat{p}(x_1, x_2, ..., x_n)} \right]^{-(1-k)}, \quad (4.6) \]

which follows from (4.1) and the commutative property of summation, one obtains

\[ G_n(k) = G_n(1 - k). \quad (4.7) \]

Eq. (7) is the FT in the \( k \) domain, and implies that the MGF is symmetric around \( k = 1/2 \). Since it is also known that \( G_n(0) = 1 \) and that the MGF is always convex, this further confirms that the mean of \( Z_n, \langle Z_n \rangle \), is always greater than or equal to zero (see [22], pp. 48-49). Eq. (4.7) also provides a symmetry among the moments of \( p/\hat{p} \). In particular, \( k = 1 \) corresponds to the harmonic mean of \( p/\hat{p} \) which is therefore always equal to 1.

A more meaningful form of the FT is obtained in terms of the probability distribution of \( Z_n, p_{Z_n}(Z_n) \). Using the property of the MGF (e.g., discrete Laplace transform) that \( L^{-1}\{G_n(k - a)\} = e^{aZ_n}L^{-1}\{G_n(k)\} \) and \( L^{-1}\{G_n(-k)\} = p_{Z_n}(-Z_n) \), where \( L^{-1}\{\cdot\} \) denotes inverse Laplace transform, it is immediate to show that

\[ p_{Z_n}(-Z_n) = p_{Z_n}(Z_n)e^{-Z_n}, \quad (4.8) \]

which implies that the negative tail of the probability distribution decays faster than the positive one. It is important to highlight that the two equivalent properties (4.7) and (4.8) require only stationarity, which is implicit in the definition of the averaging operation of the MGF, and are a direct consequence of the symmetry property (1) used in Eq. (6).
4.4 Continuous Time Series

All of the above properties remain valid for time series in continuous time and with continuous state space, with only the proviso of extending the Kullback-Leibler distance to continuous variables using differential entropies ([18], p. 231). A different and useful formulation valid at a point in time can also be obtained assuming that the \( n \) points in the sequence are separated by \( \Delta t \), and then considering the limit \( \Delta t \to 0 \). In this limit, in fact, there is a one-to-one correspondence between the joint probability density functions (pdf’s) of a sequence of \( n \) points of a time series and that (distinguished by an asterisk) of its subsequent time derivatives, at a point in time,

\[
p(x_1, x_2, ..., x_n) \sim p^* \left( x, \frac{dx}{dt}, \frac{d^2x}{dt^2}, ..., \frac{d^{n-1}x}{dt^{n-1}} \right) \bigg|_{x=x_1}.
\]  

(4.9)

Taking into account that, when the continuous time series is looked at backwards in time, the odd derivatives change sign compared to the forward ones, the fundamental relationship (1) now becomes

\[
\hat{p}^*(x, \frac{dx}{dt}, ..., \frac{d^{n-1}x}{dt^{n-1}}) = p^*(x, -\frac{dx}{dt}, ..., (-1)^{n-1} \frac{d^{n-1}x}{dt^{n-1}}).
\]  

(4.10)

Thus, in analogy to the discrete case one can define the continuous version of Equation (3) as

\[
Z_n^* = \log \frac{p^*}{\hat{p}^*},
\]  

(4.11)

where the arguments of the joint pdf’s have been dropped for conciseness. Since this new definition refers only to a point in time, it can also be used to characterize non-stationary signals.
Using conditional probabilities in Eq. (11), the measure of asymmetry given by the mean of $Z_n^*$ can be written as the mean of the asymmetries at any given level $x$,

$$
\langle Z_n^* \rangle = \int p_X(x) \langle Z_n^* | x \rangle dx,
$$

(4.12)

where $p_X(x)$ is the pdf of the stationary process and

$$
\langle Z_n^* | x \rangle = \int \ldots \int p^*(\dot{x}, \ddot{x}, \ldots | x) \log \frac{p^*(\dot{x}, \ddot{x}, \ldots | x)}{p^*(-\dot{x}, \ddot{x}, \ldots | x)} d\dot{x} d\ddot{x} \ldots,
$$

(4.13)

where, for conciseness, we used the dot notation for time derivatives (The same approach using conditional probabilities is equally valid for the discrete case). While (12) and (13) are problematic to compute for real time series, a simple measure of asymmetry involving the first time derivative of a signal can be obtained as

$$
\langle Z_2^* \rangle = \int p_X(x) \int p_X(\dot{x} | x) \log \frac{p_X(\dot{x} | x)}{p_X(-\dot{x} | x)} dx d\dot{x},
$$

(4.14)

where the integrals extend over the whole domains of $X$ and $\dot{X}$. In applications to time series, $p_X(\dot{x} | x)$ can be easily computed for different values of $x$ and then averaged out to obtain (14). Since $\langle Z_n^* \rangle$ is strictly increasing with $n$, $\langle Z_2^* \rangle$ provides a sufficient condition for asymmetry.

Similarly to the discrete case, Eqs. (7) and (8) also hold for the fluctuations of $Z_n^*$. This can be shown as before, using the MGF and the properties of the continuous (bilateral) Laplace transforms ([22], pp. 48-49). Interestingly, however, the FT is now also valid at a point in time (does not depend on the stationarity assumption) and for any given value of $x$, i.e.,

$$
p_{Z_n^* | x,t}(-Z_n; x, t) = p_{Z_n^* | x,t}(Z_n; x, t) e^{-Z_n}.
$$

(4.15)
For \( n = 2 \), Eq. (15) provides a useful relationship for the pdf of \( Z_2^* = \ln \frac{p_X(x|x)}{p_X(-x|x)} \) for a given \( x \) that, as will be seen below, can be readily verified in real time series.

Before moving to the applications, it is interesting to derive directly the previous property for a general one-dimensional Langevin equation driven by Gaussian noise, with drift \( f(x) \) and diffusion \( g(x) \). In this case, \( \dot{X}|x \) has a Gaussian distribution with mean equal to \( f(x) \) and variance \( \lim_{\Delta t \to 0} \left[ \frac{g^2(x)}{\Delta t} \right] \) \cite{22, 117, 46, 70}. It follows that \( Z_2^*|x \) is a linear function of \( \dot{X}|x \), i.e., \( Z_2^*|x = \lim_{\Delta t \to 0} \left[ 2(\Delta x/\Delta t)f(x)/(g^2(x)/\Delta t) \right] \). As a result, \( Z_2^*|x \) is still a Gaussian with mean \( \lim_{\Delta t \to 0} \left[ 2f^2(x)/(g^2(x)/\Delta t) \right] \) and variance always equal to twice the mean, which is exactly the condition for a Gaussian distribution to satisfy Eq. (4.15). It is also worth noticing that since the mean of \( Z_2^*|x \) is proportional to \( \Delta t \), \( \langle Z_2^*|x \rangle \to 0 \) for \( \Delta t \to 0 \). Since the process is Markovian this is also true for \( n \geq 2 \), and therefore detailed balance and reversibility are always satisfied. This fact is well known, as for stationary solutions of Langevin equations detailed balance is equivalent to the existence of a potential solution \cite{46, 70}.

4.5 Applications

The estimation of the relative entropy as our measures of asymmetry, \( \langle Z_n \rangle \) and \( \langle Z_n^* \rangle \), especially for high values of \( n \), raises several technical issues \cite{78} worthy of future detailed investigation. Here we use only a simple binning procedure, based on equiprobabilistic partitioning, to discretize the support of the forward and backward probability distribution, limiting the estimation to relatively coarse partitioning and small values of \( n \). A further difficulty for estimation is the occurrence of infinities. Recall that if any sequence is found in the forward direction but not in reverse, the ratio of the probabilities generates an infinity. Due to statistical errors in the estimation of small probabilities for finite data sequences, we stipulated that only finite ratios of the probabilities may contribute to the estimate, thus excluding infinities.
Figure 4.1: Asymmetry statistic, $\langle Z_n \rangle / n$, as a function of $\alpha$ (i.e., for increasingly Gaussian signals) for different block lengths ($n = 2$, circles; $n = 3$, squares; $n = 4$, diamonds). The insets show portions of time series, $x_t$, for three values of $\alpha$. Values used for simulations: $k = 0.07$, $\gamma = 0.04$, $\lambda = 0.28$.

generated by possible forbidden sequences in the reverse direction. Note that in the full classification of asymmetric series there may be those that are asymmetric only by forbidden sequences, those that have both forbidden sequences and other finite asymmetries (e.g., $Z_n$ or $Z^*_n$ finite and different from zero), and those with only finite asymmetries. Our estimation method attempts only to detect the latter two scenarios.

Following this approach, we first applied our method to stochastic time series generated by

$$x_t = (1 - \alpha) \eta_t + \alpha \xi_t$$

(4.16)

where $\eta_t$ is the (asymmetric) jump process with exponential decays of rate $k$ and exponentially distributed jumps of mean $\gamma$ occurring as a Poisson process with rate $\lambda$ [?], while $\xi_t$ is an Ornstein-Uhlenbeck process with mean and variance equal to those of $\eta_t$. Figure 1 shows $\langle Z_n \rangle / n$ estimated for different $\alpha$ for three values of $n$ and
equi-probable partition with eight states. Notice the relatively smooth decay to zero as the Gaussian noise becomes more dominant, reflecting the decreasing asymmetries in the signal. Also, \( \langle Z_n \rangle / n \) appears to increase with \( n \) before decaying to a finite limit for large \( n \) (not shown). Computations not reported here also confirm the validity of the fluctuations theorem for various \( n \) and partition number.

![Image of Po River discharge series](image)

**Figure 4.2:** Illustration of the FT applied to Po River (Italy) discharge series. a) Subset of the Po River discharge series, b) probability distribution of time derivative given \( x \) chosen as the mode of the distribution, c) the functional form of \( Z_n^2(x|\dot{x}) \), d) the probability distribution of \( Z_n^2(x|\dot{x}) \) derived from b) and c), and e) the resulting FT for the pdf of \( Z_n^2|x \) where \( x \) is chosen respectively as the mode (circles), mean (squares), and the median (crosses) of the discharge series. Note that b) and d) are shown with suitably coarse binning for the purposes of clear illustration while do not match, computationally, the values reported in c) and e) which are computed with finer resolution.

As a second application we illustrate the validity of the FT for the distributions
of time derivatives for different values of $x$, $Z^*_2|x$. To this purpose we have taken 50 years of discharge measurements from the Po River, Italy, (1935-1985) [114] and took sections at the mode, median, and mean of $x$ to derive the statistics of $Z^*_2|x$ for each. Each slope in the series, $x_{t+1} - x_t$, was assigned the coordinates of its midpoint. The ordinate axis was then binned and all slopes with midpoints lying in the bin centered at the chosen value of $x$ were selected for analysis. Figure 2b shows the resulting distribution of slopes after taking a section at the mode of the discharge series. From the distribution of slopes $Z^*_2|x$ is found by taking $\log[p(\dot{x}|x)/p(-\dot{x}|x)]$. The resulting relationship, Fig. 2c, is then used to derive the probability distribution of $Z^*_2|x$ by change of variable, Fig. 2d. The FT is illustrated in Fig. 2e by taking $\log[p(Z^*_2|x)/p(-Z^*_2|x)]$ for the mode, median, and mean of the time series.

4.6 Conclusion

We have presented a statistic measuring temporal asymmetry in time series which draws on the theory of NESS in thermodynamics and retains general applicability for both discrete and continuous cases. We have shown also that the symmetry in the pdf of $\log(p/\hat{p})$, referred to as the FT, applies generally to measured stochastic time series and arises from basic properties of its MGF.
This chapter is based on the manuscript: J.R. Rigby and Amilcare Porporato. Directional Asymmetry in DNA Sequences. Proceedings of the National Academy of Sciences, (submitted).

5.1 Introduction

The genetic code is the fundamental carrier of the information that sustains life. It has been called the "language of genes" [126] and has inspired linguistic [90] and information theoretic [35] analyses because of its cipher-like appearance and the importance of the information that it carries. Since the 1980’s, as both genomic sequences and computing power have become readily available, studies of DNA sequence composition have proliferated (e.g., [148, 58, 146]). The variety of methods applied to DNA sequences include information theoretic approaches which highlight the redundancy and long-correlations of sequences [35, 41], as well as so-called linguistic approaches which exploit the analogy with the vocabulary and grammar of natural languages [126, 90]. An intuitive property of languages is their directionality, such that reading Hamlet, for example, only introduces us to a troubled Danish prince if
it is read in the proper direction. In previous studies the directionality of DNA sequences has largely been taken for granted, likely due to the structural directionality of the asymmetric bonds in the sugar molecule producing the 5’-3’ orientation. Given the analogy with natural languages, however, it seems natural to ask whether a DNA sequence proceeding from 5’ to 3’ can be quantifiably distinguished from its reverse (3’ to 5’) on the basis of the sequence alone. Despite numerous studies examining nucleotide composition and codon context within sequences [131, 10, 94], however, the question of whether these results amount to a signature of "directionality" within the sequences themselves has not yet been asked. Proper quantification of directionality within the sequence is further motivated by recent studies examining irreversible temporal processes associated with copolymerization [6] and nucleotide substitution [132] in DNA. In particular these studies examine the role physical processes play in producing stored information in systems far from equilibrium and the subsequent manipulation of this information through site mutations over time. Such processes may leave within the genetic code a residue of the arrow of time [6].

5.2 Results

To this end we analyzed DNA sequences from four species of varying organismal complexity in order to quantify the degree to which forward and reverse sequences were distinguishable (i.e., whether the sequence showed a preferred direction). To distinguish the forward from the reverse we used a relative entropy measure, Zn, based on the joint probability distribution of the forward and backward sequence (see Methods) [142]. Thus, for \( n = 2 \), the measure \( Z_2 \) quantifies the imbalance in the probability of a pair of values occurring in one order \( (X_1 = a, X_2 = b) \), versus the reverse \( (X_1 = b, X_2 = a) \). This quantity has strong theoretical foundations in the theory of non-equilibrium fluctuations [49], time-series analysis [115], and information theory [142] and is based, for example, on whether the probability of
observing the nucleotide sequence 5'-AATCGA-3' is statistically the same as that of its reverse 5'-AGCTAA-3'. The results of this analysis are shown in Figure 1a for *E. coli* and *A. thaliana* for “words” of two to five nucleotides. In both cases the sequence is well-distinguished from its reverse even for relatively short word lengths. The values of Zn show that on average the probability of each word differs from its reverse by 1%-6% for *E. coli* and 1%-2% for *A. thaliana*, increasing with the length of the word. Additionally, we show in Fig. 1a that the distinguishability of sense direction is substantially higher within coding regions versus non-coding regions while the rate of increase in distinguishability with word length is comparable between regions. Thus the mean directionality in nucleotide sequences, while too subtle to be observed by eye, is easily distinguished within confidence limits by the Zn measure.

As a consequence of the markedly different behaviour of Zn between coding and non-coding regions, we explored the use of a local estimate of directionality as an indicator of the coding potential of a region. Figure 1b shows the performance of Z2 estimated over windows of 500 nucleotides from chromosome 1 of *A. thaliana* as an indicator of coding status. To quantify the performance we calculated two relevant conditional probabilities: (1) the probability that a region is coding given a value of Z2 above a threshold (hereafter "specificity") and (2) the probability that Z2 is above the threshold given that the estimation window includes a coding region (hereafter "sensitivity"). The figure suggests that when Z2 is statistically different from zero (standard error shown on plot) both the specificity and sensitivity are greater than 0.8 for thresholds up to approximately three times the standard error at which point sensitivity falls rapidly as the threshold is increased further. Figure 1b also shows a histogram of all Z2 local values over the entire sequence to give an idea of the variability of the measure. Comparable results were found for the same procedure performed on chromosome A of *C. glabrata*. The results exceed expectation, particularly noting the relatively few parameters (16 estimated
Figure 5.1: The asymmetry rate as a function of word length (panel a) in coding and non-coding regions for both *E. coli* and *A. thaliana*. Note that in both cases there is a significant difference between the coding and non-coding regions as well as dramatic differences in slope between species. Panel b then shows the performance of using $Z_2$ as a statistic for locating coding regions in *A. thaliana* (see Methods). The curves show the proportion of cases in which $Z_2 > t$ correctly indicated a coding region (circles) and the proportion of coding regions for which the corresponding $Z_2$ value was above the threshold ($x$'s). The histogram shows the probability distribution of all local $Z_2$ estimates from the sequence (magnified by a factor of 7 for comparison). The mean standard error across the estimates, $\sigma_z$, is also shown, as well as the region (shaded) in which the proportion for both indicators is between 0.8 and 0.9. Notice that the statistic performs well for thresholds above the standard error. Thus any region that has an asymmetry significantly different from zero is highly likely to be coding.

Quantities for $Z_2$) and the simplicity of the measure ($Z_2$ in this case considers only di-nucleotides), and suggest that directionality may be a useful and parsimonious tool in detecting coding regions in genomic sequences.

While nucleotides are the most basic alphabet of DNA, the "genetic code" refers primarily to codon/amino acid pairing. Therefore we also examined $Z_2$ for the corresponding codon and amino acid series. Considering a sequence (of coding DNA) as a string of codons, given the nucleotide sequence 5' AATCGA-3', which consists of two codons, the statistic examines the probability of the sequence producing the reverse codon/amino acid pairing, 5' CGAAAT-3' rather than the reverse nucleotide sequence. Figure 2 shows that the directionality is significantly greater in the codon
Figure 5.2: The asymmetry measured for sequences of codons and amino acids for each of the four species. Of particular interest is the apparent negative trend with organismal complexity. Note also that the asymmetry is much greater in codon sequences suggesting that significant information is coded in ordered pairs of synonymous codons for the same amino acid pair.

sequence than in the corresponding amino acid sequence (mean ratio $Z_{2}^{\text{codon}} : Z_{2}^{\text{amino}}$ after normalization by alphabet size is 3.8). This suggests that significant information is coded in the degeneracy of the genetic code over and above usage bias (as $Z_{2}$ is sensitive both to usage and relative position). The implication is that on average specific synonyms occur preferentially in the right or left hand position of adjacent amino acids in the sequence.

Finally, an apparent trend is manifest in Fig. 2 which shows a decrease in directionality with organismal complexity (i.e., from archaebacteria to plant). While the data presented here are insufficient to posit a more general phylogenetic trend, the suggestion is enticing. Such a trend has no obvious explanation and may therefore provide further insight into evolutionary relationships. Indulging in speculation for
a moment, perhaps sequence directionality acts as a proxy measure of the evolved efficiency of the genome in keeping with the claim that eukaryotes possess genetic fossils absent from prokaryotic DNA [111].

5.3 Conclusion

Life is a, perhaps the, canonical example of a non-equilibrium process maintaining itself through the production of thermodynamic entropy [125]. Central to the maintenance and replication of life is the genetic code. Motivated by recent studies that have examined the irreversibility associated with genetic processes, we applied an information theoretic measure of directionality to DNA sequences. We have shown that DNA sequences display distinguishable directionality within the structure of the sequence itself and that broad patterns exist between coding/non-coding regions, nucleotide/codon/amino acid series, and across species. Consideration of the simplest directional asymmetry (i.e., in ordering of di-nucleotides) shows a strong correlation with coding potential and thus represents a potentially useful 16-parameter, intrinsic, content statistic that does not require a training sequence (for comparisons see [54, 91, 45]). Finally, and surprisingly, Fig. 2 displays a correlation between sequence directionality and phylogeny that invites speculation on the possibility of broader patterns in phylogeny based on sequence directionality. Such relationships, if found to be general, may also link phylogeny with thermodynamic constraints on information flow and temporal irreversibility at the very heart of the evolution of living organisms.

5.4 Methods

We obtained a set of DNA sequences for four species from the NCBI database comprised of the complete genomes of Aeropyrum pernix (archeabacteria) [1], and Escherichia coli (bacteria) [4], as well as chromosome A of Candida glabrata (fungi) [3]
and chromosome 1 of Arabidopsis thaliana (plant) [2]. Coding and non-coding regions were determined using the published protein table for each species at the NCBI site. For the longer eukaryotic sequences, lengths were limited to approximately 106 nucleotides to ease computational memory requirements.

Asymmetry was measured using the Zn measure defined as [115]

\[
Z_n = \sum p_{X_1...X_n}(x_1...x_n) \log_2 \frac{p_{X_1...X_n}(x_1...x_n)}{p_{X_1...X_n}(x_n...x_1)}
\]

(5.1)

where \(X_1...X_n\) represents a vector of consecutive random variables (nucleotides, codons, amino acids). Note that the indices in the denominator of the logarithm have been reversed. In the \(n = 2\) case, the \(Z\) measure is therefore essentially equivalent to the skew in the pdf of the first difference of the series. For \(n > 2\), \(Z_n\) is a nonlinear measure of the directional asymmetry in a process. An important property of \(Z_n\) is its monotonicity, i.e., for \(Z_2 > 0\) is a sufficient indicator of directionality while \(Z_2 = 0\) is not sufficient to rule out directionality for higher \(n\). The standard error was estimated following the treatment in [121].

We measured asymmetry both for the nucleotide sequence and the corresponding non-overlapping codon and amino acid sequences within coding regions. Coding regions were delineated using start/stop sequence positions published in the protein table for each species. All remaining regions were treated as non-coding. The amino acid sequence was constructed by giving all synonymous codons the same value in the sequence. For the purposes of analysis the nucleotide bases \{A,C,G,T\} were given the numbers \{0,1,2,3\} respectively. Similarly, codons were assigned a number between 0 and 63 generated by the base-4 representation of the three-base composition (e.g., \{ATG\} = \{032\} = 14).

For local estimates of \(Z_2\) we segmented the full \textit{A. thaliana} and \textit{C. glabrata} (not shown) sequence into non-overlapping blocks of 500 nucleotides. We then estimated
$Z_2$ for each block. In *A. thaliana* 76% of non-coding regions were larger than 500 nucleotides in length based on published protein tables. The conditional probabilities used to evaluate the performance of $Z_2$ in detecting coding regions were determined in the following manner. The “accuracy”, quantified by $\text{Prob}(\text{coding}|Z_2 > t)$, was calculated by taking the set of 500-nt blocks with local $Z_2$ greater than the threshold and then counting the proportion of those that overlapped at least one coding region. Similarly, the “specificity”, quantified by $\text{Prob}(Z_2 > t|\text{coding})$, was calculated by taking the set of coding regions and counting the proportion of those that overlap at least one local $Z_2$ estimate that was above the threshold.
6

Precipitation, Dynamical Intermittency, and Sporadic Randomness

This chapter is based on the manuscript: J.R. Rigby and A. Porporato. Precipitation, Dynamical Intermittency, and Sporadic Randomness. Advances in Water Resources, (under review).

6.1 Introduction

An obvious property of rainfall time-series is that they are intermittent, displaying long periods of inactivity punctuated by relatively short periods of activity (i.e. storms). This property is one of the primary sources of uncertainty in hydrologic forecasting and remains difficult to model at many scales. It is assumed in the literature, albeit often implicitly, that the source of this intermittent behavior is fundamentally dynamical, that the governing equations of atmospheric phenomena generate intermittent precipitation. However, this assumption has not been employed actively in characterizing rainfall.

In this paper we wish to draw connections between rainfall phenomenology and
some dynamical models known to exhibit specific “types” of intermittency classified by the dynamical instability producing the intermittent phenomena. These paradigms of intermittent behavior are the result of dynamical features that may be demonstrated both in the low-dimensional setting of 1-D iterative maps as well as, and crucially for rainfall, in the infinite-dimensional setting of stochastic differential equations. Thus, while the dynamical features may be demonstrated in a low-dimensional map, the qualitative dynamical features underlying the behavior may also govern high-dimensional systems as demonstrated in the stochastic case in which, though the models are simple, the dimensionality is very high. Such behavior suggests that even in the ultra high-dimensional case of atmospheric turbulent convection, a few qualitatively simple features may be dominant in producing specific characteristics of rainfall time-series.

The body of literature on descriptive rainfall statistics is extensive. Within this literature several scaling relationships have been noted including an apparent power law in the inter-arrival times (dry periods) of rainfall, with an exponent very near $-3/2$ (e.g., [124]), and a power law in the power spectral density for scales of approximately one day [40], with an exponent equal to $-1/2$. These scaling relationships are also paradigmatic for some known intermittent dynamical systems (see below) and therefore may be clues to the dynamics governing rainfall at particular temporal scales. Furthermore, one of these types of intermittent dynamical system following the same scaling laws as rainfall, the Pomeau-Manneville Map (and its stochastic analogue), manifests a behavior intermediate between periodicity and pure randomness known as “sporadic randomness”. Below we suggest that rainfall may also be an example of sporadic randomness, further linking rainfall with known intermittent dynamical behaviors.

In Section 2 we provide a brief overview of some of the modeling approaches used to simulate rainfall intermittency to demonstrate the context of our approach and
its contribution to the literature on rainfall modeling. Then in section 3 we describe in some detail two paradigms of intermittency: Pomeau-Manneville Type-III and On-Off. We limit the analysis to these two paradigms of intermittency based on the characteristic scaling of laminar phases (time between storms), as only these two paradigms of intermittency produce the appropriate $-3/2$ exponent in the probability density function (pdf) of laminar phases seen in rainfall series (see Fig. 6.3). In Section 4 we present precipitation data from four locations with discussion of the pdf’s of the lengths of laminar phases, the pdf of rainfall intensities, and the power spectral density of rainfall intensities. Comparisons of seasonal subsets of data taken from Jun-Aug and Dec-Feb show significant disparity between winter and summer characteristics at three of the four sites. We suggest that summer and winter precipitation series may differ due to the predominance of convective and stratiform frontal events respectively, each producing a characteristic dynamical signature. Section 5 then provides a discussion of the comparison of the phenomenological signatures of known dynamical intermittency mechanisms with the observed rainfall series and the apparent Type-III signature of summer rainfall, as well as a physical interpretation of the phenomena, before a final discussion and conclusion in Section 6.

6.2 Modeling Intermittency in Rainfall: An Overview

As intermittency is a key feature of all rainfall series, it is thus treated in some fashion by nearly all rainfall models. While the literature concerning rainfall modeling is too vast to review systematically here, in this section we provide a modest overview of some of the predominant approaches to this problem in order to situate the present contribution.

One approach to modeling rainfall series employs stochastic point processes to model the occurrence and depth of rainfall at intermediate temporal scales (hourly to daily) (e.g., [147, 51, 39] and references therein). However, these models attempt
to capture the statistical properties of the occurrence, intensity, and/or depth of rainfall events without explicit reference to the underlying physics and often result in statistically simpler series than those observed. Exploration of the statistical complexity of the observed rainfall series has proceeded primarily through attention to the scaling properties of rainfall in both space and time. In particular a large body of literature using the fractal and multifractal scaling properties of rainfall (e.g., [89, 55, 144] and references therein) to quantify the characteristics of rainfall complexity has arisen out of similar work focused on fluid turbulence (e.g., [122, 133, 42] and references therein). The predominant rainfall modeling approach arising from fractal and multifractal frameworks has been that of multiplicative cascades (e.g., [42, 56, 92]). Additionally, the attention paid to scaling characteristics has generated a series of models using more traditional stochastic models that also exhibit complex, scaling [143, 100]. Collectively, these models have performed well in reproducing the observed statistical characteristics of rainfall scaling. However, in these approaches the link with atmospheric dynamics often still exists only in the physical analogy of the multiplicative cascade to atmospheric eddy cascades.

Despite the great progress represented by the development of multifractal models, as [144] state in their review, “one needs to better understand the sources of multifractality in rainfall.” While literature on this topic is not totally absent, the link between atmospheric dynamics and rainfall phenomenology has been difficult to produce in simple modeling frameworks and tends to proceed using a blend of stochastic and deterministic components (e.g., [52, 97]).

More recently rainfall has received some attention as a potential example of a self-organized critical (SOC) phenomenon. [105] demonstrate the scale-free behavior of rainfall and argue by analogy with earthquake phenomena that rainfall may be understood as a relaxational process. While this approach may provide an attractive conceptual framework for precipitation dynamics, and may be useful in cumulous
closure models [106], the characterization is not complete, nor does it yet establish the link between high-dimensional continuous dynamical systems (i.e., atmospheric dynamics) and the production of complex, intermittent phenomena.

Thus, despite the attention rainfall modeling has received from researchers, models of rainfall at a point are rarely presented in the form of a system of (stochastic) differential equations. No doubt this is at least in part because intermittency is a rather uncommon feature even of stochastic dynamics. Because of this difficulty the majority of modeling approaches account for intermittency either by modeling inter-arrival times as a stochastic variable drawn from a distribution (point processes), or by constructing rainfall series using iterative fractal/multifractal methods (cascades).

On the other hand, the very fact that intermittency is a a rather uncommon feature of dynamical systems would seem to be in favor of exploiting these phenomenological similarities between rainfall and well-studied systems. Therefore, to supplement these approaches we propose to focus on the commonalities between the phenomenology of rainfall series and certain intermittent dynamical systems with the explicit aim of better understanding the properties of the underlying dynamics.

6.3 Types of Dynamical Intermittency

While there is no general theory accounting for the presence of intermittent behavior in dynamical systems, there are several recognized “types” of intermittency distinguished both by particular mechanisms (if known) as well as the statistical phenomenology of representative series. Among models of intermittency, [110] described three types of intermittency that occur in systems as routes to chaos, classified according to the type of instability associated with a fixed point in the Poincaré map. Each type exhibits distinguishing phenomenological traits as for example, and perhaps most characteristically, in the form of the probability density function (pdf) of laminar phases.
In addition to Pomeau-Manneville intermittencies, other forms of intermittency have been identified including On-Off, and Crisis intermittency [64]. In this paper we will limit detailed description of types of intermittency to the properties of PM Type-III and On-Off intermittency as these are the only two types from the previous list that generate pdf’s of the laminar phases consistent with observed rainfall series (see below for further details). For each type of intermittency we will present both a 1-D map and an analogous stochastic model displaying the same intermittent character, reviewing in each case the relevant analytical properties for phenomenological comparison with rainfall series. It is important to note, for the sake of applying these models to rainfall, that the stochastic version is not a low-dimensional model but is rather a simple model with infinite degrees of freedom. Thus the dynamical signature of various types of intermittency should not be thought a result only of maps and low-dimensional dynamical systems.

6.3.1 Pomeau-Manneville Type III Intermittency

[110] outlined three mechanisms or routes to chaos via intermittency arising from the possible local bifurcations as the eigenvalues of a fixed point pass through the unit circle [9, 63]. When the Floquet multiplier is real and crosses the unit circle at $-1$, the system is said to exhibit Type-III intermittency. [110] demonstrated this type of intermittency using a general 1-D map of the form

$$x_{n+1} = -(1 + \epsilon)x_n + ax_n^2 + bx_n^3 + R_n$$  \hspace{1cm} (6.1)

where $a$ and $b$ are constants, which produces a signal characterized by slow divergence from the stable point, alternating in sign from positive to negative on each iteration while $R_n$ represents a mechanism for reinjecting the trajectory near the fixed point. A more popular example of this map occurring in the literature may be derived from the second iterate ($x_{n+2} = f(x_n)$) which removes the alternating sign and is given by
\[ x_{n+1} = (x_n + x_n^3) \mod 1 \]  \hspace{1cm} (6.2)

in which the modular arithmetic operator “mod 1” has been inserted as the reinjection mechanism [116, 67]. This operator simply sets \( x_{n+1} = x_{n+1} - 1 \) if \( x_{n+1} > 1 \), thereby bounding the series on the interval \((0, 1)\) and reinjecting the trajectory near zero. Figure (6.1) shows the map (1a) along with a representative time series (1b).

A continuous, stochastic version of this map

\[ \dot{x} = x^2 + \rho_t \]  \hspace{1cm} (6.3)

has been studied by [67], in which \( \rho_t \) is a stochastic reinjection mechanism (e.g., uniformly distributed jumps taken from the interval \((-1, 0)\) that occur whenever \( x = 1 \)). The laminar phases for these processes are defined as the time between successive reinjection events (see Fig. 6.1). Each “crossing” at \( x = 1 \) is followed by a random (with uniform distribution) reinjection back into the laminar phase. The trajectory is in essence a series of laminar phases separated by instantaneous “events”. The length of the laminar phases, \( \tau \), has been shown, for a more general case of the map in Eq.(6.2) in which the exponent is represented by the variable \( z \), to have a power law distribution of the form \( p(\tau) \sim \tau^{-z/z-1} \) [116]. For the Pomeau-Manneville case in which \( z = 3 \) the resulting distribution has a power law slope of \(-3/2\). Further analytic properties of the map are known including the pdf of \( x \), shown also to follow a power law according to \( \rho(x) \sim x^{-2} \) [50], as well as the power spectral density which [116] showed decays as \( \omega^{-1/2} \) as \( \omega \to 0 \). However, while these analytic results exist for the Pomeau-Manneville map, to the authors’ knowledge no claim has been made in the literature as to whether these are diagnostic features of Type-III intermittency. Of these phenomenological properties only the distribution of laminar
phases has been used to differentiate types of intermittency. Additionally, of the three intermittent routes to chaos, only Type-III exhibits laminar phases distributed as $\tau \sim -3/2$ in close agreement with observed rainfall records.

Experimental evidence of Type-III intermittency has been demonstrated in the horizontal temperature gradients of Rayleigh-Benard convection [31]. In addition, Type-III intermittency has been explored in such diverse systems as lasers, electronic circuits, and the dynamics of epileptic seizures. The distinguishing qualitative features of this process are the slow divergence from a quasi-steady state punctuated by abrupt reinjections, or relaxations, for which the inter-arrival times are distributed as a power law with exponent $-3/2$. In these terms an analogy with rainfall begins to appear in the slow accumulation of atmospheric water vapor followed by abrupt relaxation events in the form of storms.

We should emphasize that Type-III phenomenology does not necessarily imply that the system is in transition to chaos in the sense of low-dimensional dynamical systems (see [123]). The Pomeau-Manneville map was developed specifically as a model of the Pomeau-Manneville Type-III instability of a fixed point producing intermittent chaos. However, a large body of literature has developed around the map itself such as that of [12] and [67] linking the map to aspects of stochastic dynamics such as sensitivity to initial conditions, and the production of Lévy flights. The map (6.2) and especially its stochastic counterpart (6.3) have been used as canonical examples of a system exhibiting “sporadic randomness”, a concept transcending low-dimensional dynamics, characterizing a dynamical randomness that is intermediate between regular and purely random behavior [43].

Figure 1 shows representative time series from both the Pomeau-Manneville map, Eq. (6.2), and the continuous stochastic analogue, Eq. (6.3). The series are characterized by a slow growth followed by a rapid reinjection. For comparison, a static transformation of the map series is shown, $z = 1/(1 - x) - 1$, in the bottom panel.
Figure 6.1: Plot showing time series from (a) the Manneville map, (b) the stochastic model of type III intermittency, and (c) a static transformation of the series in (b) according to $z_t = \frac{1}{1-x_t} - 1$ which produces a strikingly rainfall-like appearance.

which produces a series qualitatively similar to rainfall.

**Sporadic Randomness**

[43] provides a classification of processes into “periodic”, “regular randomness” and “sporadic randomness” based on the entropic characteristics of trajectories. Periodic processes generate repeating sequences. Processes that have linear growth in complexity as a function of sequence length are classed as regularly random. This is the case for stochastic processes which have a convergent entropy rate (e.g., Markov chains, Brownian motion, etc.). For systems between periodicity and regular randomness, processes which display a strong memory while retaining a stochastic component, the complexity grows as a power law with the sequence length. The processes in Eqs. (6.2) and (6.3) both manifest this type of entropic behavior.
A simple method to measure this growth in complexity for discrete time series is to coarse-grain the data into discrete states and then to study the growth in information theoretic entropy (Shannon entropy) as a function of sequence length. The estimate of the entropy as a function of sequence length, $H(n)$, is defined as

$$H(n) = - \sum_{n} p(x_0 x_1 ... x_n) \log p(x_0 x_1 ... x_n)$$

(6.4)

where $p(x)$ represents the observed probability of state $x$ and $(x_0 x_1 ... x_n)$ represents a sequence of $n + 1$ consecutive states. Thus, $p(010)$ for a binary signal is the relative frequency of observing the sequence 010 in the data. For sporadic randomness we expect to find

$$H(n) \propto n^\alpha$$

(6.5)

with $0 \leq \alpha \leq 1$ [50]. However, it is generally difficult to estimate entropies for longer sequence lengths due to the exponential growth in possible sequences and the limited length of data series. The finite length of data series is a substantial problem that leads to significant underestimation of entropies for higher values of $n$.

6.3.2 On-Off Intermittency

So-called “On-Off” intermittency describes certain time series that remain in an “off”, or nearly constant, state for long periods of time followed by abrupt, random bursts of large deviations. This behavior was first reported by [108] both for a set of five coupled ordinary differential equations and for a piecewise linear map. Using a forced logistic map of the form,

$$x_{n+1} = a y_n x_n (1 - x_n)$$

(6.6)
Figure 6.2: Plot showing On-Off time series from (a) the random logistic map, and (b) the stochastic continuous version.

in which \( y_n \) is a uniformly distributed random variable in the interval \([0, 1]\) and \( a = 2.8 \), [63] produced on-off intermittent series and showed that the pdf of the laminar phases scales as \( \tau^{-3/2} \). [145] demonstrated that the underlying mechanism of on-off intermittency is captured in a multiplicative process of the form

\[
\dot{x} = (a + b\xi)x \tag{6.7}
\]

in which \( \xi \) is a Gaussian white noise with zero mean interpreted in the manner of Stratonovich, while \( a \) and \( b \) are positive constants such that \( a << 1 \) and \( b \approx 1 \). A reflecting boundary is imposed at \( x = 1 \) to simulate the nonlinearities of reinjection. [145] showed that the “on” phases exist on a fractal support with dimension \( 1/2 \) and have power spectrum that scales as \( \omega^{-1/2} \) for a finite scaling region whose width depends on the constants. Figure 6.2 shows representative time series from the two models.

While the laminar phase distribution and the power spectrum are consistent with the behavior of PM Type-III intermittency displayed in the Manneville map above, the pdf of the variable \( x \) scales as \( p(x) \sim x^{-1} \) for On-Off, distinguishing it from the \(-2\) exponent for the Type-III Manneville map. Unlike PM Type-III intermittency
the literature for On-Off intermittency has referred to each of these phenomenological characteristics (pdf of amplitudes, pdf of laminar phases, and power spectrum) as collectively diagnostic for the presence of on-off intermittency (e.g., [139]).

Pomeau-Manneville intermittency arises from a weakly unstable periodic orbit (in continuous time), while the mechanism underlying On-Off intermittency may be understood as deviations from an unstable manifold under the influence of a modulated bifurcation parameter [108, 63]. The system is characterized by an invariant submanifold, which may itself be of high dimension, within a higher dimensional phase space. Tuning a particular parameter leads to a bifurcation at which the invariant submanifold becomes (un)stable. For parameter values near the critical value, the system may spend large amounts of time in the neighborhood of the invariant submanifold punctuated by abrupt departures of large amplitude.

This geometric understanding puts the growth and reinjection phases in the context of dynamical phase space, such that the growth (or “on”) phase is produced by the departure of the dynamics from a quasi-stable submanifold in the phase space, while the nonlinearity of the system eventually “reinjects” the trajectory back onto the submanifold. [101] showed further that this on-off behavior is the nonhysteretic version of a more general “blowout” bifurcation that links on-off intermittency with the phenomena of riddled basins of attraction. Again it is to be stressed that there is nothing inherently low-dimensional about this scenario. The intermittency arises as a result of an unstable manifold within the phases space, with only a lower bound for the space’s dimensionality. Thus any dynamical system that lives for long periods in a relatively small region of the overall phase space may exhibit these intermittent departures depending on the local stability of the manifold.

AR(1) Processes and On-Off Intermittency

[139] studied a nonlinear transformation of a linear autoregressive process given by
\[ Z_{t+1} = C_1 Z_t + \sigma \eta \quad X_t = C_2 e^{Z_t} \]  

for which \( C_1 = 0.999 \) and \( C_2 = 0.001 \) and \( \eta \) is defined as a white noise taking values uniformly on the interval \([-1, 1]\). The process \( X_t \) displays bursting behavior as well as approximate \(-3/2\) scaling of laminar phases in analogy with On-Off intermittency, leading the authors to caution the use of these characteristics alone to diagnose On-Off behavior.

It may be noticed that the continuous analogue of the autoregressive process, using instead a Gaussian noise, is an Ornstein-Uhlenbeck process with a very small drift coefficient (on the order of \( 1 - C_1 \)). As [139] note, the scaling in the return times \((-3/2\)) then arises from the crossing properties of unbiased Brownian motion, and the bursting is created by the exponential transformation. The drift coefficient, although small, acts to keep the process bounded. Thus, the Ornstein-Uhlenbeck process can be written as

\[
\dot{y} = -ky + \xi \quad (6.9)
\]

such that, interpreting the noise in the manner of Stratonovich, the change of variable \( x = e^y \) gives the formula for geometric brownian motion [99],

\[
\dot{x} = -kx \log x + x \xi. \quad (6.10)
\]

This makes the connection with the multiplicative process defined by [145] evident. In fact, in [145] the scaling process is obtained by making the drift \( a \) tend to zero and introducing a reflecting barrier to contain the process. Here in (6.10) the same scaling behavior is obtained by having a containing potential (i.e., \( \int x \log x dx \)) to ensure stationarity but with a very small coefficient \( k \).
In summary, as [145] show in their development of the multiplicative model, the dynamical connection lies in the character of local fluctuations of the Lyapunov exponent around the unstable manifold. Treating the parameter $a$ as a local mean Lyapunov exponent subject to Gaussian fluctuations, $\lambda = a + b\xi$, we arrive at Eq. (6.7) as the linearized behavior of a trajectory in the neighborhood of the manifold. [139] model the underlying process as an AR(1) process equivalent to an Ornstein-Uhlenbeck process. By analogy with [145] this corresponds to modeling the Lyapunov exponents in this region as a stationary random process with zero mean and a finite correlation time. As the correlation $k$ goes to zero we get On-Off behavior. This is precisely the picture given by [145] when discussing the correlation of Lyapunov exponents along a trajectory. This connection further reinforces the identification of On-Off intermittency with the modulation of a parameter near the fixed point [63, 28] while also tying the phenomena to the phase space of higher dimensional systems [145].

6.4 Properties of Rainfall Intermittency

6.4.1 Data sets

To examine the intermittent properties of rainfall and their similarities with known dynamical mechanisms, we chose data sets from four locations of varying climate: Duke Forest experimental site (Durham, NC, USA), Florence (central Italy), Chiavari (northwest Italy), and Valentia (south-west Ireland). These sites represent humid subtropical, inland Mediterranean, coastal Mediterranean, and marine west coast climates, respectively. Table 1 summarizes relevant site and data characteristics. Each data set was aggregated to 1hr resolution.

For each data set we analyzed the pdf’s of rainfall intensity, laminar (dry) phases, and the power spectral density of the intensity series for comparison with analytical properties of known types of intermittency. Each probability density is estimated
Table 6.1: Description of data series used. Temperatures are the maximum of monthly mean high temperatures in June-August obtained from www.worldclimate.com.

<table>
<thead>
<tr>
<th>Name</th>
<th>Duke Forest</th>
<th>Florence</th>
<th>Chiavari</th>
<th>Valentia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>Durham, NC, USA</td>
<td>Florence, Italy</td>
<td>Chiavari, Italy</td>
<td>Valentia, Ireland</td>
</tr>
<tr>
<td>Depth Resolution [mm]</td>
<td>0.1</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>Time Resolution [hr]</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Annual Rainfall [mm]</td>
<td>1126</td>
<td>841</td>
<td>1092</td>
<td>1457</td>
</tr>
<tr>
<td>High Temp [deg C]</td>
<td>28</td>
<td>27.2</td>
<td>30.5</td>
<td>18.0</td>
</tr>
</tbody>
</table>

\(^1\) Aggregated from 1/2 hr of Chiavari. \(^2\) Aggregated from 1/60 hr \(^3\) Data for Genoa, Italy approximately 50km NW

using bin widths that approximate an equipartition of the data (i.e., approximately the same number of samples in each bin).

For comparison, we also analyzed seasonal subsets of each series to investigate differences in summer and winter precipitation which, depending on local climate, might then suggest differences in the dynamics of convective and stratiform precipitation. The seasonal data sets are comprised of observations taken between June 1 and August 31 for summer, and those taken between December 1 and February 28 for winter. Each subset was analyzed separately and the results were then pooled for probability distributions (e.g., all summer laminar periods in the record).

6.4.2 Laminar Phases

Full Record Analysis

In each rainfall series a “laminar phase” is defined as a contiguous sequence of measured zero rainfall intensity. The estimated probability density for the laminar (i.e., dry) phases of each rainfall series is shown in Fig. 6.3. For temporal scales up to about one day the series show clear power law scaling with exponent \(-3/2\). At the scale of one to three days the series starts to diverge from the \(-3/2\) scaling. Interestingly, three of the four series show a characteristic exponential shoulder forming in the log-log plot of the pdf, while the Valentia series shows a similar fall-off for large \(\tau\) but without the inflection. [141] has shown the appearance of a similar shoulder
in studying the effect of additive noise on On-Off intermittent systems.

The seasonal case (not shown) displays the same $-3/2$ scaling in the laminar phases for both summer and winter.

### 6.4.3 Power Spectrum

**Full Record Analysis**

The power spectral density (psd) estimates for each site, in which the Fourier spectrum is averaged over annual segments, are shown in Fig. 6.4. In each case a straight line is shown indicating the $-1/2$ scaling expected from arguments of Type-III and On-Off intermittency. In each of the series except Valentia, there is potential scaling for frequencies higher than about $10^{-6}\text{sec}^{-1}$. The psd for the data from Valentia is distinct from the other three sites both in that the shape of the density is slightly more complex and that the slope of the density in the high frequency region is significantly steeper than the other sites.
Figure 6.4: Log-log plots of the power spectral density for each of the four data sites. The straight line in each case indicates the $-1/2$ slope expected from Type-III and On-Off intermittency for low frequencies.

Seasonal Analysis

Power spectral density estimates were made for each summer/winter season at each site. Clearly reducing the size of the windows over which to estimate the psd (from 12 months to 3 months) will increase the variance in the estimate. As the $-1/2$ scaling appears to be concentrated in the higher frequencies (Fig. 6.4), in order to reduce the variance in the psd plots the seasonal data was divided into three-day windows, within each summer/winter season, for which the psd was estimated and
then averaged over all three-day windows. The resulting psd at high frequencies is shown in Fig. 6.5.

In Figure 6.5 there is striking agreement among all four sites for winter (right panel) and for three of the sites in summer (left panel). For summer the data from Duke Forest, Florence, and Chiavari each show slopes very near $-1/2$ in the high frequency region, particularly for periods shorter than one day. The psd for winter shows a slope very near $-1$ for all four sites. Interestingly, the summer psd for Valentia is almost identical to the winter estimate at the site.

As the slope of the summer psd estimates appears close to $-1/2$ particularly at frequencies very near the resolution of the data we performed the same seasonal analysis on a higher resolution series (1 minute) from Chiavari. Figure 6.6 shows the psd estimate, again for 3-day windows, taken from a 1-min resolution series for both summer and winter. In this case the $-1/2$ scaling is again apparent and appears to end abruptly near periods of about one hour, giving way to another possible scaling region with exponent close to $-3/2$. The winter series again shows the $-1$ slope, now extending to periods on the order of minutes.
Figure 6.6: Log-log plot of the power spectral density for the 1min resolution series from Chiavari. For periods between 1hr and 1day the spectra show scaling near $-1/2$, while for higher frequencies an apparent $-3/2$ scaling appears. The $-1$ scaling in winter is again apparent.

6.4.4 Rainfall Intensity

Full Record Analysis

Figure 6.7 shows the estimated probability density of rainfall intensities [mm/hr] for each site. In each panel lines are provided indicating the $-1$ and $-2$ slopes expected for On-Off and Type-III intermittencies, respectively. While Duke Forest, Florence and Chiavari appear to show some scaling for intensities greater than about 3 mm/hr, no site shows scaling consistent with the two intermittency models.

Seasonal Analysis

The case is quite different for the seasonal intensities. Figure 6.8 shows consistent scaling in the summer months with slope near $-2$ for three of the sites (Duke Forest, Chiavari, and Florence). For small intensities a possible $-1$ scaling is seen for the winter months at Florence, Chiavari, and perhaps Valentia. Interestingly, the dis-
tributions of intensities at Valentia for summer and winter are nearly identical (see insert), in strong contrast to the other sites.

6.5 Analysis of Results

6.5.1 Comparison with On-Off and Type-III Intermittencies

We have presented three aspects of rainfall for comparison with dynamical models of intermittency: pdf of laminar phases, power spectral density of rainfall series, and the pdf of rainfall intensity. Significant differences were found between summer and winter characteristics of both the power spectral densities and pdf of intensities for all sites except Valentia. All four sites show scaling in the pdf of laminar phases with exponent $-3/2$ up to periods of about one day. Of the known dynamical sources of
intermittency, only two types display this scaling: Type-III and On-Off.

In the case of On-Off intermittency, the literature suggests three phenomenological traits that should be present. These three are scalings in the pdf of laminar phases, pdf of intensities, and the power spectral density, with exponents $-3/2$, $-1$, and $-1/2$ respectively. While rainfall has the classic qualitative appearance of an On-Off variable, and both the laminar phase pdf and the power spectral density are consistent with On-Off over some range of scales, the scaling of the intensities clearly does not show the expected $-1$ slope.

The case for Type-III intermittency is complicated by the absence of clear phenomenological markers. While the literature on the Manneville map provides a wealth of analytical results, rarely have researchers made claims about the necessity or sufficiency of these results in diagnosing Type-III intermittency. That is, the majority of attention has been devoted to examples of Type-III intermittency in mathematical models rather than to diagnosing Type-III behavior in observed series in which the precise dynamics are uncertain. If one accepts the scaling properties for the pdf of amplitudes and the power spectral density for the Manneville map as
a general requirement for Type-III intermittency, then the rainfall data suggest that summer rainfall in three of the sites are consistent with Type-III intermittency. For summer rainfall at Duke Forest, Chiavari, and Florence the pdf of intensities tends to fall off as $-2$ for high intensity rainfall.

The spectral densities for summer show scaling very close to $-1/2$ for periods between approximately one hour and one day. The cutoff at about 1-hr is confirmed by the 1-minute resolution series from Chiavari (Fig. 6.6).

Indeed, for both the laminar phases and the power spectral densities, the summer data indicate that the dynamics most consistent with Type-III intermittency are those on temporal scales between one hour and one day. These scales correspond to the scaling region for frontal dynamics as classified by [40] who distinguished the frontal scales between 2.4 hours to 3 days from intra-storm dynamics at higher frequencies.

The notable differences between the data from Valentia and the other sites suggests further that the presence of convection may be a key factor in the scaling relationships. In contrast to the data from Duke Forest, Chiavari, and Florence the data from Valentia show almost no difference between summer and winter characteristics. Valentia’s location on the southwestern coast of Ireland distinguishes it from the other sites as both the wettest, in total annual rainfall, and coldest (see Table 1) location among the sites. In fact the average daily summer high temperature is ten degrees C less than any of the other sites. Consider also the scarcity of intense rainfall at Valentia. The proportion of events greater than 10 mm/hr at Duke Forest, Florence and Chiavari is 0.02, 0.006, and 0.03, respectively, while the proportion at Valentia is $7 \times 10^{-4}$, nearly a full order of magnitude lower than the next closest site. The low temperature combined with a relative scarcity of heavy rainfall events suggest that the data from Valentia lacks a significant convective signature.

Thus the analysis indicates that in warmer climates the characteristics of summer
rainfall follow the scaling relationships of Type-III intermittency at the daily-to-hourly temporal scales. The apparent importance of convection, combined with the clear time-scales over which the scaling relationships hold further suggests that convective dynamics such as those in multi-cell storms which have lifetimes on the order of hours [66] may be the dynamics in question.

6.5.2 Sporadic randomness

Throughout this analysis we have suggested that rainfall is driven by very high dimensional dynamics and yet displays characteristics very similar to that found in simple stochastic models of intermittency. While it is highly unlikely that rainfall is a Type-III intermittent system in the sense of Pomeau-Manneville’s work on routes to chaos, the similarities between the two phenomena suggest connections on a more general level, and especially with the stochastic counterpart introduced by Ignaccolo et al. [67].

A property that both Type-III intermittent maps and their stochastic analogues share is sporadic randomness [43]. While both exhibit “random” behavior, the randomness is not as complete as in many stochastic processes such as Markov chains and Brownian motion. This is manifest in the rate of growth of the complexity of a trajectory in time. For most stochastic processes this rate (e.g., \( h(n) = H(n) - H(n - 1) \)) converges to a constant value relatively quickly, \( \lim_{n \to \infty} h(n) = h \) where \( h \) is constant. For regular systems the rate converges to zero. However, in the case of sporadically random dynamics the entropy rate does not converge, either to zero or to a constant. Thus the trajectories are less “random” than other stochastic processes and yet more random than regular sequences. This is obviously also demonstrated in the block entropy, \( H(n) \), only in an integrated form. For periodic processes (of period \( \tau \)) the block entropy reaches a constant at \( n = \tau \) and no longer grows for \( n > \tau \). Thus, the entropy rate is zero for \( n > \tau \). For pure randomness the entropy grows linearly.
with \( n \). So, for any sequence length there is a constant level of uncertainty present in each subsequent step. For sporadic randomness the entropy grows sub-linearly but without reaching a plateau. Thus, the system is always yielding information, making the next step less uncertain, while never becoming regular.

As sporadicity has been demonstrated for several intermittent maps (e.g., [50]) and suggested as a distinguishing feature in weather systems [96], DNA sequences [84], and some linguistic texts [36], we hypothesize based on the correspondence with Type-III intermittency that rainfall may also exhibit this sporadic behavior. As estimation of entropies carries a unique set of challenges that are beyond the scope of this paper, we offer here only a preliminary analysis of the block entropy growth.

In general when estimating the entropy one must estimate the probabilities of discrete sequences with the number of possible sequences increasing as \( p^n \) where \( p \) is the number of states in the coarse-grained data series. Clearly as \( p \) and \( n \) increase the number of possible sequences rapidly surpasses the length of the data set. Since we are primarily interested in the growth of \( H(n) \) we would like to maximize the value of \( n \) at which finite size effects will begin to be significant. Thus we partitioned the rainfall data at each site into a binary series (rain/no rain), \( p = 2 \), in order to allow estimation for higher block lengths. In each case we determined the maximum block length by taking \( n_{\text{max}} = \log_2[N/100] \) so that the estimate would be sensitive to words of probability on the order of 0.01 (or equivalently, for a uniform distribution of words of length \( n_{\text{max}} \) there would be approximately 100 events of each word found in the series). The estimate of the slope of \( H(n) \) is shown in Fig. 6.9. As \( n \to \infty \) this slope should converge to a constant value if the series is sporadic. In all four cases this log-entropy rate appears to grow sub-linearly with an exponent between 0.8 and 0.9.

Thus rainfall appears to show sub-linear entropic scaling consistent with sporadic randomness. While this analysis suggests sub-linear scaling, it is probable that rain-
Figure 6.9: Plot of the block entropy, $H(n)$, (left) and the slope of the logarithm of the block entropy, $\alpha = \log\left[\frac{H(n+1) - H(n)}{\log(n+1) - \log(n)}\right]$, (right) as a function of the block length. For sporadic signals $H(n) \sim n^{-\alpha}$, thus the slope on log-log scale should converge to a constant value, $\alpha$. For each site the maximum $n$ for which the entropy was estimated is based on $n_{\text{max}} = \log_2[100N]$, where $N$ is the length of the series, to avoid finite-size effects. The log-entropy rate for each of the sites appears to reach a plateau consistent with sporadcity.

fall series exhibit memory for longer sequences than 14hrs. It is then possible that for much longer sequences the entropic growth will become linear. In such a case the transient growth might appear sporadic while for longer sequences the growth is approximately linear. And even if such a linear behavior were demonstrated at for large $n$ it is highly likely that these $n$ would be significantly larger than the time-scales over which rainfall series share characteristics with other models of intermittency. These questions remain open for future research.

6.5.3 The physical picture

The picture that emerges from the analysis suggests that the time-scales of interest with respect to Type-III and On-Off intermittencies is from daily to sub-daily, and further that the dynamics of interest are most likely those of convective storms. The qualitative picture of Type-III intermittency is that of a slow build-up/accumulation followed by a random relaxation. There is analogous to the dynamics of evapotranspiration at the regional scale and the atmospheric convection. Heating at the
Earth's surface drives both evaporation, which contributes to the accumulation of atmospheric vapor, and sensible heating which in turn lifts moisture into the upper atmosphere by convection where it may condense, form clouds, and eventually precipitate (e.g., [69]). Heating and the build up of sufficient moisture in the atmospheric boundary layer, at the regional level, as well as accumulation of moisture in the free atmosphere taking place at synoptic scales are necessary conditions for the local atmosphere to leave the stable non-precipitating state and induce strong convection and precipitation mechanisms that are responsible for the relaxation and subsequent re-injection of the system's trajectory onto a more stable state. Thus the intermittent properties are possibly linked to the growth of convective instability and the subsequent local dynamics of convection and precipitation.

6.6 Discussion and Conclusions

One of the outstanding problems in rainfall modeling is identification of the dynamical source of the highly intermittent nature of precipitation, and thus whether and how it can be modeled. Several approaches to this problem are available in the literature. In this paper we have attempted to address the problem from a novel standpoint by asking what dynamical features may be at work in producing observed features of rainfall at a point.

Toward this end we investigated the possibility that rainfall intermittency is governed by one of the known models of intermittency in the literature of nonlinear dynamics. We used the $-3/2$ scaling of laminar phases to narrow the field of possibilities to On-Off and Pomeau-Manneville Type-III intermittency. In our analysis of the scaling properties of the laminar phases, intensities, and spectral densities at four sites of varying climate, we have found evidence that the scaling properties in these variables are likely a signature of convective dynamics. The characteristics of summer rainfall at Duke Forest, Chiavari, and Florence further suggest that
these convective rainfall events may be consistent with Type-III intermittency on
time-scales consistent with multi-cell convective storms, though it is not certain how
diagnostic the phenomenological properties of the Manneville map are for Type-III
intermittency.

The defining feature of Type-III intermittency as an example of “sporadic” ran-
domness [50] is the interplay between regular, predictable dynamics and random
events. Such systems occupy a space between purely random and purely predictable
behaviors. More subtly, sporadic systems are a combination of smooth deterministic
dynamics and stochastic components [50]. Qualitatively, we may see this in the dy-
namics of convective cells within a storm, as the large-scale convection is a relatively
smooth deterministic process that is punctuated by the occurrence of rainfall. This
analysis suggests directions for future research on rainfall and convective dynam-
ics at these scales that explores in more detail this sporadicity in rainfall, possibly
employing high-resolution space-time rainfall fields.

It is hoped that this approach might yield insight into the qualitative dynamical
features that should be sought in future modeling, particularly in efficient and par-
simonious balancing of stochastic and deterministic components of models. It may
be, for example, that the several notable characteristics of rainfall (such as inter-
mittency and multifractality) may be the result of only a few dominant dynamical
features. This would accord with the speculation by [106] regarding a relatively sim-
ple structure of convective rainfall fluctuations, while linking the phenomena directly
to a dynamical model. Were such to be the case then mathematical modeling efforts
might then profit from a focus on parsimonious generation of these qualitative fea-
tures as a more organic approach to reproducing these several properties of rainfall
series.
Toward a Simple Model of Land-Atmosphere Interaction: The Constant Bowen Ratio Approximation in Daytime ABL Dynamics

7.1 Introduction

The dynamics of the atmospheric boundary layer (ABL) is the central hydrologic linkage between the soil (groundwater) and the atmosphere (clouds). As such it is a crucial part of understanding mass and energy transfer associated with the hydrologic cycle. This in turn has broad implications from atmospheric pollutant transport to the understanding of ecohydrological dynamics to parameterizations for large scale climate models. In each of these cases the pursuit of simple models that focus on exploiting the essential dynamical features of the interaction could provide fruitful results for a wide range of applications. Even the simplest land-atmosphere interaction model, however, requires no fewer than three coupled equations (for height of the ABL, temperature, and humidity) to capture the interaction. While in the simplest, lumped models these equations are first-order ODE’s, their analytical solution remains challenging.
Within ABL dynamics, the Bowen ratio represents the coupling between the atmospheric moisture, represented here by the specific humidity $q$, and the air temperature, represented by the atmospheric potential temperature $\theta$. It stands to reason that the pursuit of simplified models of ABL might benefit from a close examination of the Bowen ratio and possible approximations surrounding it. With this in mind, this chapter extends the analysis of [112] which focused on the dynamics of the atmospheric boundary layer with a constant Bowen ratio. In this chapter we consider several possible approximations concerning the Bowen ratio and compare these with exact numerical solution.

7.2 ABL Model Formulation

The basic model formulation presented here follows that of [112] and [138] for an idealized vertically, integrated “slab” model for the mixed-layer with zero-order jump at the capping inversion. Conditions are assumed to be horizontally homogeneous both within and above the mixed-layer with no large-scale subsidence. Additionally we neglect radiative effects in the mixed-layer (these may be thought to be absorbed into the surface sensible heat flux [138]). It is important also to note that the convection treated here is “dry” in the sense that no latent heat release is allowed. The profiles of specific humidity and potential temperature are assumed to be constant (i.e., instantaneous mixing) with height within the mixed layer and linear above the capping inversion. The free atmospheric profiles are given by

\[
q_f = q_{f0} + \gamma_q z \quad (7.1)
\]
\[
\theta_f = \theta_{f0} + \gamma_\theta z \quad (7.2)
\]

where $\gamma_q$ and $\gamma_\theta$ are the free atmospheric lapse rates.
7.2.1 Energy Balance

The energy balance is represented by the partitioning of the net available radiation $Q(t)$ into sensible and latent heat at the surface,

$$Q = \lambda E + H$$  \hspace{1cm} (7.3)

where $\lambda$ is the latent heat of vaporization/condensation. The available radiation is assumed given and will be represented herein by the parabolic form

$$Q(t) = Qm(1 - \left(\frac{t}{t_0} - 1\right)^2).$$  \hspace{1cm} (7.4)

The evapotranspirative flux is assumed to be proportional to the saturation deficit of the air at the surface,

$$E = g_e \rho (q_s^* - q)$$  \hspace{1cm} (7.5)

where $q_s^*$ is understood to be the saturation specific humidity evaluated at the surface temperature. Similarly the sensible heat flux is proportional to the temperature
gradient at the surface. The surface layer is assumed to be shallow so that the
gradient may be represented by the difference between the surface and mixed-layer
temperatures,

\[ H = g_h \rho c_p (\theta_s - \theta) \]  

(7.6)

where \( \theta_s \) is the potential temperature at the surface and \( \theta \) is the potential temperature
of the mixed layer. The Bowen ratio is then

\[ Bo = \frac{H}{\lambda E} = \frac{g_h c_p (\theta_s - \theta)}{g_e (q^*_s - q)} \]  

(7.7)

and for given \( q \) and \( \theta \) the energy balance described by Eqs. (7.5), (7.6) and (7.3) is
closed.

7.2.2 Dynamical Equations

The equations for conservation of energy and mass give us for \( \theta \) and \( q \),

\[ \rho c_p \frac{\partial \theta}{\partial t} = - \frac{1}{h} \frac{\partial (\theta w')}{\partial z} \]  

(7.8)

\[ \rho \frac{\partial q}{\partial t} = - \frac{1}{h} \frac{\partial (q w')}{\partial z} \]  

(7.9)

where \( \rho \) is the air density, \( c_p \) is the specific heat of air at constant pressure, \( h \) is
the height of the boundary layer, the primes indicate turbulent fluctuations and
fluxes are considered positive upward. We assume that potential temperature and
specific humidity are well-mixed with consequent linear profiles for the turbulent
fluxes. The fluxes then may be integrated through the mixed layer to give the
lumped conservation equations [47],
\[
\frac{d\theta}{dt} = \frac{(w'\theta')_s - (w'\theta')_h}{h} \quad (7.10)
\]
\[
\frac{dq}{dt} = \frac{(w'q')_s - (w'q')_h}{h} \quad (7.11)
\]

where the subscripts \(s\) and \(h\) denote the fluxes evaluated at the surface and the top of the mixed layer, respectively. Additionally, in this idealized framework the growth of the convective atmospheric boundary layer grows due to the turbulent entrainment of overlying warmer air. That is, surface heating creates a convective instability that drives entrainment at the top of the boundary layer as plumes penetrate into the overlying layer. Thus the rate of growth of the boundary layer is proportional to the flux of potential temperature at the top of the layer and inversely proportional to the size of the inversion jump \([85, 47]\)

\[
\frac{dh}{dt} = -\frac{(w'\theta')_h}{\Delta \theta}. \quad (7.12)
\]

Plugging this into the equations for \(q\) and \(\theta\) and noting that the surface fluxes are the latent and sensible heat fluxes yields

\[
\rho c_p h \frac{d\theta}{dt} = H + \rho c_p (\theta_f - \theta) \frac{dh}{dt} \quad (7.13)
\]
\[
\rho h \frac{dq}{dt} = E + \rho (q_f - q) \frac{dh}{dt} \quad (7.14)
\]

Finally, we make use of the entrainment assumption \([?]\) that the entrainment can be represented as a constant fraction (\(\beta\)) of the surface sensible heat flux. Making this assumption and ignoring the effect of the morning transition from the stable nocturnal boundary \([112]\) layer yields,
which closes the system of coupled ordinary differential equations for the daytime convective boundary layer in the absence of clouds.

### 7.2.3 Analytical Developments in Porporato (2009)

Porporato (2009) showed two important analytical developments. First, he showed that Eqs. (7.15) and (8.12) could be combined to yield an analytic solution for the potential temperature (neglecting the morning transition) as a linear function of the boundary layer height,

\begin{equation}
\theta = \gamma_\theta \frac{1 + \beta}{1 + 2\beta} h + \theta_f \quad (7.16)
\end{equation}

Figure 7.2: Exact numerical solution of the coupled ABL equations for (a) the height of the mixed-layer, (b) the potential temperature, (c) the specific humidity and (d) the Bowen ratio.
effectively reducing the problem to a pair of coupled differential equations since $\theta$ can be treated as a function of $h$.

Additionally, [112] used the common observation that the Bowen ratio is approximately constant during the midday hours in order to find an approximate solution for $q$. This solution implies that when the Bowen ratio is constant (i.e., around midday) $q$ should behave as a linear function of $\theta$ given by

$$q = \frac{\phi}{2}(\theta - \theta_f) + q_f$$

where

$$\phi = \frac{c_p}{\lambda(1 + \beta)Bo} + \frac{(1 + 2\beta)\gamma_q}{(1 + \beta)\gamma_\theta}. \quad (7.18)$$

The limitations of this latter approximation are made clearer by considering the limiting behavior of the Bowen ratio (also considered in some detail in [112]). In the morning just before sunrise, $Q$ is assumed to be near or equal to zero. From the energy balance it is clear that the Bowen ratio must approach -1 in this case. Thus, in the early morning evaporation tends to cool the earth surface until radiation can provide enough sensible heat to counteract the cooling. At the other extreme, as $Q$ becomes very large, [112] showed that the Bowen ratio approaches a limiting value given by $Bo_{\text{max}} = \frac{g_h c_p}{(g_e \lambda \epsilon)}$ in which $\epsilon_r$ is the approximate slope of the saturation specific humidity curve at a reference temperature between the surface and mixed layer temperatures. It is thus clear that the Bowen ratio passes through zero twice each day, so that the linear $q$ solution becomes singular. Thus the solution finds its domain of applicability limited between these bracketing zeros.

For constant Bowen ratio then [112] presents a complete analytical solution to the land-atmosphere interaction represented by the daytime ABL growth with surface energy partitioning into sensible and latent heat. However, the assumption of strictly
constant Bowen ratio is a significant weakness of the approach as, for example, the derivative of the specific humidity is inversely proportional to the Bowen ratio (see Eq. (7.33) below). Thus lower values of the Bowen ratio to significantly impact the dynamics of $q$. Below we build on the work of [112] using a combination of analytical and numerical methods in order to find an equally tractable representation that is not confined to a constant Bowen ratio.

### 7.3 Simplifications and Solutions

In this section we describe a set of approaches to the problem representing varying degrees of approximation. It may be helpful to keep two limiting cases in mind. We have, on one hand, the exact numerical solution to the fully coupled ABL equations presented above. On the other hand we have the solution for $q$, $\theta$ and $h$ assuming a constant Bowen ratio as presented in [112]. Between these two extremes we will suggest a set of potentially useful approximations.
7.3.1 Limiting Bowen Ratio

We begin our analysis by revisiting the limiting cases of the Bowen ratio as presented in [112]. The Bowen ratio can be written using the formulae for surface latent and sensible heat flux above as

\[
Bo = \frac{g_h c_p (\theta_s - \theta)}{g_e (q_s^* - q)}.
\]  

(7.19)

Following the Penman-Monteith combination approach we can make an algebraic modification separating the vapor deficit into two components,

\[
E = \rho g_e (q_s^* - q^* + q^* - q).
\]  

(7.20)

Again following Penman-Monteith we then approximate the difference in the saturation values between the surface and the air by linearizing the saturation specific humidity curve around a reference temperature (between surface and atmospheric temperatures). Representing the slope by \( \epsilon_r \) we have

\[
E = \rho g_e [\epsilon_r (\theta_s - \theta) + (q^* - q)].
\]  

(7.21)

Plugging this into the Bowen ratio, using \( Q = H + \lambda E \), and rearranging gives,

\[
Bo = \frac{\frac{g_h c_p}{g_e \lambda \epsilon_r} - \frac{\rho g_h c_p}{\lambda \epsilon_r} Q (q^* - q)}{1 + \frac{\rho g_h c_p}{\lambda \epsilon_r} Q},
\]  

(7.22)

from which the limiting cases are easily seen as \( Bo_{Q \to \infty} = \frac{g_h c_p}{g_e \lambda \epsilon_r} \) and \( Bo_{Q \to 0} = -1 \).

We will condense the equation by letting \( Bo_{\text{max}} = \frac{g_h c_p}{g_e \lambda \epsilon_r} \) and \( \bar{Q} = Q/(Bo_{\text{max}} \rho g_e) \) so that we have

\[
Bo = \frac{Bo_{\text{max}} \bar{Q} - (q^* - q)}{\bar{Q} + (q^* - q)}.
\]  

(7.23)
Figure 7.3: ... saturation specific humidity curve with linear approximation over the range T=288-295K.

For clarity, we reiterate that Eq. (7.23) gives an approximate Bowen ratio based on the Penman-Monteith combination approach. That is to say that it is no longer an exact representation of the Bowen ratio. However, we can compare the effect of this approximation directly by computing both the exact Bowen ratio (by solving the closed energy budget) and the approximate form in Eq. 7.23 using the exact numerical solution for \( q \) and \( \theta \) to the ABL equations (see below).

7.3.2 Simplified Specific Humidity Deficit

While the saturation specific humidity is generally a complicated exponential function, for the common range of mixed layer temperatures during the day we can reasonably approximate the curve with a linear relation (see Fig. 7.3). That is, we approximate \( q^* = q_{\theta_r}^* + \epsilon_q(\theta - \theta_r) \), where it is understood that this is evaluated at the base of the mixed layer so that \( \theta \approx T \), and \( \theta_r \) is a suitable reference temperature at which the slope \( \epsilon_q \) is taken.

Additionally, [112] we employ the constant Bowen ratio assumption for \( q \) setting the Bowen ratio in the denominator of \( \phi \) equal to the limiting value, \( \text{Bo}_{\text{max}} \). We then use this relation as an approximation for the saturation deficit, assuming that for much of the day the Bowen ratio is approximately constant (and equal to its
maximum value) so that this relation holds, while not requiring that the Bowen ratio remain constant. Thus, we avoid the singularities in $\phi$ by employing a constant Bowen ratio in $\phi$ while allowing the Bowen ratio to vary elsewhere.

Combining these two approximations yields the specific humidity deficit as a function of potential temperature,

$$q^* - q = q_{\theta_r}^* + \epsilon q (\theta - \theta_r) - \phi/2(\theta - \theta_{f0}) + q_{f0} = A + (\epsilon_q - \phi/2)\theta$$  \hspace{1cm} (7.24)

We may go one step further and relate the deficit directly to the height of the mixed layer since $\theta$ grows linearly with $h$ (Eq. (7.16)),

$$q^* - q = A + \gamma_\theta (\epsilon_q - \phi/2) \left( \frac{1 + \beta}{1 + 2\beta} \right) h$$  \hspace{1cm} (7.25)

### 7.3.3 Simplified Equation With Varying Bowen Ratio

Returning now to the Bowen ratio, we can substitute the linearized specific humidity deficit of Eq. (7.25) into Eq. (7.23) to arrive at a further approximation of the Bowen ratio as a function only of $h$ and $Q$,

$$\text{Bo} = \frac{\text{Bo}_{\text{max}} \tilde{Q} - (A + \gamma_\theta (\epsilon_q - \phi/2) \frac{1+\beta}{1+2\beta} h)}{\tilde{Q} + (A + \gamma_\theta (\epsilon_q - \phi/2) \frac{1+\beta}{1+2\beta} h)}$$  \hspace{1cm} (7.26)

Note that the assumption of approximately constant Bowen ratio was used to linearize the specific humidity deficit which has now been plugged back into an equation for the Bowen ratio. We have now an approximate Bowen ratio depending only on the radiative forcing and the height of the ABL.

We can now substitute this approximate Bowen ratio back into the equation for the growth of the boundary layer,
\[
\frac{dh}{dt} = \frac{(1 + 2\beta) H}{\rho c_p \gamma \theta h} \quad (7.27)
\]
\[
= \frac{(1 + 2\beta) BoQ}{\rho c_p \gamma \theta h (1 + Bo)} . \quad (7.28)
\]

Solving this equation then gives us both \( h(t) \) and \( \theta(t) \) since \( \theta \) is a linear function of \( h \). This equation poses significant challenges to solution, however, and does not show much promise for getting analytic solutions for both \( h(t) \) and \( q(t) \).

### 7.3.4 An Alternative Approximate Approach with Analytic Solution

Here we present an approach with one degree more of simplification from the previous section. To begin, we can rewrite Eq. (7.23) in this way

\[
Bo = \frac{Bo_{\max} \xi - 1}{\xi + 1} \quad (7.29)
\]

where \( \xi = \tilde{Q}/(q^* - q) \). Then substituting this into the term \( Bo/(1 + Bo) \) in the equation for \( h \) yields

\[
\frac{Bo}{Bo + 1} = \frac{Bo_{\max}}{Bo_{\max} + 1} - \frac{1}{\xi(Bo_{\max} + 1)} \quad (7.30)
\]

and interestingly the variable \( \xi \) occurs only once in this relation, the other terms all being constant. The question is whether this can be reasonably approximated in a useful way. The solid curve in Fig.7.4 shows the behavior of \( \xi \) according to the exact numerical solution of the system (i.e., Eqs. (8.12), (8.13), and (7.15)). Noticing that \( \xi \) has roughly the shape of \( Q(t) \), we suggest that \( \xi \) can be reasonably approximated as a constant proportion of \( Q(t) \). The dashed line shows this approximation, \( \alpha Q(t) \), where \( \alpha \) is assumed to be determined by \( \xi_0 \), i.e. \( \alpha = (\lambda \rho g Bo_{\max} (q^*_0 - q_0))^{-1} \).
Figure 7.4: The term $\xi = \tilde{Q}/(q^* - q)$ with the saturation deficit computed exactly (solid line) compared with the approximate representation $\xi \approx \alpha Q$ for $\alpha \approx 0.015$ (dashed line).

Plugging this approximation into Eq. (8.14) yields the simple ordinary differential equation

$$\frac{1}{2} \frac{dh^2}{dt} = \frac{(1 + 2\beta)}{\rho c_p \gamma_\theta} \left( \frac{B_{\text{max}} Q}{B_{\text{max}} + 1} - \frac{1}{\alpha (B_{\text{max}} + 1)} \right)$$

which can be integrated directly to give

$$h = \sqrt{\frac{2(1 + 2\beta)}{\rho c_p \gamma_\theta} \left( \frac{B_{\text{max}}}{B_{\text{max}} + 1} \int_0^t Q(u) du - \frac{t}{\alpha (B_{\text{max}} + 1)} \right) + h_0^2}.$$ 

This approach effectively bypasses all of the simplifications in the previous section by collecting the terms in question into a single variable $\xi$. We then close the system by making an assumption about the behavior of this one variable with the result that the mixed-layer height is solvable analytically.

7.3.5 Toward a Solution for Specific Humidity

We would like to have a solution to the entire land-atmosphere coupling based on these simplifications. However, the assumption that $q - q_{f0} = \phi/2(\theta - \theta_{f0})$ cannot...
be used directly to return a reasonable value of \( q \) unless the Bowen ratio is truly constant.

We begin with Eq. (8.13) for \( q \), substituting Eq. (7.15) for \( \frac{dh}{dt} \), and writing the evapotranspiration in terms of the Bowen ratio which gives

\[
\rho h \frac{dq}{dt} = \frac{Q Bo}{1 + Bo} \left[ \frac{1}{\lambda Bo} + \frac{1 + 2\beta \gamma_q}{c_p \gamma_\theta} + \frac{(1 + 2\beta)\tilde{q}}{c_p \gamma_\theta h} \right]
\]

(7.33)

where \( \tilde{q} = q_f - q_f^0 \). We again notice the term \( Bo/(1 + Bo) \) and will make the same simplification from Eq. (7.30) with \( \xi = \alpha Q \). Making this substitution and inserting the approximate analytic solution for \( h(t) \), yields an equation of the form,

\[
\frac{d\tilde{q}}{dt} = X_1(t) + X_2(t)\tilde{q}
\]

(7.34)

where

\[
X_1(t) = \left( 1 + \beta \right) \gamma_q \left( \frac{Bo_{max}}{1 + Bo_{max}} \frac{Q(t)}{h(t)} - \frac{1}{\alpha(1 + Bo_{max})} \frac{1}{h(t)} \right) + \frac{1}{\rho \lambda (1 + Bo_{max})} \left( \frac{Q(t)^2}{h(t)} \frac{Q(t)}{\alpha h(t)} \right)
\]

(7.35)

\[
X_2(t) = - \left( 1 + 2\beta \right) \frac{Bo_{max}}{c_p \gamma_\theta} \left( \frac{Bo_{max}}{1 + Bo_{max}} \frac{Q(t)}{h(t)} - \frac{1}{\alpha(1 + Bo_{max})} \frac{1}{h(t)} \right).
\]

(7.36)

Equation (7.34) can be solved generally by the use of an integrating factor despite the complexity of the functions of time. The general solution is then

\[
\tilde{q}(t) = e^{-\int_0^t X_2(u)du} \left( \int_0^t X_1(w)e^{\int_0^w X_2(u)du}dw \right)
\]

(7.37)
7.4 Results and Discussion

Figure 7.5 shows a comparison of the solutions for \( h(t) \) found using the various approximations in this paper. The solid line represents the exact numerical solution to the fully coupled system of ABL equations while the dotted line represents the solution according to the constant Bowen ratio assumption presented in [112] with the Bowen ratio set to its maximum value. The two dashed lines correspond to approximations presented in this paper. The first notable aspect of the approximate solutions is their relative agreement with the exact solution. The constant Bowen ratio approximation significantly overestimates the height of the mixed layer. Taking a representative Bowen ratio of one half of the maximum improves the estimation (not shown) but remains significantly above the exact solution.

The dashed line in Fig. 7.5 corresponds to the numerical solution of Eq. (7.15) using the approximate form of the Bowen ratio given in Eq. (7.26). Finally, the dash-dotted line represents the analytical solution in (7.32) which appears to be a good approximation to the exact solution without the ambiguity of choosing a representative Bowen ratio.

Figure 7.6 shows the same approximations for the specific humidity \( q \) with one addition. In this figure there appear two dotted lines. The bold dotted line represents the constant Bowen ratio solution which relates \( q \) linearly to both \( h \) and \( \theta \) with Bowen ratio set to its maximum value. The higher dotted line corresponds to the Eq. (7.33) integrated numerically with the Bowen ratio constant but with \( h(t) \) given by the analytic expression in (7.32). As such it is a sort of hybrid solution that incorporates the variable Bowen ratio only via the dependence on \( h(t) \). Again the agreement is strong amongst all solutions except that of the strictly constant Bowen ratio. In the latter case the curve can be shifted up by reducing the value of the Bowen ratio used as representative.
Figure 7.5: A comparison of $h(t)$ solutions. The solid curve represents the exact solution to the coupled ABL equations. Increasing degrees of approximation are then represented by the dotted/dashed lines. Dashed curve: numerical solution to Eqs. (8.14) and (7.26). Dot-Dashed curve: analytic solution given in Eq. (7.32). Dotted curve: analytic solution for the case of constant Bowen ratio (Porporato 2009 Eq. 37) with $Bo = Bo_{\text{max}}$.

In comparison, the analytic solution in (7.32) which assumes an approximated, time-dependent Bowen ratio is more faithful to the exact solution than constant Bowen ratio solution. A comparison of the various Bowen ratios is shown in Fig. 7.7. This final comparison raises an interesting question regarding the mechanism behind the nearly constant Bowen ratio. The argument made in [112] suggests that the approximately constant value of the Bowen ratio through midday is due to a limiting value. This is why the maximum value has been used here in representations of the analytic solutions of [112]. However, the value of $Bo_{\text{max}}$ for the parameters used here is close to 0.6 which is double or triple the maximum value found in any of the solutions. The only arguable parameter in $Bo_{\text{max}}$ is $\epsilon_r$ which has been taken at the reference temperature 290K for these simulations. However, in order to reduce $Bo_{\text{max}}$ by a factor of two the reference temperature would have to be raised over
Figure 7.6: A comparison of $q(t)$ solutions of varying degrees of approximation. The solid curve represents the exact solution to the coupled ABL equations. The dashed curve shows the numerical solution to Eq. 8.13 using the analytic solution of $h(t)$ given in Eq. 7.32 and the Bowen ratio given by the Penman-Monteith approximation (Eq. 7.23). The dash-dot curve shows the solution using the same analytic $h(t)$ but with the approximate Bowen ratio given in Eq. 7.29. Finally, the two dotted curves show the solution for constant Bowen ratio (taken as 1/2 of the maximum). The lower dotted curve shows the analytic relation for strictly constant Bowen ratio given in [112]. The upper dotted line shows the numerical solution to Eq. 8.13 with the analytic $h(t)$ but with constant Bowen ratio.

$300K$ which seems unreasonable. Additionally, this value of $\epsilon_r$ and therefore $B_{\text{max}}$ used in the other approximate solutions generates a Bowen ratio that is amenable to the exact solution, so it is unlikely that the error is in the selection of the reference temperature. It might be tempting to raise the maximum net radiation to simulate the limit as $Q \to \infty$, but within the dynamics of the ABL this simply results in the saturation deficit becoming strongly negative as $q$ increases in the mixed layer and $h$ grows rapidly into drier air. Thus what we may be seeing is that the constant Bowen ratio depends more on the balancing of entrainment of dry air and evapotranspiration than on a simple limiting value. This may be demonstrated by taking the time
derivative of (7.7) and setting it equal to zero, which produces the relation,

\[ \frac{d}{dt} \log(\theta_s - \theta) = \frac{d}{dt} \log(q_s^* - q), \] (7.38)

the implications of which are not immediately obvious and warrant further study. The main implication is that the constant Bowen ratio may be a dynamic effect rather than an algebraic one. This pushes the question back a step further toward investigating why this dynamical balance occurs.

7.5 Conclusion

In this chapter we have extended the work of [112] on a simple model of the land-atmosphere interaction via the “dry” convective atmospheric boundary layer. We have presented a comparison of several degrees of approximation with regard to the Bowen ratio as a means toward solving the ABL equations. Among these we in-
roduced a new approximation based on setting the variable $\xi$ equal to a constant proportion of $Q$. This then allows for the solution in Eq. (7.32) which is a new analytic solution to the daytime evolution of the atmospheric boundary layer that agrees well with the exact numerical simulations. In addition we find an interesting discrepancy between the rationale for the constant Bowen ratio approximation presented in [112] and the actual value at which the Bowen ratio appears to be constant.
8.1 Introduction

In the previous chapter we studied a simple lumped model of the dry atmospheric boundary layer following the treatments of [112] and [138] with emphasis on finding potentially useful simplifying assumptions. In this chapter we extend the work further to consider the evolution of the boundary layer with the inclusion of cloud formation. When clouds are present, however, the dynamics of the ABL become significantly more complicated. In this chapter we begin to develop a model of the evolution of the ABL through both dry and cloud-topped stages in order to study the effect of the land-atmosphere interaction on the triggering of free convection.

In this chapter we consider the cloud layer to be a homogeneous strato-cumulus layer with well-defined vertical boundaries at the level of condensation (see below for details) and the top of the ABL. The presence of clouds significantly complicates the picture of the atmospheric boundary layer through the introduction and dynamical importance of radiative fluxes and phase changes [30]. Whereas in the dry ABL the
significance of phase changes and radiative fluxes are limited to the earth surface, in the cloud-topped boundary layer (CTBL) cloud-top long-wave radiation and vapor condensation play significant roles in the thermodynamics of the CTBL at height. Cloud-top longwave radiation acts to cool the air at the top of the ABL thereby feeding the turbulence in the CTBL through negative buoyancy production. This long-wave radiation is concentrated in a relatively shallow layer at the top of the cloud on the order of 10 meters [30]. On the other hand, cooling at the cloud top tends to condense vapor and release latent heat which tends to lessen the instability at the cloud top. The turbulent entrainment of free atmospheric air tends to make the boundary layer drier and warmer suppressing the formation of clouds. In addition there are the processes of cloud microphysics, shortwave absorption and wind shear that all play important roles in the CTBL.

Because of these complexities, in contrast to the dry ABL, there is currently no consensus on the modeling of the CTBL particularly with regard to parameterizing the turbulent entrainment at cloud top [30, 135, 140]. While considerable research has focused on this entrainment parameterization as well as the modeling of the CTBL and the dry ABL separately, much less energy has gone into developing simplified models with a tractable entrainment scheme that aim to model the evolution of a dry ABL into a cloud-topped layer in a single framework. The development of such an integrated approach, despite simplifications, may yet yield insight into essential interplay of land-atmosphere interactions and their role in convective rainfall initiation. In this chapter we begin the development of such an integrated framework by combining the model for the dry ABL in the previous chapter with a significantly simplified form for the dynamics of the CTBL. We then use this model to investigate, for example, the coupling of surface conditions with the likelihood of initiation of free convection, and the strength of that convection.
8.2 Atmospheric Variables and Profiles

8.2.1 The Free Atmospheric Temperature Profile

As in the previous chapter we assume horizontally homogeneous conditions (hydrostatic atmosphere). As the vertical profile of temperature at higher altitudes will play an important part in determining the convective available potential energy below, we will take a moment to consider in more detail the assumption of a linear potential temperature profile in the free atmosphere. Using the temperature profile of the U.S. Standard Atmosphere (linear in the troposphere) [95], we have the atmospheric pressure distribution,

\[ P(z) = P_0 \left( \frac{T_0 - \gamma_T z}{T_0 - \gamma_T z} \right)^{\frac{g}{R \gamma_T}}, \tag{8.1} \]

where \( g \) is the acceleration due to gravity, \( R \) is the gas constant, \( \gamma_T \) is the atmospheric lapse rate \( (\gamma_T \approx 0.0065) \), and \( T_0 \) represents the surface temperature taken here as \( T_0 = 288K \). We will assume this to be the pressure distribution of the atmosphere (both within the mixed-layer and in the free atmosphere) throughout this paper.

From the pressure distribution and the definition of potential temperature we can derive the profile of \( \theta(z) \),

Qualitative description of moist convection, cloud processes, etc.

\[ \theta = T \left( \frac{P_b}{P} \right)^{\frac{R}{cp}} \]

\[ = (T_0 - \gamma_T z) \left( \frac{T_0}{T_0 - \gamma_T z} \right)^{\frac{\gamma_d}{\gamma_T}} \]

\[ = T_0^{\gamma_d/\gamma_T} T(z)^{1-\gamma_d/\gamma_T}, \tag{8.2} \]

where \( \gamma_d \) represents the dry adiabatic lapse rate \( \gamma_d = g/c_p \approx 0.01 \). We can now evaluate the effect of the assumption of linear profiles in \( \theta \) in [112] and the previous
chapter. From this equation we see that the slope of the $\theta$ profile is dependent on $z$. A first approximation for a linearized slope might then be $d\theta/dz|_{z=0} = (\gamma_d - \gamma_T)$. Equivalently the slope may be evaluated at an intermediate point (e.g., $z = 1000\text{m}$) to represent an average linearized slope over the range relevant to ABL dynamics. Thus, the assumed linear profile for $\theta$ is equivalent to assuming a power-law form of the free atmospheric temperature profile. Figure 8.1 illustrates the assumption of linear potential temperature in the free atmosphere using the Standard U.S. Atmospheric temperature profile in the troposphere. For our purposes here the linear approximation is probably sufficient, noting also that the linear assumption underestimates the atmospheric temperature in the higher troposphere (see below for further discussion).

The free atmospheric profiles of potential temperature and specific humidity are then, as in the previous chapter,

$$
\begin{align*}
\theta_f &= \gamma_d z + \theta_f^0 \\
q_f &= \gamma_q z + q_f^0,
\end{align*}
$$

(8.3)

where $\gamma_\theta \approx \gamma_d - \gamma_T$. 

Figure 8.1: Illustration of the profile of $\theta$ assuming hydrostatic conditions (dashed line) and the linear approximation (solid line) through the troposphere.
8.2.2 Conserved Quantities in the Cloud-Topped Atmospheric Boundary Layer

In the cloud-topped boundary layer because of latent heat release the potential temperature $\theta$ and specific humidity $q$ are no longer conserved variables. However, much of the framework from the previous chapter on the ABL can be preserved by using the equivalent potential temperature $\theta_e$ and total water content $q_t$ as the dynamic variables. The total water content of the air is simply the sum of the vapor and condensed phases $q_t = q_l + q$, where $q_l$ is the liquid water content of the air. For the purposes of ABL modeling we may use the linearized form of $\theta_e$ such that $\theta_e = \theta + \lambda/c_p q$ [47].

Because the potential temperature and specific humidity profiles are assumed linear, it follows that both $\theta_e$ and $q_t$ are linear as well with the free atmospheric profile of $\theta_e$ given by

$$\theta_{ef} = \theta_f + (\lambda/c_p)q_f = (\theta_{f0} + (\lambda/c_p)q_{f0}) + (\gamma_\theta + (\lambda/c_p)\gamma_q)z$$

(8.4)

Because the free atmosphere is assumed to be unsaturated we have that $q_{tf} = q_f$, where $q_{tf}$ represents the profile of $q_t$ in the free atmosphere. Our free atmospheric profiles are then,

$$\theta_{ef} = \gamma_\theta z + \theta_{ef0}$$

$$q_{tf} = \gamma_q z + q_{f0} = q_f$$

(8.5)

8.3 Thresholds of Moist Convection

When moist processes are taken into account in the dynamics of the boundary layer, three thresholds corresponding to specific heights above the surface become important: the lifting condensation level (LCL), the level of free convection (LFC), and
the level of neutral buoyancy (LNB). In this section we will discuss each of these thresholds.

8.3.1 Lifting Condensation Level

The lifting condensation level (LCL) is the level at which a parcel of air at the top of the boundary layer with potential temperature $\theta$ and specific humidity $q$ would saturate if lifted adiabatically (see Fig. 8.2. Since during adiabatic lifting $\theta$ remains constant the LCL is determined by $q_t$ of the boundary layer and the Clausius-Clapeyron equation for the saturated specific humidity,

$$q^*(\theta) = q^*[T_{adiabat}(z_{LCL})] = q_t \quad (8.6)$$

Once the mixed-layer reaches the LCL strato-cumulous clouds begin to form. Within the cloud layer the lapse rate follows the saturated adiabatic lapse rate which is less than the dry adiabatic lapse rate because of latent heat release. Thus, once the ABL reaches the LCL further entrainment will act to decrease the strength of the capping inversion as latent heat release warms the cloud layer (see Fig. 8.3).
Figure 8.3: Schematic representation of the lifting condensation level and the level of free convection. In the unsaturated ABL the lapse rate is approximately adiabatic while in the free atmosphere the lapse rate is sub-adiabatic (i.e., stable). Above the LCL the lapse rate follows the saturated adiabat (dash-dot line). The level of free convection is the height at which a parcel lifted adiabatically first becomes warmer than the surrounding atmosphere.

8.3.2 Level of Free Convection

As the boundary layer grows by further entrainment in a saturated cloud layer the strength of the capping inversion is eroded with the possibility that the temperature jump at the inversion is negligible. At this point the boundary layer is convectively unstable since the adiabatic lifting of a parcel above this level would result in its being warmer than the surrounding air (because of latent heat release) so that it would continue to rise. This level at which the saturated adiabat intersects the free atmospheric lapse rate is termed the level of free convection. At this point the “lid” on the convective boundary layer, in which the convection has hitherto been driven by surface heating, has been removed and free convection begins as saturated air accelerates rapidly through the atmosphere, resulting in the towering convective cumulous clouds associated with thunderstorms.
8.3.3 Level of Neutral Buoyancy and CAPE

Saturated air will continue to rise until the surrounding air is warmer. This occurs when the saturated adiabat (which is curved) again crosses the free atmospheric temperature profile (which in general can be much more complicated than our simple linear example). At the level when the saturated adiabat crosses the free atmospheric temperature profile parcels are neutrally buoyant. Above this level, the level of neutral buoyancy, adiabatically lifted parcels are cooler than the environment because, as the air has dried during ascent, the marginal release of latent heat through further lifting is less than the difference between the dry adiabatic lapse rate and the free atmospheric lapse rate.

The distance between the LFC and the LNB and particularly the temperature defect between the moist adiabat at the free atmospheric temperature play an important part in the strength of the resulting convection. An index that is used to characterize the strength of the convection is the vertically integrated buoyant force due to the temperature difference between the two levels. This quantity, with units of energy, is termed the convective available potential energy or CAPE and is given by [65],

$$\text{CAPE} = \int_{z_{LFC}}^{z_{LNB}} g \left( \frac{T_{\text{adiabat}} - T_f}{T_f} \right) \, dz$$  \hspace{1cm} (8.7)

where \(g\) is the acceleration due to gravity, \(T_{\text{adiabat}}\) is the temperature along a saturated adiabat and \(T_f\) is the temperature of the free atmosphere. Typical values of CAPE for a thunderstorm are approximately 500 \(m^2s^{-2}\) while severe storm conditions over land may have CAPE in excess of 2500-3000 \(m^2s^{-2}\) [65].
8.4 Cloud-Topped Boundary Layer Dynamical Model

In the absence of condensation/clouds the simplified ABL evolves exactly as in the previous chapter. When the top of the mixed layer reaches the LCL, the variables $q$ and $\theta$ are no longer conserved. In this case the equivalent potential temperature, $\theta_e$ and the total water mixing ratio $q_t$ are conserved and constant through the mixed layer. In addition, once the LCL is reached several new physical processes are brought into the mix by the presence of clouds including long-wave radiation from droplets at the cloud top, solar absorption by the cloud, and reduction of net available radiation to the surface.

Approaches to the modeling of the cloud-topped boundary layer have treat cloud top radiation dominates the dissipative effect of solar absorption within the cloud layer.

8.4.1 Energy Budget

The surface energy budget proceeds as in the dry ABL model. For the energy budget we assume the net radiation $Q(t)$ is given. If cloud shading is incorporated into the model it may be done through the $Q(t)$ function. In that case $Q$ becomes dependent on the cloud albedo. For now we neglect cloud shading and assume $Q$ is only a function of time. The latent heat and sensible heat are given by

$$
\begin{align*}
E &= g_e \rho (q^* - q), \\
H &= g_h \rho c_p (\theta_s - \theta), \\
Q &= \lambda E + H,
\end{align*}
$$

(8.8)

where the subscript “s” denotes conditions at the surface. With $Q(t)$ given and the Clausius-Clapeyron equation for $q^*$, we have for the energy balance three equations in five unknowns ($\theta$ and $q$ being the other unknowns). These equations for $H$ and $E$ can be converted into terms of the conserved variables by noting first that $q_t = q$
at the surface (assuming the surface is not saturated) and then using the linearized relation between $\theta_e$ and $\theta$,

$$E = g_e \rho(q_{ts} - q_t),$$

$$H = g_h \rho c_p(\theta_s - \theta_e + (\lambda/c_p)q_t).$$

(8.9)

In this equation $\theta_s$ is not written in terms of $\theta_e$ since it is solved “offline” within the energy budget since $q_{ts} = q_{ts}(\theta_s)$. That is, the energy budget is a closed system for given $q$ and $\theta$.

**8.4.2 ABL Growth**

Following [47, ?] assuming linear fluxes, the equations for $\theta_e$ and $q_t$ in the mixed layer become

$$\frac{\partial \theta_e}{\partial t} = \frac{(w' \theta'_e)_0 - (w' \theta'_e)_h}{h} - \frac{\Delta R}{\rho c_p h}$$

$$\frac{\partial q_t}{\partial t} = \frac{(w' q'_t)_0 - (w' q'_t)_h}{h}$$

(8.10)

The fluxes at the surface are determined by the surface energy budget. Additionally we have,

$$\frac{\partial h}{\partial t} = \frac{-(w' \theta'_e)_h}{\Delta \theta_e} = \frac{-(w' q'_t)_h}{\Delta q_t}$$

(8.11)

Plugging in for the surface fluxes $(w' \theta'_e)_0 = (w' \theta'_e)_0 + \lambda/c_p(w' q'_t)_0 = H/\rho c_p + \lambda/\rho c_p E$ and $(w' q'_t)_0 = E/\rho$, we have

$$\rho c_p h \frac{\partial \theta_e}{\partial t} = Q + \rho c_p \Delta \theta_e \frac{\partial h}{\partial t} - \Delta R$$

(8.12)

$$\rho h \frac{\partial q_t}{\partial t} = E + \rho \Delta q_t \frac{\partial h}{\partial t}$$

(8.13)
For the growth of the boundary layer height we have

\[
\frac{\partial h}{\partial t} = -\frac{(w'\theta_e')_h}{\Delta \theta_e}.
\]  

(8.14)

### 8.4.3 Entrainment Condition

To close the system of equations (8.9), (8.12), (8.13) and (8.14) the entrainment flux at the top of the boundary layer must be provided in Eq. (8.14). In the absence of clouds the closure scheme is again the assumption that the entrainment is proportional to the surface sensible heat flux [112, 138]. However, in the presence of clouds a variety of complex closure schemes have been developed (e.g.,[27, ?]) based on the buoyancy flux through the boundary layer. The gamut of closure schemes has been categorized by [118] into “Eulerian” and “process” partitioning. Here we describe briefly a canonical example of each approach as well as a third simplified approach which ignores condensation effects in the cloud (the so-called “dry cloud”). Finally, we describe the closure adopted for our model and place it in relation to the other three.

In each case the closure scheme proceeds by assuming that the turbulent kinetic energy budget can be reduced to the form \( B = D \) where \( B \) is buoyant production and \( D \) is dissipation. The buoyancy term is further partitioned into production and consumption terms, \( B = P - N \). Eulerian and process partitioning differ primarily in the method of accounting for these terms. The primary closure assumption is then that the dissipation can be represented as a fraction of the production, or \( D = (1 - A)P \) where \( A \) is an empirical parameter. Then, as the production and consumption terms are functions of the entrainment rate, the problem is closed and may be solved for the entrainment rate.
Deardorff (1976) presents a canonical example of what [118] terms Eulerian partitioning in order to parameterize the entrainment at the top of the CTBL. In this scheme the buoyancy flux is partitioned into production and consumption terms. The assumption of Eulerian partitioning is that the entrainment flux is given by a fraction of the vertically integrated net buoyancy flux. As presented in [47] this assumption is given by

\[
(w' \theta_v) h = \frac{2 \beta}{h(1 - \beta)} \int_0^h w' \theta_v dz
\]  

(8.15)

where \( \theta_v \) is the virtual potential temperature which can be represented by a linearized version as \( \theta_v = \theta(1 + 0.61q - q_l) \). This, for \( \beta = 0.2 \), reduces to the familiar condition for the dry ABL if the buoyancy flux is linear and continuous through the boundary layer. In general however, the buoyancy flux is piecewise linear with a jump at the cloud base due to the increase in buoyancy flux from condensation. The parameterization for \( \overline{w' \theta_v} \) itself becomes quite involved and is beyond what we require here (see [27] for the full parameterization).

**Process Partitioning**

Process partitioning, as presented for example in [134] treats each physical process acting on a fluid parcel as though it were acting independently of the others [118]. Thus surface heating, surface evaporation, and radiative cooling are each treated as contributing to production or consumption of turbulent kinetic energy independently of one another. The result of these assumptions is that production and consumption are linear functions of the entrainment rate (see [134] for details), in contrast to the Eulerian partitioning of [27] which has a more complicated relationship between the terms. Process partitioning can therefore be easier to employ. However, process
partitioning carries with it several questionable properties, the most obvious of which is the very assumption that physical processes operate on parcels independently of one another. Another difficulty is that the scheme is more sensitive to consumption terms than to production terms [118]. Thus reallocating a productive term as a negative consumptive term or vice versa may change model results considerably.

**The Dry Cloud Closure**

Lock and MacVean (1999) used large eddy simulations of radiating dust clouds to parameterize the entrainment rate. In the scheme the entrainment rate for the dust cloud was represented as being proportional to a linear combination of the surface sensible heat flux and the cloud-top radiative flux, or

\[-(w'\theta')_h = C_1(w'\theta'_s + g/\theta_0 \Delta F)\]  (8.16)

where \(\theta_0\) is a reference temperature and \(\Delta F\) is the cloud-top radiative flux represented as a discontinuity in the radiative flux at the cloud top.

This closure, without the surface flux term, was used by in a model of the well-mixed cloud-topped boundary layer and shown to perform within the limits of the predictions of the more complex parameterizations with regard to the growth of the CTBL.

**A Combined Simplified Approach**

For our closure scheme we desire both the simplicity of the scheme as well as the property of the more complicated partitioning schemes that the CTBL entrainment reduces to the dry ABL entrainment assumption. In order to preserve these traits we follow in assuming that the buoyancy flux is a linear combination of the contributions due to the surface buoyancy flux and the radiative cooling at the cloud top (assuming discontinuous inversion). In our own notation,
\[-(w'\theta')_h = \beta_1 (w'\theta')_s + \beta_2 \frac{\Delta R}{\rho c_p}\]  

(8.17)

Note that when \( \Delta R \to 0 \) the condition reduces to the dry ABL entrainment condition where \( \beta_1 = \beta = 0.2 \). Finally, writing the condition in terms of \( \theta_e \) (i.e., \( \theta' = \theta'_e - (\lambda/c_p)q' \)) yields,

\[-(w'\theta'_e)_h = \beta_1 \left[ (w'\theta'_e)_s - (\lambda/c_p)(w'q')_s \right] + \beta_2 \frac{\Delta R}{(\rho c_p)} - (\lambda/c_p)(w'q')_h = \beta_1 \left[ (w'\theta'_e)_s - (\lambda/c_p)(w'q')_s \right] + \beta_2 \frac{\Delta R}{(\rho c_p)} - (\lambda/c_p)H \]  

(8.18)

where Eq. 8.11 has been substituted in the third term on the left for the flux of water vapor at the top of the boundary layer. Additionally the flux of liquid water at the top of the boundary layer is exactly zero due to instantaneous evaporation/condensation so that \( q'_t = q' \). Rearranging to solve for the flux term yields,

\[-(w'\theta'_e)_h = \frac{\beta_1 H + \beta_2 \Delta R}{\rho c_p (1 - \frac{\lambda \Delta q}{c_p \Delta \theta_e})}\]  

(8.19)

Substituting Eq. (8.19) back into Eq. (8.14) then gives us the rate of growth of the boundary layer,

\[\frac{\partial h}{\partial t} = \frac{\beta_1 H + \beta_2 \Delta R}{\rho c_p (\Delta \theta_e - (\lambda/c_p)\Delta q_t)}\]  

(8.20)

Clearly, when \( \Delta R \to 0 \) we return to the dry boundary layer equation

\[\frac{\partial h}{\partial t} = \frac{\beta_1 H}{\rho c_p \Delta \theta}.\]  

(8.21)
8.4.4 Relationship Between ABL Height and Inversion Strength

Following the appendix of [112] we derive the relationship between the inversion strength \( \Delta \theta_e = \delta = \theta_{ef} - \theta_e \) and the height of the boundary layer \( h \). Taking the time derivative of the definition of \( \delta \) we get the kinetic equation,

\[
\frac{d\delta}{dt} = \gamma \theta_e \frac{dh}{dt} - \frac{d\theta_e}{dt}
\]

and recall that the boundary layer grows as

\[
\delta \frac{dh}{dt} = - (\theta_e' w')_h.
\]

We can replace \( d\theta_e/dt \) above with

\[
h \frac{d\theta_e}{dt} = (\theta_e' w')_s - (\theta_e' w')_h - \Delta R/(\rho c_p).
\]

Substituting the closure condition for \( (\theta_e' w')_s \) and \( \delta dh/dt \) for the flux at the top of the boundary layer gives,

\[
h \frac{d\delta}{dt} = \gamma \theta_e h \frac{dh}{dt} - \left( 1 + \frac{1}{\beta_1} \right) \delta \frac{dh}{dt} + \left( 1 + \frac{\beta_2}{\beta_1} \right) \frac{\Delta R}{\rho c_p},
\]

which again reduces to the case of [112] for \( \Delta R \to 0 \) as \( \delta \) may then be solved in terms of \( h \) (implicit in time) yielding a linear relationship.

8.4.5 Model Summary

Combining the results of the previous sections we have the system
\[
\rho c_p h \frac{\partial \theta_e}{\partial t} = Q + \rho c_p \Delta \theta_e \frac{\partial h}{\partial t} - \Delta R
\]
\[
\rho h \frac{\partial q_t}{\partial t} = E + \rho \Delta q_t \frac{\partial h}{\partial t}
\]
\[
\frac{\partial h}{\partial t} = \frac{\beta_1 H + \beta_2 \Delta R}{\rho c_p (\Delta \theta_e - (\lambda/c_p) \Delta q_t)}
\]

where the jumps in \(\theta_e\) and \(q_t\) are given by

\[
\Delta \theta_e = \theta_{ef} - \theta_e
\]
\[
\Delta q_t = q_{tf} - q_t,
\]

and the profiles in the free atmosphere are assumed to be linear

\[
\theta_{ef} = \gamma_{\theta_e} h + \theta_{e0}
\]
\[
q_{tf} = \gamma_{q_t} h + q_{t0},
\]

and \(q_t = q + q_l\) where \(q_l\) represents the mass of liquid water per unit mass of air. Of course, \(q_l > 0\) only in the cloud layer.

8.5 Simulation and Preliminary Results

Figure 8.4 shows the results of simulating the ABL combined model. Initially the ABL is cloudless and evolves according to the model of the previous chapter with an identical entrainment condition. At the LCL the entrainment rate receives a boost from the cloud-top radiation. The boost is seen clearly in comparison with the dry ABL solution in the figure. The enhanced growth of the boundary layer increases the rate at which drier air is incorporated into the mixed layer thus retarding the growth of the mixed-layer specific humidity.

Additionally, we show the evolution of the three thresholds (LCL, LFC, and LNB) during the day up until the initiation of free convection. Note that the distance
Figure 8.4: Numerical simulation of the daytime boundary layer evolution: (a) shows the height of boundary layer (solid line), the solution for the dry ABL (dashed line), and the cloud base at the LCL (dotted line); (b) shows the equivalent potential temperature of the mixed-layer; (c) shows the total water mixing ratio in the mixed-layer (solid line) and the saturation specific humidity curve (dashed line); (d) shows the evolution of the LNB (solid line), LFC (dashed line) and the LCL (dotted line) during the day.

Figure 8.5: The temperature for a parcel lifted along an adiabatic path from the surface (solid line) and the free atmospheric temperature profile (dashed line).
Figure 8.6: Instantaneous profile of the temperature profile with the jump at the top of the boundary layer (solid line) and the extrapolated free atmospheric profile for comparison (dashed line): (a) before reaching the LCL and (b) at the LFC.

Figure 8.7: The partitioning of water into liquid and vapor phases at the top of the boundary layer.

Figure 8.8: The evolution of CAPE during daytime hours up to the initiation of free convection.
between the LFC and LNB grows approximately linearly in time. Figures 8.5 and 8.6 show a sample parcel path as well as the vertical temperature profiles for two times during the day. We also show in Fig. 8.7 the water balance at the top of the mixed-layer during the daytime evolution.

In Fig. 8.8 we show the evolution of the CAPE during the day. Interestingly CAPE is approximately linear through the day. Thus, in the absence of variable environmental lapse rate, the model suggests that storm intensity is proportional to the time it takes to reach convective initiation.

8.6 Conclusions

This chapter describes a simplified model of ABL dynamics that integrates the development of the dry ABL prior to reaching the LCL with the development of the cloud-topped ABL up to the level of free convection. The dry ABL model follows that described in the previous chapter. In the presence of clouds we follow [104] and [?] by using a simple closure model for the cloud-topped boundary layer that takes the entrainment rate to be a linear combination of the surface latent heat flux and the radiative flux at the cloud-top. Only the most preliminary results have been presented here. Future work will focus on investigating the role of land surface conditions in determining the timing of cloud cover and free convection as well as the strength of resulting convection (represented by CAPE).
Intermittency and irreversibility are two fundamental characteristics of the hydrologic cycle. While most flows, particularly at the hourly to daily time-scales, are basically continuous in time, precipitation introduces a strongly intermittent component into the cycling of water in the atmosphere. “Intermittent” is understood here to describe processes that may be qualitatively divided into “on” and “off” states. Precipitation time-series, for example, exhibit long periods of zero rainfall interrupted by relatively brief bursts of highly fluctuating precipitation intensities (see Chapter 6).

In addition to the intermittent behavior exhibited by precipitation, the cycling of water itself is due to the irreversible (entropy producing) thermodynamic processes of the Earth’s surface and atmosphere as they are forced by solar radiation. Thus intermittency and irreversibility are inherently linked in the land-atmosphere system and perhaps nowhere as tightly as in the dynamics of the atmospheric boundary layer.

In this work we have sought to advance the understanding of these two properties as being inherently linked in the dynamics of the hydrologic cycle. We may
describe the methodology of this approach heuristically as a balance of stochastic and
deterministic approaches that reflect the intermittency of processes such as rainfall
(stochastic) with the irreversibility associated with deterministic evolution of an ex-
ternally forced system. Again using rainfall as an example, the literature on rainfall
modeling can be broadly divided into approaches that seek to model phenomenology
of rainfall with approaches that seek to model physically the initiation and dynam-
ics of rainfall events. The former are most often purely mathematical constructions
drawn from the theory of stochastic processes which seek to mimic rainfall series.
The latter, on the other hand, generally consist of very high-dimensional numerically
solved systems of equations. Thus, in the stochastic case we often seek to mimic the
intermittency of rainfall, while in the physical case we attempt to simulate the de-
terministic trajectory of rainfall events.

In contrast to such approaches, we suggest a hybrid approach that seeks to capture
the deterministic trajectory through judicious use of idealized models, particularly
when those models are solvable analytically, coupled with appropriate stochastic
features. That is, in many of these high-dimensional systems the dynamics may, in
specific regions of the phase space and at particular scales, be governed primarily by
only a few variables while in other regions the influence of particular variables may
be more evenly spread across the degrees of freedom. In the case of a few variables
dominating the dynamics we may employ relatively simple deterministic models for
the trajectory through phase space. Similarly, where the trajectories in phase space
become more sensitive to many degrees of freedom we may employ stochastic terms
to simulate this high-dimensionality.

While the previous description is purely heuristic, it suggests a reasonable method
for investigating these complex environmental systems. Furthermore, while we move
out of a deterministic framework in order to model those high-dimensional regions
using stochastic terms, we should not forget entirely the picture of a trajectory in
phase space. We thus need tools for analyzing these high-dimensional dynamics that help recognize the underlying causality. As a step in this direction we developed a measure of temporal asymmetry for application to stationary time-series (chapter 5). This measure effectively quantifies the degree of statistical asymmetry in a series. In terms of information theory, this measure is a relative entropy that describes the degree to which a series is distinguishable from its reverse. An apropos example of an asymmetric series is the time-series of soil moisture at the daily time-scale which is characterized by a series of nearly vertical jumps followed by approximately exponential decay. This is a visually striking case of irreversibility. It is important to be mindful of the source of that asymmetry: precipitation. It is the intermittent character of precipitation that introduces a temporal asymmetry in the soil moisture series. In this respect we may again say that the irreversibility is a reflection of the causal link between rainfall and soil moisture.

While we have provided a measure of irreversibility, we have not yet probed the precise link between this measure and dynamical systems. What, for example, is the simplest dynamical system that uses only the Wiener process in its stochastic components that displays temporal asymmetry in its dynamics? Is there one? If so, how does the degree of irreversibility in the dynamics of a single variable of the system relate to the dynamics of the whole? This latter question reflects the fact that in environmental research we generally measure a single variable within a much larger, coupled system (and that variable still may not be a state-variable). In the future, therefore, we require some theoretical account of how irreversibility may be generated in these simple stochastic models, and, conversely, what the measurement of irreversibility in an environmental series implies about the underlying dynamics. This is beyond the scope of this dissertation but is certainly a line of research for the future.
Bibliography


Biography

James Rigby was born in Jackson, MS on March 23, 1980. He earned a Bachelor of Science degree in Physics from the University of Mississippi in 2003 graduating summa cum laude. In 2004 he attended Oxford University (UK) in the School of Geography where he worked in a geochronology research group studying the methodology of optical luminescence dating. In 2005 he received an M.Sc. in Hydrology for Environmental Management from Imperial College London (UK) graduating with Distinction and receiving the Victor Appleby Prize for his thesis on a coupled model of soil moisture-biomass dynamics in water-limited ecosystems.

Published Papers

