



Norwegian University of  
Science and Technology

# Maximum Entropy and Maximum Entropy Production in Macroecology

**Ida Andrea Braathen Sognnæs**

Master of Science in Physics and Mathematics

Submission date: April 2011

Supervisor: Asle Sudbø, IFY

Co-supervisor: John Harte, UC Berkeley, USA







## Abstract

The Maximum Entropy Theory of Ecology (METE), developed by John Harte, presents an entirely new method of making inferences in ecology [1]. The method is based on the established mathematical procedure of Maximum Information Entropy (MaxEnt), developed by Edwin T. Jaynes, and is used to derive a range of important relationships in macroecology [2]. The Maximum Entropy Production (MEP) principle is a more recent theory. This principle was used by Paltridge to successfully predict the climate on Earth in 1975 [3]. It has been suggested that this principle can be used for predicting the evolution of ecosystems over time in the framework of METE. This idea is at the very frontier of Harte's theory. This thesis investigates the hypothesis that the information entropy defined in METE is described by the MEP principle.

I show that the application of the MEP principle to the information entropy in METE leads to a range of conceptual and mathematical difficulties. I show that the initial hypothesis alone cannot predict the time rate of change, but that it does predict that the number of individual organisms and the total metabolic rate of an ecosystem will continue to grow indefinitely, whereas the number of species will approach one.

I also conduct a thorough review of the MEP literature and discuss the possibility of an application of the MEP principle to METE based on analogies. I also study a proof of the MEP principle published by Dewar in 2003 and 2005 in order to investigate the possibility of an application based on first principles [4, 5]. I conclude that the MEP principle has a low probability of success if applied directly to the information entropy in METE.

One of the most central relationships derived in METE is the expected number of species in a plot of area  $A$ . I conduct a numerical simulation in order to study the variance of the actual number of species in a collection of plots. I then suggest two methods to be used for comparison between predictions and observations in METE.

I also conduct a numerical study of selected stability properties of Paltridge's climate model and conclude that none of these can explain the observed MEP state in nature.



## Preface

This project has been carried out under the supervision of John Harte at UC Berkeley in Fall 2010 and beginning of Spring 2011. Harte is a professor in the Energy and Resources group at Berkeley, an interdisciplinary graduate program performing studies in areas of clean energy, climate science, ecosystems and biodiversity, energy systems, international development, technology and society, and water policy. The group aims at placing specialized knowledge into a larger integrated perspective guided by the goal of a sustainable environment and a just society. Harte's background is from physics and he worked as an assistant professor in the field of theoretical particle physics for several years before he changed track and became one of the six core faculty members of the Energy and Resources Group in Berkeley. Today, Harte pioneers the field of theoretical ecology, working on the nature and causes of patterns in the distribution and abundance of species and how ecosystem responses to climate change may result in feedbacks to climate. Most recently, Harte has developed an entirely new approach to macroecology based on the principle of Maximum Information Entropy (MaxEnt). This work is being published in the book *Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics* coming out on Oxford Press in June 2011, which I have contributed to through my work.

The Maximum Entropy Theory of Ecology (METE), presented in Harte's book, describes a new method for inferring important ecological distributions from sparse data, based on the MaxEnt principle. More specifically, it derives expressions describing the distribution of energy among individual organisms and numbers of individuals within different species in an ecosystem. It also describes spatial distributions of species and individuals.

MaxEnt leads to predictions at a particular instance in time only. Hence, METE only makes static predictions. In the last chapter of his book Harte suggests a potentially powerful, yet unexplored idea: applying the principle of Maximum Entropy Production (MEP) as a method for arriving at a dynamical theory of ecology. This is what I was set out to do. My project is thus at the very frontier of METE.

At the outset of my work Harte provided a set of partial differential equations (PDE's) based on an initial understanding of the MEP principle. These equations were meant to give the time derivatives leading to the MEP state of an ecosystem. My project consisted of verifying the mathematical derivation of these equations, solving them and interpreting the results. The first step in the process was therefore to learn and understand METE. Second, I spent a significant amount of time studying PDE's and possible methods for solutions. It was not until after this I

discovered that the set of equations I was given could not actually be solved, even numerically. This puzzled me for a while before I realized that these equations did not even correspond to the maximum of the entropy production, in other words, they did not describe the MEP state. Because this was mostly an intuition to begin with, and Harte was convinced of both the solvability of the equations and that they expressed the correct maximum condition, it took me a while to believe that my own viewpoint was actually the correct one. At the end, I managed to solidify my argument and to convince him of my conclusions. At this point, my project came to a halt. It was apparent that the initial idea that formed the basis of my project did not hold water. Because my project was at the very frontier of the field (which in itself is tiny) there were no related alternative routes I could learn from or switch to and I was left with little to continue my work on. In the mean time, Harte suggested that I look into the stability issues of one model where the MEP principle had been successful. I spent some time doing this, but soon concluded that stability properties could not explain what we were looking for, namely an explanation of the prevalence of the MEP state. In light of the situation I decided to get to the bottom of the question regarding whether the MEP principle could be applied to METE at all, since this has never been rigorously demonstrated. I therefore studied numerous alternative applications and conducted a thorough literature review on the topic in search of a useful analogy. The complexity and the difficulty of these papers required me to study Bayesian statistics in general and the MaxEnt principle in particular in great depth, as well as non-equilibrium thermodynamics. I also had to refresh and advance my knowledge in equilibrium thermodynamics, statistical mechanics and atmospheric physics. After all of this, I came to the conclusion that none of the previous applications of the principle could provide a valid analogy to my work. I therefore had to look at the very proof of the MEP principle itself to investigate a possible application based on first principles. In total, the unexpected turn of my project required me to learn a huge amount of theory, read a large amount of background material, and apply a wide range of knowledge from statistics, math and physics. Due to the interdisciplinary nature of the Energy and Resources Group and the uniqueness of my topic, I had no guidance on the most challenging parts of my thesis.

In the very beginning I also spent some time investigating the statistical certainty of one of the most central predictions of METE. This involved the calculation of the variance of a long and difficult expression for a specific probability distribution. I developed a simulation procedure for this, which is described in this thesis. This project was only meant to constitute a smaller part of my thesis, and I left it when it turned out to require a background in statistics beyond my training and more time than I had available. I did, however, return to this topic after my in depth investigation of MEP and MaxEnt, and I have included the results of my simulations as well suggestions for future research on this topic.



Despite the challenges along the way and the frustration of being the only person working on a very difficult topic, I believe that my work has provided a very important contribution to the MaxEnt theory of ecology. The in depth analysis that I provide is necessary to get to the bottom of the MEP principle as applied to METE, and it is with disappointment that I conclude that the initial idea should be abandoned. My work on the MEP principle, however, has also shed light on METE in general and the interpretation of the information entropy as defined in the theory in particular. My thesis also sheds light on the MEP principle itself and its prospect for predicting the behavior of information entropy not only as defined in METE, but in general. Moreover, my work on the variance has provided specific suggestions for the further development of METE. My thesis is a significant body of work that provides both important background theory and a critical analysis and suggestions for future work that can be used by the graduate students working in Harte's lab and elsewhere. It has made a significant contribution to Harte's book *Maximum Entropy and Ecology*, which I have been given due credit for in the final textbook.

I would like to thank John Harte for inspiration and ideas and Justin Kitzes for a valuable discussion on the variance simulations. I would also like to thank my friend Jessica Goddard who have kindly helped me proofread the thesis.

Berkeley, April 27, 2011  
Ida Sognmæs



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Models</b>	<b>4</b>
2.1	Biology versus Physics . . . . .	5
2.2	Models in Ecology . . . . .	7
<b>3</b>	<b>The Maximum Entropy Principle (MaxEnt)</b>	<b>7</b>
3.1	Formulation of the Problem . . . . .	8
3.2	Shannons Information Entropy . . . . .	9
3.3	Solving the problem: MaxEnt as statistical inference . . . . .	11
3.4	Maximum Likelihood . . . . .	12
3.5	Jaynes's Philosophy . . . . .	13
3.6	Applications of MaxEnt . . . . .	15
<b>4</b>	<b>MaxEnt and Equilibrium Thermodynamics</b>	<b>16</b>
4.1	Conventional Equilibrium Thermodynamics and Statistical Mechanics	16
4.2	Connection to Information Entropy . . . . .	20
<b>5</b>	<b>The Maximum Entropy Theory of Ecology (METE)</b>	<b>23</b>
5.1	Basics of METE . . . . .	23
5.2	Metrics in Macroecology . . . . .	24
5.3	Deriving METE . . . . .	26
5.4	Resulting metrics . . . . .	29
5.5	Evaluation of the model . . . . .	30
<b>6</b>	<b>A study of variance</b>	<b>32</b>
6.1	Discussion . . . . .	43
<b>7</b>	<b>A Maximum Entropy Production Theory of Ecology?</b>	<b>45</b>
7.1	Initial Problem Formulation . . . . .	46
7.2	Solution to the Initial Problem . . . . .	47
7.3	Meaning of zero partial derivatives . . . . .	51
7.4	Maximum rate of entropy production, $dS_I/dt$ . . . . .	52
7.5	Entropy, $S_I$ . . . . .	55
7.6	Discussion . . . . .	59
<b>8</b>	<b>Non-Equilibrium Thermodynamics</b>	<b>61</b>
<b>9</b>	<b>Maximum Entropy Production</b>	<b>63</b>
9.1	Paltridge's climate model . . . . .	65
9.2	Other applications . . . . .	69

<b>10 A study of stability in Paltridge's models</b>	<b>71</b>
10.1 Theory of Stability . . . . .	71
10.2 Stability in Paltridge's model . . . . .	72
10.3 Discussion . . . . .	77
<b>11 Discussion of MEP analogies to METE</b>	<b>77</b>
<b>12 Dewar's proof</b>	<b>81</b>
12.1 Discussion . . . . .	88
<b>13 Conclusions</b>	<b>91</b>
<b>14 Suggestions for future work</b>	<b>94</b>
<b>References</b>	<b>99</b>
<b>A APPENDIX</b>	<b>100</b>





---

# 1 Introduction

Entropy and its production have been at the core of various ideological discussions concerning grand questions like the evolution of the world and the course of time. Several notable scientists, including Clausius, Boltzmann, Gibbs and Onsager dedicated their time to studying the concept [6]. The principle of Maximum Entropy (MaxEnt) was originally motivated by statistical physics, which attempts to relate macroscopic, measurable properties of physical systems to a description at the atomic or molecular level. Combining information theory with Bayesian logic in the 1950s, Edwin T. Jaynes showed that the principle can be used as a general method for statistical inference and can be applied to a wide variety of problems [2]. Since then, the method has been applied to an increasing variety of fields. In the book *Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics* coming out in June 2011, John Harte presents an entirely new application of the principle of MaxEnt to the field of macroecology. Macroecology is the study of diversity, abundance and distribution of individual organisms and their energy use in ecosystems across spatial and temporal scales. There are many existing theories and models that describe one or more ecological relationships independently of each other. The Maximum Entropy Theory of Ecology (METE), on the other hand, derives numerous macroecological relationships from one principle alone, namely the MaxEnt principle. Several of these relationships are crucial in the study of biodiversity and the evaluation of extinction risks for threatened species and they provide important tools for studying ecological responses in a time of considerable human impact on the biosphere.

METE predicts important ecosystem parameters at an instance in time. It does not, however, say anything about how ecosystems change over time. Unlike static models in ecology there are very few ecological theories explaining the dynamic behavior of ecosystems. A principle termed the Maximum Entropy Production principle (MEP) was shown by Garth W. Paltridge to predict the climate on Earth with a high degree of accuracy in 1975 [3]. The principle has since been used to predict and describe the behavior of complex systems in a wide range of fields [6]. The MEP principle enables predictions of macroscopic dynamic behavior without having to consider the detailed internal behavior of systems. Harte suggested that it could also be used to predict the dynamical behavior of ecosystems the way such systems are defined in METE. If this was successful it would be extremely powerful because it could be used to predict a range of relationships for which there are currently no alternative models.

As mentioned in the preface, my project ended up following a somewhat unexpected route. The thesis is therefore written in the order of natural progression, along the lines of discovery. It fluctuates between parts containing background the-

ory, applications and discussions and present points and conclusions along the way.

Due to the unique interdisciplinary focus of my thesis and the fact that it is not written in a well-defined field, I have included a significant amount of the background theory in order for the reader to be able to follow the applications and the analysis. I begin with a review of scientific models in general and a discussion of the difference between physics and ecology, in particular, in Chapter 2. Next, I describe the principle of Maximum Entropy (MaxEnt), which is at the very core of both METE and MEP. The derivation is based on Bayesian statistics, which belongs to a different school of thought than conventional frequentist statistics. It also uses the notion of information entropy, defined by Claude E. Shannon in 1948 [7]. In Chapter 4 I show the connection between information entropy and thermodynamic entropy, and how the Boltzmann distribution can be derived using the MaxEnt formalism. In Chapter 5 I show how METE is derived and include the results that are important for this thesis. This chapter also defines and explains the basic entities which will be used for the remainder of the thesis.

Chapter 6 contains the first application. One of the most important predictions of METE is a relationship expressing the expected number of species in a plot of area  $A$ . The purpose of Chapter 6 is to predict the variation of the number of species in different plots in order to determine whether the expected value is a good guess or not. This chapter also discusses general challenges regarding the comparison of predictions from METE with real observations. It suggest two new methods for handling these challenges.

At the outset of my Master's project I was given three partial differential equations that were based on an initial MEP hypothesis within the context of METE. My initial project consisted of solving these equations and interpreting the results. The calculations and the subsequent discussion of these equations (equations (7.4) - (7.6)) are presented in Chapter 7. I show that these equations do not answer the original question and that the initial MEP hypothesis is problematic. In order to better understand the principle itself I conducted a review of the literature on the subject, which is presented in Chapter 9. In order to understand the alternative applications of the principle in the literature, it was necessary to look into aspects of non-equilibrium thermodynamics, which are presented first, in Chapter 8.

One of the most important applications of the MEP principle was the climate study by Paltridge in 1975. In Chapter 10, I investigate aspects of the stability in Paltridge's model. Chapter 11 is dedicated to a discussion regarding the MEP principle based on the review in Chapter 9 and the model in Chapter 10. I argue that the initial application of the MEP principle to METE cannot be justified by an analogy alone. Another option is to base the application on first principles. In



---

Chapter 12 I present a proof of the MEP principle published by Dewar in 2003 and 2005 followed by my final discussion regarding the MEP principle and its application to information theory [4, 5]. In Chapter 13 I present the conclusion of my thesis, and in Chapter 14 I presents some possible ideas for the future.

The chapter on the variance, Chapter 6, and the stability in Paltridge in Chapter 10, are based on numerical calculations in MATLAB and provide quantitative results. A larger part of my thesis, however, is dedicated to the review of the theory behind MaxEnt and MEP as well as thermodynamics and statistical mechanics, and the subsequent discussions. Due to the complexity of the topic and the lack of any related examples in the literature, most of my analyses are highly conceptual, rather than quantitative.

The same symbols and variable names might be used do denote different quantities in different parts of the thesis. It should be clear, however, in each case what is meant by the notation.

## 2 Models

Jaynes wrote that "...the logic of science is universal; the same principle of reasoning that work in statistical mechanics will work as well in astronomy, geophysics, biology, medical diagnosis and economics." [8, page 2]

This thesis deals with the application of a method based on the Maximum Entropy (MaxEnt) principle to the field of ecology. The method has been successfully used in physics (statistical mechanics) but the idea of applying it to problems in ecology is new [1]. Whenever a method is introduced in a new field, there is a need to go through some fundamental questions related to similarities and differences between the present field and the field (physics) that has previously benefited from the method. This is a natural first step because an analysis at an early stage can reveal if there are basic differences indicating that the method will have a low probability of success, in spite of the similarities between the fields. Therefore, I have in the following included a review of models in general, and a discussion of the difference between physics and ecology in particular.

Physics, or "nature" in ancient Greek, is one of the oldest academic disciplines. Maxwell opens his book "Matter and Motion" from 1876 with: "Physical science is that department of knowledge which relates to the order of nature, or, in other words, to the regular succession of events" [9, p. 9]. Physics has to do with the position and movement of matter. According to Young and Freedman physics is first and foremost an experimental science. The central focus is to observe phenomena in nature and try to find patterns and principles relating them. When patterns are found, we call them physical theories or, when they are very well established and of broad use, physical laws or principles [10]. Physics again is only one subfield of the wider category of science, defined by The Merriam-Webster online dictionary as: "knowledge or a system of knowledge covering general truths or the operation of general laws especially as obtained and tested through scientific method" [11].

Models are of central importance in the practice of science. We use models to express relationships and phenomena in nature. Significant parts of scientific investigation are carried out on models rather than on reality itself because by studying a model we can discover features of and ascertain facts about the system the model seeks to represent. Models are usually not derived solely from theory, nor solely from empirical data, but from a combination of the two [12]. One type of model often used in physics is the idealized model. An idealization is a deliberate simplification of something complicated with the objective of making it more tractable. Frictionless planes, point masses, infinite velocities and isolated systems, are all well-know examples from physics [12]. Approximations, even though they are similar to idealizations, relate instead to the actual mathematical procedure of

computing an answer [12].

Another important set of tools in science are analogies. At the most basic level, two things are analogous if there are certain relevant similarities between them. It can be similarity in properties or resemblance of relations between parts of two systems. If two systems are analogous, we might be able to use a model derived for one as a description of the other. The point is that the similarities have to be relevant for the case we are investigating [12].

When building a scientific theory it is generally expected that it meets three criteria [1]:

- **Falsifiability:** Whereas a theory can never be proven, there should be clear criteria by which it can be disproven.
- **Comprehensiveness:** The predictions should be applicable across a wide variety of conditions and phenomena.
- **Parsimony.** A theory should be lean. It should explain a lot with a little, i.e. the ratio of distinct testable predictions to assumptions should be large.

A theory can meet these three criteria in very different ways, to different degrees, and there is no clear cut way to determine when a theory is good, and when it is bad. One theory might predict more, less accurately, and another one predict less, more accurately. Which theory is "best" depends on our research question.

## 2.1 Biology versus Physics

This thesis is concerned with a method for finding patterns and principles that relate phenomena in macroecology. Instead of dead matter, as in physics, macroecology deals with complex living organisms and communities. If these organisms are animals they even possess some degree of free will. Such systems cannot be fully described by looking at forces and energy only. Furthermore, ecosystems often exhibit strong interactions between both individuals and between species. Unlike certain physical systems, the behavior of an ecological system cannot be reduced to a superposition of the behavior of each organism alone.

I will look at some of the most important differences between physics and biology in this section. To clarify the distinction between the different subfields:

**Biology** : the science of life and of living organisms, including their structure, function, growth, origin, evolution, and distribution. It includes ecology as a subdivision [13].

**Ecology** : the study of living organisms in relation to their environment. Includes macroecology as a subdivision [13].

**Macroecology** : the study of diversity, abundance and distribution of organisms as well as energy use in organisms across spatial and temporal scales [14] [1].

Biologists, in contrast to physicists, tend to focus on the uniqueness of their objects of study. The special case is often considered more interesting than the general patterns. The goal is description rather than prediction. Physicists, on the other hand, tend to look for general principles, and ideal cases. One exceptions to this generalization is Darwin's theory of evolution, which says that despite the uniqueness of every individual organism, and every species, they are all guided by the same, universal principle: the principle of natural selection. Harte calls for a synthesis of the Newtonian and Darwinian worldview in order to expedite progress in the field of ecology [15].

A large part of the difference between physics and biology lies in the culture of the two fields, but there are also factual, fundamental differences. A number of reasons can be pointed out as to why it is difficult to treat ecosystems in the same way as we treat a non-living physical system [1]. For example

- No two ecosystems are alike. This greatly limits the possibility of making controlled experiments. A physicist on the other hand can use similar photons, or similar electrons, to conduct the same experiment in a controlled manner repeatedly. A physicist can therefore more easily control for and eliminate external variables and sources of variability.
- Ecosystems, and species, undergo change both from natural causes and from human activity and are therefore in a state of constant flux. Molecules have always been the same. Species can go extinct.
- Accurate ecological measurements, such as a complete census of the organisms that live in an ecosystem, are impossible to obtain. We never know quite how accurate our data collection is, and what might have been missed. Hence, there are no well-defined measures of error in macroecology.
- System boundaries are often ill-defined. An ecosystem is always affected by its surroundings to a greater or lesser extent.
- Conducting experiments on ecosystems at large spatial scales and over long time periods is not feasible. Whereas it is possible to obtain information at smaller scales, and over shorter time periods, information about big patterns are difficult to obtain.

All of the above factors complicate the study of ecology, but they do not mean that we cannot say anything general about ecosystems. Mathematical principles in biology and ecology are rare, but they do exist, as we will see in the next section.

## 2.2 Models in Ecology

In a theory of macroecology, we seek to explain or predict the distribution and abundance of individual organisms and species within an ecosystem in space and over time. A distinction is often made between mechanistic theories and statistical theories. A mechanistic theory of ecology can include any of the following major processes operating in nature: predation, competition, mutualism, commensalism, birth, death, migration, dispersal, speciation, disease and resistance to disease, a range of different reproductive strategies, organism behaviors and social dynamics, fluctuations in the environment and so on [1]. We can regard the individuals within a species as identical, or we can include trait differences into the model. An example of a simple mechanistic model is the Lotka-Volterra model which describes species interaction. Darwin's theory of evolution is also a mechanistic model. Among statistical models, we have Coleman's random placement model, which states that the organisms residing in a specific area are randomly situated. A negative binomial distribution, a fractal model, and a Poisson clustering model have also been suggested for this purpose [1]. A more familiar statistical model is the explanation of the Gaussian variation of height among children.

Yet another family of ecological theories are based on optimization principles. Extremum principles are well known from physics, where we often minimize the potential energy, the action integral, or some other quantity. In more complex systems, like ecosystems, we often want to optimize a set of functions, which tend to give conflicting results [1].

Harte notes that none of the existing models of macroecology provide a comprehensive framework of macroecology. The models usually depend on ad hoc statistical assumptions and need to be supplemented with empirical knowledge in order to answer the conservation questions posed in the beginning of this chapter [1]. The Maximum Entropy Theory of Ecology avoids these problems. We will see that this theory is also based on an extremum principle, but that the function that is maximized is completely different from existing extremal functions in ecology.

## 3 The Maximum Entropy Principle (MaxEnt)

A dice with 6 faces is tossed 1000 times. Let us assume that we know the average number of dots in all of these tosses is but that the number of dots in each toss is unknown. With this knowledge, we want to estimate the frequencies of each of

the six possible outcomes [2]. To this problem orthodox statistics has no answer. If the average is 3.5, most people, however, would have a strong intuition for the uniform distribution. This is usually justified by arguing that there is no reason to prefer any one outcome over another. If instead the average number of dots in the 1000 tosses is 4.5, the frequencies cannot be equal. What can we say about the frequency distribution in this case? This chapter introduces a principle, The Maximum Entropy Principle (MaxEnt), that allows us to make inferences in cases like this, where information is incomplete.

### 3.1 Formulation of the Problem

We have some quantity,  $x$ , capable of assuming discrete values  $x_i$  ( $i = 1, \dots, n$ ). Each value is associated with some unknown probability  $p_i$ . All we know is the expectation value of some function  $f(x)$

$$\langle f(x) \rangle = \sum_{i=1}^n p_i f(x_i) \quad (3.1)$$

We also assume the normalization condition

$$\sum p_i = 1 \quad (3.2)$$

We now ask: on the basis of the given information, what is the expectation value of some other function,  $g(x)$ ? The expectation value,  $\langle g(x) \rangle$ , can be calculated if we know the  $p_i$ 's, but the above information does not determine these. To determine  $n$  unknowns, we would need  $n$  equations, or  $(n-2)$  more equations than what we have. In the dice example, the two probability distributions  $\{\frac{1}{2}, 0, 0, 0, 0, \frac{1}{2}\}$  and  $\{\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\}$  both give an average of 3.5. A number of different sets of probabilities,  $\{p_i\}$ , can usually give the same  $\langle f(x) \rangle$ . Is there any criterion for picking one probability distribution over another when both satisfy the given information?

An early attempt at a criterion of choice was Laplace's "Principle of Insufficient Reason". In the dice example, this amounts to picking the uniform distribution,  $p_i = \frac{1}{6}$  for all  $i$ , because there is no reason to think otherwise (corresponding to most people's intuition). Unfortunately, in more complex cases, Laplace's principle appears quite arbitrary, and can be shown to generate paradoxes [2]. Presently, this way of formulating problems has been largely abandoned, and probability theory has developed in two different directions. The "objective" school of thought regards the probability of an event as an objective property of that event, verified by observation of actual frequencies [2]. The "subjective" school of thought regards probabilities as expressions of human ignorance [2]. It interprets the probability of an event as a formal expression of our expectation that the event will occur based on the available information. The goal in this school of thought is not to evaluate some physical probability, but to inform plausible conclusions in cases where there

is not enough information available to lead to certain conclusions. This is what is meant by expressing human ignorance. The test of a good subjective probability distribution is whether it correctly represents our state of knowledge regarding the value of the variable  $x$ . This knowledge is expressed as the probability of each of the possible  $x_i, p_i$ . The two schools yield identical mathematical formulations, but the concepts themselves refuse to be united. The subjectivist view stands as the broader one, since it is always possible to interpret frequency ratios as expressions of human ignorance [2]. Furthermore, the subjectivist will admit many questions as legitimate which the objectivist consider meaningless. The dice problem posed at the beginning of this section is one such example, and one therefore has to adopt a subjectivist view to even consider it.

### 3.2 Shannons Information Entropy

Our partial understanding of the processes which determine the value of  $x$  can be represented by assigning probabilities,  $p_i$  ( $i = 1, \dots, n$ ), corresponding to the different outcomes,  $x_i$ . For the case of the dice, when the average is 4.5, we assign higher probabilities to outcomes with more dots, and lower probabilities to outcomes with fewer dots. We want to do this in a way that does not assume more information than what is given by the average value. That is, we want to pick a distribution that does not imply more certainty about the outcome than we actually have. A mathematical expression for uncertainty was provided by Shannon with his development of information theory in 1948 [7]. This uncertainty measure is presented below.

We seek a quantity  $H(p_1, \dots, p_n)$  which measures in a unique way the amount of uncertainty represented by the probability distribution,  $\{p_i\}$ , such that increasing values of  $H$  corresponds to increasing uncertainty. Shannon shows that three conditions of consistency must determine the function  $H(p_1, \dots, p_n)$  [16]:

1.  $H(p_1, \dots, p_n)$  is a continuous function of the  $p_i$ 's.
2. If all  $p_i$  are equal ( $p_i = \frac{1}{n}$ ), the quantity  $A(n) = H(\frac{1}{n}, \dots, \frac{1}{n})$  is a monotonically increasing function of  $n$ . In other words, more possibilities means that we are more uncertain than when there are fewer possibilities [16].
3. The composition law. We can group the probabilities ( $p_1, \dots, p_n$ ) into probabilities of combined events ( $w_1, \dots, w_r$ ) where  $r < n$ , such that  $w_1 = (p_1 + \dots + p_k)$ ,  $w_2 = (p_{k+1} + \dots + p_{k+m})$  and so on. The amount of uncertainty in the composite event is then  $H(w_1, \dots, w_r)$ . This will be different from  $H(p_1, \dots, p_n)$ . The conditional probabilities of the events,  $(x_1, \dots, x_k)$ , when the first composite event,  $w_1$ , takes place are  $(\frac{p_1}{w_1}, \dots, \frac{p_k}{w_1})$ . We can use the conditional probabilities to express  $H(p_1, \dots, p_n)$  with respect to the composite events as:

$$H(p_1, \dots, p_n) = H(w_1, \dots, w_r) + w_1 H\left(\frac{p_1}{w_1}, \dots, \frac{p_k}{w_1}\right) + \dots + w_r H\left(\frac{p_1}{w_r}, \dots, \frac{p_k}{w_r}\right) \quad (3.3)$$

First, we encounter the uncertainty,  $H(w_1, \dots, w_r)$ , related to the probabilities of the groups of events  $(w_1, \dots, w_r)$ . Next, we encounter an additional uncertainty with the probability  $w_1$  if this is where  $x_i$  belong,  $w_2$  if this is where  $x_i$  belong, and so on. The point is that it should not matter how the choices are broken down. We must obtain the same uncertainty for a distribution  $\{p_i\}$  independently of whether we express this directly or via the conditional probability of the groups,  $(w_1, \dots, w_r)$ , and their uncertainty. Equation (3.3) tells us that the uncertainties are additive [16]. If we go from the  $w_i$ 's to the  $p_i$ 's we increase our uncertainty because we have included more choices [17].

From these conditions Shannon arrives at the unique function

$$H(p_1, \dots, p_n) = -K \sum_i p_i \ln p_i \quad (3.4)$$

where  $K$  is a positive constant [7]. The function  $H$  is called the *entropy* or the *information entropy* of the distribution  $\{p_i\}$  [7]. It is the unique function satisfying the imposed conditions (1-3) [16]. It is important to distinguish between information entropy, which is a property of any probability distribution, and the experimental entropy of thermodynamics, which is a property of a thermodynamic state [16]. We will see later, however, that the concepts are closely related and that thermodynamic entropy is in fact the information entropy of a particular probability distribution.

By the definition of a probability function, all probabilities have values between 0 and 1, ( $0 \leq p_i \leq 1$ ). Each term in the sum of expression (3.4) is therefore negative and consequently,  $H \geq 0$ . The minimum possible value of the information entropy corresponds to complete certainty and has the value zero. This is obtained when the probability of one event is 1 and the probability of all the others are 0. The maximum entropy corresponds to maximum uncertainty and is obtained when all probabilities are identical ( $p_i = \frac{1}{n}$ ). In the dice example, the distribution  $\{\frac{1}{2}, 0, 0, 0, 0, \frac{1}{2}\}$  has a significantly lower information entropy ( $H = 0.69K$ ) than the distribution  $\{\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\}$  ( $H = \max = 1.79K$ ). Both distributions give rise to an average of 3.5, but the last distribution expresses a higher uncertainty than the former.

It can be useful to think of the expression in (3.4) in a slightly different manner. If we rewrite equation (3.4) as  $H(p_1, \dots, p_n) = K \sum_i p_i \ln(1/p_i)$  we see that it represents a weighted average over  $\ln(1/p_i)$ . The quantity  $(1/p_i)$  can be interpreted as the amount of "surprise" in an event  $x_i$ . If the probability of an event is low, the



surprise is large if that event actually occurs. Consequently, a lot of information is conveyed by telling you that  $x_i$  happened. Thus, information entropy is essentially a measure of the "average surprise" or "average uncertainty" of a random variable [18]. A more flat distribution, for which  $H$  is large, has a high "average uncertainty", whereas a more peaked distribution, with small  $H$ , has a small "average uncertainty". We are more likely to guess the actual outcome in the latter case than in the first case.

### 3.3 Solving the problem: MaxEnt as statistical inference

Shannon provided a unique, unambiguous criterion for the "amount of uncertainty" represented by a probability distribution. The measure,  $H$ , increases with more options and is additive for independent sources of uncertainty. It agrees with our intuitive notions that a flat distribution represents more uncertainty than a sharply peaked one. We now have a means of solving the problem stated in section 3.1. To infer a probability distribution from incomplete information we choose the probability distribution which has maximum information entropy subject to that information. In other words, the probability distribution with the largest value for  $H$  of the ones that result in the known average value (the information). This is the most unbiased probability assignment we can make, because it is based solely on the information given and makes no additional assumptions not supported by that information [2]. If we chose any distribution with a lower entropy than the maximum possible, this would mean that we favor some outcomes over others without this being supported by the information (they all comply with this already).

The mathematical procedure amounts to maximizing (3.4) subject to (3.1) and (3.2). From now on,  $K$  will be set to 1. The maximization is carried out by use of Lagrange multipliers  $\lambda_i$ , the details of which are shown by Jaynes [2]. For a general problem, with  $m$  constraints:

$$\langle f_r(x) \rangle = \sum_i p_i f_r(x_i) = F_r \quad (3.5)$$

where  $r = 1, \dots, m$ , the result is

$$p_i = \frac{1}{Z} \exp(-[\lambda_1 f_1(x_i) + \dots + \lambda_m f_m(x_i)]) \quad (3.6)$$

where

$$Z(\lambda_1, \dots, \lambda_m) = \sum_i \exp(-[\lambda_1 f_1(x_i) + \dots + \lambda_m f_m(x_i)]) \quad (3.7)$$

The constraints in equation (3.5) do not have to take the form of sample averages.  $F_r$  can be any data whatsoever, including single measurements. Equation (3.5) simply means that  $p_i$  must agree with the available data in the sense that the

measured value of the constraint,  $F_r$ , is recoverable from  $p_i$  as the expectation value  $\langle f_r \rangle$ . The expectation value,  $\langle f_r \rangle$ , is then the estimator of  $f_r$  with minimum expected square error [5].

The Lagrange multipliers,  $\lambda_r$ , are constants that can be determined by substituting (3.6) into the constraint and normalization condition (3.5) and (3.2). This yields the result

$$\langle f_r(x) \rangle = -\frac{\partial}{\partial \lambda_r} \ln Z, \quad (3.8)$$

The entropy of the probability distribution (3.6), which by definition is the maximum entropy distribution, is found by substituting (3.6) into (3.4), which yields

$$H_{max} = \ln Z + \lambda_1 \langle f_1(x) \rangle + \dots + \lambda_m \langle f_m(x) \rangle \quad (3.9)$$

The maximum information entropy,  $H_{max}$ , is uniquely defined by the set of Lagrange multipliers,  $\lambda_r$ , or equivalently, by the set of constraints,  $F_r$ . It is a matter of convenience whether we choose  $\mathbf{F} = (F_1, \dots, F_m)$  or  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_r)$  to describe the imposed constraints. Equation (3.9) also gives [19]:

$$\lambda_r = \frac{\partial H_{max}}{\partial \langle f_r(x) \rangle} \quad (3.10)$$

The Lagrange multipliers have an important meaning. Not only do they define the probability distribution,  $p_i$ , and the value of the maximum information entropy,  $H_{max}$ , they also show how the value of the information entropy changes in the neighborhood surrounding  $\mathbf{F}$ . They measure the importance of the constraints. In any variational problem, adding a redundant constraint cannot change the solution. Any redundant constraints we might include in the statement of the MaxEnt problem will drop out automatically because it will give a Lagrange multiplier equal to zero [19]. A highly relevant constraint (corresponding to highly relevant information) will give a large  $\lambda_r$ .

The MaxEnt procedure can be regarded as a more rigorous extension of the principle of insufficient reason presented by LaPlace, where the probability distribution is uniquely determined as the least biased one. It assigns positive probabilities to every situation that is not excluded by the given information. Proofs and derivations of the principle can be found in Jaynes' papers [2, 19, 16].

### 3.4 Maximum Likelihood

We have established the favored status of the MaxEnt probability distribution as the least biased distribution consistent with the given information. The question

remains now to determine in what sense and how strongly distributions of lower entropy are ruled out.

The rationale of choosing the most uniform distribution was understood already by Bernoulli and Laplace [19]. They calculated multiplicities, such as the multinomial coefficient

$$W = \frac{N!}{N_1!N_2!\dots N_n} \quad (3.11)$$

which gives the number of ways of distributing  $N$  distinguishable elements into  $n$  categories, each with  $N_i$  elements, ( $i = 1, \dots, n$ ). The coefficient was used to find the distributions,  $\{N_i\}$ , that could be realized in the largest number of ways. The Stirling approximation reads

$$\lim_{N \rightarrow \infty} \frac{\ln W}{N} = - \sum_i \left(\frac{N_i}{N}\right) \ln\left(\frac{N_i}{N}\right) \quad (3.12)$$

The right hand side of equation (3.12) is in the Shannon entropy form. Even though we have equality only in the limit,  $N \rightarrow \infty$ , this relationship shows the connection between entropy and multiplicity. Distributions of higher multiplicity have a larger entropy. This means that they can be realized in more ways. The connection is important for the Maximum Entropy Principle, because it provides a justification for why MaxEnt probability distributions are more likely to be observed than other distributions. When  $N$  becomes very large the relative preference,  $W_2/W_1 \sim \exp[N(H_2 - H_1)]$ , becomes so large that exceptions to the MaxEnt distribution are (practically) never seen [19]. In thermodynamics,  $N$  is on the order of Avogadro's number ( $6 \times 10^{23}$ ), and  $W_2/W_1$  becomes extremely large. This is why we call the Second Law of Thermodynamics a law (violations are still possible, but they are extremely unlikely).

Jaynes also presented the Entropy Concentration Theorem in order to show that the majority of the distributions allowed by our constraints have entropies near the maximum value [16]. This means that most possible probability distributions are arbitrarily close to the MaxEnt distributions. Jaynes therefore notes that to choose a distribution with an entropy away from the maximum value would amount to ignoring the vast majority of all the possibilities allowed by the data and concentrate our attention instead to a small and unrepresentative subclass of them [16].

### 3.5 Jaynes's Philosophy

The MaxEnt probability distribution agrees with everything that we know and avoids assuming anything that we do not know. It is the best prediction we are

able to make based on our incomplete information [8]. The main goal of the MaxEnt procedure is to predict one or more unknown macroscopic quantities  $\langle g(x) \rangle$  from one or more known macroscopic quantities (given in the form of constraints). Inferring the  $p_i$ 's is a step along the way in order to relate known macroscopic quantities with unknown macroscopic quantities. The unknown macroscopic quantities are calculated as

$$\langle g(x) \rangle = \sum_i p_i g(x_i) \quad (3.13)$$

This represents the best guess to the value of  $G = \langle g(x) \rangle$  given the information that we have. In some cases, however, this is not a very good guess. Obviously, a flat probability distribution does not tell us much and Jaynes points out that the theory makes definite predictions only when, and to the extent that, it leads to sharp distributions [2]. Even then, it is necessary also for  $g(x_i)$  to be smooth over the domain of  $x$ . The overwhelming majority of the outcomes with considerable weight (larger values of  $p_i$ ) must in other words give rise to the same macroscopic behavior. Without this extra condition, statistical mechanics would have no experimental validity [2].

The MaxEnt distribution depends on the constraints,  $\mathbf{F}$ . It also depends on the prior information that determines the set of allowed outcomes and corresponding function values  $f_r(x)$  (e.g. quantum mechanics in statistical physics) [5]. Before the Maximum Entropy Principle can be used, the problem domain needs to be set up. In cases involving physical systems, this means that the various states in which the system can exist need to be identified, and the parameters involved in the constraints known. For instance, in statistical mechanics, the energy, electric charge, and other quantities associated with each of the quantum states is assumed known. If the problem is not set up correctly, even a peaked MaxEnt distribution will not predict experimental results correctly. This, however, is one of the most useful aspects of the MaxEnt theory. Experimental evidence that a definite prediction is incorrect, does, according to Jaynes, give evidence of the existence of new laws of nature. The failures of classical statistical mechanics for instance was solved by quantum theory [2].

The difference between the subjectivist and the objectivist schools of probability were introduced in section 3.1. The objectivist (or frequentist) viewpoint tells us that probability is an inherent property of the real world, equal to the sampling frequency. The subjectivist (or Bayesian) viewpoint states that probability is a property of our state of knowledge about the real world [20]. Both schools use probability theory, but where the objectivist school deals with frequencies and measurements, probabilities in the subjectivist school are interpreted as a form of extended logic [21]. We know that  $x$  will attain one of the values  $x_i$  and give each

of the different  $p_i$ 's a number between 0 and 1, with a higher number representing a greater belief that this event will happen, and a lower number representing a belief that this event will probably not happen. When the outcome is learned, these parameters can be adjusted to 0 or 1, resulting in a zero uncertainty ( $H = 0$ ). We use probability as a means of coping with our lack of knowledge. We require that the probabilities obey the fundamental axioms of probability theory and we can apply all the common rules of probability theory to them [21]. The  $p_i$  however do depend on our state of knowledge and are therefore subjective. Since two observers may have different knowledge (they might have measured different quantities) this implies that the probabilities and all the quantities derived from them are observer dependent [21].

According to the subjectivist viewpoint, the probability distributions in statistical mechanics used by Maxwell, Boltzmann and Gibbs are not properties of nature but descriptions of incomplete human information about Nature. They yield the best predictions possible, which in this case are extremely good. Probability distributions in this sense are not "right" or "wrong", but some distributions are better predictions than others because they contain more relevant information. "Random" in the subjectivist view simply means "unknown" [22]. It is possible to imagine, for instance, that if we knew the angle, the height and the initial force vector in the toss of a dice, we could determine the outcome. Because these initial variables are unknown to us, we say instead that the outcome is random (at least for a fair dice).

### 3.6 Applications of MaxEnt

The first use of MaxEnt can be dated back to Laplace, who used the similar principle of indifference to infer a probability distribution for Saturn's mass [1]. Information Theory, as introduced by Shannon in the late 1940's, is used by engineers to design and analyze communication systems and by neuroscientist to quantify the amount of information conveyed by a neuron or a population of neurons [18]. The MaxEnt algorithm was developed by Jaynes as a general algorithm for predicting all the results of equilibrium thermodynamics. He also applied it to a range of other fields including image reconstruction, spectral analysis and inverse problems. Later, he applied the method to non-equilibrium statistical mechanics. Today, MaxEnt is a widely used variational method for the analysis of both complex equilibrium and non-equilibrium systems. It is being increasingly employed in a variety of fields such as nuclear magnetic resonance spectroscopy, x-ray diffraction, electron microscopy, image reconstruction, ecology and even economics to infer from incomplete data sets [23] [1].

In the next chapter I provide a review of MaxEnt in equilibrium thermodynamics.

This chapter shows the important connection between the MaxEnt formalism and statistical mechanics. In Chapter 5 I will present how MaxEnt can be applied to macroecology in the form of METE, providing enormously useful predictions about the state of ecosystems.

## 4 MaxEnt and Equilibrium Thermodynamics

Harte presents the MaxEnt theory of Ecology in *Maximum Entropy and Ecology* as an analogy to thermodynamics. I have mentioned earlier how thermodynamics can be derived using the MaxEnt principle. In order to say something about how far we can take this analogy and to better understand the application of MaxEnt from a physics perspective, I here include a review of the maximum entropy method in equilibrium thermodynamics. This will clarify the distinction and the connection between information entropy and thermodynamic entropy, which will become important in the analysis of the Maximum Entropy Production principle.

### 4.1 Conventional Equilibrium Thermodynamics and Statistical Mechanics

The systematic study of macroscopic behavior started in the 19th century, and the laws discovered then formed the subject of "thermodynamics". Clausius defined the thermodynamic, empirical entropy,  $S$ , via the identity  $dS = \delta Q/T$  circa 1850. In the second half of the century, the theory of the atomic constitution of all matter gained acceptance and macroscopic systems began to be analyzed from a more microscopic point of view [24]. Around the turn of the century, statistical mechanics was developed [24]. In the 1870s, Boltzmann interpreted Clausius entropy as the logarithm of the number of microstates,  $W$ , in which a given macrostate can be realized,  $S = k_B \ln W$ , where  $k_B$  is Boltzmann's constant. The second law of thermodynamics could thus be interpreted as stating that the observed macrostate is the most probable one, i.e. the one consistent with the largest number of microstates. In 1902 Gibbs managed to get to Boltzmann's results by minimizing the quantity  $\sum_r P_r \ln P_r$  with respect to the microstate probabilities  $P_r$  and subject to the appropriate constraints on energy and particle number for a closed thermodynamic system. Gibbs called this quantity,  $\langle \ln P_r \rangle$ , the "average index of probability of phase" [20]. In 1948 Shannon introduced the notion of information entropy as a measure of the amount of uncertainty associated with a probability distribution [7]. In the 1950's Jaynes made the connection between Shannon's information entropy, Gibbs's algorithm and Boltzmann's work. He showed that the thermodynamic entropy,  $S$ , is the information entropy of the distribution  $P_r$ . And from this point on, he realized that the MaxEnt algorithm is a general recipe for statistical inference that can be applied to any field, not just thermodynamics [20].

Entropy is closely linked to the reversibility of processes. The second law of thermodynamics states that entropy cannot decrease in an isolated thermodynamic system (no interaction with the surroundings). If an isolated system undergoes a process during which the entropy remains constant, this process is reversible. In any other case, the process will lead to an increase in entropy, and the process is irreversible. If the system on the other hand is open (can exchange both matter and energy with its surroundings), we can add energy from the outside and use this to decrease the entropy in the system. In this case, however, the entropy of the environment has to increase by at least the same amount, such that the entropy of the total system (the isolated system including both the system and its surrounding) does not decrease [1]. For any thermally insulated system undergoing a quasi-static process (i.e. an adiabatic process)  $\delta Q$  is zero by definition and Clausius relationship tells us that  $dS = 0$ . Thus, a quasi-static adiabatic process is reversible. If we slowly decrease the volume of a thermally insulated gas, the energy of the systems will change by the amount that we use to compress the gas, but the entropy will remain constant. We can retrieve the energy we used, by letting the gas expand slowly back to its original volume (for instance through potential energy of a spring). We will see below how reversibility is related to the number of states accessible to the system. If a thermally insulated system instead undergoes a non-quasi-static process, the entropy will in general increase, making the process irreversible. Mixing is an example of this [24]. In short, for reversible processes the entropy of the system *and* the surroundings remain constant, and the two can be restored to their initial states without loss of energy [25].

Thermodynamics is a macroscopic theory of matter. It mainly looks at the energy conversion between heat and mechanical work. Statistical mechanics on the other hand, derives the macroscopic properties by looking at the microscopic constituents. It provides a way of dealing with systems involving very large numbers of particles. The concept is simple. By applying statistical rules to the microscopic mechanics (classical and quantum) of the microscopic constituents, all the well-known general rules of the macroscopic properties of thermodynamics arise. Statistical mechanics demonstrates the statistical nature of all the macroscopic parameters of classical thermodynamics. Because we are applying statistics to numbers on the order of Avogadro's number ( $10^{23}$ ) the predictions are extremely accurate [24].

Essential to the analysis of statistical mechanics are the following components [24]:

1. Specification of the states of the system. It is necessary to enumerate the possible microstates, or equivalently the degrees of freedom,  $f$ , of the system under the given constraints. Classically, the microstate of the system is specified by identifying the momentum,  $p$ , and the coordinate,  $q$ , of each and every

particle in the system. For  $N$  particles, a microstate is usually represented as a point in  $6N$ -dimensional space, called state space or phase space and corresponding to  $6N$  degrees of freedom. Quantum mechanically, the specification of the state is equivalent to determining all the quantum numbers of the wave function. These specifications are complete because, in the classical case, the laws of classical mechanics are such that knowledge of  $q$  and  $p$  at any given time permits prediction of  $q$  and  $p$  at any later time. In the quantum case, knowledge of the wave function allows for calculation of all physical quantities as well as prediction of the state at all later times [24].

2. Statistical ensemble. Instead of focusing on single experiments and single outcomes, we analyze an ensemble of many identical systems. In a statistical description the representative ensemble contains all microstates that are consistent with the specified available knowledge about the system, i.e. the constraints (e.g. total energy). We can then calculate the probability of a particular outcome from this ensemble [24].
3. The basic postulate about *a priori* probabilities. This postulate says that an *isolated* system in equilibrium is equally likely to be in any of its accessible microstates. The validity of the postulate can be determined by making theoretical predictions based on the postulate and checking that these predictions are confirmed by experiments. A system will in the course of time make transitions between all its various accessible states as a result of small interactions between constituent particles (ergodic hypothesis) [24].
4. Probability calculations. Probability theory lets us calculate the probability of the outcome of an experiment as the fraction of all microstates with this outcome over all accessible states as well as expectation values [24].

A microstate refers to a complete determination of all the degrees of freedom of a system. A macrostate on the other hand is some kind of function of all the particles. An example is the total kinetic energy, which is the sum of each particles kinetic energy. If we swap the kinetic energy between two of the particles, the total energy is unchanged, resulting in the same macrostate for two different microstates. One macrostate usually corresponds to a large number of microstates. Boltzmann's expression,  $S = k_B \ln W$ , state that more likely macrostates have higher entropy, because they are compatible with more microstates,  $W$ . In statistical mechanics all the macroscopic properties are derived as statistical averages over the relevant ensemble. When macroscopic variables change it means that the respective ensembles change [24].

A probability distribution of fundamental importance in statistical mechanics is the Boltzmann distribution, or the canonical distribution. The canonical ensemble is an ensemble consisting of closed (can exchange heat, but not matter) systems



in contact with a heat reservoir. All the systems are taken to be in equilibrium with the heat reservoir. This means that each system has the same temperature. The total energy shared between the system and the reservoir is conserved and the average energy of all the systems is  $\bar{E} = \langle E \rangle$ , where the brackets indicate the average over the ensemble. We want to know the probability,  $P_r$ , of finding the system in any *one* microstate of energy  $E_r$ . Note that, in general, several  $E_r$  correspond to the same energy,  $E$ , such that the probability of finding the system in any one *energy state* is  $P_r$  times the number of states with this energy. By using the conservation of the total energy shared between the system and the heat reservoir, and some mathematical approximations, it can be shown that the probability of state  $r$  is [24]:

$$P_r = \frac{1}{Z} e^{-\beta E_r} \quad (4.1)$$

where  $\beta = 1/(kT)$ , where  $T$  is the temperature, and

$$Z = \sum_{r=1}^M e^{-\beta E_r} \quad (4.2)$$

such that the distribution is normalized. This is the canonical distribution and it tells us the probability that a closed system at temperature  $T$  is in one particular microstate,  $r$  [24]. The probability decreases rapidly with the energy of the system,  $E_r$ , because this corresponds to a lower energy for the heat reservoir, and thus fewer available states for the heat reservoir.  $\beta$  is a very large positive number, on the order of  $10^{20}$ . The probability of finding the system in any one energy state is

$$P(E) = \frac{1}{Z} \Omega(E) e^{-\beta E} \quad (4.3)$$

where  $\Omega(E)$  is the number of states in the system that has energy in the small range between  $E$  and  $E + \delta E$ . Because  $\Omega(E)$  is a rapidly increasing function of  $E$  (more rapidly the larger the closed system),  $P(E)$  is a highly peaked distribution. The larger the system is, the sharper the maximum of  $P(E)$  is. Average values over the canonical ensemble can be found by either summing the variable over  $P_r$  over all the states (where more states correspond to higher energies), or  $P(E)$  over all the energies [24]. The sharpness of the distribution assures that the variable we are averaging over will also be peaked.

The thermodynamic entropy can be expressed using the Clausius relationship and an expression for thermodynamic work in terms of  $P_r$  as [24]

$$S \equiv k_B (\ln Z + \beta \bar{E}) \quad (4.4)$$

## 4.2 Connection to Information Entropy

Looking closely at the expression for the thermodynamic entropy, (4.4), we notice that it looks very similar to the expression for maximum information entropy, (3.9), presented in section 3.3. In fact, if we set  $\lambda_1 = \beta$  and  $\langle f_1 \rangle = \bar{E}$  and we use  $K = k_B$  instead of  $K = 1$  in equation (3.4), the two expressions are equivalent. Jaynes worked out the connection between thermodynamic entropy and information entropy when he showed what the Gibbs algorithm meant by applying Shannon's work to Boltzmann's insight [20]. Jaynes proposed MaxEnt as a universal method for constructing the microscopic probability distributions of statistical mechanics, and demonstrated how all the results of both equilibrium and non-equilibrium statistical mechanics are derivable consequences of this principle [19].

An alternative derivation of the canonical distribution, clearly illustrating the relationship to Jaynes' maximum entropy method, is as follows. We have an ensemble of  $N$  systems, of which  $n_r$  are in state  $r$ . The constraint  $\sum_r n_r = N$  must then hold. The probability,  $P_r$ , that a system in the ensemble is in state  $r$  is then  $P_r = n_r/N$ . Using this, the previous constraint reduces to the normalization condition on  $P_r$ . We then impose one extra constraint on the ensemble, that the average energy is  $\bar{E}$ , which means that  $\sum_r P_r E_r = \langle E \rangle = \bar{E}$ , where  $E_r$  is the energy of a system in state  $r$ . The number of ways of putting a distinct ensemble in the different states,  $n_r$ , is given by the multiplicity of  $N$ ,  $W_{\{n_r\}} = N!/(n_1!n_2!\dots n_M!)$  for  $M$  states. From the Stirling approximation we know that maximizing  $W_{\{n_r\}}$  is equivalent to maximizing  $\sum_r P_r \ln P_r$  and the result of the maximization thus yields the MaxEnt distribution  $P_r = 1/Z e^{-\beta E_r}$ , which depends on the constraint,  $\bar{E}$ . As we see, this is equivalent to the canonical distribution (4.1) [24]. The identification of temperature, free energy, etc. can then be derived from this distribution. Among the results are  $\beta = 1/k_B T$  and  $S = -k_B \sum_r P_r \ln P_r$ , where  $K = k_B$  is the Boltzmann constant and  $T$  is the temperature [2]. The thermodynamic entropy is therefore identical to the information entropy of the canonical distribution, which is the MaxEnt distribution of a particular system for which the average energy is given.

The connection between the entropy,  $S = k(\ln Z + \beta \bar{E})$ , and the information theoretical expression for entropy can also be shown directly

$$\begin{aligned}
S &= k \left( \ln Z + \beta \bar{E} \right) \\
&= k \left( \ln Z + \beta \sum_r P_r E_r \right) \\
&= k \left( \ln Z - \sum_r P_r \ln (Z P_r) \right) \\
&= k \left( \ln Z - \ln Z \sum_r P_r - \sum_r P_r \ln P_r \right) \\
&= -k \sum_r P_r \ln P_r
\end{aligned}$$

where we have used that  $Z P_r = e^{-\beta E_r}$  and  $\sum_r P_r = 1$  [24].

The sharpness of the Boltzmann distribution with respect to  $E$  (not with respect to  $E_r$ ) assures that any one system under the given constraints is most likely to have the energy  $\bar{E}$ . No ergodicity or a priori equal probabilities are assumed in this derivation. It is simply shown using the MaxEnt algorithm that the overwhelming majority of all states will have this energy. It is only our best guess, but in this case that is an extremely good guess.

The relationship between  $S = k(\ln Z + \beta \bar{E})$  and the expression  $S = -k_B \ln W$  proposed by Boltzmann can be derived as follows. We have that  $Z = \sum_r e^{-\beta E_r} = \sum_E \Omega(E) e^{-\beta E}$ .  $W$  is the number of microstates with a certain macroscopic value for some macroscopic variable, in this case energy. Thus,  $W(E) = \Omega(E)$ . Since the summand is extremely peaked around the average value of the energy,  $\bar{E}$ , only the values in a narrow range  $\Delta \bar{E}$  make a significant contribution to the sum. We can therefore write  $Z = \Omega(\bar{E}) e^{-\beta \bar{E}} (\Delta \bar{E} / \delta E)$ , where the fraction is the number of intervals  $\delta E$  contained in  $\Delta \bar{E}$ . If we take the logarithm of the last expression, simplify it and combine it with the expression for  $S$  above, we arrive at  $S = -k_B \ln \Omega(\bar{E}) = -k_B \ln W$  [24].

The macroscopic state of a system is defined by specifying the external parameters of the system and any other conditions to which the system is subject [24]. The external parameters are those macroscopically measurable independent parameters (e.g. volume, strain tensor, gravitational potential) that are known to affect the value of the different energy levels of the system ( $E_r = E_r(\alpha_1, \alpha_2, \dots)$  where  $\alpha_i$  are the external parameters) [2]. Several connections can be made between the thermodynamic behavior and the canonical probability distribution via the information theoretical expression for the entropy. Since the entropy remains constant in a (quasi-static) adiabatic process, the  $P_r$ 's must also remain constant. A change in external parameters will change the total energy, because it changes the energy of each particle, but not the entropy, as long as the process is adiabatic and quasi-

static. Changes of this sort are reversible, and are referred to as work. If instead, heat is allowed to flow over the system boundary, the probability distribution itself can change. Heat is therefore represented by change in the  $P_r$ 's. The energy of each state is unchanged in this situation, and the change in total energy comes from the change in the average due to the change in the probability distribution. It is worth noting that reversible processes can also involve heat and therefore changes in entropy for the system, but only in the case where the combined entropy of the system and its environment remains constant. Change in energy for a system, in general, is due to a combination of work and heat ( $dE = \delta Q - \delta W$ ), where work is the amount of change in the mean energy of the system due to changes in the  $E_r$ 's and heat is the change in energy due to changes in the  $P_r$ 's [24].

Statistical mechanics before Jaynes's time was (and usually still is) constructed based on the equations of motion, supplemented by additional hypotheses of ergodicity, metric transitivity, or equal a priori probabilities (see the 4-point list in the previous section). The mathematical expressions concerning maximization of entropy in Gibbs's work are given status as side remarks not essential to the theory and the identification of entropy is usually made only at the end, by comparing the resulting equations with the laws of phenomenological thermodynamics. Jaynes, however, showed that the logic can be turned around and that it is possible to take entropy as the starting concept. The fact that the canonical probability distribution is the one giving the maximum of the entropy subject to the constraints is the essential fact justifying the use of this distribution for inference. The fundamental statistical postulate about a priori equal probabilities of each state in an isolated system leads to the same results as Jaynes's MaxEnt derivation, but the interpretation is different. In Jaynes's case the uniform probabilities are the MaxEnt distribution when no other information is given. Equal probabilities are simply our least biased guess. It is not a postulate about nature, but a consequence of our lack of knowledge [24]. The MaxEnt derivation of the canonical distribution and the results then show that the rules of statistical mechanics can be justified independently of any physical argument. In fact, Jaynes interpreted statistical mechanics as a form of statistical inference, rather than a physical theory. His method leads to a conceptual and mathematical simplification of the results and frees the theory from physical hypotheses. The principles and mathematical methods thus become available for treatment of many new physical problems [2].

The only place where "subjective" statistical mechanics (the MaxEnt method) makes contact with the laws of physics is in the enumeration of the different possible states in which the system might be. This lack of "physics" in the derivation makes it apparent how little content the argument really has. The reason why we are still able to make very accurate predictions based on statistical mechanics is the large number of degrees of freedom, caused by the large number of particles.

---

This leads to the sharpness of the canonical distribution, which again causes the probability distributions of the usual macroscopic quantities to possess a single extremely sharp peak (which is also why it doesn't matter whether we consider the median, the average, or the most probable estimate when calculating macroscopic variables: they all coincide).

Martyushev et al. points out that Jaynes's approach is viewed as the simplest and most convenient method for construction of equilibrium statistical (classical and quantum) thermodynamics. By using Jaynes' formalism it is possible to take the information principle as initial (instead of physical properties) and derive statistical physics from it [6]. In the next chapter, we will see how the same method, that has been used to successfully derived all the known results of thermodynamics, can be used in ecology.

## 5 The Maximum Entropy Theory of Ecology (METE)

Macroecology is highly relevant for conservation biology, which is aimed at preserving biological diversity in a world where ecosystems are affected both by natural events and human activity [26]. Some of the most fundamental questions conservation biologists seek the answers to are: a) how we can estimate species diversity at large scales from small-scale census data, b) how we can infer abundance from sparse presence/absence data, c) how we can estimate the number of species that will be lost under habitat loss and d) how to determine the most likely associations of habitat characteristics with species presence [1].

I will here describe the Maximum Entropy Theory of Ecology (METE) presented by John Harte in his book *Maximum Entropy and Ecology: A theory of Abundance, Distribution, and Energetics* coming out on Oxford University Press in June 2011. As a part of my work, I reviewed the draft of this book and provided conceptual and mathematical feedback that helped shape the final result. METE provides both the background for my work on the variance in Chapter 6, and the starting point for my work on the Maximum Entropy Production Principle (MEP) in ecology. I here explain what METE is and show how it provides answers to the questions in conservation biology presented above.

### 5.1 Basics of METE

METE is a simple theory. The point is not to give detailed predictions, but to predict the most central tendencies, similar to the way the equations of free fall give a good first prediction for the trajectory of a ball. The deviation of the actual trajectory from the prediction tells us how important other factors are, and if we want more accurate predictions, we can include friction. Just like the failure of the

ideal gas law lead to the understanding of dipole-dipole forces between molecules, failures of METE can be used to understand mechanisms in ecosystems better. Failure also drives science forward. METE in in a sense similar to a null hypothesis in statistics.

Analogous to thermodynamics, METE is defined by its state variables. State variables are properties of the system that need to be specified to implement the theory, but whose determination lies outside of the theory itself. The state variables in METE are:

$A_0$  : the area occupied by our ecosystem

$S_0$  : the number of species within this area

$N_0$  : the total number of individuals in all the species in this area

$E_0$  : the total metabolic rate for all the individuals.

Area,  $A_0$ , is chosen as the first state variable because it is the obvious measure of the physical scale of the system (analogous to volume in thermodynamics).  $S_0$  is chosen because of the central role that species richness plays in ecology and in macroecological metrics. Species here is not restricted to species in the usual sense (taxonomic species), but can mean any defined set of groups of individuals, like a family or a trait group. The number of species will vary greatly depending on the type of habitat. For instance, a rainforest has a lot more species than a desert, but we do not necessarily include all the species that exist in  $A_0$  into  $S_0$ . Sometimes we might only be interested in plants. Total abundance,  $N_0$ , and total metabolic energy rate,  $E_0$ , are chosen as the remaining two state variables because they scale additively and increase linearly with area. The individual organism and its energy requirement are also of fundamental importance in biology. The last two state variables share a close analogy to the number of molecules and the total internal energy in thermodynamic systems [27]. Individuals are used in the usual sense of the word, but it could perfectly well mean a cluster or another grouping of individuals. The number of species and the number of individuals are dimensionless, and the total metabolic rate has the dimension of power (energy over time). An ecosystem, in this thesis, is taken to mean a set of state variables,  $S_0$ ,  $N_0$ ,  $E_0$  and  $A_0$ .

With the definition of the state variables at hand, we can start looking at the important metrics of macroecology.

## 5.2 Metrics in Macroecology

A macroecologist acquires data about real ecosystems through censusing. The data is then analyzed in order to answer research questions. Observed data can be

used to produce graphs and obtain metrics. Metrics simply refer to the functions used to express relationships between data. Censusing requires a certain amount of resources, depending on the scale of the data collection. We are not able to collect data for instance, for the entire Amazon. It is therefore in our interest to be able to say as much as possible from as little data as possible. If the metrics obtained from smaller scales are believed to hold generally they can be used to say something about ecosystems at larger scales, for which we have no data.

I will present a subset of the metrics presented by Harte: the ones relevant to my work. In the next chapter, I show how these metrics are derived theoretically in the framework of METE.

The metrics are divided into two categories: species-level metrics and community-level metrics. Species-level metrics describe properties within one species, whereas community-level metrics describe properties of a collection of species. The ecosystem can be of any size, from a pond, to the entire Amazon. Each metric takes the form  $f(X|Y)$ , where  $X$  denotes the usual independent variable, and  $Y$  denotes the quantities that the value of  $f$  for each  $X$  is conditional on (which can also be treated as the independent variable). When  $f$  is a probability distribution we recognize this notation as the conventional notation for conditional probabilities.

#### **The Spatial-Abundance Distribution, $\Pi(n|A, n_0, A_0)$**

This is a species-level metric denoted by  $\Pi(n|A, n_0, A_0)$ . For a species that has abundance  $n_0$  in area  $A_0$  this metric tells us the probability that we will find  $n$  individuals of this species in the area  $A$  where  $A \leq A_0$ . The shape of a real spatial abundance distribution compared to the theoretical  $\Pi(n|A, n_0, A_0)$  can reveal whether individuals within a species are randomly placed, spread out, or clustered.

#### **The Species-Abundance Distribution (SAD), $\Phi(n|S_0, N_0)$**

In most ecosystems some species are rare, whereas others are abundant. For instance, you will usually find a lot more mice than bears in a forest. The distribution of common versus rare species is one of the most widely studied metrics in macroecology. The Species-Abundance Distribution (SAD) is a community-level metric. It is a probability distribution, denoted by  $\Phi(n|S_0, N_0)$ , telling us the probability that a species has abundance  $n$ . This is a community-level metric because it deals with the collection of a number of species. The probability distribution depends on the number of individuals,  $N_0$ , the number of species,  $S_0$ , and the area they inhabit,  $A_0$ . Note the different meaning of  $n$  here (total abundance of a species in  $A_0$ ), and in the previous metric (number of individuals in  $A$ ). The expected abundance must be  $N_0/S_0$  such that

$$\langle n \rangle = \sum_{n=1}^{N_0-S_0+1} \Phi(n)n = \frac{N_0}{S_0} \quad (5.1)$$

### The Species-Area Relationship (SAR), $\bar{S}(A|S_0, N_0, A_0)$

The Species-Area Relationship is a community-level metric denoted by  $\bar{S}(A|S_0, N_0, A_0)$ . Unlike the previous two metrics, this is not a probability distribution. The SAR tells us how many species we expect to find in an area  $A$  and the bar is used to denote average value. Typically,  $A_0$  is divided into half-plots such that ( $A = A_0, A_0/2, A_0/4, A_0/8, \dots$ ). All the smaller plots are therefore contained within the larger plots, and  $\bar{S}(A)$  is a monotonically increasing function of  $A$ . When computing empirical values for this relationship we count the number of species in each plot of a certain size (for example in the 8 plots of size  $A = A_0/8$ ) and compute the average. We then repeat this for each plot size. Different mechanisms (e.g. environmental conditions, competition, dispersal) can be used to explain the specific shape of the SAR when we compare actual distributions to theory.

The probability that a species with abundance  $n_0$  is present in an area  $A$  is equal to 1 minus the probability of absence. Using the Spatial-Abundance Distribution defined above this is:  $[1 - \Pi(0|A, n_0, A_0)]$ . The average number of species in  $A$  is then

$$\bar{S}(A) = \sum_{\text{species}} [1 - \Pi(0|A, n_0, A_0)] \quad (5.2)$$

i.e. the sum of the probabilities of presence for each species. This sum can only be taken if we know the abundances of all the species,  $\{n_0\}$ . More commonly we do not know the abundances and use instead

$$\bar{S}(A) = S_0 \sum_{n_0=1}^{N_0-S_0+1} [1 - \Pi(0|A, n_0, A_0)]\Phi(n_0) \quad (5.3)$$

where  $\Phi(n_0)$  is the Species-Abundance Relationship defined above.  $S_0\Phi(n_0)$  is the expected number of species with abundance  $n_0$ . This is multiplied by the probability of presence for a species with abundance  $n_0$  to find the total probability of presence for a species with this abundance. The sum is then taken over all the abundances to get the expected number of species in  $A$ .

### 5.3 Deriving METE

The Maximum Entropy Theory of Ecology (METE) represents an entirely new method of arriving at different ecological metrics, some of which are presented above. As the name reveals, the method is based on applying the principle of Maximum Entropy to ecology. This gives METE the advantages of saying a lot



using few variables, namely the state variables defined in section 5.1. As shown in section 3.5, we need to define the fundamental entities (e.g. gas molecules in thermodynamics) and events (microstates in thermodynamics) before we can apply the MaxEnt algorithm. The state variables,  $S_0$ ,  $N_0$ ,  $E_0$  and  $A_0$ , will act as our constraints. The only criteria for choosing the fundamental entities in a MaxEnt application is that they are unambiguously defined in a manner that allows specification of the numerical values of the constraints. There is no a priori way of justifying the choice of state variables or the fundamental entities, and the success of the choice is ultimately determined by comparison of the predictions with empirical data.

Central to METE are two different probability distributions derived from two separate MaxEnt applications. The first probability distribution,  $R(n, \varepsilon)$ , yields all the metrics describing the distribution of energy and abundances. The second distribution is the Spatial-Abundance Distribution,  $\Pi(n)$ , introduced in the previous section.

### The Ecosystem Structure Function, $R(\mathbf{n}, \varepsilon)$

$R(n, \varepsilon)$  is a joint probability function defined over the species and the individuals in  $A_0$ .  $R(n, \varepsilon)$  is discrete over the abundances,  $n$ , and continuous over the metabolic rates of individuals,  $\varepsilon$ . Note that  $n$  in this setting refers to total abundance of a species, not number of individuals. The Ecosystem Structure Function describes how abundances are distributed among species, and how metabolic rates are distributed among individuals. More specifically:

$R \cdot d\varepsilon$  is defined as the probability that if a species is picked at random, it has abundance  $n$ , and, that if an individual is picked at random *from that species* (with abundance  $n$ ), it has metabolic rate in the interval  $(\varepsilon, \varepsilon + d\varepsilon)$ .

The minimum metabolic rate is defined to be  $\varepsilon_{min} = 1$ . The normalization condition on  $R(n, \varepsilon)$  reads

$$\sum_{n=1}^{N_0-S_0+1} \int_{\varepsilon=1}^{E_0-N_0+1} d\varepsilon \cdot R(n, \varepsilon) = 1 \quad (5.4)$$

The upper limits in the sum and the integral are simply the maximum possible energy available to an individual, and the maximum number of individuals available to a species. The upper limit on  $n$  is  $N_0 - S_0 + 1$  because each species that is present has to have at least one individual (or else it is obviously not present). In most applications however,  $N_0 \gg S_0$ , and setting the limit to  $N_0$  yields a good approximation. The upper limit on the integral is  $E_0 - N_0 + 1$  because the minimum metabolic rate per individual is one, and hence the maximum energy available to one individual is the total minus the minimum needed by the other

individuals. In most cases,  $E_0 \gg N_0$ , and we can set the last limit to  $E_0$ . I will use these approximate limits in all of the following expressions.

The state variables,  $S_0$ ,  $N_0$  and  $E_0$  determine the constraints in the form of averages in our first maximum entropy application. The three state variables define two constraints, the average abundance per species  $N_0/S_0$ , and the average over species of the total metabolic rate  $E_0/S_0$ :

$$\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} d\varepsilon \cdot n \cdot R(n, \varepsilon) = \frac{N_0}{S_0} \quad (5.5)$$

$$\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} d\varepsilon \cdot n \cdot \varepsilon \cdot R(n, \varepsilon) = \frac{E_0}{S_0} \quad (5.6)$$

We assume nothing about how the individuals are distributed over the species (if the species tend to have similar abundances, or if some tend to have a very low abundance, and others a very high abundance) or how the energy is distributed among the individuals. The only information we have are the state variables,  $S_0$ ,  $N_0$  and  $E_0$ . From this incomplete information (in the sense that it is not enough to determine the actual distribution of energy or abundances among species), we seek to infer the shape of  $R(n, \varepsilon)$ . In order to do this we apply the MaxEnt algorithm and maximize the continuous form of the information entropy

$$I_R = - \sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} d\varepsilon \cdot R(n, \varepsilon) \cdot \ln(R(n, \varepsilon)) \quad (5.7)$$

subject to the constraints, (5.5) and (5.6), and the normalization condition (5.4).

The goal of METE is not simply to determine the distribution for  $R(n, \varepsilon)$  itself. Rather,  $R(n, \varepsilon)$  is used to derive several other distributions. The Species-Abundance distribution has more direct use than  $R(n, \varepsilon)$  and is found as

$$\Phi(n) = \int_{\varepsilon=1}^{E_0} d\varepsilon \cdot R(n, \varepsilon) \quad (5.8)$$

### The Species-Level Spatial Distribution $\Pi(n)$

The Species-Level Spatial Distribution,  $\Pi(n|A, n_0, A_0)$ , is derived from a separate application of the MaxEnt algorithm. Here,  $n$  denotes number of individuals in a species, not the abundances (as in  $\Phi(n)$ ). The normalization condition is

$$\sum_n^{n_0} \Pi(n) = 1 \quad (5.9)$$

where I have left out the conditional variables in the expression,  $A$ ,  $A_0$  and  $n_0$ . Only one of the state variables,  $A_0$ , are used in the derivation of this distribution, but we need to know the abundance of the species. The average number of individuals in area  $A$  must be proportional to the area,  $n_0 A/A_0$ , which gives us the constraint equation

$$\sum_n^{n_0} n \cdot \Pi(n) = n_0 \frac{A}{A_0} \quad (5.10)$$

In order to obtain the distribution,  $\Pi(n)$ , we maximize the information entropy

$$I_{\Pi} = - \sum_n^{n_0} \Pi(n) \ln(\Pi(n)) \quad (5.11)$$

subject to (5.9) and (5.10).

#### 5.4 Resulting metrics

Using the notation presented in section 3.3 in the derivation of  $R(n, \varepsilon)$  we have  $f_1 = n$ ,  $f_2 = n\varepsilon$ ,  $\langle f_1 \rangle = N_0/S_0$  and  $\langle f_2 \rangle = E_0/S_0$ . Substitution directly into equation (3.6) and (5.13) gives the results

$$R(n, \varepsilon) = \frac{1}{Z(\lambda_1, \lambda_2)} e^{-\lambda_1 n} e^{-\lambda_2 n \varepsilon} \quad (5.12)$$

and,

$$Z(\lambda_1, \lambda_2) = \sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} d\varepsilon \cdot e^{-\lambda_1 n} e^{-\lambda_2 n \varepsilon} \quad (5.13)$$

The Lagrange multipliers are expressed in their exact form by plugging (5.12) into the constraint equations (5.5) and (5.6). These expressions, however, are lengthy and cannot be solved for analytically. For most combinations of the state variables, and as long as  $S_0 > 4$ , the following expressions yield good approximations to the Lagrange multipliers [1]. The larger  $S_0$  is, the better the approximations are.

$$\lambda_2 \approx \frac{S_0}{E_0 - N_0} \quad (5.14)$$

$$\frac{S_0}{N_0} \approx \beta \ln \left( \frac{1}{\beta} \right) \quad (5.15)$$

$$Z \approx \frac{\ln \left( \frac{1}{\beta} \right)}{\lambda_2} \quad (5.16)$$

where  $\beta = \lambda_1 + \lambda_2$ . From the above results and approximations Harte derives the resulting Species-Abundance distribution [1]

$$\Phi(n) \approx \frac{1}{\ln\left(\frac{1}{\beta}\right)} \cdot \frac{e^{-\beta n}}{n} \quad (5.17)$$

For the Species-Level Spatial Abundance Distribution, with  $f_1 = n$ , and  $\langle f_1 \rangle = n_0 A/A_0$  we get [1]:

$$\Pi(n) = \frac{1}{Z_{\Pi}} e^{-\lambda_{\Pi} n} \quad (5.18)$$

where

$$Z_{\Pi} = \sum_{n=0}^{n_0} e^{-\lambda_{\Pi} n} = \frac{1 - e^{-\lambda_{\Pi}(n_0+1)}}{1 - e^{-\lambda_{\Pi}}} \quad (5.19)$$

In order to get the Lagrange multiplier,  $\lambda_{\Pi}$ , we insert (5.18) back into the constraint, (5.10). The Lagrange multiplier depends on  $n_0$ ,  $A$  and  $A_0$ , which gives the aforementioned dependence of  $\Pi(n)$  on these variables. In the special case where  $A = A_0/2$  we obtain  $\Pi(n) = 1/(1 + n_0)$ , telling us that any distribution of the individuals into half plots is equally likely [1].

## 5.5 Evaluation of the model

The goal of METE is to derive numerous macroecological metrics from one fundamental principle, the MaxEnt principle [27]. The MaxEnt algorithm is an application of logic. It is a rigorously proven mathematical procedure for inferring the most likely probability distribution when our knowledge about that distribution can be incorporated as a set of constraints on the distribution. Physical or biological processes are incorporated into METE only in the form of the constraint equations, which themselves make no assumptions about any mechanisms. Furthermore, there are no adjustable fitting parameters. The method itself is therefore absent of any known relationships in ecology. Yet, METE provides a theoretical framework predicting all the central tendencies for the entire range of metrics at once. Each metric alone might still be better predicted by alternative models, but the purpose of METE is to provide a unified, theoretical model of ecology [27].

The Spatial-Abundance Distribution  $\Pi(n)$  can be found for any species as long as we know its abundance  $n_0$  in  $A_0$ . This distribution is separate from the distributions derived from  $S_0$ ,  $N_0$  and  $E_0$  via  $R(n, \varepsilon)$  (including  $\Phi(n)$ ). The purpose of METE, however, is to be able to derive all the metrics from the state variables,  $S_0$ ,  $N_0$ ,  $E_0$  and  $A_0$ , alone, without knowledge of the abundances  $\{n_0\}$ . We can do this by using  $\Phi(n_0)$  to predict the abundances and then apply the Spatial-Abundance

Distribution,  $\Pi(n)$ , to these abundances, as in equation (5.3). The Species-Area Relationship,  $\bar{S}(A)$ , is then completely determined by the measured value of  $S_0$ ,  $N_0$  and  $E_0$  only. This is a unique feature of METE compared to other models in macroecology. For the predictions to be valid, both  $\Phi(n_0)$  and  $\Pi(n)$ , have to be valid [27]. Only when we are testing the predictions of  $\Pi(n)$  is it useful to know the actual abundances [27].

Of fundamental importance to the theory is the ability to make predictions at spatial scales for which we do not have data.  $\Phi(n_0)$  can be found at any spatial scale  $A$  finer than  $A_0$  by using  $N_A = N_0(A/A_0)$  and  $S_A = \bar{S}(A)$  as constraints. However, we are usually not interested in predictions at finer scales because we usually have the data for this if we have the data at the larger scale. This is only really used to test the predictions. What we are really interested in is to predict the state variables and the distributions at larger scales. This enables us to make predictions based on smaller censusing plots regarding the total surrounding area. An upscaling of the predictions based on Bayes' rule is possible for all our distributions. The specific technique is described in Harte's book [1].

The different ecological metrics are important because they help answer the central questions in conservation biology, a) - d) stated at the beginning of this chapter. The Species-Area Relationship can be used to scale up values from a smaller plot where we are capable of conducting a complete census, to a larger plot, thereby providing answers to question a), how to estimate species diversity at different scales. Diversity however, does not only regard presence or absence of species, but also their relative abundances. This can be obtained from the Spatial Abundance Distribution, which can also provide good answers to b), estimating abundances from sparse data. If a species has a very low abundance it is generally at a higher risk of extinction and if we combine the SAR and the SAD we can say something about how many species are at risk of extinction in a certain area, answering c). Whether a species is unique to the area of question is also important in order to measure the risk of extinction, and this is another metric that can be obtained from METE (not included here). The shape of the graphs of the different metrics, that is, how they deviate from predictions, can give us hints about d), the habitat characteristics.

In 2008, Harte et al. compared the various macroecological metrics derived from METE with observed data from a variety of spatially explicit vegetation data sets [27]. A vast amount of data is also compared to theory in the book, *Maximum Entropy and Ecology* [1]. It is shown that the equal-allocation prediction for half-plots match several data sets well and that the Species-Abundance Distributions show a good fit with all data [1]. Considering that no parameters are available to adjust the predicted SAR's, Harte et al. state that these predictions also fit quite

reasonably with data [27].

Since the MaxEnt procedure gives us only the best guess based on the information we have, the predictions obtained from it are only as good as the information itself. The power of the principle however, is that the logic can be turned around, and we can test the predictions against empirical data in order to let systematic discrepancies between predictions and data lead us onto unknown processes (this assumes of course, that the already included constraints are measured correctly). METE is a null theory because it makes falsifiable predictions assuming no explicit mechanisms and its value derives in part from the nature of its failures. Harte points out that failure in the predictions are expected for systems in rapid change, i.e. systems where the state variables change over relatively short time intervals. Also, the SAR's are likely to fail if the theory is used to scale up from a homogenous habitat, to a heterogenous habitat, because there will be more species if our area consist of several different habitats [1]. Harte et al. therefore suggest a large-scale heterogeneity as an additional constraint in METE [27].

Lastly, it is worth commenting on the choice of the Ecosystem Structure Function,  $R(n, \varepsilon)$ . The choice of state variables have their arguments in section 5.1 but the exact definition of  $R(n, \varepsilon)$  might seem somewhat arbitrary. The truth is that it is. Harte tested a range of different versions before arriving at this function, and found that this gave the best predictions. In other words, it is chosen because it works.

The next chapter is devoted to my investigation of the distribution of number of species.

## 6 A study of variance

The sharpness of the probability distributions arising in statistical mechanics is one of the reasons why it works as a description of nature. It is entirely possible that all the molecules in a room are found on one side of the room, but this is extremely unlikely - so unlikely that we call it impossible in our day to day language. The entropy concentration theorem states that the vast majority of possible distributions have entropies near the maximum value, hence, most distributions allowed by the constraints are very close to the MaxEnt distribution [28]. Consequently, it is extremely unlikely that we will find a system in a distribution very different from the MaxEnt distribution, which in the case of the gas molecules and the room, would be the even distribution of molecules over the entire room.

Under what conditions do we get such sharp distributions? In statistical mechanics, the extremely large number of particles, which is on the order of Avogadro's

---

number ( $\sim 10^{23}$ ) assures this. In our case we are dealing with a number of species,  $S_0$ , on the order of  $10^2$ - $10^4$  and the number of individuals,  $N_0$ , is on the order of  $10^3$  -  $10^5$ . These numbers are large, but certainly not as large as Avogadro's number.

In order to answer the question of how well METE works as a description of nature, we have to look more closely at the probability distributions predicted by the theory. One of the most important metrics obtained from METE is the Species-Area Relationship,  $\bar{S}(A)$ , which tells us the expected number of species in an area  $A$ . When  $A < A_0$  this metric is obtained from equation (5.2) if we know the abundances,  $\{n_0\}$ , and from equation (5.3) if we do not know the abundances. The metric can also be used to predict the number of species in a larger area for which we have no information by using the upscaling technique presented by Harte [1].

The Species-Area Relationship tells us the expected number of species in  $A$ , but it does not tell us the probabilities of finding a specific number of species in  $A$ . If we want to know whether the expectation value is a good prediction or not, we need to know what the probability distribution for finding  $s$  number of species in  $A$ ,  $P_A(s)$ , looks like. This probability distribution is unknown, but can be expressed implicitly as

$$\bar{S}(A) = \langle s \rangle = \sum_s s P_A(s) \quad (6.1)$$

where  $s \in (0, S_0)$  when  $A < A_0$ . We can find none or all the species in  $A$ , or any number in between. If  $P_A(s)$  is a flat distribution, the expected number of species is not a good prediction in the sense that we will often find a different number of species in  $A$  (but the expectation value will still give the lowest square error in a series of guesses). If  $P_A(s)$  is a peaked distribution on the other hand, the expectation value is a good guess because it is uncommon to find a different number of species than the expected value. A good measure for the certainty of finding the expected number of species is the variance of  $P_A(s)$

$$\sigma_S^2 = \sum_s (s - \langle s \rangle)^2 P_A(s) = \langle s^2 \rangle - \langle s \rangle^2 \quad (6.2)$$

In order to compute the variance we need to know the expression for  $P_A(s)$ . I will illustrate how this can be found with a simple example. Say we have two different species, 1 and 2, with abundances  $n_{0,1} = 1$  and  $n_{0,2} = 4$ , in  $A_0$ . We want to know the probability of finding  $s$  species in a quarter plot,  $A = A_0/4$ , that is,  $P_{A_0/4}(s)$ , where  $s \in (0, 2)$ . The probability of finding  $n$  individuals of species 1 in  $A$  is given by  $\Pi_1(n|A, n_{0,1}, A_0) = \Pi(n|1/4, 1, 1)$  and of species 2 in  $A$  by  $\Pi_2(n|A, n_{0,2}, A_0) = \Pi(n|1/4, 4, 1)$ . The distribution of the two species are independent of each other, that is,  $\Pi_1(n)$  is independent of  $\Pi_2(n)$ .

We start out with the case where both species are absent,  $s = 0$ . The probability of absence for species 1 is  $\Pi_1(0)$  and for species 2 it is  $\Pi_2(0)$ . Since the two are independent of each other, the probability that both species are absent is simply:

$$P(s = 0) = \Pi_1(0)\Pi_2(0) \quad (6.3)$$

The situation where  $s = 1$  can be satisfied in more than one way. Either species 1 is present, or species 2 is present. The probability that  $s = 1$  is therefore the sum of the probability that species 1 is present and 2 is absent, and that species 1 is absent and 2 is present. Whereas species 1 only has one way of being present (having its one individual in  $A$ ), species 2 has four ways of being present (it can have one, two, three or four individuals in  $A$ ). We get:

$$P(s = 1) = \Pi_1(1)\Pi_2(0) + \Pi_1(0)[\Pi_2(1) + \Pi_2(2) + \Pi_2(3) + \Pi_2(4)] \quad (6.4)$$

$$= \Pi_1(1)\Pi_2(0) + \Pi_1(0)[1 - \Pi_2(0)] \quad (6.5)$$

where we have used that the probability of presence for any species is  $[1 - \Pi(0)]$ . When  $s = 2$  both species have to be present. We have again one way for species 1 of being present, combined with 4 ways of species 2 being present. The result is:

$$P(s = 2) = \Pi_1(1)[1 - \Pi_2(0)] \quad (6.6)$$

The range of  $s$  and the number of combinations for each value increases with the number of species,  $S_0$ , but the logic is the same. It is just a matter of multiplying and adding probabilities of presence and absence for every species. As  $S_0$  increases however, the number of combinations of presence and absence in the range  $1 \leq s \leq S_0 - 1$  becomes very large very fast. For instance, if  $S_0 = 100$  and we want to know  $P_A(s = 50)$ , the number of different ways 50 out of 100 species can be present is given by the binomial coefficient,  $100!/50!50! \approx 10^{29}$ . If, in addition,  $\{n_0\}$  is unknown, we have to take into account every possible distribution of abundances. In the example above, where  $S_0 = 2$  and  $N_0 = 5$ , the only two possible abundance distributions are  $\{1, 4\}$  and  $\{2, 2\}$ , but the number of combinations becomes increasingly larger with more species and more individuals. If we didn't know the actual abundances we would have to multiply all the probabilities obtained above, with the probability of observing the distribution  $\{1, 4\}$ . This can be found using the Species-Abundance Distribution,  $\Phi(n)$ . We would then have to repeat the procedure for the distribution  $\{2, 2\}$ . We see how this gets very complicated, very fast, for large numbers of  $S_0$  and  $N_0$ .

In order to compare theory with data, I look at the case where the distribution of abundances,  $\{n_0\}$ , is known.  $\Pi(n)$  can be expressed in a simple form when



---

$A = A_0/2$ , but even in this case, deriving a compact expression for  $P_A(s)$  is either very difficult, or impossible. In order to study the probability distribution,  $P_A(s)$ , for all  $A < A_0$ , I chose instead a numerical procedure. The method is inspired by a discussion with Justin Kitzes, a graduate student at UC Berkeley, who has been working on similar calculations for other metrics in METE.

The method is based on random draws from the different  $\Pi_i(n)$ 's where  $i$  denote the species number. If we draw a sufficient amount of times, the frequency distribution of the different  $n$ 's will approach the theoretical distribution (by the central limit theorem). I looked at the case where  $A = A_0/256$  in order to compare with two available data sets, but the method can be applied for any  $A < A_0$ . The simulation procedure is as follows:

1. For species,  $i$ , draw randomly from  $\Pi_i(n) = \Pi(n|1/256, n_{0,i}, 1)$ .
2. Decide whether species  $i$  is present (assign 1) or absent (assign 0).
3. Repeat for all species,  $i = (1, \dots, S_0)$ , present in  $A_0$ .
4. Count the total number of species present in  $A$  by summing the assigned values.
5. Repeat steps 1. - 4. a sufficient number of times,  $T$  (explained below).
6. Compute  $\langle s \rangle$  and  $\sigma_s^2$  from the simulated values.
7. Plot the frequencies at which  $s$  species are found.

The simulations were done in MATLAB and the code can be found in APPENDIX A. The computations in step 6. are done by using equations (6.1) and (6.2) with  $P_A(s) = f(s)/T$ , where  $f(s)$  is the number of times we get  $s$  species in  $T$  repetitions.

The abundances used in the simulations correspond to the actual abundances observed in Little Blue Ridge, a californian serpentine grassland, in 1998 and 2005 [29]. The data sets describe the spatial distribution of over 37 000 plants covering a  $8\text{m} \times 8\text{m}$  plot. The number of individuals from each species was recorded for each of the 256 subplots of area  $A = 0.25\text{m}^2$ .

$\Pi(n)$  gives us the expected fraction of cells with  $n$  individuals. If we multiply this with  $A_0/A$  we get the expected number of cells with  $n$  individuals. The sum  $\sum (A_0/A)\Pi(n)n$  then gives the total number of individuals,  $n_0$  (as it should, from the constraint equation (5.10)). I have plotted six theoretical distributions,  $(A_0/A)\Pi(n|1, n_0, 256)$ , for the abundances  $\{10792, 6990, 1418, 112, 50, 1\}$ , together with the actual values found in the real data for these species in Figure 1 and 2 below. The vertical axis tells us the predicted and observed number of cells with  $n$  individuals, which is given on the horizontal axis. We see that the central trend of the predictions (an exponential decrease) are also found in the real data. However, there are a few obstacles to directly comparing  $(A_0/A)\Pi(n)$  with the

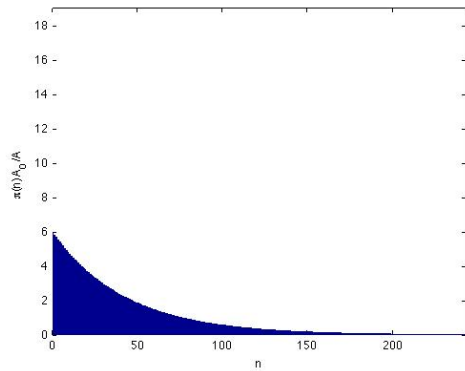
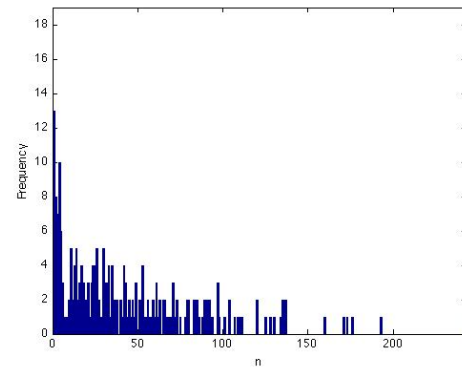
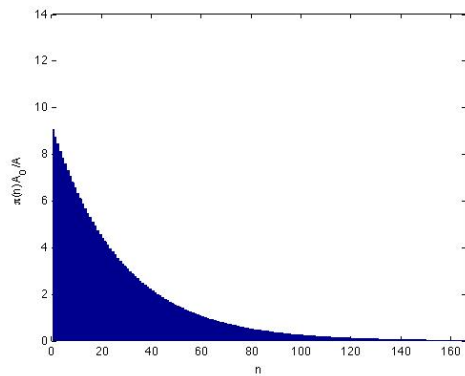
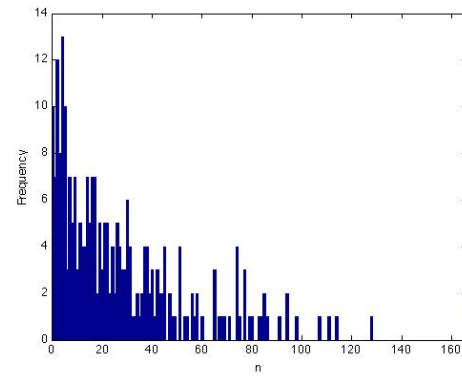
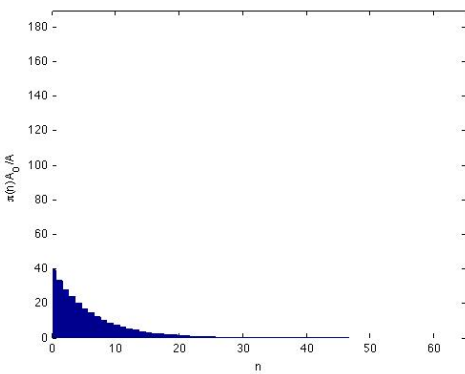
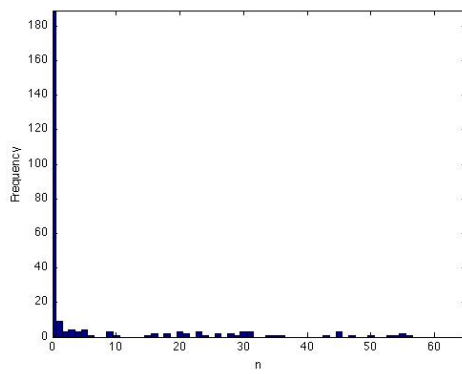
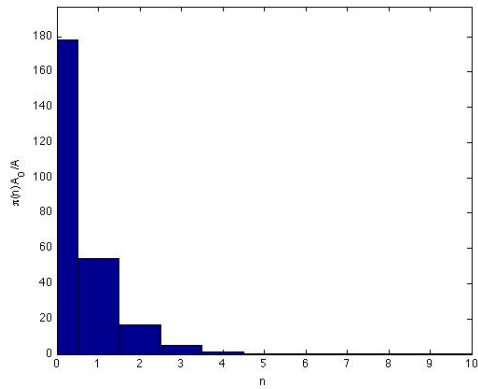
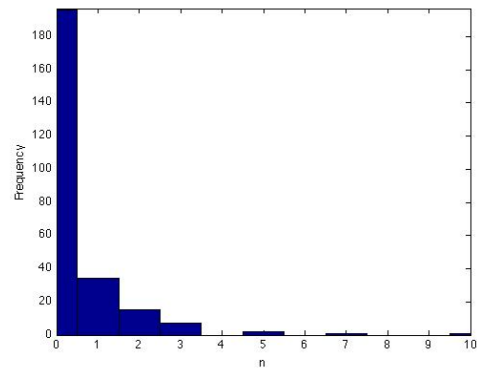
(a)  $n_0 = 10792$ , predicted(b)  $n_0 = 10792$ , observed(c)  $n_0 = 6990$ , predicted(d)  $n_0 = 6990$ , observed(e)  $n_0 = 1418$ , predicted(f)  $n_0 = 1418$ , observed

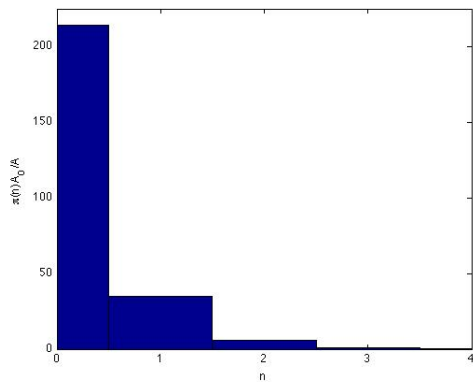
Figure 1: Predicted number of cells with  $n$  individuals,  $\Pi(n)A_0/A$ , on the left and observed number of cells with  $n$  individuals on the right. Data from Little Blue Ridge, 1998 [29].



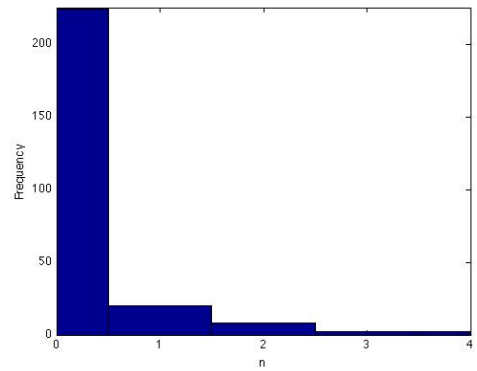
(a)  $n_0 = 112$ , predicted



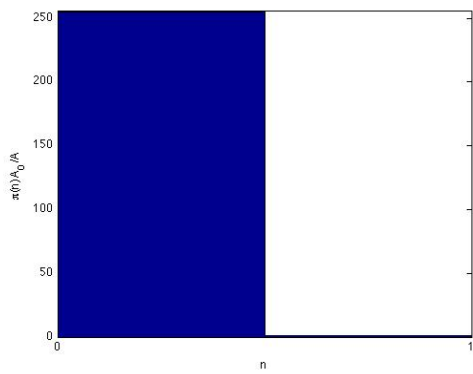
(b)  $n_0 = 112$ , observed



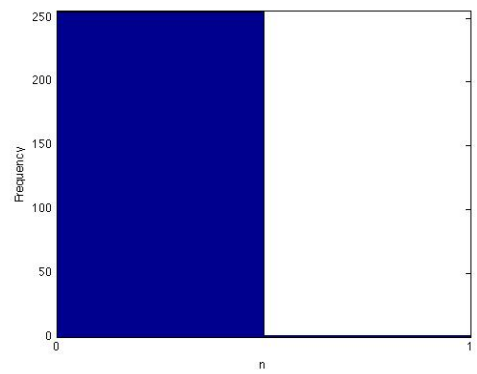
(c)  $n_0 = 50$ , predicted



(d)  $n_0 = 50$ , observed



(e)  $n_0 = 1$ , predicted



(f)  $n_0 = 1$ , observed

Figure 2: Predicted number of cells with  $n$  individuals,  $\Pi(n)A_0/A$ , on the left and observed number of cells with  $n$  individuals on the right. Data from Little Blue Ridge, 1998 [29].

observed data. In the real data sets we have a discrete number of cells for each  $n$ .  $\Pi(n)$  however, is a continuous function and when we multiply it with the number of cells of area  $A$ ,  $(A_0/A)$ , we do not (in general) get a discrete number of cells for each value of  $n$ . This can be seen in Figure 1(a), where most of the predicted number of cells lie between the discrete numbers on the vertical axis. Obviously, we will not find 4.5 cells with 10 individuals, which is what is predicted. The real data will therefore never look exactly like the theoretical predictions. If we instead of 4.5 cells have 4 cells with 10 individuals, we must increase one of the other rectangles to account for the 5 "missing" individuals, for instance by increasing the value of the number of cells with 5 individuals by one. Whereas we expect the actual number of cells in this case to be either 4 or 5, finding the expected number of cells where the predictions are far below one becomes increasingly more difficult. I will discuss this issue further below.

For now we can state that the theory predicts the general trends found in the real data. I have included a representative collection of distributions for a range of abundances. One thing to notice is that the actual number of cells with zero individuals is generally larger than what is predicted by the theory. The number of cells with zero individuals given by theory and the number found in the real plots is shown in Figure 3 for all the  $S_0$  species in the real data set. Only in two cases are the actual values below the predicted values. This is a sign that individuals within species tend to cluster. Clustering will give more individuals in some cells, and fewer in others. This mechanism is not included into the MaxEnt procedure (which is free of mechanisms).

The discrepancy between predictions and real data is especially apparent for the species with 1418 individuals, shown in Figures 1(e) and 1(f). Whereas theory predicts the number of zero cells to be 39, a total of 188 zero cells are observed for this species. Consequently, there are more cells with many individuals than what is predicted by the theoretical distribution. For the species with low abundances, the real distributions are close to the predictions. For the species with 112 individuals, the predicted number of cells with  $n > 4$  lies between 0 and 1, so we expect only a few of these to be 1 and a few to be zero. We notice the somewhat arbitrary looking tail in the real data shown in Figure 2(b). A possible method for predicting the actual tail is presented below. For  $n_0 = 1$  real data has to be equal to the prediction. For the species with only 1 individual, the only possible distribution is one cell with that 1 individual, and the rest of the cells,  $(A_0/A - 1)$ , with zero individuals.

The purpose of the simulations is to find the distribution of number of species in a cell of area  $A = A_0/256$ . We do this by drawing randomly once from each species and calculating the number of species  $T$  times. As stated above, the frequency

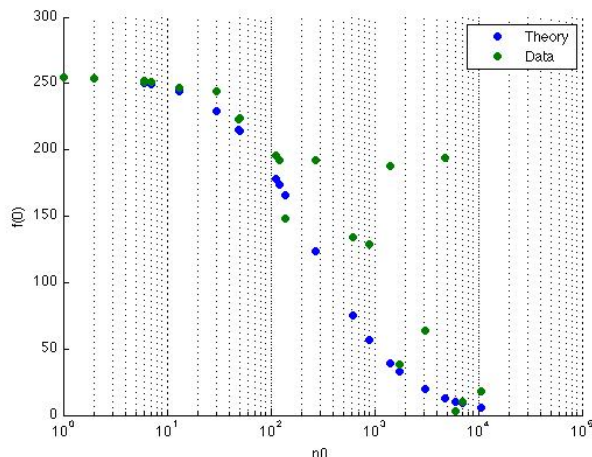


Figure 3: Number of cells with zero individuals for species with  $n_0$  individuals in  $A_0$ . Blue dots show the predictions from  $\Pi(0)A_0/A$  and green dots show the actual values. Data includes all the species in Little Blue Ridge, 1998 [29].

distribution for the number of individuals from the species with abundance  $n_0$  will approach the theoretical distribution,  $\Pi(n|A, n_0, A_0)$ , as  $T \rightarrow \infty$ . If we choose  $T = 256K$ , we can look at the simulation procedure as  $K$  simulations of the entire  $A_0$  plot, which consists of 256 smaller plots of area  $A = A_0/256$ . Each of the  $K$  simulations will give a different distribution of individuals from all the species over  $A_0$ . In one of the  $K$  simulations, a cell with 5 individuals of the species with  $n_0 = 112$  might coincide with the cell where the one individual from the species with  $n_0 = 1$  is found, whereas in most cases it will not. The frequency of any given  $s$  in the  $K$  simulations ( $T$  trials) is then taken to give the probability,  $P_A(s) = f(s)/T$ . There is a subtle problem with this simulation procedure. For each  $K$ , that is, for each time we draw 256 times, these draws are independent of each other. This means that in 256 trials, there is no guarantee that the total number of individuals drawn from each species will equal the observed abundances. The actual abundance of each species is only the expected abundance after 256 random draws. For species with few individuals, the relative difference will be larger (for instance, there's a fair chance that instead of 1 individual we will end up with 0 or 2 individuals for a species that has abundance 1 in 256 draws). This problem, however, can be avoided by choosing a large number for  $K$ . The chance of getting a larger than observed abundance in some cases is then weighted out by the chance of getting a lower than observed abundance in other cases. Instead of presenting a formal proof for why this is true, I argue that this must hold true because the same results are reproduced every time a simulation with  $K = 4000$ , or  $T = 256K = 1,024,000$  trials are performed. This is what is meant by a

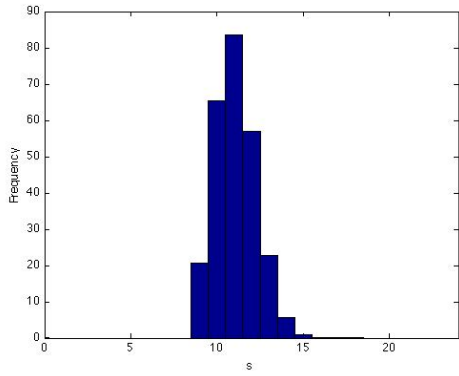
sufficient number of trials in step 5. in the description of the simulation procedure.

Figure 4 shows my simulations of  $P_{A_0/256}(s)$  together with the real data for the abundances found among the 24 species in 1998 and in the 28 species found in 2005 in Little Blue Ridge. In 1998 the average number of species in the real data was 7.56 with a standard deviation of 2.96 (variance 8.78). The simulations give an average of 11.06 and a standard deviation of 1.19 (variance 1.41). In 2005 the average number of species in the real data was 8.73 with a standard deviation of 2.95 (variance 8.73). The simulations give a mean of 11.55 and a standard deviation of 1.33 (variance 1.77). The predicted averages are higher than observed, and the predicted variances are considerably lower. We therefore predict a sharper distribution than what is observed in nature. According to the predictions, we can make more certain guesses than we can in reality. The lowest number of species observed in any cell in the 1998 data was 1, whereas the predicted minimum for a cell in this case is 9. The largest number observed the same year was 15, and the largest number predicted was 18. In 2005, the lowest observed number of species was also 1, and the prediction is 8. The largest number of species observed was again 15, whereas the prediction is 19. Note that also in this case there is a problem with the discretization. The number of cells predicted to have more than 15 species are in both cases less than zero.

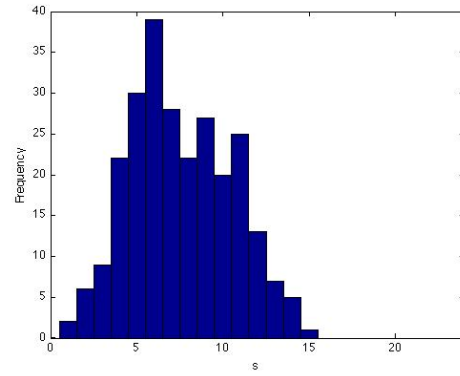
How can we interpret the discrepancy between predictions and actual data? What does  $\Pi(n)$  tell us, and what does  $P_A(s)$  tell us?  $\Pi(n)$  is the continuous MaxEnt distribution of individuals over  $(A_0/A)$  cells. It tells us the probability of finding  $n$  individuals in an area  $A < A_0$  for a species with abundance  $n_0$  in  $A_0$ . If we are looking at a second subplot in the same  $A_0$ , however, the probability of finding  $n$  individuals in this case is different if we have gained knowledge about how many individuals were in the first subplot. This information can be taken into account by saying that the species we are looking at has abundance  $(n_0 - n^{(1)})$ , where  $n^{(1)}$  is the number of individuals found in the first subplot, in the new total area  $(A_0 - A)$  and using this to calculate the new  $\Pi(n)$ . The probability of a specific ordered distribution of individuals over  $(A_0/A)$  cells numbered from 1 to  $(A_0/A)$ , where  $n^{(i)}$  denotes the number of individuals found in cell  $i$  is

$$P(n^{(1)}, n^{(2)}, \dots, n^{(A_0/A)}) = \Pi^1(n^{(1)}) \cdot \Pi^2(n^{(2)}) \cdot \dots \cdot \Pi^{A_0/A}(n^{(A_0/A)}) \quad (6.7)$$

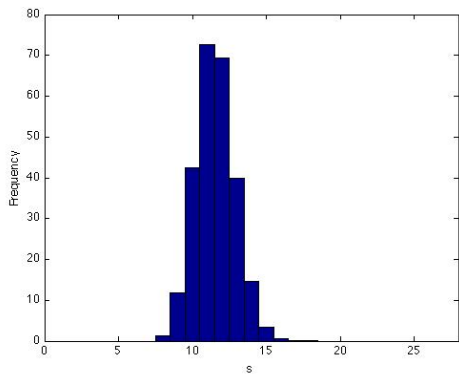
where each distribution  $\Pi^i(n^{(i)})$  is conditional on  $n^{(i)}$  in all the previous ones. Note that all of the  $\Pi^i(n^{(i)})$ 's, here described the same species.  $\Pi^1(n^{(1)})$  is equal to the original,  $\Pi(n^{(1)}|A, n_0, A_0)$ , but  $\Pi_2(n^{(2)}) = \Pi(n^{(2)}|A, (n_0 - (n^{(1)})), (A_0 - A))$ . In general



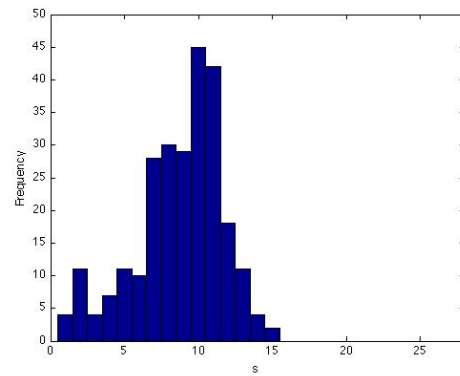
(a) Number of cells with  $s$  species, simulated, 1998



(b) Number of cells with  $s$  species, observed, 1998



(c) Number of cells with  $s$  species, simulated, 2005



(d) Number of cells with  $s$  species, observed, 2005

Figure 4: Number of cells with  $s$  number of species, from simulation on the left and observed on the right.  $A/A_0 = 1/256$ . Data from Little Blue Ridge, 1998 ( $S_0 = 24$  and  $N_0 = 37182$ ) and 2005 ( $S_0 = 28$  and  $N_0 = 60346$ ) [29].

$$\Pi_i(n^{(i)}) = \Pi(n^{(i)}|A, (n_0 - \sum_{i=1}^{i-1} n^{(i)}), (A_0 - (i-1)A)) \quad (6.8)$$

In our case, however, we are not interested in the order of the sequence. We want to know the probability of finding a certain number of plots with  $n^{(1)}$ ,  $n^{(2)}$  and so on. The probability,  $P(n^{(1)}, n^{(2)}, \dots, n^{(A_0/A)})$  however depends on the order and is in general not equal to the probability of  $P(n^{(A_0/A)}, \dots, n^{(2)}, n^{(1)})$ . The probability of a certain outcome  $\{n^{(A_0/A)}, \dots, n^{(2)}, n^{(1)}\}$  is therefore found by summing over the probability for all permutations. In this way we can calculate the probability of all the different distributions from our theory. This avoids the problem of the discretization of  $\Pi(n)$  and uses instead the theoretical predictions in order to evaluate the probability of all possible actual outcomes. This provides a quantitative measure of the likelihood of a real distribution, based on our theory, in the form of a probability distribution of distributions. This can then be compared to several real distributions (preferably more than two), in order to evaluate the predictive power of our theory. If the theory is correct, and we have enough observations, the relative frequencies of the actual distributions should approach the predicted probabilities of the different distributions.

I also suggest a second way to compute the likelihood of different actual distributions. Our predicted distribution,  $\Pi(n)$ , has a certain information entropy,  $I_\Pi$ , associated with it. By definition, this is the maximum possible entropy, compatible with our constraints. We (usually) need to change  $(A_0/A)\Pi(n)$  in order to arrive at an actual distribution with a discrete number of cells for each  $n$ . We can calculate the information entropy of any discretized distribution (by dividing by  $A_0/A$  again to attain a normalized probability distribution). The information entropy is a measure of how certain or uncertain we are when guessing how many individuals a random cell will contain, based on that distribution. This value will necessarily be less than the value for the continuous version of the theoretical distribution. I suggest that among the possible discrete versions of the theoretical distribution, we should pick the one with the highest entropy as the prediction for the actual data. Furthermore, the information entropy can be calculate for all the observed distributions. This value can then be compared with the discrete prediction as a measure of how likely the observed distribution is, given that the theoretical predictions are true. The lower the value of the entropy of the observed distribution is, the less likely this observation is. If all of the observed distributions have a significantly lower information entropy than the predicted distribution, this is a sign that there are unknown constraints acting on the distribution.



## 6.1 Discussion

The  $\Pi(n)$  distribution acts as a null-distribution: we assume that nothing matters for the spatial distribution of the individuals within a species. The derivation of  $\Pi(n)$  assumes only the average value of  $n$ . The discrepancies between real data and predictions can then be used to point out which factors do matter. If the pattern, on the other hand, is found to be very close to the predicted pattern, this tells us that, yes, indeed, as long as the average number of species is satisfied, the rest is random. Many factors can affect the spatial distribution of species. Some species will cluster, and some species will disperse. Two competing plant types will be found in different locations whereas two mutualist plant types will be found together. In our case the real distributions have more plots with zero individuals and more plots with many individuals than the theoretical distributions. This tells us that some sort of clustering mechanism is at play. In accordance with the MaxEnt philosophy, comparison of data with predictions is used either to verify the validity of the theoretical model, or to point towards new mechanisms. In order to do this however, we need a defined procedure of comparison.

We want to determine the certainty of the predictions derived in METE. In statistical mechanics, the Boltzmann distribution gives an extremely peaked distribution for the energy of the canonical ensemble. The energy of a system in contact with a heat reservoir (and in equilibrium), is most certainly equal to the expected value. Any other macroscopic value we might be interested in will also be very peaked, and therefore very certain. If we wanted to predict the velocity of a particle in an ideal gas, however, this is less certain. The distribution of velocities is given by the Maxwell-Boltzmann velocity distribution, which is much less peaked than the distribution for the systems total energy (and becomes increasingly more spread at higher temperatures). If we are really talking about an ideal gas (minimal interaction between molecules), we would expect the molecules to be distributed according to this distribution and the more molecules we have, the closer the real distribution should be to the prediction. Still, we can always use the theory to calculate the probability for observing any other distribution. Neither  $\Pi(n)$  nor  $R(n, \varepsilon)$  are peaked distributions.  $\bar{S}(A)$  is more so, but still not on the level of statistical mechanics.

We want our predictions from METE to be as certain as possible, but the question is what it is that we want to be peaked in METE? Statistical mechanics is used to predict unknown macroscopic quantities from known macroscopic quantities and the Boltzmann distribution is used only as a step along the way. The purpose of METE, however, is qualitatively different. Instead of predicting macroscopic variables from macroscopic constraints, we predict microscopic variables from the macroscopic constraints. We are much more interested in the distributions telling

us about individual behavior in METE than we usually are in statistical mechanics. It does not make sense for the Species-Area Relationship to be extremely peaked, because we will always expect a natural variation in the number of species over different plots. Since we are not really using our derived MaxEnt distribution to predict any macroscopic quantities of the sort that measure the entire system ( $\bar{S}(A)$  still measures cells within the system) we are more interested in how well we can predict the Species-Area distributions,  $\bar{S}(A)$ , the Spatial-Abundance distribution,  $\Pi(n)$ , and the Species-Abundance distribution,  $\Phi(n)$ . However, we do not have large enough numbers of individuals or cells to expect the observed  $n$  to be identical to  $\Pi(n)$ , even if the theory is correct. With smaller numbers the chances of observing something different than what is expected is larger. The same holds for a small thermodynamic system: its molecules are less likely to be distributed according to the Maxwell-Boltzmann distribution. Theory predicts a non-zero probability for all of the observed distributions in this Chapter. The likelihood of observing any distribution can be determined using the methods that I suggest at the end of the previous section. If we know the likelihood of each outcome, this can tell us when there is sufficient reason to suspect that the theory is incorrect, and that there are unknown mechanisms.

The two simulated distributions for  $P_A(s)$  differ quite a lot from the observed distributions. According to the predictions, we should expect zero plots with less than 8 or 9 species, whereas we find plots with as little as one species in the real data. The simulations were only compared with two data sets, which is of course very little, but such a clear discrepancy is not likely to be random. It is a sign that the theoretical MaxEnt predictions do not contain all the relevant constraints in order to fully describe the ecosystems. This is also not expected from such a crude theory. Furthermore, the censusing itself will always introduce errors. For instance, the numbers could be more correct when there are few individuals in a plot than when there are many, because it is easier to spot all of them. And lastly, it could be a sign that something is wrong with the simulations themselves.

The discrepancy between the predicted  $P_A(s)$  and the observed frequencies, given that the simulations are correct and that the errors in the data are insignificant, can come from two different sources. It can come from the fact that the individuals within the species are not correctly predicted by the theoretical distributions, or it can be due to interaction between different species (which the theory assumes to be zero). It would be interesting to do a simulation using the real distributions of individuals from each species instead of the theoretical distributions. This could be done by randomly combining actual abundances from the observed plots for each species and count the number of species present. By repeating this we will arrive at a distribution for the amount of times the different number of species show up in a cell. If this distribution is very different from the observed distribu-

---

tion, then this tells us that the species might not be independent of each other, which is one of the assumptions in METE and in the simulations. In any case, this distribution should be closer to the observed distribution than the ones obtained from the simulations in the previous section because we have removed the error due to the fact that the individuals within each species are not predicted fully by the theoretical distributions.

When it comes to the entropy method presented at the end of the previous section, it is necessary to find out what the range of entropy values mean, and how to compare them. Even though I suggested the method in order to compare  $\Pi(n)$  with observations, it should be applicable also to  $P_A(s)$ .

This chapter works as a starting point for defining a clear procedure for comparing predictions with data. Time limitations prevented me from developing any of the suggested methods further myself, but I suggest both of them as possible approaches in the future work on METE. When a more rigorous procedure for comparison is established, more data sets should be used to test the predictions of METE.

## 7 A Maximum Entropy Production Theory of Ecology?

Harte opened up a completely new approach to macroecology with his application of the Maximum Entropy principle to ecosystems. METE provides a new way of inferring important ecosystem parameters from incomplete census data taken at an instant in time. What the theory does not do, however, is say anything about change in the parameters over time. Ecosystems are never really in equilibrium. The number of species and individuals constantly changes. Harte suggests that some of the failures of the MaxEnt predictions are due to fast changing state variables in some of the ecosystems for which we have empirical data. In analogy to the local equilibrium hypothesis in thermodynamics, METE should hold when the distributions of individuals within species ( $\Phi(n)$ ) and over area ( $\Pi(n)$ ) adjust themselves faster than the time rate of change of the state variables. If the state variables change slowly enough they should be able to "drag" the distribution patterns along, such that the distributions are always true for a given set of state variables.

Harte has proposed the Maximum Entropy Production (MEP) principle as a way to obtain values for the time rate of change of the state variables in an ecosystem. The investigation of this hypothesis is the main topic of this thesis. Harte presented the basic principle of Maximum Entropy Production, and some of what it implies for the expressions in METE in Chapter 11.3 in *Maximum Entropy and Ecology*.

This work provided the basis for my application of the principle, and the statement of the initial problem.

### 7.1 Initial Problem Formulation

Harte uses Paltridge's climate studies (1975) and Dewar's proof (2005) to present the idea of the Maximum Entropy Production in *Maximum Entropy and Ecology*. In Harte's words: "MEP states that a physical system far from equilibrium, undergoing an irreversible transition from one macroscopic state to another, will most likely choose a path in phase space that maximizes the rate of production of entropy" [1, Ch. 11.3]. Harte also writes about Dewar's proof that: "His approach was to derive from MaxEnt a probability distribution for the number of paths in phase space leading from an initial state to each possible future state. He then shows that the distribution achieves its maximum when the final state is the one in which the maximum possible amount of entropy is produced in the transition from initial to final state" [1, Ch. 6.3.5]. Even though Dewar's proof is not only complicated but also not regarded as rigorous, it can be assumed that his proof will be completed, and that MEP will be accepted as a general law of far-from-equilibrium thermodynamics. According to Harte - at the outset of my work - Dewar's proof also suggested that MEP should apply not only to the production of thermodynamic entropy, for which the principle has been shown to be successful by Paltridge, but also to the production of information entropy.

This initial understanding of the MEP principle in general and Dewar's proof in particular motivated Harte to suggest a specific application of the principle to METE. This application was based on the hypothesis that the MEP principle also applies to the information entropies in METE. In order to investigate this hypothesis I was asked to study the implications of assuming that the information entropy of the Ecosystem Structure Function,  $I_R$ , defined by equation (5.7), and restated here

$$I_R = - \sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} d\varepsilon \cdot R(n, \varepsilon) \cdot \ln(R(n, \varepsilon))$$

is being produced at the maximum possible rate.  $R(n, \varepsilon)$  was given by (5.12) as

$$R(n, \varepsilon) = \frac{1}{Z(\lambda_1, \lambda_2)} e^{-\lambda_1 n} e^{-\lambda_2 n \varepsilon}$$

That is, we want to know the consequences of the assumption that  $dI_R/dt = \max$ . Harte decided to denote the information entropy,  $I_R$ , instead by  $S_I$ , in order to stay closer to thermodynamics convention. I have therefore used  $S_I$  throughout my application and analysis. From now on, I will also use  $S$ ,  $N$  and  $E$  without the zero subscript, to express that we are talking about state variables that are functions of time (the state variable  $A$  is not involved in the information entropy,  $S_I$ ). We

still assume that  $S > 4$  in order to use the approximate results for the Lagrange multipliers, equation (5.14) - (5.16). By assuming  $E \gg N$  we can simplify even more by taking  $\lambda_2 \approx S/E$ . For convenience, we also define the variable

$$\omega \equiv \ln \left( \frac{1}{\beta} \right) \quad (7.1)$$

where  $\beta = \lambda_1 + \lambda_2$ .

In the draft of the book, Harte presented the value of the entropy as

$$S_I = 1 + \ln \left( \frac{E}{N} \right) + \omega + \omega^{-1} \quad (7.2)$$

where  $\omega$  depends on  $S$  and  $N$  through  $\beta$ , and the time derivative of the entropy as

$$\frac{dS_I}{dt} = \omega^{-1} \frac{1}{N} \frac{dN}{dt} + \frac{1}{E} \frac{dE}{dt} - (1 + \omega^{-1}) \frac{1}{S} \frac{dS}{dt} \quad (7.3)$$

The principle of Maximum Entropy Production was taken to mean that expression (7.3) would be maximized with respect to the state variables,  $S$ ,  $N$  and  $E$ . I started out with an even more specific problem, and the initial problem statement of my project was:

1. To verify the expression for the entropy (7.2) and the time rate of change of entropy (7.3).
2. To work out the consequences of MEP by setting  $\partial(dS_I/dt)/\partial S = 0$ ,  $\partial(dS_I/dt)/\partial N = 0$  and  $\partial(dS_I/dt)/\partial E = 0$ , and solve for the time derivatives. The partial differential equations we get from doing this, equations (7.4) - (7.6), were also already derived by Harte.

From this point I began my investigation.

## 7.2 Solution to the Initial Problem

The entropy, (7.2), can be derived by plugging  $R(n, \varepsilon)$ , as defined by (5.12), directly into the expression for the information entropy,  $I_R$  given in (5.7). A second, and faster method, is to directly substitute the constraints and the Lagrange multipliers into the expression for  $H_{max}$  given in (3.9). In both cases we use the assumption that  $e^{-S_0} \ll 1$  in order to use the approximate results for the Lagrange multipliers presented in section 5.4.

From  $H_{max}$  in (3.9) we obtain

$$\begin{aligned} S_I &= \ln Z + \lambda_1 \langle f_1(x) \rangle + \lambda_2 \langle f_2(x) \rangle \\ &= \ln Z + \lambda_1 \frac{N}{S} + \lambda_2 \frac{E}{S} \end{aligned}$$

From equation (5.14) we have

$$\lambda_2 \approx \frac{S}{E - N}$$

Combining this with (5.15) and using  $\beta = \lambda_1 + \lambda_2$  we can write

$$\lambda_1 = \beta - \lambda_2 \approx \frac{S}{N} \frac{1}{\ln \frac{1}{\beta}} - \frac{S}{E - N}$$

From equation (5.16) and (5.15) we get

$$\begin{aligned} Z &\approx \frac{\ln \left( \frac{1}{\beta} \right)}{\lambda_2} \approx \frac{\frac{S}{N} \frac{1}{\beta}}{\frac{S}{E}} \approx \frac{E}{N} \frac{1}{\beta} \\ \ln Z &\approx \ln \frac{E}{N} + \ln \frac{1}{\beta} \end{aligned}$$

Combining the above equations we obtain

$$\begin{aligned} S_I &\approx \ln \frac{E}{N} + \ln \frac{1}{\beta} + \left( \frac{S}{N} \frac{1}{\ln \frac{1}{\beta}} - \frac{S}{E - N} \right) \frac{N}{S} + \left( \frac{S}{E - N} \right) \frac{E}{S} \\ &= \ln \frac{E}{N} + \ln \frac{1}{\beta} + \frac{1}{\ln \frac{1}{\beta}} - \frac{N}{E - N} + \frac{E}{E - N} \\ &= \ln \frac{E}{N} + \ln \frac{1}{\beta} + \frac{1}{\ln \frac{1}{\beta}} + \frac{E - N}{E - N} \\ &= 1 + \ln \frac{E}{N} + \omega + \omega^{-1} \end{aligned}$$

where we have used (7.1). This finishes the derivation of (7.2), which was the first part of point 1. in the project statement. I also verified equation (7.3) by computing the partial derivatives of equation (7.2). The calculations are somewhat lengthy because  $\omega$  is only an implicit function of the state variables  $S$  and  $N$ , but other than that they are fairly straight forward. The derivations are included in Appendix A.

Setting the partial derivatives of the entropy production,  $dS_I/dt$ , equal to zero as prescribed in point 2. of the project statement results in three coupled differential equations

$$(1 + \omega) \frac{\partial \dot{S}}{\partial S} = \frac{1}{(\omega - 1)} \frac{\dot{N}}{N} + \omega \frac{S}{E} \frac{\partial \dot{E}}{\partial S} + \frac{S}{N} \frac{\partial \dot{N}}{\partial S} + \frac{(\omega^2 - 2)}{(\omega - 1)} \frac{\dot{S}}{S} \quad (7.4)$$

$$\frac{(\omega^2 - 1)}{\omega} \frac{N}{S} \frac{\partial \dot{S}}{\partial N} + \frac{\dot{N}}{N} = (\omega - 1) \frac{N}{E} \frac{\partial \dot{E}}{\partial N} + \frac{(\omega - 1)}{\omega} \frac{\partial \dot{N}}{\partial N} + \omega^{-1} \frac{\dot{S}}{S} \quad (7.5)$$

$$(1 + \omega) \frac{E}{S} \frac{\partial \dot{S}}{\partial E} + \omega \frac{\dot{E}}{E} = \omega \frac{\partial \dot{E}}{\partial E} + \frac{E}{N} \frac{\partial \dot{N}}{\partial E} \quad (7.6)$$

where the dots denote time derivatives, (e.g.  $\dot{S} = dS/dt$ ). These equations were also provided in the draft of Harte's book. I also verified these equations, and again the calculations are lengthy, but straight forward. The technique is similar to the derivation of (7.3), which is shown in Appendix A. Since the derivations are not of any particular interest to my thesis I have not included them. For compactness I will from now on adopt the notation  $\partial \dot{S}/\partial S = \dot{S}_S$  for the partial derivatives. According to Harte - at the beginning of my work in August 2010 - the solution to these equations describes the behavior of the state variables,  $S$ ,  $N$  and  $E$  in time. That is, they define the time derivatives of the state variables yielding the Maximum Entropy Production state of the ecosystem. The goal of my work was therefore to predict the evolution from knowledge of the initial state variables  $S_0$ ,  $N_0$  and  $E_0$ , based on these equations. We see that the equations contain the time derivatives,  $\dot{S}$ ,  $\dot{N}$  and  $\dot{E}$ , all their partial derivatives,  $\dot{S}_S$ ,  $\dot{S}_N$ ,  $\dot{S}_E$ ,  $\dot{N}_S$ ,  $\dot{N}_N$ ,  $\dot{N}_E$ ,  $\dot{E}_S$ ,  $\dot{E}_N$  and  $\dot{E}_E$  as well as the state variables,  $S$ ,  $N$  and  $E$ , themselves. The complicated structure of the equations combined with the fact that  $\omega$  is an implicit function of  $S$  and  $N$  suggests that an analytical solution is, if not impossible, very difficult to obtain. There is no general theory for the solvability of partial differential equations (PDE's), and none of the classical PDE's that have known solutions (e.g. the wave equation) are close enough to tell us anything about the solution of (7.4) - (7.6). The choice therefore fell on a numerical approach.

By closer inspection of the above equations it is clear that in order to obtain numerical solutions one has to know not only the initial value of the state variables, but the initial values also of the time derivatives,  $\dot{S}_0$ ,  $\dot{N}_0$  and  $\dot{E}_0$ . We can then find the values at the next time step,  $S_1$  and  $\dot{S}_1$ , from the linear Taylor expansions

$$\begin{aligned} S_1 &= S_0 + \dot{S}_0 dt && \text{(known from initial conditions)} \\ \dot{S}_1 &= \dot{S}_0 + \dot{S}_S dS + \dot{S}_N dN + \dot{S}_E dE \end{aligned}$$

where the partial derivatives are taken for the initial conditions. Because  $dS =$

$S_1 - S_0 = \dot{S}_0 dt$  (and similarly for  $N$  and  $E$ ) we can rewrite the above equation to get

$$\dot{S}_1 = \dot{S}_0 + \dot{S}_S \dot{S}_0 dt + \dot{S}_N \dot{N}_0 dt + \dot{S}_E \dot{E}_0 dt$$

which again gives

$$\begin{aligned} S_2 &= S_1 + \dot{S}_1 dt \\ &= (S_0 + \dot{S}_0 dt) + (\dot{S}_0 + \dot{S}_S \dot{S}_0 dt + \dot{S}_N \dot{N}_0 dt + \dot{S}_E \dot{E}_0 dt) dt \\ &= (S_0 + 2\dot{S}_0 dt) + (\dot{S}_S \dot{S}_0 + \dot{S}_N \dot{N}_0 + \dot{S}_E \dot{E}_0) dt^2 \end{aligned}$$

Hence,  $S_2$  depends on the six initial conditions (for the state variables and their time derivatives) and the three partial derivatives of  $\dot{S}$ . Similar expressions apply to  $N_2$  and  $E_2$ . The time step,  $dt$ , has to be chosen sufficiently small in order for the linear approximation to be a good one. We then have  $S_t = S(\dot{S}_S, \dot{S}_N, \dot{S}_E)$ ,  $N_t = N(\dot{N}_S, \dot{N}_N, \dot{N}_E)$  and  $E_t = E(\dot{E}_S, \dot{E}_N, \dot{E}_E)$  where the partial derivatives are taken at the previous time step for which the state variables and their time derivatives are defined. This procedure will then give the time evolution of the state variables, given their initial values and their initial time derivatives.

The partial derivatives of the time derivatives are found by equations (7.4) - (7.6). However, we have nine unknowns (three partial derivatives for each of the three time derivatives of the state variables), and only three equations. We are therefore not able to calculate the partial time derivatives needed for the numerical procedure from these equations alone. This realization came as a surprise in the project. In order to find a solution we would have to decrease the number of unknowns, or increase the number of equations. Additional equations regarding the relationships between the different time derivatives could be expected to hold, e.g. an increase in total metabolic rate would probably correlate with an increase in the number of individuals. Incorporating relationships like these into the framework, however, would be counter productive to the initial goal of creating a simple theory that does not require additional hypotheses. I therefore chose not to look further into this. The other alternative is to only look at the case where some of the state variables are constant in time. In order to reduce the number of unknowns down to three we can set two of the three state variables to constants, and look at the variations of the third one only. The result of doing this leads to ordinary differential equations that can be easily solved for in each of the three cases (where only one state variable is allowed to change). I have not included the results of doing this, due to the argument in the following section.

The partial differential equations, (7.4) - (7.6), were at the outset of the project though to be solvable and the analysis of the solution was meant to constitute the



main part of my thesis. A lot of time was therefore taken to study the theory behind PDE's and techniques for analyzing and solving these types of equations in general. When I later found out that there is not enough information to solve this set of equations, the inability to derive more interesting results inspired me to look more closely at the maximization of the rate of the entropy production,  $dS_I/dt$ , itself, and what this really means.

### 7.3 Meaning of zero partial derivatives

The Maximum Entropy Principle suggests that  $dS_I/dt$  is maximized. In the initial problem statement this maximum was meant to be found by setting all the partial derivatives of  $dS_I/dt$  with respect to the three state variables equal to zero, and solve for the time derivatives. I will now argue that this procedure does not actually give the correct maximum condition for the state variables and their derivatives.

Say we have a function of three variables,  $f = f(x, y, z)$ . In order to find the maximum of  $f$ , we set all the partial derivatives to zero,  $\partial f/\partial x = \partial f/\partial y = \partial f/\partial z = 0$ , and look for the value(s) of  $(x, y, z)$  that satisfy this condition. At points where all the partial derivatives disappear we have either a local or global maxima or minima. This is the conventional and straight forward maximization problem. Ours, however, is different. Instead of a function of three variables, we have a function of six variables,  $f = f(x, y, z, \dot{x}, \dot{y}, \dot{z})$ , where the dots denote time derivatives. We postulate that the time derivatives are functions of  $(x, y, z)$ , that is,  $\dot{x} = \dot{x}(x, y, z)$  and similarly for  $\dot{y}$  and  $\dot{z}$ . Hence, we can again write  $f = f(x, y, z)$  and take the partial derivatives of  $f$  with respect to the three variables. Since the function  $f$  still contains the time derivatives (their functional expressions are unknown), the partial derivatives have to be found using the chain rule  $(\partial f/\partial x)_{y,z} = (\partial f/\partial x)_{y,z,\dot{x},\dot{y},\dot{z}} + (\partial f/\partial \dot{x})_{x,y,z,\dot{y},\dot{z}}(\partial \dot{x}/\partial x)_{y,z} + (\partial f/\partial \dot{y})_{x,y,z,\dot{x},\dot{z}}(\partial \dot{y}/\partial x)_{y,z} + (\partial f/\partial \dot{z})_{x,y,z,\dot{x},\dot{y}}(\partial \dot{z}/\partial x)_{y,z}$  where the subscripts denote the variables that are held constant under the partial derivation. Equivalent expressions apply for  $y$  and  $z$ . Again, we require  $(\partial f/\partial x)_{y,z} = (\partial f/\partial y)_{z,x} = (\partial f/\partial z)_{x,y} = 0$ . As opposed to the previous case, we now have a dynamical system with non-zero time derivatives, meaning that even if we find a point  $(x^*, y^*, z^*)$  where all the partial derivatives disappear, the system will not stay here because it in the next time step will evolve according to the time derivatives at this point. If, however, the zero conditions on the time derivatives are not used to find the point(s)  $(x^*, y^*, z^*)$ , but rather to find a set of time derivatives  $(\dot{x}^*, \dot{y}^*, \dot{z}^*)$  as functions of  $(x, y, z)$ , the situation is different.  $(\dot{x}^*, \dot{y}^*, \dot{z}^*)$  will then define the set of time derivatives that are such that the partial derivatives of  $f$  will be zero for all  $(x, y, z)$ . In other words, the time derivatives will be functions of  $(x, y, z)$  with the property that when substituted into the expression  $f = f(x, y, z, \dot{x}, \dot{y}, \dot{z})$  to arrive at the expression  $f = f(x, y, z)$ , all the variables,

$(x, y, z)$  will actually disappear from the expression. This is because, in order for the partial derivatives with respect to  $x$ ,  $y$  and  $z$  to be zero everywhere, the function  $f$  cannot depend on them. In general, for any function  $g$ , we have that  $dg/dx = 0$  (everywhere)  $\implies g \neq g(x)$ .

This is the exact case that we had in the previous section, where instead of  $f(x, y, z)$  we had  $S_I(S, N, E)$ . The zero conditions for the partial derivatives are expressed in equation (7.4) - (7.6). I therefore conclude that the original procedure does not actually give the maximum rate of entropy production,  $dS_I/dt$ , but instead the condition where  $dS_I/dt = \text{constant}$ . This realization prompted me to investigate what the actual condition for finding the maximum rate of entropy production should be.

#### 7.4 Maximum rate of entropy production, $dS_I/dt$

The statement "find the maximum of  $dS_I/dt$ " requires further clarification. Maximum with respect to what? What is known? What is free to vary? Do we have to make any other assumptions, and in case we do, what are they?

We cannot be maximizing with respect to  $S$ ,  $N$  and  $E$  because these are given as initial conditions and are not supposed to be found from the maximum condition. What we want to know are the time derivatives,  $\dot{S}$ ,  $\dot{N}$  and  $\dot{E}$  and hence they must be what is "free to vary" and what we should be maximizing with respect to.

If we look at the expression for  $dS_I/dt$  in (7.3) from a strictly mathematical viewpoint the state variables and their time derivatives are independent of each other. Furthermore  $S$ ,  $N$  and  $E$  all denote positive quantities. Harte also shows that  $\beta < 1$  and thus  $\omega = \ln(1/\beta) > 0$  [1]. This means that the terms in front of  $\dot{N}$  and  $\dot{E}$  in equation (7.3) are positive, and that the term in front of  $\dot{S}$  is negative. If the time derivatives of the state variables are regarded as independent of the state variables themselves, and the time derivatives are free to take on any value that will give a maximum for  $\dot{S}_I$ , the maximization problem is unbounded. For any value of  $S$ ,  $N$  and  $E$ , the maximum of  $\dot{S}_I$  will correspond to  $\dot{N} \rightarrow \infty$ ,  $\dot{E} \rightarrow \infty$  and  $\dot{S} \rightarrow -\infty$ . In reality, the maximum rate of change of the state variables will obviously be limited by physical processes. If one has information about what these limits are, say  $M_{\dot{N}}$ ,  $M_{\dot{E}}$  and  $M_{\dot{S}}$ , the maximization problem stated this way only tells us that the time derivatives will always take on the maximum (for  $N$  and  $E$ ) and minimum (for  $S$ ) values.

This result is neither very interesting, nor does it make much sense that the time derivatives will always attain their maximum values completely independently of the values of the state variables. We therefore postulate that: the time derivatives

of the state variables must be functions of the state variables themselves, that is,  $\dot{S} = \dot{S}(S, N, E)$  (and similarly for the other two state variables). This postulate is based on the following arguments.

- i)  $S$ ,  $N$  and  $E$  are the three defining variables. They tell us everything we can know about the system and there is nothing else the time derivatives can depend on.
- ii) It makes physical sense. For instance the rate of speciation,  $\dot{S}$ , is likely to depend on the number of species or individuals. Very low  $S$  and  $N$  should correspond to a lower chance of speciation than for very high  $S$  and  $N$ .
- iii) If we do not assume a relationship between the time derivatives and the state variables, the above maximization gives unphysical results. To an extent, it is a trial and error procedure.

Thus, we should not be maximizing with respect to the time derivatives as independent variables. We are not trying to find a point (where point refers to a numerical value of the time derivatives) where  $dS_I/dt$  is maximum, but a set of functions that make  $dS_I/dt$  a maximum for any  $(S, N, E)$ . What would these functions look like? And how do we find them?

If we look at the variation of  $E$  only and  $\dot{S} = \dot{N} = 0$ , we have  $dS_I/dt = (1/E)\dot{E}$ . According to the initial procedure we should take the derivative with respect to  $E$  and set this expression equal to zero. This gives the solution  $\dot{E} = CE$ , where  $E$  is a constant of integration. In this case,  $dS_I/dt = C$ . It is the constant entropy production solution. We can compare this to another expression for the time derivative,  $\dot{E} = E^2$ . The entropy production in this case is  $dS_I/dt = E$ . Since  $E$  is positive  $\dot{E}$  is positive and  $dS_I/dt = E$  is an increasing function both of  $E$  and of time. If  $E_0 < C$  we have a situation where the constant solution gives the highest entropy production to begin with, but is overtaken by the other function when  $E = C$ . After this point this function will give a larger and larger entropy production.

It is tempting to write the new maximum condition as

$$\frac{\partial \dot{S}_I}{\partial \dot{S}} = \frac{\partial \dot{S}_I}{\partial \dot{N}} = \frac{\partial \dot{S}_I}{\partial \dot{E}} = 0 \quad (7.7)$$

where all the time derivatives are regarded as functions of  $S$ ,  $N$  and  $E$ . Due to the linear nature of (7.2) however, the time derivatives themselves fall out in this procedure and we obtain results like  $\partial \dot{S}_I / \partial \dot{E} = 1/E = 0$ , where the time derivatives disappear (and give meaningless results). Again, the maximum time rate of change of the entropy is found at the maximum and minimum time rate of change for the state variables. From the expression for  $dS_I/dt$  given in (7.3) we see that  $\dot{E} = E^2$

gives a higher rate of entropy production than  $\dot{E} = E$  (as long as  $E > 1$ , which it always is). The exponent can be increased more to give an even higher rate of entropy production. Even though the time derivative is expressed as a function of a state variable, the maximization problem remains unbounded. The problem statement is similar to asking what dimensions of a box will maximize the volume, without specifying an equation constraining the surface area. The maximization of  $dS_I/dt$  as formulated by the MEP principle here gives only trivial solutions. More equations are required if we want to arrive at an interesting solution.

The discussion regarding the maximum condition, however, has another component to it as well. If the time rate of change of the metabolic rate is large in the first step, the ratio  $1/E$  will be much smaller in the next step. This gives rise to a significantly smaller entropy production in the next step. There is a trade-off between large derivatives, and large terms in front of the derivatives if we take time into account. This again raises the question of what a maximum entropy production fundamentally means. I have found three different interpretations:

- i) Maximum at time  $t$
- ii) Asymptotic maximum
- iii) Integrated maximum

Version i) is the one discussed above. At every instance in time, we seek the time derivatives resulting in the largest entropy production. Version ii) can also be exemplified using the different functions for  $\dot{E}$  presented above. Whereas  $\dot{E} = CE$  leads to a constant entropy production,  $\dot{E} = E^2$  leads to an entropy production that increases as  $E$  increases. Whereas the constant entropy production can be larger to begin with, the entropy production that is proportional to  $E$  will be greater asymptotically. The asymptotic version of the maximum entropy production principle implies that the second time derivative function is the MEP function. Version iii) however, yields a somewhat different concept. In this version we look at the entropy produced at each step in time, and maximize the sum. In the discrete case, each step yields the entropy production  $\Delta S_{I,i}/\Delta t_i$ . The sum,  $\sum_i \Delta S_{I,i}/\Delta t_i$ , is then maximized as the system evolves from the initial  $S_I^{(0)}$  to the final  $(S_I^{(0)} + \Delta S_I)$  during the time interval  $\Delta t$ . In this case, a collection of few but large  $\Delta S_{I,i}$  over short  $\Delta t_i$  followed by small  $\Delta S_{I,i}$  over the rest of the  $\Delta t_i$ 's will give a larger sum than a collection where all  $\Delta S_{I,i}/\Delta t_i$  are equal. An interesting result of this version is that the procedure automatically connects the time derivatives to the state variables. This is analogous to the way position and velocity are automatically linked when we minimize the action integral in analytical mechanics. If we use this maximization procedure, we do not need to postulate that the time derivatives are functions of the state variables themselves. This is an automatic result of the maximization procedure itself.

Both ii) and iii) involve the future state of the system. There are different ways in which the result of such principles could be computed, for instance by using some of the techniques from analytical mechanics. However both of these principles have a teleological character. Why would the system choose a lower entropy production now, in order to arrive at a higher entropy production in the future, or in sum? The last two principles seem too speculative and I therefore leave the discussion of these on this note. From the conclusion that the maximum condition on (7.3) alone does not provide sufficient information to determine the time derivatives of the state variables I decided instead to look more closely at what the entropy,  $S_I$ , really means. What conclusions, if any, can we draw from the principle that the production of this quantity should be maximized?

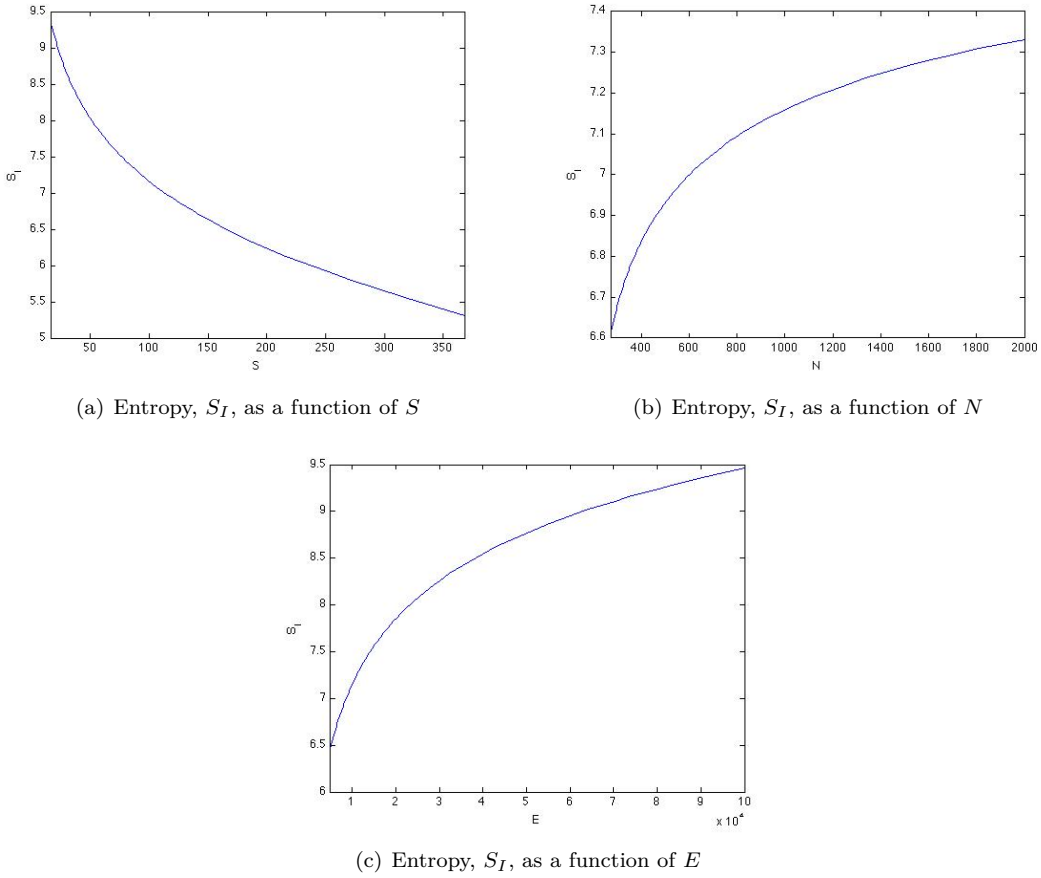
## 7.5 Entropy, $S_I$

So far, we have been looking at the local properties of the entropy production. An important question is, if the entropy keeps increasing, where will the system end up? Is there a maximum in the entropy  $S_I$ ?

For any point in phase space (where phase space is defined by  $S$ ,  $N$  and  $E$ ) the entropy has a specific value, given by equation (7.2). Several combinations of the state variables will give the same entropy. In fact, if we look back at the MaxEnt theory, it tells us that the value of the information entropy is always a function of the constraints. This means that only the two ratios  $N/S$  and  $E/S$  can matter. Different combinations of the ratios can also give the same entropy. Equation (7.2) defines the surfaces of constant entropy. The requirement that the state variables be positive, and the required relationships between them for the approximations to be good, puts limits on the allowed part of phase space.

Figure 5 shows the change of entropy with respect to one out of the state variables at a time, with the other two held constant. I did the calculations in MATLAB, using the approximations presented in Chapter 5. The range of values are chosen such that these approximations are good ( $S \gg 1$ ). We see that the entropy increases with the number of individuals and the total metabolic rate, but decreases with the number of species. This corresponds to the conclusions from the last section. The general shapes of the graphs did not change when I changed the values of the two constant state variables. There is no maximum in the entropy, and assuming that the principle of maximum entropy production holds for  $S_I$  our system will approach  $S = 1$  (since this is the minimum number of species),  $N \rightarrow \infty$  and  $E \rightarrow \infty$  (where we require  $N < E$  always because  $\varepsilon_{min} = 1$ ).

Instead of writing the time rate of change, we can write instead the infinitesimal


 Figure 5: Entropy as a function of  $S$ ,  $N$  and  $E$ .

rate of change

$$dS_I = S_{I,S}dS + S_{I,N}dN + S_{I,E}dE \quad (7.8)$$

where the  $S_{I,S}$  denotes  $\partial S_I/\partial S$ . Whereas (7.3) is an exact equation, (7.8) expresses the first order Taylor approximation to the actual change in entropy,  $\Delta S_I$ , when moving from  $(S_0, N_0, E_0)$  to  $(S_1, N_1, E_1)$ . The smaller the differences,  $dS$ ,  $dN$  and  $dE$  are, the better the approximation is. We can find the trajectory of a system evolving in the direction of maximum entropy production, but from the discussions in the previous sections we do not have enough information to say anything about the actual time this will take. MEP only states that the system will evolve along this path in the shortest possible amount of time.

The direction of maximum increase is given by the gradient of  $S_I$ . The magnitude of the gradient gives the rate of change of the entropy in this direction. The

gradient of a general scalar function  $f(x, y, z)$  in three variables is  $\nabla f = f_x \hat{x} + f_y \hat{y} + f_z \hat{z}$ , where subscript denote partial derivatives and the hat denotes the unit vector in the respective dimension. The gradient is easily found from equation (7.3) to be

$$\nabla S_I = S_{I,N} \hat{N} + S_{I,E} \hat{E} + S_{I,S} \hat{S} \quad (7.9)$$

$$= \omega^{-1} \frac{1}{N} \hat{N} + \frac{1}{E} \hat{E} - (1 + \omega^{-1}) \frac{1}{S} \hat{S} \quad (7.10)$$

The terms in front of each of the unit vectors are the linear approximations describing how much the entropy changes when each state variable changes by one. Since  $\beta$  is a decreasing function of  $N$  and an increasing function in  $S$ , it follows from (7.1) that  $\omega^{-1}$  is too.  $S_{I,N}$  is therefore a decreasing function with respect to  $N$  and an increasing function with respect to  $S$ .  $S_{I,E}$  is obviously a decreasing function of  $E$  and  $S_{I,S}$  is a decreasing function of  $N$ , but ambiguous when it comes to  $S$ , since  $\omega^{-1}$  is an increasing function of  $S$  but  $1/S$  is a decreasing function of  $S$ . From my calculations in MATLAB it turns out that  $S_{I,S}$  is an increasing function of  $S$  as well. Since all the terms have the same direction of dependency the magnitude of the gradient is a decreasing function of  $N$  and  $E$  and an increasing function in  $S$ . This can also be seen directly in the graphs of the entropy versus the state variables in Figure 5, where the terms in the gradient are seen as the derivatives. As the system evolve along the gradient (in a direction of increasing  $N$  and  $E$  and decreasing  $S$ ), the rate of change of entropy per unit change in phase space, becomes less and less. If the time derivatives are constants, the rate of change of entropy will therefore decrease.

When we take the gradient of the conventional metric space, all the distances are well defined. We can measure distances not only along each dimension  $x$ ,  $y$  and  $z$  separately, but the distance,  $d$ , of vectors with non-zero values in each of these dimensions can be found from  $d = \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}$ . If we follow the gradient in such a space, it corresponds to the shortest distance, for the largest increase in the scalar function (for instance temperature). In our case no such distance is defined. The sum of  $(\Delta S)^2 = 2^2$  and  $(\Delta N)^2 = 10^2$  for instance has no combined measure. The only distances we can measure, are separate distances in each dimension, i.e. a certain change in the number of species, a certain change in the number of individuals and a certain change in the total metabolic rate. It is therefore questionable if the concept of following the gradient makes any sense. If  $S$  decreases with one,  $N$  increases with 10 and  $E$  increases with 50, they only do so separately.

Values of the entropy for different ecosystems are given in Table 1 below. Again, I calculated the values in MATLAB by using the approximations stated in the

derivation of (7.2). Table 1 shows that different state variables will give the same value for the entropy, if they give the same ratios,  $F_1$  and  $F_2$ . A system evolving according to a rule of constant entropy could evolve from system No. 1 to system No. 2 by increasing all the state variables. If we increase only the number of individuals however, as in system No. 4, the first constraint will increase and we will get a higher entropy. Similarly, if we increase the total metabolic rate only, as in system No. 5, the second constraint will increase, and again, the entropy increases. This is again consistent with the graphs showing the entropy dependence on the different state variables. If we increase both the number of individuals and the total metabolic rate, the increase in entropy is equal to the increase in entropy due to the two effects individually.

Table 1: Values of the entropy,  $S_I$ , for selected ecosystems

<i>No.</i>	<i>S</i>	<i>N</i>	<i>E</i>	$F_1 = \frac{N}{S}$	$F_2 = \frac{E}{S}$	$S_I$
1	100	1000	10000	10	100	7.16
2	101	1010	10100	10	100	7.16
3	500	5000	50000	10	100	7.16
4	100	5000	10000	50	100	7.51
5	100	1000	50000	10	500	8.77
6	100	5000	50000	10	500	9.12
7	600	6000	60000	10	100	7.16
8	500	6000	60000	12	120	7.39
9	100	2000	8400	20	84	7.16

In thermodynamics it is common to look at the effect of mixing two systems. System No. 7 corresponds to a sum of system No. 1 and system No. 3, if, and only if, none of the two ecosystems contain the same species. The entropy of the combined system is not twice the entropy of each system alone, in fact, the entropy of the combined system is exactly the same as the two systems separately. System No. 8 corresponds to a sum of system No. 1 and No. 3 when all the species in system No. 1 is already contained in system No. 3. We see that the entropy of this system is slightly higher than the two systems alone, but far lower than the sum of the two. This can be explained by the increase in both of the constraints. No. 9 has the same entropy as No. 1, 2, 3 and 7 but different constraints.  $F_1$  is higher whereas  $F_2$  is lower.



## 7.6 Discussion

$S_I$  is the maximum entropy found by maximizing  $I_R$  with respect to the probability distribution,  $R(n, \varepsilon)$ . This maximization is done subject to the state variables,  $S$ ,  $N$  and  $E$ , which define our constraints. This procedure creates a least biased guess, i.e. the most flat distribution consistent with the constraints. From this we infer how individuals are distributed within species, and how energy is distributed over individuals. Entropy in the MaxEnt procedure is not a function of the state variables, but a function of the possible sets  $\{R(n, \varepsilon)\}$  of probability distributions satisfying the constraints. Of course, the numerical value of  $S_I$  ends up being dependent on the state variables, but these are imposed by our data. In the application of the MEP principle above, however, we are not maximizing the entropy with respect to the probability distribution, but maximizing the already maximized entropy with respect to the state variables. What does this really mean?

We can imagine two different ideal gases, in two different containers, under different conditions, and with different entropies. If both gases are closed systems in contact with a heat reservoir the entropy in each case can be found by maximizing the information entropy of the canonical distribution. The maximum entropy will depend on the state variables for the gas, but we do not maximize the entropy with respect to these, because these are not free to vary, but imposed externally. It would not make any sense to say that a gas would change its state variables in order to increase its entropy. For instance, it could not just change its number of particles. Based on this logic, it makes little sense to say that an ecosystem will change its state variables in order to increase its entropy. It seems illogical that one system will evolve "into another" system, described by a different set of constraints. Can we really compare the entropy of two systems with different sets of state variables in this way? It seem like the entropy must in some way be a relative measure, subject to the constraints of any one system.

Of course, the analogy between thermodynamic entropy and the entropy measure in METE has its limits, and a thorough analysis of the extent to which it holds is beyond the scope of this thesis. But it seems intuitive, at least, that the extensive parameters of the system should remain constant under the entropy maximization. In our case,  $N$  and  $E$  are extensive, but  $S$  is neither extensive nor intensive. If we hold  $N$  and  $E$  constant, the MEP principle predicts a decrease in the number of species. It is possible to imagine a situation also, where the energy is constant, but the distribution of metabolic rates over the individuals changes such that  $N$  can change in a way that does not alter the total metabolic rate of the system. In this case, the principle predicts an increase in  $N$  and a decrease in  $S$ . That is, for a limited amount of resources (defined as the total metabolic rate), the energy will be utilized by a few species with many individuals. The system will be limited by

the fact that  $N < E$  and  $S > 0$ , and be predicted to stop at  $N = E$  and  $S = 1$  (the approximations used to derive the entropy dependence on the state variables, however, are only really valid for  $S > 4$ ).

The entropy,  $S_I$ , depends on the two ratios of the state variables only. Even though we have three state variables, we only have two constraints. On the other hand,  $N$  and  $E$  define the upper limits of the sums and the integrals presented in Chapter 5. The values of these therefore result in different distributions. Distributions with identical ratios but different values of the state variables will thus differ by scale, but not by shape or by entropy value.

There is another way to look at the question of what a maximization of entropy with respect to the state variables means, which is based on information entropy. Shannon's second condition tells us that the information entropy for a flat distribution is a monotonically increasing function in the number of possibilities (see section 3.2). If we change the state variables in METE in such a way that the number of possibilities increases, then this should also increase the value of the entropy, because there are more "possible states". This is also consistent with the interpretation of the value of the entropy as the number of states with non-negligible probability (section 4.2). Since  $N$  and  $E$  are the upper limits of the sum and the integral, increasing these variables will increase the number of abundances and energy levels we have to include. In fact,  $d\varepsilon R(n, \varepsilon)$  measures the probability of picking a species with abundance  $n$  and that an individual within this species has metabolic rate in the interval  $(\varepsilon, \varepsilon + d\varepsilon)$ . If we increase the number of individuals, there will be more possible abundances, and therefore more possible abundances for the species that we pick. And a larger total metabolic rate increases the number of possible intervals for the metabolic rate that one species can have. In light of Shannon's theory it becomes clear that a maximization of the entropy with respect to the state variables amounts to increasing the number of possible states.

The application of the MEP principle to the information entropy in METE, based on the initial hypothesis, implies that the number of individuals and the total metabolic rate should increase, whereas the number of species should decrease. However, there is not enough information in the problem statement for us to be able to either express the state variables as functions of time, or for us to predict the time rate of change of the entropy. The only place where we might be able to obtain some kind of information involving time is from the metabolic rate. It is possible to imagine that this quantity, since it already involves time, could be used as some kind of characteristic time of the system. However, there are other reasons for why I did not pursue this. The most recent discussion questions whether it makes logical sense for our system to change its state variables in order to increase its value of the entropy. We could get around this argument, however,

---

by saying that the MEP principle is not a causal statement. It does not say that a system develops such as to maximize its entropy production because it "wants to". The MEP principle could simply state an effect of a yet to be discovered, underlying logic. This is the case with MaxEnt. Systems do not maximize their entropy, because they "want to", but we happen to find them in the MaxEnt state because this is the most likely state. In the same way, MEP could just be the effect that we observe. Even so, the predictions from this chapter do not seem correct. If they were true, why have we seen an increase in the number of species on the planet through evolution, and why are there so many species present in most ecosystems? In order to get any further at this point, I decided to look more closely at the MEP principle itself. What does it actually mean, and in what cases has it been successful?

In order to understand the literature on the subject, it was necessary first to obtain a background in non-equilibrium thermodynamics.

## 8 Non-Equilibrium Thermodynamics

The field of non-equilibrium thermodynamics provides us with a general framework for the macroscopic description of irreversible processes, in many ways similar to other branches of macroscopic physics like fluid dynamics or electromagnetic theory [30]. Equilibrium for an isolated system usually means that the probability of finding the system in any one state is independent of time. In other words, the representative ensemble is constant in time, meaning that the  $P_r$ 's are constant and all the macroscopic parameters describing the system are time-independent [24]. Non-equilibrium systems on the other hand are characterized by the presence of unbalanced potentials, or driving forces, causing fluxes and non-uniform distributions of state variables. In an equilibrium state all fluxes and forces within the system vanish.

Stationary states are characterized by state variables that are independent of time. The concept therefore includes both equilibrium and non-equilibrium states, depending on the boundary conditions imposed on the system [30]. In the case where fluxes are present, but the incoming fluxes are equal in magnitude to the ones going out, the properties of the system will (after sufficient time) not change in time, and the system is said to be in a non-equilibrium stationary state [31]. The local equilibrium hypothesis is essential in the treatment of a vast amount of non-equilibrium systems. If the system is divided into sufficiently small cells, this hypothesis states that all the laws from equilibrium thermodynamics should hold for each of these cells. This assures that the state variables, including the entropy, are well-defined [31].

In non-equilibrium thermodynamics the balance equation for the entropy plays a central role. The change in entropy of a system can be written as the sum of two terms: the entropy supplied to the system by the surroundings (through flows of heat or matter over the boundary),  $d_e S/dt$ , and the entropy produced within the system,  $d_i S/dt$ . For a system with entropy  $S$  at time  $t$ :

$$\frac{dS}{dt} = \frac{d_e S}{dt} + \frac{d_i S}{dt} \quad (8.1)$$

We use the convention that entropy is positive if supplied to the system and negative if transferred from the system to the surroundings [31].  $T(d_i S/dt)$  is sometimes called the rate of dissipation [31]. For isolated systems the second law of thermodynamics tells us that the entropy can never decrease,  $dS/dt \geq 0$  [31]. For system that exchange energy and/or matter with the surroundings the second law takes the more general form [31]

$$\frac{d_i S}{dt} \geq 0 \quad (8.2)$$

where equality refers to reversible processes (or an equilibrium condition), and inequality to irreversible processes. For closed and open systems, there is nothing in the way for the total entropy of the system to increase,  $dS/dt < 0$ , because  $d_e S/dt$  can take on any value [31]. Non-equilibrium stationary systems degrades the energy they receive in order to maintain steady state [31]. The entropy of a stationary system does not change in time, but in the non-equilibrium case the internal entropy production must then be positive. The entropy transport must therefore be equal in magnitude to the entropy production, but negative,  $d_e S/dt = -d_i S/dt$ .

The three terms in (8.1) can be written as

$$\frac{d_e S}{dt} = - \int_{\Omega} \mathbf{J}^s \cdot \mathbf{n} \, d\Omega, \quad (8.3)$$

$$\frac{d_i S}{dt} = \int_V \sigma^s \, dV \quad (8.4)$$

$$\frac{dS}{dt} = \int^V \rho \frac{ds}{dt} \, dV \quad (8.5)$$

where  $\mathbf{J}^s$  denotes the entropy flux: the amount of entropy crossing the boundary,  $\Omega$ , per unit area and per unit time [31].  $\mathbf{n}$  is the unit normal pointing out of the system [31].  $\sigma^s$  denotes the entropy produced per unit volume and per unit time inside the system,  $s$  is the entropy per unit mass and  $\rho$  is the mass density [30]. Combining (8.3)-(8.5) with the balance equation (8.1) one can derive the local balance relation [31]

---


$$\rho \frac{ds}{dt} = -\nabla \cdot \mathbf{J}^s + \sigma^s \quad (8.6)$$

where, in accordance with the second law, since the entropy production cannot be negative,  $\sigma^s \geq 0$  [31].

The entropy source strength,  $\sigma^s$ , can serve as a basis for the systematic description of the irreversible processes in a system. It is often expressed as a sum of products of a flux, and a thermodynamic force,  $\sigma^s = \sum_i J_i X_i$ . The flux,  $J_i$ , is characteristic of an irreversible process, and the thermodynamic force,  $X_i$ , is related to the non-uniformity of variables in the system (e.g. the gradient of the temperature). These expressions are usually supplemented with phenomenological equations expressing the relationships between the causes (the forces) and the effects (fluxes) [31]. These are often linear functions of the form  $J_i = \sum_k L_{ik} X_k$ , where  $L_{ik}$  denote the phenomenological coefficients. The entropy production can then be written as  $\sigma^s = \sum_i J_i X_i = \sum_{i,k} L_{ik} X_i X_k$  [30]. Under equilibrium conditions,  $\sigma^s = 0$ , and all thermodynamic forces, and fluxes are zero.

## 9 Maximum Entropy Production

We saw in section 4.2 that no physical relationships are necessary in order to derive statistical mechanics and all of its results. For predicting the course of time-dependent phenomena however, Jaynes states that knowledge of the equations of motion is needed [2]. An application of the MEP principle to METE was presented in Chapter 7. As an answer to the question regarding the validity of the initial hypothesis I provide a thorough review of the MEP principles content and meaning, including the most well known and successful application to climate studies, which was performed by Paltridge.

Attempts to find some universal function, whose extremum determines the development of a system, have been made in the field of physics for all time. Extremum principles often have the advantage of greatly simplifying calculations. Variational or extremum principles in physical systems can be applied to find the state of the system and its stability properties, to describe fluctuations, to derive equations of motion and to find constraints on the direction of processes and evolutions [32]. Examples of successful principles include Fermat's principle in optics and Hamilton's principle of least action in mechanics. In equilibrium thermodynamics, maximization of entropy for isolated systems or minimization of Gibbs's free energy for systems at constant temperature and pressure are important examples of successful variational principles. In non-equilibrium thermodynamics, however, such principles are limited [31]. Prigogine introduced the principle of minimum entropy production in 1947 to describe the evolution of non-equilibrium dissipative

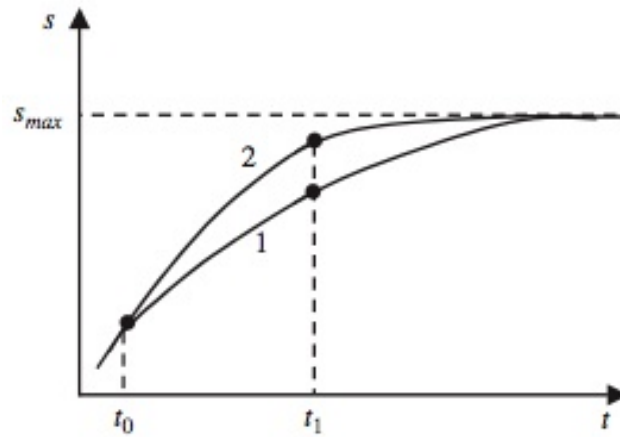


Figure 6: Variation of entropy  $s$  with time  $t$  for two possible trajectories of the development (from Martyushev et al., 2006 [6, p. 7])

structures. The Maximum Entropy Production principle (MEP) is known much less than Prigogine's principle, but has been used by several scientists throughout the 20th century to deal with both general theoretical issues in thermodynamics and statistical physics and to solve specific problems [6]. In short, The Maximum Entropy Production (MEP) principle states that [6]:

a non-equilibrium system develops so as to maximize its entropy production under preset constraints.

The second law of thermodynamics tells us the direction of evolution. The MEP principle on the other hand not only states that an isolated system tends to the state with maximum entropy, but that it does so in the shortest amount of time, or equivalently, along the shortest possible path [6]. This is illustrated in Figure 6, where the principle states that path 2 is preferred to path 1 because it yields a larger entropy production between  $t_0$  and  $t_1$ . The principle therefore adds to the second law by including not only the direction of evolution, but also the rate of evolution [6].

In a way similar to MaxEnt, the MEP principle has great potential in systems where there is little information available to characterize the system's state [33]. Unlike the theory of MaxEnt, which is a well-established and widely used method for statistical inference, the MEP principle is neither a well-known nor a well-established theory [5]. Paltridge himself was not able to explain why the principle works as a description of the earth's climate system and due to a lack of a theoretical justification for most of the time since Paltridge's studies, MEP has been regarded as nothing more than a curiosity by many. Dewar, however, recently

published a proof of the principle which, even though it is not regarded as rigorous, again has motivated an expansion of the principles into other fields. In search of the correct way to apply the principle in the framework that METE provides, I have reviewed the literature on the principle from Paltridge's work in the 1970s through later applications to other problems and up to the recent proof published by Dewar along with the subsequent responses. Paltridge's work is central not only in the development of the MEP principle itself, but because it is presented as an analogy to the application of the MEP principle in METE in Harte's book, *Maximum Entropy and Ecology* [1].

I will present Paltridge's studies in more detail in the coming section, but first I will comment on the difference between MEP and the minimum entropy principle, developed by Prigogine. This is the best known extremum principle in non-equilibrium thermodynamics. From the name alone, this principle might sound contradictory to the MEP principle. However, it applies to a different situation. Prigogine's minimum entropy principle holds only in the linear case, in which case the system's state is already determined. In this case there is only one steady state solution, and the minimum entropy production principle simply states that this state is the one with minimum entropy production compared to any non-steady condition the system might be pushed to [34, 3]. What this means will become more clear after reading the coming chapter on Paltridge's model. The point is that Prigogine's principle does not add anything when it comes to solving for the steady state. The principle is merely a reformulation of the conservation laws [31]. The principle of maximum entropy production on the other hand applies to problems where the boundary conditions are not fixed, and where a multitude of steady-states are possible [34]. The MEP principle therefore works as a selection principle that complement the conservation laws and make it possible to determine the steady state realized by the system [35].

## 9.1 Paltridge's climate model

The MEP principle first gained popularity after a particularly successful application by Paltridge in 1975 in the prediction of Earth's climate. His model is powerful because it allows one to make accurate predictions without the need for detailed consideration of the complex internal dynamics of the climate system [34]. Paltridge treats the earth as a classically closed thermodynamic system in steady state. The system is not in equilibrium and as a consequences work has to be performed in order to maintain the steady state. This is done by a range of earth system processes, the most important one being the atmospheric and oceanic circulation, which degrades the energy from the sun. The global circulation transports heat from warmer (equatorial) to colder (polar) regions and in doing so produces thermodynamic entropy [33].

The central constituents in Paltridge's treatment are the net incoming shortwave radiation from the sun, the heat flux from warmer to colder regions, and the net outgoing longwave radiation. The incoming radiation is taken as an external constraint imposed by the sun, whereas the outgoing radiation is dependent on the temperature via the Stefan-Boltzmann black-body relationship,  $\sigma T^4$ . In Paltridge's model from 1975 both the incoming and outgoing radiation is also affected by the albedo (more specifically, cloud cover), but I will not discuss this dependence here. Because the solar input per area is larger at the equator than at the poles, the temperature at the equator is larger than at the poles. If there was no heat flux between the equator and the poles, the temperature of each region would be determined by the steady-state requirement for that region alone. A positive heat flux from the equator to the poles, however, increases the temperatures near the poles, and reduces the temperatures near equator, such that the temperature difference becomes smaller. The maximal heat flux corresponds to the value at which the temperature difference between equator and pole is zero. At this point, no more heat will flow. The temperature difference is therefore a decreasing function of the heat flux, where every combination yields a possible steady state for the system. The observed value of the heat fluxes and the temperatures found on Earth lies somewhere in between the limits. Paltridge's goal was to predict this, and the albedo, at the different latitudes. In order to choose one pair of values among all the possible pairs, Paltridge applied the principle of Maximum Entropy Production to the earth system. According to this principle the preferred steady state of the system should be the state that maximizes entropy production [3]. At the time this had been shown to hold true for certain small-scale convective heat transfer processes. By applying the MEP principle to Earth, Paltridge was able to predict meridional profiles of surface temperature, cloud cover, atmospheric and oceanic energy fluxes, and atmospheric and oceanic dissipation in very good agreement with observation [3].

More specifically Paltridge created a simple two-dimensional model where the entire atmosphere is divided into 10 boxes, each one characterized by uniform values of the temperature, and each one subject to the steady state assumption. The energy balance for each box is shown in Figure 7. As mentioned above the maintenance of the non-equilibrium steady state requires a positive internal entropy production from irreversible processes,  $\dot{S}_i > 0$ , where we have introduced the dot to denote time derivatives. In order for the total entropy of the system to be constant, as it should for a steady state, the export of entropy over the system boundary must equal the total rate of internal entropy production,  $\dot{S}_i = -\dot{S}_e$ . That is, the entropy production due to the global circulation, which delivers heat from warmer to colder regions, must be exported to the universe. According to Paltridge, the MEP principle states that the internal entropy production,  $\dot{S}_i$ , is



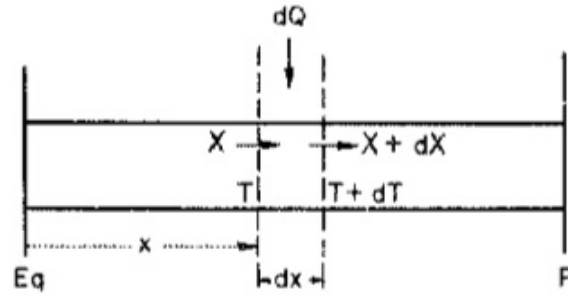


Figure 7: Energy balance of a box in Paltridges model.  $X$  is the meridional heat flux,  $T$  is the temperature and  $dQ$  is the net radiative input (from Paltridge, 1978 [3, p. 928]).

maximized, which in a stationary case is identical to a minimum in the entropy exchange,  $\dot{S}_e$  [34].

In each box (Figure 7) we get a net flowrate of entropy [3]:

$$d\dot{S}_e = \frac{X}{T} - \frac{(X + dT)}{(T + dT)} + \frac{dQ}{T} \quad (9.1)$$

Using the energy balance,  $dQ = dX$ , and the entropy balance,  $d\dot{S}_i = -d\dot{S}_e$ , Paltridge rewrites the above equations as [3]:

$$\dot{S}_i = - \int_{T_{Eq}}^{T_p} d\dot{S}_e = - \int_{T_{Eq}}^{T_p} \frac{X}{T^2} \quad (9.2)$$

The total rate of dissipation is found by multiplying this expression by  $T$  and the MEP principle is equivalent to a maximum dissipation principle [3].

In the linear case, the flux would be proportional to the temperature gradient,  $X = k(dT/dx)$ , where  $k$  is some constant transfer coefficient. This is the case of linear phenomenological relationships between the flux and the thermodynamic force as presented in section 8. This case is related to Prigogine's minimum entropy principle, which was discussed at the end of the previous section. Paltridge instead assumes that there are sufficient degrees of freedom in the dynamics and thermodynamics of the system to allow any steady state satisfying energy balance and boundary conditions [3]. He therefore considers what he claims is a more realistic non-linear situation, where  $k$  is unconstrained. Each steady state corresponds to a different value of the entropy production,  $\dot{S}_i$ , where zero value occurs only in the limiting cases of zero flux and zero temperature difference.

For a classic closed steady state system we can also express the internal entropy

production through the entropy exchange as

$$\dot{S}_i = -\dot{S}_e = - \int \frac{dQ}{T_a} dA \quad (9.3)$$

where  $dQ$  is the net rate of radiant energy input to the area element  $dA$  of boundary surface.  $T_a$  is the temperature at the boundary, i.e. at the top of the atmosphere [3]. The net incoming radiation in each box is non-zero because some of the incoming energy is transported from warmer to colder regions. Negative values of  $dQ$  at mid-latitudes give rise to positive values at higher latitudes, such that the net radiant energy into the total system is zero. The largest positive values of  $dQ$  are found at the equator and the largest negative values are found at the poles.

Paltridge calculates the total entropy production in his 10-box model using the discrete version of (9.3):

$$\dot{S}_i = -\dot{S}_e = - \sum_{i=1}^{10} \frac{(F_S - F_L)_i}{T_{ai}} \quad (9.4)$$

where  $F_L$  and  $F_S$  denote the longwave and shortwave flows of radiant energy into the top of each box and  $T_{ai}$  is the temperature at the top of the atmosphere [3]. Due to energy balance, the net incoming radiant energy in any one box equals the net outgoing heat flux from that box to the neighboring boxes. From this, Paltridge finds the unique set of fluxes and temperatures leading to a maximum in  $\dot{S}_i$  [3]. In addition to the simple, two-dimensional 10-box model Paltridge also included a three-dimensional 400-box model. Broad agreement with observation was found both with the 10-box model, and the 400-box model [3]. In 2007 Paltridge et al. used the MEP principle to investigate the effect of cloud cover on feedbacks to climate change [36]. They then found that cloud feedback will reduce the temperature response to doubled  $CO_2$  slightly at high altitudes and slightly amplify it at low altitudes, but they point out that the study only can provide qualitative results with high uncertainty [36].

There are subtleties in Paltridge's method that I have chosen not to include (e.g. in the assignment of average temperature, and assumptions about fluxes between the ocean and atmosphere), but none of these factors should affect the general results in any significant way. The more interesting critique of Paltridge's work is directed towards the choice of the entropy production that is maximized. Ozawa et al. show that only a small part of the total entropy production on earth is included in Paltridge's analysis [35]. The entropy production associated with the direct absorption of solar radiation at the surface, in the atmosphere, and from the surface to the atmosphere is not taken into account. This has been a central objection to Paltridge's studies.

Ozawa et al. studies the earth as a heat engine operating between thermal reservoirs at two different temperatures, the temperature at the equator and at the poles. Earth receives radiant energy from the sun at a very hot temperature (5800K), transports heat from warmer to colder regions via the atmosphere and the oceans, and reemits radiant energy to outer space at a very low temperature (4K) [35]. The amount of radiation entropy can be found as the flux of radiant energy divided by the brightness temperature,  $dS_{rad} = \delta Q_{rad}/T_{br}$ . The brightness temperature of the radiation from the sun is that of the sun,  $T_{br} = T_{sun}$ . The total entropy produced by earth is then  $\dot{S}_{tot} = \int_A (F_L/T_a - F_S/T_{sun}) dA$ , which due to the high temperature of the sun is much larger than the entropy production calculated by Paltridge [35]. Paltridge's expression for entropy production corresponds to the rate of entropy production due to turbulent dissipation only. In reality, the turbulent part of the entropy production contributes to only about 5% of the total rate. However, Ozawa et al. conclude that it is nonetheless this term that tends to be maximized in the climate system [35]. They argue that this is because the absorption only depends on the material under consideration and cannot be altered by changing fluxes. Radiation can therefore be viewed just as an energy source for the climate system, that is, as a constraint. Ozawa et al. therefore conclude that the entropy production due to direct absorption of solar radiation is irrelevant to the maximized entropy associated with the turbulent flux from equator to poles [35].

## 9.2 Other applications

The application of the MEP principle in the framework of METE was motivated by successful applications of the principle in a variety of fields. Paltridge's climate studies provides the most important example of a successful MEP application and the original hypothesis presented in Chapter 7 was based on transferring the principle from thermodynamic entropy production in the Earth system to production of the information entropy defined in METE. The analogy between our ecosystem and the climate system, however, has its limitations, which will be discussed in greater detail in Chapter 11. The lack of both a recognized theoretical foundation and an established MEP procedure increases the need of a valid analogy to base our approach on. In order to get an impression of the general applicability of the principle and to look for an application that is closer to what we have in METE, I conducted an extensive literature review on the subject. I present the most relevant alternative applications here.

Paltridge's predictions of Earth's climate using the MEP principle was initially dismissed as a coincidence by some [33]. This, however, has become harder to claim, after it was shown by Lorenz et al. in 2001 that also Mars and Titan operate in MEP states [35, 33]. Related to the case for the global climate are

Suzuki and Sawada, and Chen and Wangs studies of Bernard-type convection cells in the 1980s [6]. They obtained multiple steady states under the same boundary conditions but they found that these states were not equally stable against perturbations. In fact, the steady states tended to shift towards states of higher entropy production following a perturbation, thereby suggesting that the MEP state is in some way a preferred state [35].

Since Paltridge's time, promising results have also been obtained for non-equilibrium systems in a range of other studies, including crystal growth morphology, kinetic features of solids and bacterial metabolism and photosynthesis [20, 6]. Martyushev et al. also show how the principle can be used to deduce the Kirchoff law for an electric circuit, to characterize the behavior of a body in a radiation field, to describe the evolution of structures in a plasma, to describe the behavior of a periodic array of Josephson Junctions and to determine what chemical reactions will take place among a set of possible alternatives [6]. The list of successful applications is long and serves as a strong motivation for an expansion of the principle into other fields of research. Martyushev et al. also state that the principle shows the greatest promise in astrophysics and biology, in particular ecology [6].

Paltridge's studies suggests that the Earth system maximizes entropy production. The relevant entropy production in his case comes from the turbulent atmospheric and oceanic flows. Organisms and ecosystems however, also make a contribution to the entropy production on Earth and an interesting questions is whether these systems are also guided by the MEP principle. Several energy-based principles have been suggested in order to explain how ecosystems organize themselves. One of these, first presented by Lotka in 1922, is a principle that says that ecosystems evolve to maximize the energy flux through the system under the given constraints. He also stated that species which utilize the available energy most efficiently (all other things being equal) will increase their population and as a consequence the total flux of energy through the system. Lotka's principle was further developed as the maximum power principle by Odum and Pinkerton in the 1950s which says that the available energy degrades at the maximum possible rate [33]. Schneider and Kay suggested that ecosystems attain states of maximum dissipation in the 1980s. This is really equivalent to a maximum entropy production principle, but they avoided this term as the entropy is not rigorously defined far from equilibrium [37]. All of the above principles are only slightly different versions of the same concept, and concerns rates of entropy production when the boundary conditions are not fixed [33].

MaxEnt has been used successfully as a completely statistical method to infer from incomplete data in economics in a way that is separate from all thermodynamic principles. I therefore looked to economics again for applications of the

---

MEP principle in Kleidon and Lorenz' *Non-equilibrium Thermodynamics and the Production of Entropy* where I found that Ruth applies concepts from both equilibrium and non-equilibrium thermodynamics in order to describe and understand economic activity [38]. The MEP principle in economics is thus strongly related to thermodynamic entropy.

## 10 A study of stability in Paltridge's models

Paltridge's model predicted the empirical values of temperature, heat fluxes and cloud cover very well. In an attempt to justify the MEP state he looked at fluctuations in the heat fluxes and claimed that these will push the system towards the point of maximum entropy production over time. Martyushev et al. however point out that his argument depends on new assumptions that are not themselves justified [6]. Could it be that the reason why the system is found in the MEP state, is that this state is in some way more stable than other steady states? Following a suggestion by Harte, I here present a brief investigation of this question. The aim is not to obtain any accurate values, but to look at the qualitative behavior of the model in the different steady states.

### 10.1 Theory of Stability

An equilibrium is defined to be stable if all sufficiently small perturbations around this point damp out in time [39]. By linearizing about the fixed points, we can say something about the rate of decay or growth of perturbations. We consider the system

$$\dot{x} = f(x, y) \tag{10.1}$$

$$\dot{y} = g(x, y) \tag{10.2}$$

where  $(x^*, y^*)$  is a fixed point such that

$$f(x^*, y^*) = 0 \tag{10.3}$$

$$g(x^*, y^*) = 0 \tag{10.4}$$

$$\tag{10.5}$$

We then let  $u = x - x^*$  and  $v = y - y^*$  denote small perturbations around the fixed point. We want to know whether these perturbations grow (unstable equilibrium) or decay (stable equilibrium) over time. In other words, we want to know the value of  $\dot{u}$  and  $\dot{v}$  at the fixed point. This is given by

$$\begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix}_{(x^*, y^*)} \begin{pmatrix} u \\ v \end{pmatrix} + O(u^2, v^2, uv) \quad (10.6)$$

where  $O(u^2, v^2, uv)$  denote the quadratic terms. These are small and in most cases safe to neglect [39]. The matrix with the partial derivatives is called the Jacobian, and it is to be evaluated at the equilibrium point. For the equilibrium to be asymptotically stable, all the eigenvalues of the Jacobian matrix must have negative real parts [40].

## 10.2 Stability in Paltridge's model

Instead of a detailed 10-box model including variables for cloud cover, we use a simple three-box model with one box for the equator, and two for the poles, as shown in Figure 8. The two poles are treated as identical. It is assumed that the number of boxes will not alter the results qualitatively. Average temperatures are  $T_e$  for the equator and  $T_p$  for the two poles. When the system is not in a steady state, the total energy of the boxes will increase or decrease by an amount equal to energy in minus energy out. It is assumed that all of this energy is stored as heat. The rate of change of the temperature in each box can then be written as  $dQ_e/dt = c_e dT_e/dt$  at the equator and  $dQ_p = c_p dT_p/dt$  at the poles, where  $dQ/dt$  is the rate of energy into the system minus the rate of energy out of the system.  $c_e$  and  $c_p$  denote the heat capacities which, for convenience, are set equal to 1. The energy balance equations for equator and pole, in units of Watts per meter are then

$$dT_e/dt = \Omega_e - 2Q - \sigma T_e^4 \quad (10.7)$$

$$dT_p/dt = \Omega_p + Q - \sigma T_p^4 \quad (10.8)$$

where  $\sigma$  is Stefan-Boltzmann's constant,  $Q$  is the energy flux from the equator to the poles, and  $\Omega_e$  and  $\Omega_p$  represent the net incoming solar radiation at the equator and the poles respectively. In the stationary state, the temperatures do not change and the left hand sides of equation (10.7) and (10.8) are zero. Combining them in this case gives

$$\Delta T = \frac{1}{\sigma^{1/4}} [(\Omega_e - 2Q)^{1/4} - (\Omega_p + Q)^{1/4}] \quad (10.9)$$

where  $\Delta T = T_e - T_p$ . Equation (10.9) defines the curve of stationary states in variables of the thermodynamic force,  $\Delta T$ , and the flux,  $Q$  [6]. The entropy production in each of the stationary states is different. The MEP principle as used by Paltridge implies that the steady state chosen by the system should be

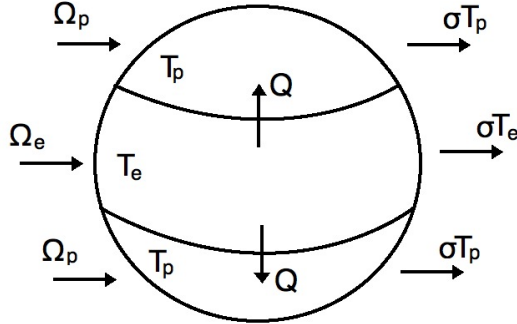


Figure 8: Energy input and output in the 3-box model.  $\Omega_{e,p}$  denote the incoming solar radiation,  $T_{e,p}$  denote the temperature of the region and  $\sigma T_{e,p}$  denote the outgoing black-body radiation.  $Q$  is the heat flux from equator to poles.

the one maximizing the entropy production. We find the entropy production,  $\sigma_s$ , as a function of the heat flux and the temperatures by using a 3-box version of equation (9.4), with  $(F_S - F_L) = \Omega - \sigma T^4$  for the equator and the poles resulting in

$$\sigma_s = \frac{2Q}{T_p} - \frac{Q}{T_e} - \frac{Q}{T_e} = 2Q \frac{\Delta T}{T_p T_e} \quad (10.10)$$

In order to solve for the maximum of  $\sigma_s$  one can use the steady state versions of (10.7) and (10.8) to substitute for the temperatures and solve with respect to  $Q$ . The MEP temperatures,  $T_e^*$  and  $T_p^*$  are then found by plugging the MEP flux,  $Q^*$ , back into the steady state equations.

For any given  $Q$ , the dynamical behavior of the temperatures is described by equation (10.7) and (10.8). If  $Q$  is a given constant, the steady state values of  $T_e$  and  $T_p$  will be uniquely defined, and the dynamical behavior of each one is decoupled. It is easy to see that this equilibrium must be stable, because a small perturbation in any of the temperatures will give a time derivative of the temperature in the opposite direction, leading the system back to the equilibrium. This however, is not the case we are interested in. More often, a linear relationship between fluxes and forces is assumed,

$$Q = k(T_e - T_p) \quad (10.11)$$

where  $k$  is a proportionality constant. In many systems  $k$  is found as a phe-

nomenological parameter that describes the underlying relationship between the fluxes and the gradients in the system. If we know  $k$  we can substitute this expression for  $Q$  in (10.7) and (10.8). This gives a set of coupled equations which can be solved for the temperature in the steady state. Therefore, a known  $k$  will define the entire system. We will instead look at the case where the value of  $k$  is given by the criterion that this is the value resulting in the MEP state of the system. Then this  $k$  will define the relationship between heat flux and temperature difference by equation (10.11). In fact, the set of variables,  $k$ ,  $Q$ ,  $T_e$ ,  $T_p$  and  $\Delta T$  are all determined when one of them is given and define a unique set of equilibrium values.

I used MATLAB to compute the temperatures along with the values of  $Q$  and  $k$  for all the possible steady states between  $Q_{min} = 0$  and  $Q_{max} = (\Omega_e - \Omega_p)/3$  (found by setting  $T_e = T_p$ ). I then computed the eigenvalues of the Jacobian in all of these states, and plotted the values as functions of  $k$  and  $Q$ . I also plotted the value of the entropy production as a function of  $Q$ , in order to study whether the maximum entropy state had any particular features compared to the other steady states.

The values of the incoming radiation are chosen to get values in a range that is reasonable for a simple model of Earth's climate, but have no physical significance beyond this. I used  $\Omega_e = 600 \text{ W/m}^2$  and  $\Omega_p = 200 \text{ W/m}^2$ .  $\sigma = 6 \times 10^{-8} \text{ W/(m}^2\text{K}^4)$  was used as Stefan-Boltzmann's constant. All the plots were tested with varying parameters, and no qualitative changes were observed.

In the steady state the temperatures are given by

$$T_e = [(\Omega_e + Q)/\sigma]^{1/4} = [(\Omega_e + k(T_e - T_p))/\sigma]^{1/4} \quad (10.12)$$

$$T_p = [(\Omega_p - 2Q)/\sigma]^{1/4} = [(\Omega_p - 2k(T_e - T_p))/\sigma]^{1/4} \quad (10.13)$$

The temperatures and  $k$  are plotted as functions of  $Q$  in Figure 9. We see that for  $Q = 0$  the temperature difference is at maximum.  $T_e$  is maximum (316 K) and  $T_p$  is minimum (240 K). As  $Q$  increases, the temperature difference decreases. Equator and poles reach the same temperature (273 K) when  $Q = Q_{max}$  ( $133 \text{ W/m}^2$ ).  $k$  increases faster and faster as  $Q$  increases (because the temperature difference decreases). As the temperature difference goes to zero,  $k = Q/(T_e - T_p)$  goes to infinity.

The stability of the system at the different equilibrium points is found by looking at the eigenvalues of the Jacobian. If we call  $dT_p/dt = f$  and  $dT_e/dt = g$  we can write the Jacobian as



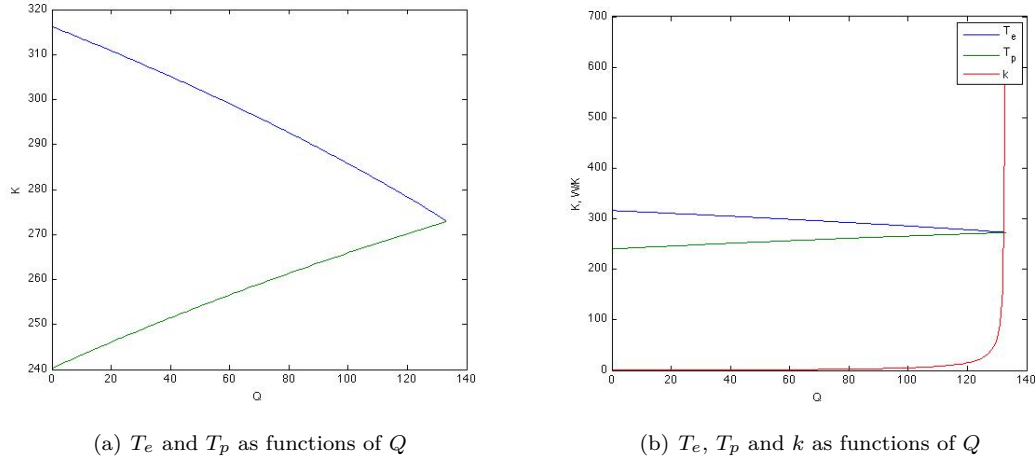


Figure 9:  $T_e$ ,  $T_p$  in Kelvin,  $K$ , and  $k$  in  $W/K$  plotted against  $Q = k(T_e - T_p)$  for  $0 < Q < 0.95Q_{max}$ .

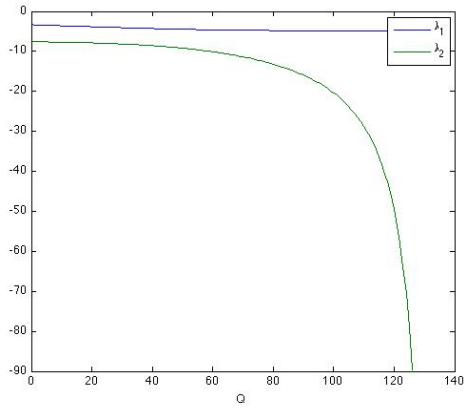
$$J = \begin{pmatrix} \frac{\partial f}{\partial T_p} & \frac{\partial f}{\partial T_e} \\ \frac{\partial g}{\partial T_p} & \frac{\partial g}{\partial T_e} \end{pmatrix} \quad (10.14)$$

The eigenvalues are then given by the equation

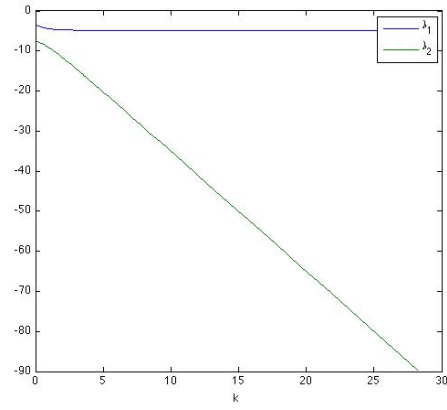
$$\left(\frac{\partial f}{\partial T_p} - \lambda\right)\left(\frac{\partial g}{\partial T_e} - \lambda\right) - \frac{\partial f}{\partial T_e} \frac{\partial g}{\partial T_p} = 0 \quad (10.15)$$

The two eigenvalues were computed in MATLAB and are plotted as functions of  $Q$  and  $k$  in Figure 10. We see that the eigenvalues are real and negative for all the possible steady states (between  $Q = 0$  and  $T_e = T_p$ ).  $\lambda_1$  is roughly constant for all steady states, whereas  $\lambda_2$  decreases with  $Q$  (and  $k$ ). The more negative the eigenvalues, the faster the return to equilibrium after a perturbation. The states at larger values of  $Q$  (and  $k$ ) are in this sense more stable. It can also be shown directly from the expression of the eigenvalues that they are real and negative for any values of the temperatures.

The entropy production is plotted versus  $Q$  in Figure 11. The maximum value of the entropy production,  $(ds/dt)^* = 0.07 W/(m^2K)$  is found at  $Q^* = 68 W/m^2$ . This corresponds to  $T_e^* = 297K$  and  $T_p^* = 259K$  and  $k^* = 1.8$ .



(a)  $\lambda_1$  and  $\lambda_2$  as functions of  $Q$



(b)  $\lambda_1$  and  $\lambda_2$  as functions of  $k$

Figure 10: Eigenvalues,  $\lambda_1$  and  $\lambda_2$ , shown as functions of  $Q$  and  $k$ .

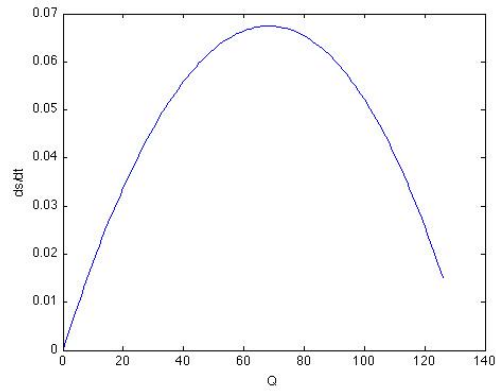


Figure 11: Entropy production,  $ds/dt$ , as a function of  $Q$ .

### 10.3 Discussion

The above results show that the MEP state is not qualitatively different from the other states when it comes to the stability of the system. The match between Paltridge's predictions and the empirical values can therefore not be explained on the grounds of stability. Even though I have only discussed perturbations in the temperature these perturbations also involve perturbations in  $Q$ , because the heat flux is linked to the temperature difference through  $k$ . If  $k$  itself could increase (corresponding to a larger flux per temperature difference), this will simply lead the system to a new steady state, with a higher  $T_p$  and a smaller  $T_e$ . Our equations only give direct information about the dynamical behavior of the temperatures. It is possible to imagine perturbations in the net incoming solar radiation also. In reality this depends on the albedo, the reflection coefficient of the surface. A high albedo means that a large portion of the solar radiation is reflected back into space and a low albedo means that a larger fraction will be absorbed by Earth. In order to study this effect, we could rewrite the radiation terms as  $\Omega = \Omega^0(1 - a(T))$  where the albedo,  $a(T)$ , depend on the temperature of the box. Typically, a higher temperature at the poles will lead to ice melt, which will reduce the albedo. This again will increase the net incoming radiation, which again will increase the temperature. Including the albedo into the equations might therefore render the system unstable (for all possible steady states) because it adds a positive feedback mechanism. The albedo is also linked to cloud cover, a variable that is responsible for one of the larger sources of uncertainty in climate modeling. More cloud cover not only reduces the incoming radiation due to a higher reflection coefficient, it also reduces the outgoing radiation by reflecting the longwave radiation from the surface of Earth back down to the surface.

I looked briefly at the albedo and found no evidence that the MEP state became any more different from the other states when this was included. From the above stability analysis, there seems to be no sign of a qualitative difference between the MEP state and the other states, rendering the MEP state unique in any way. Even if we change the model a little, it doesn't seem plausible that a qualitative difference in regard to stability would present itself. This statement requires more investigation in order to be established as true. I ended my investigation of the stability with the above results.

## 11 Discussion of MEP analogies to METE

The implications of assuming that the rate of production of the information entropy,  $S_I$ , is maximized were investigated in Chapter 7. This investigation was based on an early interpretation of the MEP principle as a principle stating that

a system will evolve according to the path of maximum entropy production. The review of the MEP principle in Chapter 9 provides us with a better background necessary for a discussion of the applicability of the MEP principle to METE in more detail.

The applications mentioned in Chapter 9 include planetary climates, convection cells, maximum power in ecosystems and studies of the economy among other examples. All of these applications are, however, related to thermodynamic entropy. For now, therefore, we can still only assume that the principle will hold also for information theory, without proving this.

The definition of the MEP principle stated by Martyushev et al. and given in the beginning of Chapter 9 states that: a non-equilibrium system develops so as to maximize its entropy production under preset constraints. This definition and Figure 6 indicates that the principle refers to the evolution of a non-equilibrium system as it approaches equilibrium, where the equilibrium corresponds to the state where the entropy of the system attains its maximum possible value. This brings up the question of how to look at the equilibrium concept in the framework of METE. An equilibrium state is characterized by state variables that are constant in time. By this definition, our system is not in equilibrium (if it was, the application of MEP would become meaningless, because the goal is to say something about the dynamical behavior). The MEP principle then tells us that the system should approach its equilibrium at the fastest possible rate. The analysis in Chapter 7 and the graphs in Figure 5, however, show that the entropy in our case is unbounded. The system does not have an equilibrium in this respect. Another way to look at equilibrium is in analogy to a quasi-static thermodynamic process. Here, the state variables are changing, but the change is slow enough for us to say that the system is still in equilibrium at every step in the process. This means that the entropy is maximized at every set of the state variables, and that the system is in equilibrium because it can be described by the MaxEnt distribution. It seems more plausible to say that something like this is the case for our system. The system is then always in equilibrium in the sense that the information entropy,  $I_R$ , takes on the maximum value and that  $R(n, \varepsilon)$  is described by the MaxEnt distribution for every set of  $S$ ,  $N$  and  $E$ . The value of the maximum entropy then changes according to the state variables. A third way to look at equilibrium is to look at whether there are fluxes present. We do indeed have a flux of energy going through the system in the form of the metabolic rate. The question, however, is whether this flux really describes a flux in our representation of the system. Even though the metabolic rate represents a flux from a thermodynamic perspective, there is nothing in the theory of METE that separates the state variable  $E$  from  $S$  and  $N$ . That is,  $E$  is simply not defined as a flow (it has no spatial component and no direction). We conclude that there are no fluxes in METE. According to

---

this view, our system is also in equilibrium.

Paltridge's work was intended as the main analogy for the application of MEP to METE. We can now evaluate the strength of this analogy. The climate system is in the non-equilibrium steady state for which the thermodynamic entropy is maximized. The difference between thermodynamic entropy and information entropy is set aside for now. For stationary states, the state variables are independent of time, and we have already stated that this is not the case for our ecosystems. We have also established that it makes little sense to regard our system as a non-equilibrium system. On these grounds, the analogy is starting to look weak.

The presence of fluxes is central to the entropy production in Paltridge's case, where the entropy depends on the product of the temperature difference and the flux as shown in equation (10.10) for the simple 3-box model. This also corresponds with Chapter 8 where I showed that thermodynamic entropy can be expressed in terms of products of fluxes and forces. If either the heat flux is small or the temperature difference is small, the entropy production is small in Paltridge's case. The maximum is found somewhere in the middle, as shown in Figure 11. Fluxes are central also to the other applications mentioned in section 9.2. These fluxes need not be heat but can be electrons or chemical components, but the entropy production is always expressed as a product of fluxes and forces. In all the applications, it is the thermodynamic entropy that is maximized. All of the examples describe non-equilibrium systems, and most describe stationary processes, even though some describe the evolution of systems.

There is another difference between Paltridge's application and ours. Because the Earth is in a steady state, its entropy is by definition constant. Paltridge is only maximizing the internal entropy production of the system. This entropy is exported to the universe, such that the total entropy of the system remains constant. In our case, we are simply maximizing the one entropy measure that we have,  $S_I$ . Is this the correct part of the entropy to maximize? Why would this entropy increase at all, if the entropy of the climate system does, in fact, not? Moreover, even the part of the internal entropy that is maximized by Paltridge is not straight forward, and has been questioned by several authors [35]. Ozawa et al. concluded that it is the irreversible part of the entropy production that should be maximized, and that this is equivalent to the turbulent part not including the direct absorption. Can we determine a part of the entropy in our system that is analogous to an irreversible entropy production? If  $S_I$  then represents the total entropy and is set to a constant this does, however, not imply constant state variables, because the same value for the entropy can be obtained from several combinations of the constraints. At this point the comparison gets confusing.

Any non-equilibrium system has a defined equilibrium state. For Earth, this would be the state where it is in equilibrium with the surrounding universe. And if we look closely, the Earth is not really in a stationary state. For any system that uses energy from the surroundings, this energy will at some point run out. The temperature of the sun will ultimately decrease, whereas the temperature of space will increase (yet in this case by no significant amount, because the energy of the sun is small in comparison to the rest of the universe), until there is no longer a difference, and no longer a potential for work. Without an external energy source Earth cannot be kept out of equilibrium, and will in the end equilibrate with its surroundings. All the entropy produced by Earth is exported to the universe, and increases the entropy there. According to Ozawa et al. the contribution to the entropy production from direct absorption is determined by the boundary conditions and therefore only acts as a constraint [35]. Thus, the maximum in the turbulent entropy production seems to coincide with the maximum in the total entropy production for the isolated system: Earth and universe combined. Viewed in this way it seems like we are back where we started: for an isolated system, the entropy will increase at the fastest possible rate. Can we apply this to our system?

Our system, however, is not isolated. It is constantly interacting with its surroundings, where it gets its energy from, and where it gets new individuals from. It might therefore only be regarded as a subsystem, equivalent to Earth, for which the entropy is actually constant. Because the possible stationary states for the climate system are restricted by the physical constraints of the incoming solar radiation and the maximum heat flow, the total rate of entropy production in the combined universe and Earth system is bounded. This is important. If it was not for this we would not be able to determine a maximum entropy production state. As stated earlier, we have not imposed any bounds on  $S_I$ . Of course, any ecosystem cannot use more energy than what is delivered from the sun, but setting this as the upper limit would only lead to the conclusion that any ecosystem will eventually use all the available energy, a conclusion that does not make much sense. A principle for the total or partial entropy production of a system is only interesting given that we have identified constraints that limit the available states of the system.

In a case where the total thermodynamic system approaches equilibrium, the thermodynamic entropy reaches its maximum subject to the constraints on the total system. The representative ensemble in the equilibrium state will be described by the MaxEnt distribution. In the case of a closed system in thermal contact with a heat reservoir this is the canonical distribution. But an isolated system can be approximated by the canonical distribution to a good degree. The relaxation of the total system to equilibrium corresponds to the probability distribution approaching the MaxEnt distribution. It is the distribution, not the constraints on

---

the total system, that changes during this relaxation process. In the application in Chapter 7 the state variables defining the constraints themselves were changed in order to change the entropy, a method that was questioned in the discussion. In the case for Paltridge, however, the temperature difference and the heat flux are chosen so as to maximize the internal entropy production. These quantities are in many ways similar to a form of constraints, but in this case, the problem is bounded, and we can actually find a solution.

My analysis in Chapter 7 and review in Chapter 9 gives little hope for an application of the MEP principle to  $S_I$  based solely on an analogy with Paltridge's work or any of the other applications. Even if we ignore the difference between thermodynamic entropy and information entropy, the logic from Paltridge is not easily transferred to our problem. And in the end, we cannot escape the obvious difference between thermodynamic entropy on one hand, and information entropy on the other. The analysis up till now has showed the complexity of the method but left us with few conclusions when it comes to the application of the MEP principle to information theory.

## 12 Dewar's proof

There exists no established algorithm for how to apply the Maximum Entropy Production principle to a given system. My work was initially guided by the analogy with Paltridge's studies, but it is shown that this analogy is limited. In order to extend the principle and apply it in a way that is valid in the framework of METE, a deeper understanding of the principle is needed. To see what the principle actually says, in which cases it holds, what the underlying assumptions are and how it should be applied, I went through the most fundamental work on the principle. This was published in three papers by Dewar in 2003 and 2005, in which he derives the MEP principle, as well as the Fluctuation Theorem and the principle of Self-Organized Critically, all based on the MaxEnt formalism [4, 5, 20]. His derivation lead to renewed interest in the MEP principle, but is still not generally accepted as a rigorous proof. The derivations are difficult to follow in detail and have been met with criticism [41, 35, 42]. I have not included this critique because it is outside the scope of my thesis, but I have included my own critical remarks at the end. One of the problems, also noted by Ozawa, is that Dewar does not provide any examples [35]. An example would have made it easier to both believe the derivations, understand the principle, and see how it can be applied in general. An improved proof is underway from Dewar, and the current one should be regarded only as incomplete. I present the parts that I believe are most related to my work with METE. I have included both a conceptual explanation and a more detailed mathematical derivation.

Dewar's derivation of the Maximum Entropy Production principle is really an abstraction of Jaynes's MaxEnt algorithm where the notion of microstate (the state  $i$  that we associate the probability  $p_i$  with) is switched with the notion of a microscopic phase space path,  $\Gamma$ . These paths are determined by the microscopic equations of motion, and will be explained in more detail below. Just as in the usual MaxEnt case,  $p_\Gamma$  will denote the probability of microscopic path,  $\Gamma$ . The path information entropy is then defined as [20]

$$S_\Gamma = - \sum_{\Gamma} p_\Gamma \ln p_\Gamma \quad (12.1)$$

The more specific derivation is described as follows. Even though it is in terms of physics, the concepts can be transferred to non-physical situations, as we will see later. We have an open thermodynamic system with volume,  $V$  and boundary,  $\Omega$ . The system consists of several constituents ( $i = 1, \dots, m$ ) undergoing mutual transformations. In this system  $u(\mathbf{x}, t)$  denote the internal energy density and  $\rho_i(\mathbf{x}, t)$  the mass density of constituent  $i$  at position  $\mathbf{x}$  and at time  $t$ .  $\mathbf{f}_u(\mathbf{x}, t)$  denotes the internal energy flux density and  $\mathbf{f}_i(\mathbf{x}, t)$  denotes the mass flux density. These definitions enables us to apply local conservation laws for energy and mass in the differential form at each point  $(\mathbf{x}, t)$ . The macroscopic state vectors are defined as  $\mathbf{d} = (u, \{\rho_i\})$  and  $\mathbf{F} = (\mathbf{f}_u, \{\mathbf{f}_i\})$ . The normal components of  $\mathbf{F}$  on the boundary  $\Omega$  are denoted by  $\mathbf{F}^n$  [4]. For any quantity,  $X$ , the time average over the time interval  $\tau$  is given by

$$\bar{X} = \frac{1}{\tau} \int_0^\tau X(t) dt \quad (12.2)$$

and the expectation value over the path probability distribution  $p_\Gamma$  is given by

$$\langle X \rangle = \sum_{\Gamma} p_\Gamma X_\Gamma \quad (12.3)$$

where  $X_\Gamma$  is the value of  $X$  for the path  $\Gamma$ . Dewar's procedure amounts to maximizing the path information entropy,  $S_\Gamma$  given in (12.1), subject to the external constraints. The external constraints are typically surface flux inputs, externally imposed gradients, energy and mass balance in the stationary case, and any other *a priori* information, like local energy and mass conservation [4]. The complete maximization procedure, however, differs slightly from the usual MaxEnt algorithm, and is done in two steps. In the first step,  $S_\Gamma$  is maximized in the usual way with respect to  $p_\Gamma$  subject to



---


$$\sum_{\Gamma} p_{\Gamma} = 1 \quad (12.4)$$

$$\sum_{\Gamma} p_{\Gamma} \mathbf{d}_{\Gamma}(\mathbf{x}, 0) = \langle \mathbf{d}(\mathbf{x}, 0) \rangle \quad (\mathbf{x} \in V) \quad (12.5)$$

$$\sum_{\Gamma} p_{\Gamma} \overline{\mathbf{F}}_{\Gamma}(\mathbf{x}) = \langle \overline{\mathbf{F}}(\mathbf{x}) \rangle \quad (\mathbf{x} \in \Omega) \quad (12.6)$$

The first constraint is just the normalization condition. The second constraint denotes fixed initial configurations of internal energy and mass density within volume  $V$  at time  $t = 0$ . The third constraint denotes fixed time-averaged configurations of internal energy and mass flux densities on the boundary  $\Omega$  over the time interval from  $t = 0$  to  $t = \tau$ . If the system is in a steady state  $\langle \mathbf{d}(\mathbf{x}, 0) \rangle$  and  $\langle \overline{\mathbf{F}}(\mathbf{x}) \rangle$  are sufficient to describe the macroscopic state of the system and its interactions with the surroundings throughout the interval  $\tau$ . The real significance of these two constraints appear first in the next step. Maximization of  $S_{\Gamma}$  in the first step yields

$$p_{\Gamma} = \frac{1}{Z} \exp A_{\Gamma} \quad (12.7)$$

where the path action,  $A_{\Gamma}$ , is a functional of the Lagrange multipliers,  $\lambda(\mathbf{x})$ . As described in the MaxEnt theory in section 3 the Lagrange multipliers are uniquely defined by the constraints, in this case  $\langle \mathbf{d}(\mathbf{x}, 0) \rangle$  and  $\langle \overline{\mathbf{F}}(\mathbf{x}) \rangle$ . The Lagrange multipliers and the constraints give two equivalent sets that alone define everything else. By using the local conservation laws for mass and energy, followed by some mathematical calculations, Dewar arrives at an expression for the path action as a function of the time averaged thermodynamic entropy production,  $\sigma_{\Gamma}$

$$A_{\Gamma} = -\frac{1}{2} \int_V \frac{H_{\Gamma(0)} + H_{\Gamma(\tau)}}{k_B T} + \frac{\tau \sigma_{\Gamma}}{2k_B} \quad (12.8)$$

where  $H_{\Gamma(0)}$  and  $H_{\Gamma(\tau)}$  are the end point contributions of the non-equilibrium generalization of the Hamiltonian,  $T$  is the temperature and  $k_B$  is Boltzmann's constant. By substituting equation (12.8) into (12.7) we see that the entropy production,  $\sigma_{\Gamma}$ , is a key determinant of  $p_{\Gamma}(\lambda)$ . The probability of a path increases exponentially with the entropy production,  $\sigma_{\Gamma}$ , of that path. This entropy production,  $\sigma_{\Gamma}$ , is important to distinguish from the information entropy,  $S_{\Gamma}$ . Even though both of them are ultimately shown to be maximized by Dewar,  $\sigma_{\Gamma}$  refers to the internal entropy production due to irreversible processes, whereas  $S_{\Gamma}$  is the information entropy of the path probability distribution defined by equation (12.1). Dewar also shows that  $\sigma_{\Gamma}$  is expressed as products of forces and fluxes. If we reverse the path, the first term in (12.8) remains the same (it is the sum of the end-point contributions), whereas the second term changes sign, because  $\sigma_{\Gamma}$  changes sign.

The ratio of  $p_\Gamma$  over the probability of the reversed path,  $p_{\Gamma^-}$ , is thus  $e^{-\tau\sigma_\Gamma/k_B}$  which becomes exponentially small as the time interval,  $\tau$ , increases or as the system size increases (larger  $\sigma_\Gamma$ ). According to Dewar this shows how probabilities of violations of the second law become exponentially small, and he presents this as a derivation of the fluctuation theorem [4].

By combining the normalization condition (12.4) with the expression for  $p_\Gamma$  in (12.7) Dewar shows that [4]

$$S_{\Gamma,max}(\lambda) = - \sum_{\Gamma} p_\Gamma(\lambda) \ln p_\Gamma(\lambda) = \ln Z(\lambda) - \langle A(\lambda) \rangle \quad (12.9)$$

The above results conclude the first step. Step two then goes as follows. Both the constraints,  $\langle \mathbf{d}(\mathbf{x}, 0) \rangle$  and  $\langle \overline{\mathbf{F}^n}(\mathbf{x}) \rangle$ , are really unknown averages that were only artificially fixed during step one. In the second step,  $S_{\Gamma,max}(\lambda)$  is maximized with respect to  $\lambda(\mathbf{x})$  (which is related to the constraints) subject to the *remaining* constraints. The remaining constraints are the external constraints of the system as a whole, like steady state constraints. The climate system in Paltridge's study for instance, is constrained by the steady state requirement, the incoming solar radiation and the minimum temperature difference between the equator and poles. Otherwise, the boundary conditions are free to vary such that the entropy production of the system can be maximized. It corresponds to a case where the variables,  $\mathbf{F}$ , which are the heat fluxes between the atmospheric boxes, are free variables. As mentioned above, when  $\langle \mathbf{d}(\mathbf{x}, 0) \rangle$  and  $\langle \overline{\mathbf{F}^n}(\mathbf{x}) \rangle$  are already known, the steady state is already determined. This is the case of fixed boundary conditions, or a given linear flux-force relationship, where Prigogine's minimum principle instead should hold. The Maximum Entropy Production principle is equivalent to the second step in Dewar's derivation. He also shows Self-Organized Criticality to be a special case of MEP [4]. The MEP principle is powerful because only the external constraints end up mattering. We do not have to know anything about what's going on internally. All macroscopic quantities reproduced under the external constraints can then be calculated as expectation values over  $p_\Gamma(\lambda^*)$ , where  $\lambda^*$  denotes the MEP solution for the Lagrange multipliers [4]. This gives the macroscopic entropy production,  $\sigma$ , as the expectation value

$$\sigma = \langle \sigma_\Gamma \rangle \equiv \sum_{\Gamma} p_\Gamma(\lambda^*) \sigma_\Gamma \quad (12.10)$$

It is not clear in Dewar's derivations what exactly is meant by a path. A more explicit example of Dewar's paths was provided by Bruers in 2007 [42]. He looks at a system consisting of  $l$  sites, all with a real variable  $n_i(t)$  ( $i = 1, 2, \dots, l$ ) that depend on the discrete time  $t = 0, 1, \dots, \tau$ . At every time step there is a random flux between the sites. The flux,  $f_{ij} = -f_{ji}$ , from  $i$  to  $j$  depends on a real constant

---

parameter  $c_{ij} = c_{ji}$  such that  $f_{ij} = \pm c_{ij}$  where the sign is stochastic. A microscopic path  $\Gamma$  is then a specific set of values  $c_{ij}$  or  $-c_{ji}$  for every time step and every  $i$  and  $j$ . One path therefore constitutes a description of all microscopic fluxes between all the sites, at all time steps, over a specified time interval. The pathspace is the set of all possible paths, which we now see defines a finite, discrete set, which can therefore be counted. The total flux into  $i$  at a specific time can be clearly defined as

$$n_{i,\Gamma}(t+1) - n_{i,\Gamma}(t) = - \sum_j f_{ij,\Gamma}(t). \quad (12.11)$$

which is just a way of expressing conservation [42]. By writing  $\mathbf{d}(\mathbf{x}, t)$  instead of  $n_i(t)$  and  $\mathbf{F}(\mathbf{x}, t)$  instead of  $f_{ij}$  we arrive at Dewar's case. According to Bruer's then, a path simply means the exact description of  $\mathbf{F}(\mathbf{x}, t)$  and  $\mathbf{d}(\mathbf{x}, t)$  over time and space, where we can discretize all dimensions in order to make the pathspace countable. In order to make it finite, we need some kind of limit on the maximum value of the fluxes and the densities.

To summarize the two steps and provide a more conceptual description: let  $A$  denote the macroscopic values that we have real knowledge about, acting as the external constraints in the problem. Let the macroscopic quantities predicted as expectation values over the path distribution be collectively denoted by  $B$ . In Dewar's case,  $\langle \mathbf{d}(\mathbf{x}, 0) \rangle$  and  $\langle \overline{\mathbf{F}}^n(\mathbf{x}) \rangle$  belong to  $B$ . The MaxEnt algorithm is then applied in two steps. In the first step  $B$  acts as a temporary constraint used to maximize  $S_\Gamma = S_\Gamma(B)$  with respect to the path probabilities. Local conservation laws are built into this step. This step results in the flattest probability distribution for the paths, subject to  $B$  and  $A$ . The entropy is then a measure of the number of paths with non-negligible probability, when  $B$  and  $A$  are true. In the second step the path entropy,  $S_\Gamma(B)$ , is maximized with respect to  $B$  subject to the rest of the constraints which are given by  $A$ . After this step we therefore find the  $B$  that yields the largest number of paths with non-negligible probability subject to  $A$  only. Analogous to the MaxEnt logic, the system is most likely to be in the macrostate that corresponds to the largest number of paths.  $A$  does not include any microscopic knowledge or any fixed boundary conditions. The logic is that if  $A$  is sufficient to reproduce  $B$ , then the algorithm will correctly predict the observed  $B$  [20]. In other words, it has to do with reproducibility. If we could conduct an experiment to investigate the MEP principle, the microscopic paths would differ each time we set up the experiment, because this would be outside of our control. We are not interested in the microscopic details however, we are only concerned with the prediction of macroscopic behavior. If the macroscopic behavior is reproducible under the given constraints,  $A$ , this means that this behavior is characteristic of each of the vast number of microscopic paths compatible with the constraints in  $A$ . Again, we recognize the idea from the MaxEnt theory. Maxi-

mizing  $S_\Gamma$  subject to these constraints will discard all the microscopic information which by this logic is irrelevant to the reproducible behavior [4]. The expectation values calculated from the path distributions are "only" statistical inferences. They are not the result of the system somehow sampling different paths. Obviously, the system can only follow one path, and an ergodic interpretation would make no sense in this case [20]. The success of the predictions are both dependent on having correctly identified the constraints,  $A$ , and on the importance of these constraints. But just like with MaxEnt, failure of the predictions can inform us of new constraints[5].

As mentioned at the beginning, Dewar's method is analogous to the counting of microstates. The link between the MaxEnt algorithm in statistical mechanics and the Maximum Entropy Production principle is shown more formally by Dewar in 2005, where he shows that  $p_\Gamma$  correspond to the Gibbs grand-canonical distribution in the equilibrium limit,  $\overline{F^n}(\mathbf{x}) = 0$  [5]. It is the presence of non-zero macroscopic fluxes in the non-equilibrium state makes paths rather than microstates the central object of interest [20].

The MEP principle is based on the notion that a system will be in the macrostate corresponding to the largest number of paths. Paltridge, however, maximizes the thermodynamic entropy production, not the path information entropy. A crucial part of Dewar's proof is the part where he shows that the maximization of path entropy is equivalent to the maximization of thermodynamic entropy. This has to do with reversibility. The path action,  $A_\Gamma$ , consists of a reversible part and an irreversible part. Dewar considers the number of paths  $\Gamma$  that contribute to the mean behavior, and which therefore have reversible action  $\langle A^{rev}(\lambda) \rangle$  and irreversible action  $\langle A^{irr}(\lambda) \rangle$ . There are equally many paths with reversible action  $\langle A^{rev}(\lambda) \rangle$  and irreversible action  $-\langle A^{irr}(\lambda) \rangle$ . The number of paths contributing to the mean behavior is then the number of paths with irreversible action equal to  $\langle A^{irr}(\lambda) \rangle$ , that is,  $W(\langle A^{irr}(\lambda) \rangle)$ . Dewar ignores fluctuations about the mean behavior to arrive at  $Z(\lambda) = \sum_\Gamma \exp(A_\Gamma(\lambda)) \approx W(\langle A^{irr}(\lambda) \rangle) \exp \langle A(\lambda) \rangle$  [4]. Combining this result with equation (12.9) yields

$$S_{\Gamma,max}(\lambda) \approx \ln W(\langle A^{irr}(\lambda) \rangle) \quad (12.12)$$

Dewar uses this with the assumption that  $W(\langle A^{irr}(\lambda) \rangle)$  is an increasing function of  $A^{irr}$ , to show that the second step in his procedure is equivalent to maximizing the mean entropy production rate  $\langle \sigma_\Gamma(\lambda) \rangle$  with respect to  $\lambda(\mathbf{x})$  subject to the remaining constraints. He notes that it is the irreversible, material entropy production that should be maximized and therefore that Paltridge's application of entropy maximization only to the heat fluxes (not the rest of the radiation, which is reversible) is consistent with his results.

---

We saw in section 8 that thermodynamic entropy can be expressed as a series of products of fluxes and forces. The macroscopic fluxes,  $\langle \bar{\mathbf{F}}^m(\mathbf{x}) \rangle$ , are some of the central predictions of the MEP principle. Dewar states that MEP applies to the entropy production of those macroscopic fluxes that are free to vary under the imposed constraints, and corresponds to selection of the most probable macroscopic flux configuration [5]. The quantities used to define the constraints, or any of the other relevant information in the derivation, need not be physical quantities however. Dewar presents his results without any reference to physics, and without any physical interpretation in order to show that they are generic properties of MaxEnt which apply to a certain class of problems involving statistical inference. The general character of Jaynes' procedure, which is what is used, might make the principle applicable also to dynamical systems such as economies and biological populations [5]. Dewar states that if the reproducible states of any system can be described by quantities  $\langle \mathbf{d}(\mathbf{x}, 0) \rangle$  obeying local balance equations, and by fluxes,  $\mathbf{F}(\mathbf{x})$  (and possibly sources), then it will be possible to use the general result for  $p_{\Gamma}$  to arrive at a generalized entropy production. This entropy production will then be a function of generalized fluxes and forces. Dewar points out that all the corrolaries of his results (the Fluctuation Theorem, and Self-Organized Criticality in addition to MEP) may then be expected as generic features of such systems. [4]. He calls for extended applications of the principle, in order to let the results speak for themselves [20].

As mentioned in the beginning, Dewar's derivation of the MEP principle has been criticized in a few papers. I will not present this in detail, but include a summary of some of it, for reference. Bruers' points out that a mixing of different MEP principles in the literature has lead to confusion. He claims that Paltridge's MEP hypothesis is a different principle, with different assumptions and different applications than the principle derived by Dewar [32]. In 2007 Bruers claimed to find at least six different principles related to MEP: the least dissipation, the near-equilibrium (linear) minimum entropy production principle, the near-equilibrium (linear) MEP principle, the far-from-equilibrium (non-linear) non-variational MEP principle, the far-from equilibrium variational MEP principle and the optimization minimum entropy production principle. These principles are subject to different assumptions, regions of validity, constraints and applications [32]. Bruers claim that Dewars derivation in 2003 can be used to derive the minimum entropy principle rather than the maximum entropy principle [42]. Bruers also state that the MEP principle used by Paltridge remains an unproven hypothesis with a lot of controversy and unsolved questions about the necessary conditions, requirements and ranges of application [42]. He points to the need of a theoretical non-trivial example where far-from equilibrium variational MEP applies because this could help to understand more deeply the functioning of atmospherical, hydrological, biological, ecological or other systems [32].

Martyushev et al. point out that attempts to derive the MEP principle, including Dewar's, so far have been unconvincing because they require introduction of additional hypotheses, which themselves are less evident than the proved statement [6]. They claim that the MEP principle cannot be proved, and that its success will ultimately depend on experimental and empirical results [43]. Grinstein et al. on the other hand, claim that some of the steps in Dewar's derivation are flawed [41].

Even though Dewar's work has been critiqued by some, it has generally been met with interest and again brought attention to the MEP principle. We have yet to see whether Dewar will reply to the critique with a more rigorous proof, or whether his argument has some fundamental problems. I will use the presentation in this section to analyze the potential of the MEP principle in the Maximum Entropy Theory of Ecology.

### 12.1 Discussion

My application of the MEP principle in Chapter 7 was based on a specific interpretation of Dewar's proof. His proof was understood as regarding the number of paths leading from an initial state to a final state and stating that this is maximum when the final state is the one that results in the maximum entropy production during the transition. Dewar's proof was also taken to suggest that MEP holds for the production of information entropy as well as thermodynamic entropy. In the previous section I presented Dewar's proof in more detail. I will now discuss whether the proof and its implications can be transferred to the entropy in METE.

Dewar shows that the maximization of the path entropy,  $S_{\Gamma}$ , which is the information entropy of the path distribution in (12.1), is equivalent to maximization of the irreversible thermodynamic entropy,  $\sigma$ . This argument reveals the connection between path information entropy and thermodynamic entropy. The initial interpretation of Dewar's proof was related to paths in state space (e.g.  $(p, V, N)$  for an ideal gas or  $(S, N, E)$  in METE). The meaning of paths, however, is shown to be something more specific than this in Dewar's case. A path refers to a defined microscopic evolution of the system and is described in terms of fluxes of conserved quantities in space and over time. Exactly what these are has to be determined in the particular case. In any case, a steady state is described by many possible paths. A steady state, however, remains at the same point in state space. This is an important distinction. It is true that Dewar's proof connects thermodynamic entropy with information entropy, but I have shown that this is not just any information entropy. It is the information entropy of the probabilities of the paths.

Under closer inspection it becomes clear that the derivation of the MEP princi-

ple is really only a different application of the MaxEnt principle. Instead of a regular, state space we operate within a path space. There are usually several microscopic paths corresponding to one macroscopic behavior. The principle then simply states that the macroscopic behavior of a system, which is now dynamic in nature, is most likely in accordance with the overwhelming majority of all the possible microscopic paths. In the case for the climate there must be more paths resulting in a medium temperature difference and a medium flow, than a maximal gradient and zero flow, or zero gradient and maximal flow. Analogous to MaxEnt the enumeration of the possible phase space paths is an essential part of the application and will determine whether the method will give accurate predictions. Just like the different states in the usual MaxEnt application have to be finite, so do the paths in the MEP application. If the path space is infinite we cannot assign non-zero probabilities to all of the possibilities (or else the total probability will be infinite). The method then amounts to finding the most likely macroscopic flux configuration. As long as the paths are well defined, the same logic that holds for a steady state (flux in equals flux out) should apply for a system that evolves from one state to another via different possible microscopic paths. The principle should therefore be able to predict the evolution of a system from an initial state to a final state, where the final state is the one with the most paths leading to it. This is close to our initial interpretation of the proof, but the meaning of it is now much clearer.

I earlier pointed to the difference between maximizing the entropy with respect to the probability distribution and with respect to the constraints and questioned the validity of the last procedure. In Dewar's proof we see that both types of maximizations are done. In the first step the entropy is maximized with respect to the path probabilities and in the second step it is maximized with respect to the constraints,  $\mathbf{F}(\mathbf{x}, t)$  and  $\mathbf{d}(\mathbf{x}, t)$ . However, it still holds for any maximization procedure that the problem needs to be bounded if we want to obtain physical results.

It became clear already from the study of the different applications of the MEP principle that these are all closely related to the presence of fluxes. In Paltridge's studies we see clearly how the entropy production is a product of the heat flux and the temperature difference. Dewar's proof shows this connection in greater detail. The fluxes are here multiplied with forces in order to give the entropy,  $\sigma_T$ , either in the usual thermodynamic form or in a more generic form.

Our entropy,  $S_I$ , is far from a path entropy the way it is defined by Dewar or Bruers. It is quite obvious that  $R(n, \varepsilon)$  does not define the probability of paths. It simply defines the probability of picking a certain species, and a certain individual from this species. Moreover, it is only the path information entropy that is maximized in Dewar, not the production of it. The question is whether a similar connection that is made between thermodynamic entropy and path entropy by

Dewar can be made between a path entropy and our information entropy. Then, we could say that  $dS_I/dt$  should be maximized. There is nothing in Dewar's proof suggesting this. Our information entropy is not some kind of generic form of a thermodynamic entropy, defined as a product between forces and fluxes. There is therefore nothing in Dewar's proof that suggests that the evolution with the largest number of microscopic paths is equivalent to the evolution of the state variables resulting in the maximum production of  $S_I$ . Even though thermodynamic entropy can be interpreted as information entropy (of the canonical distribution), this does not alone seem to imply that MEP holds for all information entropies. There must be something special about the form of the thermodynamic information entropy and its connection to nature, fluxes and forces. It must be this connection that is unique, not the term entropy.

An essential step in Dewar's derivation, stated but not shown explicitly in the previous section, is the application of local conservation laws. These are used to relate the fluxes,  $\mathbf{F}(\mathbf{x}, t)$ , with the quantities,  $\mathbf{d}(\mathbf{x}, t)$ , and to derive the relationship between thermodynamic entropy and path entropy. Even though the application of the principle removes the microscopic details of the problem, microscopic laws are still crucial in the derivation. The constraints in Dewar's proof are of a specific type. One of them denotes a conserved quantity and the other one the corresponding flux. The derivation seem to be valid only if the system is described by conserved quantities. The constraints that define  $I_R$  (or  $S_I$ ), number of individuals per species,  $N/S$ , and total energy per species,  $E/S$ , are not conserved. If our defining quantities were conserved between one step in time and the next and we had some known expression for sources and sinks, one instance in time could be related to another. We can see how this could lead to dynamical relationships. But we do not have any microscopic "laws" expressed in our description of the system. If we did, we could use this to infer macroscopic behavior.

MaxEnt is a theory that gives a lot from a little. We want MEP to be the same. Paltridge was able to predict very accurate values of temperature and cloud cover from very few variables. It seems intuitive, however, that with no dynamical relationships in the underlying description of the problem (in the form of paths, or more specifically some kind of fluxes or other time dependent phenomena) no time can be extracted from a theory. We can get a lot from a little, but we cannot get something from nothing.

Lastly, there is one aspect of Dewar's proof which is unclear, that I want to point out. It has to do with the interpretation of the constraints. It is stated that  $\langle \mathbf{d}(\mathbf{x}, 0) \rangle$  and  $\langle \overline{\mathbf{F}^n}(\mathbf{x}) \rangle$  represent the observed initial configuration of mass and energy, and time-averaged fluxes on the boundary respectively. These are the unknowns that we seek to know in a MEP problem and these are found in the



---

second step of the procedure (for instance the net fluxes into the different boxes in Paltridge's case). The entropy production,  $\sigma$ , is also found as the path average. What is puzzling me is the connection between the path averages and the observed quantities. Clearly, a system only ever follows one path in the evolution over time. If the path probability distribution is extremely peaked, or the resulting macroscopic behavior is peaked, then the average over the paths will correspond to the overwhelming majority of the paths. In this case, most of the paths do have the macroscopic behavior we are looking for, and the prediction that the observed behavior will equal the path average is accurate. However, if the path distribution is not peaked, the average behavior is not a good prediction. In this case, many real paths will result in a different behavior. I cannot see how sharpness is automatically satisfied and even though Dewar explicitly states that he ignores fluctuations about the mean behavior in his derivation, there is no discussion of this aspect in his proof. In classical mechanics we can get around this with the ergodic hypothesis. During a measurement a system will fluctuate between several states. The system is then expected to visit different states according to the representative ensemble, such that the ensemble average is a good prediction for the measured value. In the case where we instead have paths, the ergodic hypothesis makes no sense, since there is no fluctuation between paths during the evolution of a system. The requirement that all the paths have the average behavior must therefore be stricter in this case, if we want the predictions to have any validity. The average always provides the best guess in the sense that it gives the least square error in a series of guesses, but this is not satisfactory if we want to use the theory to describe nature. Again, it is problematic that Dewar provides no actual examples. He states that we are only looking for reproducible behavior, but he does not discuss when this is what will be predicted by the averages.

If the theory works, it should predict reproducible behavior. Every time we set up an experiment, we should then observe the same macroscopic behavior (with an overwhelming certainty). Even if this logic can be transferred to a theory of ecology, a dynamical theory of ecology will be much more difficult to test than a similar physical theory. We cannot reproduce evolutions of ecosystems over time. We can compare the evolution of two similar systems, but this will always introduce external variability.

## 13 Conclusions

The starting point of this thesis was the hypothesis that the Maximum Entropy Production principle applies to the information entropy  $S_I$ . The originally purpose was to investigate a possible solution of the partial differential equations, (7.4) - (7.6) and express the results mathematically and graphically. The goal was to give clear predictions that could later be tested with real data in order to verify

or falsify the hypothesis.

My project ended up taking a quite different route. In Chapter 7, I showed that the set of PDE's are not actually solvable. Then, I showed that these PDE's do not actually describe the maximum entropy behavior. Instead, equations (7.4) - (7.6) impose a constant entropy condition on the time derivatives of the state variables  $\dot{S}$ ,  $\dot{N}$  and  $\dot{E}$ . I showed that the maximization question itself is more complex than initially assumed. The maximization condition on  $\dot{S}_I$  does not itself imply that the time derivatives are functions of the state variables themselves. This has to be added as an additional postulate. In any case, the problem is unbounded and a maximization of  $\dot{S}_I$  results in infinite time derivatives. I discussed different types of maximum conditions and argued that the system cannot behave in a way that makes it possible to predict the future. Only a maximization at each point in time therefore makes sense. The initial MEP hypothesis alone is not sufficient to say anything about the time rate of change of the entropy,  $S_I$ . The gradient gives the direction of largest increase, but I show that the meaning of following the gradient in our case is questionable. In any case, the initial MEP hypothesis predicts that the number of individuals,  $N$ , and the total metabolic rate,  $E$ , will increase as fast as possible and that the number of species,  $S$ , will decrease as fast as possible. The principle therefore predicts that all ecosystems eventually will end up in a state where  $N$  and  $E$  are growing towards infinity and  $S$  approaches one. This result does not comply with reality and serves as a strong indication that the initial hypothesis is wrong. I also questioned the maximization of the entropy with respect to the state variables, when these act as constraints imposed on the system. In light of Shannon's axioms this amounts to increasing the entropy by including more possibilities. In total, the initial MEP hypothesis lacks the predictive power we wanted it to have and leads to questionable results in cases where they can be obtained. My application and discussion in Chapter 7 showed that the initial hypothesis is not fruitful. These conclusions lead Harte to remove this hypothesis and the subsequent equations from his book. My analysis of the initial hypothesis, however, still sheds light on the meaning of the entropy  $S_I$  (or  $I_R$ ) which is of central importance to METE.

I conducted a thorough review of the MEP principle in order to determine whether there is sufficient reason to expect a MEP behavior in METE based on analogies. This review uncovered the complexity inherent in the principle itself and in the literature surrounding it, particularly in regards to what part of the entropy is maximized. I showed that the other applications in the literature are related to thermodynamic entropy. Furthermore, most of them are also related to stationary non-equilibrium systems, whereas I argue that METE is better viewed as a non-stationary system close to equilibrium, undergoing quasi-static change. Other applications are characterized by the presence of fluxes and I argue that the cur-

---

rent description of METE does not contain any fluxes. Based on Chapters 9 and 11, I conclude that the existing applications are only weak analogies to METE. The MEP hypothesis in METE cannot be justified based on success in these applications.

As a side project, suggested by Harte, I conducted an analysis of stability in Paltridge's model. I used MATLAB to calculate the possible steady states in a simple 3-box model of Earth's climate, and looked at the stability properties of these states. I showed that the MEP state is no different from the other steady states in regards to stability with respect to perturbations in the temperatures or the heat flux. The prevalence of the MEP state can therefore not be explained on these grounds.

In lack of a valid analogy I investigated instead whether an application of the MEP principle to METE could be justified based on first principles. In Chapter 12 I show how Dewar's proof states that the maximization of the thermodynamic entropy production is equivalent to the maximization of the path entropy. I argue that our information entropy bears no resemblance to the path entropy. And even if it did, it is not the production of information entropy that is maximized in Dewar's proof, but the static information entropy. The derivation of the MEP principle really amounts to a particular application of the MaxEnt principle. Dewar notes that the entropy need not be thermodynamic, but can be any generic entropy defined as a product of fluxes and forces. I argue that our entropy,  $S_I$ , cannot be regarded as any such generic entropy. I conclude that there is no reason to believe that the production of our information entropy will be maximized, based on Dewar's proof.

Based on the initial application, the review of other applications and Dewar's proof I conclude that the MEP principle as initially suggested in METE should be abandoned. There are no examples of successful applications of the principle to any cases similar to our. Dewar's proof is not fully complete so strictly speaking we cannot make a final conclusion regarding the applicability of the MEP principle to the information entropy in METE, but my investigation leaves little room for hope. If MEP were to hold for information entropies in general, this would require a completely different proof.

I also conducted a study of the variance of the number of species in subplots. I used MATLAB in order to compare two real data sets with theoretical predictions obtained from the Spatial-Abundance Distribution,  $\Pi(n)$ . I showed that the general patterns in the predictions,  $\Pi(n)$ , match the real data, but that the non-discrete nature of the predictions complicate the comparison. I developed a simulation procedure in order to compare the theoretical distribution of species in different

subplots with the observed values. Whether these predictions are said to be close or far from the real data will depend on the ecological research question. I point out however, that if MaxEnt is to work as a null theory in ecology we need a rigorous way of testing the closeness between predictions and observations. I therefore stress the need for a defined method for 1) discretizing the theoretical prediction in order to compare with real data, and 2) comparing the discretized theoretical distribution to the real distributions. I suggest two methods for doing this, where the first one is to calculate the probability of different distributions from the theory directly to obtain the relative probabilities of different distribution, and the second one is to look at the entropy of the discretized predicted distribution versus the entropy of the observed distributions.

A theory can be established by empirical evidence even in the absence of a rigorous proof. One approach to new principles is therefore to apply them and let the results speak for themselves. But there is usually something justifying such an approach and often this is a successful application to a similar case. The application of the MEP principle to  $S_I$  does not give promising results and the cases where it has been successful cannot be said to be very similar to ours.

My Thesis has investigated one approach to a dynamical form of METE and showed why it does not work. This has shed light on the interpretation of the information entropies in METE in general. Moreover, I have also reviewed the MEP principle, and discussed its wider application to generalized information entropies. I have also done work on the variance of one of the most important metrics in METE, the Species-Abundance distribution,  $\bar{S}(A)$ . I have showed the need for a defined method of testing the theory, and made two suggestions for this. My thesis provides a solid background for the grad students in the Energy and Resources Group at Berkeley interested in these topics. My work has contributed to the book, *Maximum Entropy and Ecology*, coming out on Oxford University press in June 2011 and I hope that someone can use the insight I have provided in order to find a more suitable procedure to arrive at a dynamical theory of macroecology in the tradition of MaxEnt.

## 14 Suggestions for future work

The MEP principle, at first hand quite simple-looking, turned out to be much more complex and multi-faceted when investigated more closely. The confusion surrounding the principle is apparent in the literature on the subject. In trying to deconstruct the meaning of the initial hypothesis and looking for alternative applications I got closer and closer to the very fundamentals, both of MEP and of METE itself. It was not until the very end that I had a good enough under-

---

standing of the theory to be able to propose alternative applications of the MEP principle in macroecology. It is finally clear what the MEP principle can do and what it cannot do. I here present some ideas for the future of both METE and MEP.

The most concrete suggestion is to develop in detail either of the two methods I suggested for comparison between predictions and observations in METE, in Chapter 6. These methods can provide valuable criteria when testing the theory.

The next two suggestions are motivated by statistical mechanics. In thermodynamics work is related to a change in the external parameters and does not lead to a change in the probabilities of the different states. The entropy is therefore unaffected and quasi-static processes involving only work are reversible. Heat, on the other hand, is related to a change in the probabilities, and is therefore accompanied by a change in the entropy of the system (which does not necessarily imply irreversibility, if the system is not isolated). It would be interesting to look for an analogy in the entropy measures in METE. Can we define a form of change that has to do with changing probabilities, and another one that has to do with the change in the external parameters? Is it possible to derive any relationships from this, informing us about ecological "forces"?

Temperature appears as the inverse of the Lagrange multiplier in the MaxEnt derivation of the Boltzmann energy distribution. It would be interesting to look at whether the inverse of the Lagrange multipliers in METE can be associated with a generalized ecological "temperature".

Finally, the last two suggestions are related to the MEP principle. Dewar's proof indicates that a system with sufficient degrees of freedom (not fixed boundary conditions) will be characterized by fluxes that maximize the thermodynamic entropy production. A completely different approach could therefore be to apply the MEP principle to the *thermodynamic* entropy production of a collection of organisms. This, however, would be a very different project giving very different results from what was intended by the original MEP hypothesis and it is a step away from the framework of METE. Nonetheless does it stand out as a more viable approach, if one wants to use the MEP principle on ecosystems.

The other alternative is to define some form of path entropy for our system. In order to do this we would have to specify and enumerate the possible configurations for the fluxes through the ecosystem. This could be heat flux, which would give a thermodynamic entropy, but it could also be any other flux that we wish to use. The maximization of the path entropy could then be conducted according to Dewar's two-step process. If Dewar's proof is correct, however, we can skip

this procedure and maximize instead the production of the generic entropy of the system, defined as a series of products of the fluxes and the forces determining the system.

METE is a brand-new theory with a great potential for making significant contributions to the field of ecology. I hope some of my suggestions above will be of value when it comes to further development within this field.

## References

- [1] J. Harte, *Maximum Entropy and Ecology: On the Inference of Patterns in Nature*. Oxford press, June 2011.
- [2] E. T. Jaynes, "Information theory and statistical mechanics," *The Physical Review*, vol. 106, pp. 620–630, May 1957.
- [3] G. W. Paltridge, "The steady-state format of global climate," *Quarterly journal of the Royal Meteorological Society*, vol. 104, 1978.
- [4] R. C. Dewar, "Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states," *Journal of Physics A: Mathematical and General*, vol. 36, no. 3.
- [5] R. C. Dewar, "Maximum entropy production and the fluctuation theorem," *Journal of Physics A: Mathematical and General*, 2005.
- [6] L. M. Martyushev and V. D. Seleznev, "Maximum entropy production principle in physics, chemistry and biology," *Physics Reports*, vol. 426, pp. 1–45, 2006.
- [7] C. E. Shannon, "A mathematical theory of communication," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 2001.
- [8] E. T. Jaynes, "Notes on present status and future prospects," in *Maximum Entropy and Bayesian Methods* (C. R. Smith, G. J. Erickson, and P. O. Neudorfer, eds.), Kluwer Academic Publishers, 1992.
- [9] J. C. Maxwell, *Matter and Motion*. D. Van Nostrand, 1878.
- [10] H. D. Young and R. A. Freedman, *University Physics with Modern Physics*. Addison Wesley, 11 ed., 2004.
- [11] "Merriam-Webster Online Dictionary." <http://www.merriam-webster.com/dictionary/science>, Feb 2011.
- [12] F. Roman and S. Hartman, "'Models in Science', The Stanford Encyclopedia of Philosophy." <http://plato.stanford.edu/archives/sum2009/entries/models-science/>, Jan 2009.

- [13] “Aquarena Wetlands Project Glossary.” <http://www.bio.txstate.edu/~wetlands/Glossary/glossary.html>, April 2011.
- [14] J. H. Brown and B. A. Maurer, “Macroecology: The division of food and space among species on continents,” *Science*, vol. 243, March 1989.
- [15] J. Harte, “Toward a synthesis of the newtonian and darwinian worldviews,” *Physics Today*, October 2002.
- [16] E. T. Jaynes, *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.
- [17] R. Kleeman, “Information Theory and Predictability.” Lecture Notes from NYU: <http://www.math.nyu.edu/faculty/kleeman/syllabusinfo.html>, Jan 2011.
- [18] B. A. Olshausen, “NPB 163/PSC 128 Information processing models in neuroscience and psychology.” Lecture Notes from UC Berkeley: <http://redwood.berkeley.edu/bruno/npb163/>, Sep 2010.
- [19] E. T. Jaynes, “On the rationale of maximum-entropy methods,” *Proceedings of the IEEE*, vol. 70, September 1982.
- [20] R. C. Dewar, “Maximum entropy production and non-equilibrium statistical mechanics,” in *Non-equilibrium thermodynamics and the production of entropy: life, earth and beyond*, ch. 4, pp. 41–55, Springer Verlag, 2004.
- [21] P. Penfield, “6.050J / 2.110J Information and Entropy.” Lecture Notes from MIT: <http://www-mtl.mit.edu/Courses/6.050/notes>, Jan 2003.
- [22] E. T. Jaynes, “Probability theory as logic.” Presented at the Workshop on Maximum Entropy and Bayesian methods, Dartmouth College, New Hampshire, August 14, 1989. Kluwer Publishers., 1989.
- [23] J. R. Banavar, A. Maritan, and I. Volkov, “Applications of the principle of maximum entropy: from physics to ecology,” *Journal of Physics: Condensed Matter*, vol. 22, January 2010.
- [24] F. Reif, *Fundamentals of Statistical and Thermal Physics*. Waveland Press, 1965.
- [25] J. S. Dugdale, *Entropy and its Physical meaning*. Taylor and France Ltd., 1996.
- [26] M. E. Soulé, “What is conservation biology?,” *BioScience*, vol. 35, Dec 1985.
- [27] J. Harte, T. Zillio, E. Conlisk, and A. B. Smith, “Maximum entropy and the state-variable approach to macroecology,” *Ecology*, 2008.

- [28] E. T. Jaynes, “Concentration of distributions at entropy maxima,” in *Papers on probability, statistics and statistical physics* (R. D. Rosenkrantz, ed.), pp. 315–335, Kluwer Academic Publishers, 1983.
- [29] J. Green, J. Harte, and A. Ostling, “Species richness, endemism, and abundance patterns: tests of two fractal models in a serpentine grassland.,” *Ecology Letters*, vol. 6, pp. 919– 928., 2003.
- [30] S. R. DeGroot, *Non-Equilibrium Thermodynamics*. North-Holland Publishing Company, 1962.
- [31] G. Lebon, J. David, and J. C. Vázquez, *Understanding non-equilibrium thermodynamics: foundations, applications, frontiers*. Springer, 2007.
- [32] S. Bruers, “Classification and discussion of macroscopic entropy production principles,” *ArXiv Condensed Matter e-prints*, 2007.
- [33] A. Kleidon and R. Lorenz, “Entropy production by earth system processes,” in *Non-equilibrium Thermodynamics and the Production of Entropy: Life, Earth, and Beyond* (A. Kleidon and R. Lorenz, eds.), ch. 1, Springer Verlag Berlin Heidelberg, 2005.
- [34] G. W. Paltridge, “Stumbling into the mep racket: A historical perspective,” in *Non-equilibrium Thermodynamics and the Production of Entropy: Life, Earth, and Beyond* (A. Kleidon and R. Lorenz, eds.), ch. 3, Springer Verlag Berlin Heidelberg, 2005.
- [35] H. Ozawa, A. Ohmura, R. D. Lorenz, and T. Pujol, “The second law of thermodynamics and the global climate system: a review of the maximum entropy production principle,” *Reviews of Geophysics*, vol. 41, no. 4, 2003.
- [36] G. W. Paltridge, G. D. Farquhar, and M. Cuntz, “Maximum entropy production, cloud feedback, and climate change,” *Geophysical Research Letters*, vol. 34, 2007.
- [37] S. E. Jørgensen, *Integration of ecosystem theories: A Pattern*. Springer, 3d ed., August 2002.
- [38] M. Ruth, “Insights from thermodynamics for the analysis of economic processes,” in *Non-equilibrium Thermodynamics and the Production of Entropy: Life, Earth, and Beyond* (A. Kleidon and R. Lorenz, eds.), ch. 18, Springer Verlag Berlin Heidelberg, 2005.
- [39] S. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering*. Perseus Books Group, 1 ed., 2001.
- [40] N. Boccara, *Modeling Complex Systems (Graduate Texts in Contemporary Physics)*. Springer, 2 ed., 2010.



- [41] G. Grinstein and R. Linsker, “Comments on a derivation and application of the ‘maximum entropy production’ principle,” *Journal of Physics A: Mathematical and Theoretical*, vol. 40, pp. 9717–9720, 2007.
- [42] S. Bruers, “A discussion on maximum entropy production and information theory,” *Journal of Physics A: Mathematical and Theoretical*, vol. 40, no. 27, 2007.
- [43] L. M. Martyushev, “The maximum entropy production principle: two basic questions,” *Phil. Trans. R. Soc. B*, vol. 365, no. 1545, pp. 1333–1334, 2010.

**A APPENDIX**

Using equation (5.15)

$$\frac{S}{N} \approx \beta \ln \left( \frac{1}{\beta} \right)$$

we compute

$$\begin{aligned} \frac{\partial}{\partial N} \left( \beta \ln \frac{1}{\beta} \right) &= \frac{\partial}{\partial N} \left( \frac{S}{N} \right) \\ \frac{\partial}{\partial \beta} \frac{\partial \beta}{\partial N} \left( \beta \ln \frac{1}{\beta} \right) &= \frac{\partial}{\partial N} \left( \frac{S}{N} \right) \\ \frac{\partial \beta}{\partial N} \left( \ln \frac{1}{\beta} - 1 \right) &= \left( -\frac{S}{N^2} \right) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial}{\partial S} \left( \beta \ln \frac{1}{\beta} \right) &= \frac{\partial}{\partial S} \left( \frac{S}{N} \right) \\ \frac{\partial}{\partial \beta} \frac{\partial \beta}{\partial S} \left( \beta \ln \frac{1}{\beta} \right) &= \frac{\partial}{\partial S} \left( \frac{S}{N} \right) \\ \frac{\partial \beta}{\partial S} \left( \ln \frac{1}{\beta} - 1 \right) &= \left( \frac{1}{N} \right) \end{aligned}$$

Using  $\omega = \ln \frac{1}{\beta}$  (equation (7.1)) we get

$$\begin{aligned} \frac{\partial \omega}{\partial S} &= -\frac{1}{\beta} \frac{\partial \beta}{\partial S} \\ &= -\frac{1}{\beta} \left( \frac{1}{N} \right) \left( \ln \frac{1}{\beta} - 1 \right)^{-1} \\ &= -\frac{1}{\beta} \left( \frac{1}{N} \right) (\omega - 1)^{-1} \end{aligned}$$

and

---


$$\begin{aligned}
\frac{\partial \omega}{\partial N} &= -\frac{1}{\beta} \frac{\partial \beta}{\partial N} \\
&= -\frac{1}{\beta} \left( -\frac{S}{N^2} \right) \left( \ln \frac{1}{\beta} - 1 \right)^{-1} \\
&= -\frac{1}{\beta} \left( -\frac{S}{N^2} \right) (\omega - 1)^{-1}
\end{aligned}$$

Using equation (7.2) for  $S_I$  we obtain

$$\frac{\partial S_I}{\partial E} = \frac{1}{E}$$

$$\begin{aligned}
\frac{\partial S_I}{\partial N} &= -\frac{1}{N} + \frac{\partial S_I}{\partial \omega} \frac{\partial \omega}{\partial N} \\
&= -\frac{1}{N} + \left( 1 - \frac{1}{\omega^2} \right) \frac{\partial \omega}{\partial N} \\
&= -\frac{1}{N} + \left( 1 - \frac{1}{\omega^2} \right) \left( -\frac{1}{\beta} \left( -\frac{S}{N^2} \right) (\omega - 1)^{-1} \right) \\
&= -\frac{1}{N} + \frac{1}{\omega^2} (\omega^2 - 1) \left( \frac{\omega}{N} (\omega - 1)^{-1} \right) \\
&= -\frac{1}{N} + \frac{1}{\omega^2} (\omega^2 - 1) \left( \frac{\omega}{N} (\omega - 1)^{-1} \right) \\
&= -\frac{1}{N} + \frac{1}{\omega} (\omega + 1) \frac{1}{N} \\
&= \frac{1}{N} \left( -1 + \frac{1}{\omega} (\omega + 1) \right) \\
&= \frac{1}{N} \frac{1}{\omega}
\end{aligned}$$

where we again have used (5.15), and

$$\begin{aligned}\frac{\partial S_I}{\partial S} &= \frac{\partial S_I}{\partial \omega} \frac{\partial \omega}{\partial S} \\ &= \left(1 - \frac{1}{\omega}\right) \frac{\partial \omega}{\partial S} \\ &= \frac{1}{\omega^2} (\omega^2 - 1) \left(-\frac{1}{\beta} \left(\frac{1}{N}\right) (\omega - 1)^{-1}\right) \\ &= \frac{1}{\omega^2} (\omega + 1) \left(-\frac{1}{\beta} \left(\frac{1}{N}\right)\right) \\ &= \frac{1}{\omega^2} (\omega + 1) \left(-\frac{\omega}{S}\right) \\ &= \left(1 + \frac{1}{\omega}\right) \left(-\frac{1}{S}\right)\end{aligned}$$

Taking the time derivative of equation (7.3) yields

$$\frac{dS_I}{dt} = \frac{\partial S_I}{\partial N} \frac{dN}{dt} + \frac{\partial S_I}{\partial E} \frac{dE}{dt} + \frac{\partial S_I}{\partial S} \frac{dS}{dt}$$

which, combined with the above results for the partial derivatives of  $S_I$  give equation (7.3), which we seeked to derive.