2009

Population Dynamics In A Model Closed Ecosystem

Doeke Romke Hekstra

POPULATION DYNAMICS IN A MODEL CLOSED ECOSYSTEM

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Doeke Romke Hekstra

June 2009

POPULATION DYNAMICS IN A MODEL CLOSED ECOSYSTEM

Doeke Romke Hekstra

The Rockefeller University 2009

For almost any species in any environment, it is nearly impossible to predict its fitness from molecular knowledge. If fitness is not to be a mere tautology, reproducible measurements of the survival and reproduction of populations are needed over many generations. Laboratory microbial ecosystems afford the short time and length scales required for such measurements. Their conventional implementations, batch cultures with period refreshment of growth medium or chemostats with continuous refreshment, have a number of disadvantages, such as the introduction of additional frequencies, selection for surface growth and the distortion of chemical interactions.

In closed ecosystems free energy is instead supplied as light, allowing for simpler, replicable protocols and a consistent interpretation of interactions, independent of their mode or timescale. Here, I describe a model closed ecosystem consisting of three single-celled microbes, *Escherichia coli*, *Chlamydomonas reinhardtii* and *Tetrahymena thermophila* and show that these species can coexist for hundreds of days under closure. Using a custom built *in situ* fluorescence microscopy set up, the densities of these three species can be measured automatically and noninvasively over months with low classification error and large dynamical range. When kept under identical boundary conditions, these ecosystems reproducibly diverge in composition, with characteristic divergence times of ~20 days for *T. thermophila*, ~40 days for the other two species, and an approximately linear increase of an aggregate divergence measure over the first ~60

days. For two ecosystems, densities were measured continuously under constant conditions and their dynamics shown to be nonstationary for all three species >100 days after closure. As a consequence, conventional time series methods assuming stationarity are inadequate and wavelet analysis is proposed as an alternative.

Species-species interactions are further investigated using oscillations in illumination intensity. Densities of *C. reinhardtii* and, surprisingly, *E. coli* respond to modest perturbations of light intensity. Variation of the modulation frequency strongly implicates the circadian clock of *C. reinhardtii* in its response. The nonlinearity of the *E. coli* response suggests that it depends on *C. reinhardtii* density or spatial distribution rather than directly responds to the modulation of illumination. Further improvements in the detection of interactions are proposed.

Foar Heit en Mem,

myn foarbyld yn tawijing.

ACKNOWLEDGEMENTS

Frequently throughout this thesis, I will use the pronoun 'I', rather than 'we'. This thesis project has been, as many have, both a very individual experience and a very social one. First of all, I want to thank my thesis advisor, Stanislas Leibler. Stan has been intensively involved in almost every aspect of the work presented. He provides a rare combination of creativity and a drive to be rigorous, and encouraged me from the start not to look under the light post, but to venture into the unknown. More generally, he has created a lab with a free flow of suggestions, people with diverse backgrounds and an atmosphere in which no question is too naïve to ask.

Many people contributed to specific parts of this project. Michael Krieger and Łukasz Kowalik took part in the earliest stages of the project. John Chuang provided me with a lot of advice and materials for cloning, programming and any computer–related questions. Sri Ram worked with me on creating 'non-sticky' *E. coli* and helped me much with electronics and programming. Yifan Liu (Allis Lab, The Rockefeller University, now at the University of Michigan) was very kind to help me design and construct fluorescent strains of *T. thermophila* and provided me with wildtype strains. Zak Frentz helped in the preparation of ecosystems and he, Jack Merrin and Seppe Kuehn provided helpful advice on electronics and optics.

In general, the lab members and many visitors to the lab provided sharp questioning and lively discussion. I want to thank current and former lab members, Roy Kishony, Nathalie Questembert-Balaban, Calin Guet, Edo Kussell, Remy Chait, José Vilar, Ali Kinkhabwala, Giovanni Sena, Wenying Shou and Clément Nizak, Sri Ram, Jack Merrin, Olivier Rivoire, Zak Frentz and Seppe Kuehn for their fruitful comments, company and

many reflections on life and science. Among the visitors I want to especially mention the ones that stayed for a while: Andrew Murray, Garud Iyengar, Remi Monasson, Simona Cocco and Juliette Ben Arous.

Less frequent, but all the more intense, were the faculty committee meetings, in which, along with Stan, my committee members helped me focus on the central questions. I want to thank the chair, John McKinney and the members, Roald Hoffmann, Albert Libchaber, and now Anthony Dean. I look forward to meeting them once more!

It is a privilege to do science in such an amazing place, and with such great people. I have benefitted from two educational systems. The first, in the Netherlands, provides a good education to anyone. The second one, in the United States, is less egalitarian, but, at its best, provides great opportunity. Rockefeller University is a great example of this. Specifically, I want to acknowledge support from Cornell University and the Training Program in Chemical Biology, led by Tim Ryan.

Finally, there are the people who made my stay in New York not just a stay but a life. There is the 'sushi club' including Nicolai, Umut, Holger, Doro, Carl, Andrea, Taulant and Mary, with good conversation and friendship. Likewise, Margherita, Amrita, Oliver and Megan provided me with the balance and perspective needed.

Finally, of course, there is Ellina, my soon-to-be wife. She is grounded, loving and hard working, always committed to making the best of life, even when it doesn't come easy. With her by my side, there has been and will be no reason to be afraid of the unknown.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

Biology Abbreviations

aa	amino acid
cam	chloramphenicol
cat	chloramphenicol acetyltransferase
DMSO	dimethyl sulphoxide
EGFP	enhanced green fluorescent protein
EYFP	enhanced yellow fluorescent protein
H3.2	histone 3.2
LB	Luria-Bertani medium
NB	nutrient broth medium
nt	nucleotide
ORF	open reading frame
ori	origin of replication
PCR	polymerase chain reaction
SPP	supplemented proteose peptone
TAP	tris-acetate-phosphate medium
TGB	thioglycollate broth medium
YPD	yeast-peptone-dextrose medium

other abbreviations

ACF	autocorrelation function
CCD	charge-coupled device (camera)
CCF	cross correlation function
c.d.f.	cumulative density function
DAC	digital to analog converter
DC	directed current
ELWD	diode-pumped solid state (laser)
DPSS	extra large working distance
LED	light emitting diode
LWD	large working distance
MI	mutual information
MOS	metal-oxide semiconductor
OD	optical density
NA	numerical aperture
r.m.s.	root mean square
TTL	transistor–transistor logic (0/5 V)

Symbols

$A,B,\ldots$	species indices (see below)
$a_0, \ldots$	general parameters
d	depth of focus; dimension
f	focal length; frequency; fitness
H	uncertainty
$i, j, \ldots$	general indices
I	light intensity; electrical current mutual information
M	magnification
n	population density; refractive index
N	total population; number
r	growth rate (base e)
R	resistance
s	linear object size; a general signal
S	power spectral density
t	time
T	temperature
w	light sheet width; scaled wavelet
W	wavelet transform
V	voltage; volume
x	unobserved or true variables
γ	denotes ratios
δ	light sheet thickness
θ	general model parameters
λ	wavelength
μ	mean
σ	standard deviation
τ	time scale, exposure time
ψ	mother wavelet
ω	frequency
$Ø$	aperture diameter

Species mentioned in this thesis and symbols, colors used for graphs

Chlamydomonas reinhardtii	A, green
Escherichia coli	B, red
Tetrahymena thermophila	C, blue
Euglena gracilis	A, green

Chemical names and symbols are according convention.

Plasmid names follow Lutz and Bujard, *Nucleic Acids Research* 25 (6) 1203-1210 (1997)

Chapter 1 Introduction

Self-replication and its consequences

Much of (molecular) biology is concerned with the intricate details of how organisms self-replicate, from the bacterial cell cycle to development in plants and animals. But, the mere fact that organisms are capable of self-replication introduces questions beyond chemistry. Charles Darwin was one of the first to grasp the consequences of self-replication[1] [12].

> "In the next chapter the Struggle for Existence amongst all organic beings throughout the world, which inevitably follows from their high geometrical powers of increase, will be treated of. […] As many more individuals of each species are born than can possibly survive; and as, consequently, there is a frequently recurring struggle for existence, it follows that any being, if it vary however slightly in any manner profitable to itself, under the complex and sometimes varying conditions of life, will have a better chance of surviving, and thus be *naturally selected*."

Self-replication then places two kinds of constraint on the diversity of life. First, the apple never falls far from the tree. In other words, evolutionary innovation is random but limited. Such limited random variation can have idiosyncratic effects: the order and timing of mutations within replicate populations can be different, setting different populations off on different futures.

[1] Along with the implication that species on earth could share common ancestry.

In a more general ecological context I will call such effects *historical*, understood as the long-lasting, idiosyncratic effect of random change, be it genetic, chemical or numerical (for example the random timing of birth and death).

The second constraint posed by evolution on the diversity of life comes from natural selection. New variants will have the potential to compete with ancestral and alternative variants for generally scarce resources. Some offspring will be better suited for an environment than others and have a higher chance of survival and continued self-replication, that is, some will have higher *fitness*.

This thesis is focused on the question how the action of random events and natural selection can be characterized in multispecies ecosystems.

Inferring constraints

A molecular description of the action of natural selection and historical effects may be instructive, but does not adequately describe their mode of action [13]. In addition, there are only rare examples in which fitness, or at least relative fitness between strains, can be predicted from molecular information.

In one example, detailed thermodynamic knowledge about residues critical for catalysis in isopropylmalate dehydrogenase allowed Lunzer *et al*. [14] to predict the catalytic efficiency of various mutant enzymes and, since the action of the enzyme was essential under the conditions examined, to estimate the fitness of mutants.

Probably the most successful models to date in predicting fitness from molecular information have been ones based on optimality: the assumption that, within certain constraints, the fitness of organisms has been maximized given what is physically

possible. This approach has been most extensively pursued by Palsson and colleagues for bacterial metabolism [15]. Knowledge of the reaction network of metabolic pathways together with the growth requirements of *E. coli* allowed for fairly good prediction of growth rates in minimal media (errors 5-10%) [16], but yielded only qualitative predictions for *E. coli* mutants and other species [15]. For the case of the *E. coli* lac operon, Dekel and Alon [17] demonstrated that, while lac operon expression in the wild-type strain was not optimal for the conditions studied, *E. coli* achieved optimal expression levels within a few hundred generations[2], suggesting optimality assumptions can be reasonable. The examples are instructive but their precision is limited. In addition it seems exceedingly difficult to extend these methods to situations more complex than single well-characterized species under exponential growth in simple media.

Molecular biology, in summary, doesn't get us far in determining who will and who will not survive. Biological systems are very complex: they are extremely heterogeneous in their molecular constituents and spatial organization. Moreover, they are heterogeneous in a far-from-random way, so average properties are close to meaningless [18]. And, even if adequate rate constants and spatial localization were known for all molecular processes, these processes are often nonlinear and thus hard to forecast on the time scale of the life time of organisms.

As mentioned, self-replication can introduce historical effects. These effects are hard to predict for the same reasons. In addition, they depend on small probabilities of individual events. That is, the lives of organisms are idiosyncratic: their composition and their experiences in their environment are heterogeneous and variable.

[2] As inferred from measured costs and benefits of lac operon expression.

The aim of the work in this thesis is to advance an alternative method for studying the determinants of survival. In essence, we will work towards a phenomenological method: making inferences about the action of natural selection and history from observed population dynamics in the laboratory. For example, fitness is often formulated as the expected long term per capita growth rate [19]:

$$f_a = \langle r_a(t) \rangle = \lim_{T \to \infty} \frac{1}{T} \left\langle \log \frac{n_a(T)}{n_a(0)} \right\rangle = \left\langle \frac{1}{n_a} \frac{\mathrm{d}n_a}{\mathrm{d}t} \right\rangle \tag{1.1}$$

where n_a is the population density of a species a, r_a its instantaneous growth rate, and T a time window. The average is, in theory, a weighted average over possible environments. One of the simplest models describing the effect of interspecies interactions on growth rate is the Lotka-Volterra model. It describes the dependence of the instantaneous growth rate on the density of itself and other species, n_b, by proportional coefficients A_{ab}:

$$\frac{1}{n_a} \frac{\mathrm{d}n_a}{\mathrm{d}t} = r_a^0 - \sum_b A_{ab} n_b \,, \tag{1.2}$$

with r_a^0 the growth rate when all species densities are low, and A_{ab} the interaction coefficient between species a and b. From equations and 1.1 and 1.2,

$$f_a = r_a^0 - \sum_b A_{ab} \langle n_b \rangle \tag{1.3}$$

with

$$A_{ab} = -\frac{\partial}{\partial n_b} \left(\frac{1}{n_a} \frac{\mathrm{d}n_a}{\mathrm{d}t} \right) \tag{1.4}$$

This is a simple model, and numerous extensions have been proposed (see, e.g. [20]), including, for example, delayed interactions and nonlinear density-dependence. Equations 1.3 and 1.4 show what we hope to achieve though: the inference of contributions, like species-species interactions A_{ab}, to fitness, f_a, from population dynamics time series $\{n_a(t), n_b(t)\}$. Such a problem, the inference of system properties from system behavior, rather than the prediction of system behavior from system properties, is called an *inverse problem*.

Outline of this Thesis

In this thesis I argue that the use of closed ecosystems, that is, ecosystems closed to material transport, addresses two major obstacles facing us when attempting to infer species-species interactions and stochastic effects from time series. The first obstacle is that any biological system has an enormous number of potentially relevant dimensions. The second obstacle is reproducibility. I will introduce both of these challenges here.

First, *any biological system has an enormous number of potentially relevant unobserved dimensions*. This manifests itself in equation 1.4: partial derivatives are never directly observed, but at best total derivatives. That is,

$$\frac{\partial}{\partial n_b} = \frac{\mathrm{d}}{\mathrm{d}n_b} - \sum_x \frac{\partial x}{\partial n_b} \cdot \frac{\partial}{\partial x} \tag{1.5}$$

with x additional variables characterizing the system. The traditional response to this challenge is to measure any conceivably relevant variable (pH, temperature, glucose levels, specific phenotypes), and consider other variables irrelevant. As argued in Chapter 2, closed ecosystems suggest a different approach. Under material closure, a number of

unknown variables, material fluxes across the system boundaries, is reduced to 0 and the remaining boundary conditions, light and temperature, can be controlled. As a consequence, any remaining chemical change is solely due to interactions between organisms and to the laws of chemistry. Hence, *these chemical changes are part of the very interaction coefficients estimated*, rather than external factors that need to be controlled for or kept constant (e.g., in a chemostat). The remaining two classes of variables not observed in time series of density over time, are spatial and phenotypic. Both spatial heterogeneity and phenotypic variation are inevitable properties of biological systems, and rather than minimize them, we hope to make inferences about them from observed time series.

The second obstacle is reproducibility. I have described the near-impossibility of predicting fitness. *If fitness cannot be predicted, we need to be able to measure it reproducibly, or it has little place in biology as an explanatory concept*. Difficulties in measuring fitness arise from a number of causes: systems are often poorly defined, not only 'in the field' but also in laboratory experiments; replication, if at all, is often insufficient; and time series are often too short and poorly sampled, that is, the ecosystem studied can have intrinsic time scales faster than the measurement time scale. In Chapter 3 I introduce a closed ecosystem, based on work by Kawabata and colleagues [21], for which the boundary conditions (light and temperature) and initial conditions (chemical and biological composition) can be controlled precisely. In Chapters 4 and 5 I demonstrate that time series of species densities can be measured for this system which are reproducible, in a statistical sense, over months. The system consists of microbial

species, offering the advantage of fast generation times, such that time series over many generations, not just long times, are obtained.

The measurement method is described in Chapter 4. It is based on selective plane illumination microscopy [22] to acquire time series of the population densities in a small observation volume within a closed ecosystem over long periods of time. I demonstrate that the method has a large dynamical range and low classification error. As such, strictly speaking, the system studied is a small (0.01-0.1 mm^3) open subsystem of a larger (5 cm^3) closed ecosystem.

Chapters 5 to 7 describe the actual use of the model closed ecosystem introduced in Chapter 3 to study historical effects and species-species interactions.

In Chapter 5, the divergence in composition between replicate ecosystems is studied over periods of 10 and 14 weeks. The results show the role of history, that is, of random events with lasting consequences. I demonstrate that ecosystems started from the same initial conditions and kept under constant light and temperature diverge reproducibly. The total system divergence is nearly linear over time, while individual species densities diverge in species-specific ways. Secondly, differences between ecosystems arising in the first few weeks of the experiment persist for several weeks.

In Chapter 6, fluctuations in individual time series under constant boundary conditions are examined. After the development of a null model for fluctuations in the obtained time series due to measurement noise, the fluctuations in time series obtained by continuous measurement of population densities in two ecosystems are studied. It is shown that the

power spectrum of these fluctuations is ‘open-ended’: the power in fluctuations increases with time scale, at least up to the duration of the experiment.

In Chapter 7, I will examine periodic perturbations in the boundary conditions of the model ecosystem, specifically the intensity of illumination. These perturbations leave an interesting signature in the population dynamics observed. In particular, the results suggest a strong interaction between the algae and bacteria.

In the concluding chapter I summarize the main findings of this thesis and suggest ways to make progress on the inverse problem, the inference of the effects of history and natural selection and their underlying causes from population density time series.

Chapter 2 Review of Closed Ecosystems

Abstract

Closed ecosystems were originally developed with the prospect of applications in space travel and colonization. Most of this work has followed an engineering approach, but small aquatic ecosystems, on which I will focus, were also developed, especially by the group of Clair Folsome. These ecosystems contained species communities persistent over years, or hundreds or thousands of generations and showed efficient material cycling. In addition, samples from natural ecosystems were successfully subcultured into chemically defined medium. Their persistence and the tight control over initial and boundary conditions they allow, suggest that closed ecosystems are ideal systems for the study of adaptation, species-species interactions, and the effects of ecosystem properties on population dynamics. Finally, to fulfill this promise further technical development is necessary and feasible.

Introduction

Laboratory experiments are essential to ecology. Natural ecosystems suffer drawbacks which make interpretation and quantitative replication of experiments virtually impossible: they often have poorly defined boundaries; their species composition is not fully known; and their shear spatial and chemical complexity generally precludes replication and accurate measurement.

Laboratory ecosystems, or microcosms, are small ecosystems held in containers[3] [23]. They offer a chance to study communities of species in their entirety, can be replicated, and ideally allow for the acquisition of long time series of chemical and biological dynamics. Closed ecosystems take the microcosm approach to its extreme: complete material closure. In this chapter, I review studies of small aquatic closed ecosystems. The much larger literature on microcosms will be discussed only when relevant. Some of the terminology of the ecology literature is summarized in the glossary (Table 2.1).

Both the words 'closed' and 'ecosystem' need clarification. As in thermodynamics, closure implies no material can enter or leave a system, but energy can (e.g., as heat or light) [2]. This is the sense in which I will use it. Some researchers [3] have understood closure as the ideal of complete recycling of the elements in an ecosystem, closed nutrient loops. This kind of closure is hard to prove in practice. Likewise, any assemblage of species and its abiotic environment will be considered an ecosystem regardless of how 'well' it functions.

[3] In practice, the term microcosm applies to systems <100 L (roughly the size of a typical aquarium), while larger systems are more appropriately called mesocosms.

Table 2.1: Glossary of pertinent ecological terms
Chemostat: culturing system for one or more species in which the growth medium is continuously refreshed with the goal of keeping its chemical composition constant.
Batch Culture: culturing system in which growth medium is supplied only at the start.
Persistence: continued (co)existence of the species considered over many generations.
Stability: an ecosystem is said to be stable if (a) there are very limited fluctuations in system parameters considered, or (b) the system can return to its original state after perturbation with respect to some parameter(s).
Self-selection: formation of a persistent community of species from a larger pool, typically over a period of a few months [9, 10].
Eutrophication: an increase in available nutrients or biomass in an ecosystem.
Primary production: production of new biomass directly coupled to photosynthesis
Destructive sampling: sampling by methods which require opening the ecosystem, such as for haemocytometry and plating.
Dormancy: continued existence in a state of minimal or suspended metabolic activity and growth.

Now, what would an ideal closed ecosystem look like? It would allow a known set of species to coexist over many generations within well-defined, sealed boundaries. I will call such coexistence over many generations (i.e., not just survival in a dormant stage), persistence. Its boundary conditions, especially light and temperature, can be controlled, with light serving as the only continuously available source of free energy. Likewise, its initial conditions are defined: its chemical medium and the numbers of individuals of each species present at closure. As I will describe, significant progress has been made towards this ideal.

Closure of ecosystems offers several methodological advantages over open microcosms and chemostats. First, once an ecosystem has been closed, there is no further risk of contamination. Closure also prevents exchange of gases, such as O_2, NH_3 and H_2S, as

well as volatile organic molecules, which are an integral part of interactions between organisms [24], and evaporation, and with it the need for periodic refreshment or addition of medium. In addition, it prevents gradual eutrophication of the ecosystem due to ongoing nitrogen and carbon fixation as observed in open microcosms [1, 25]. In short, closure of ecosystems allows for a conceptually consistent interpretation of interactions between organisms: *any* molecular exchange of an organism with its environment contributes to its interactions with other organisms and potentially affects its fitness. Finally, microcosm and chemostat experiments have been dismissed because of their artificiality (criticism reviewed by Jessup *et al.* [26]). Any microcosm to some extent creates a new environment and species assemblage. Closed ecosystems give us the opportunity to study both adaptation, and after sufficient time, communities that can persist, suggesting their composition is not arbitrary.

I will describe what is known about the construction and properties of closed ecosystems and discuss how study of closed ecosystems can provide information on species-species interactions, their change over time and the interplay between the properties of constituent species and the properties of the entire ecosystem.

A brief history

Closed ecosystems research started in the 1950s in the US and the USSR (at the Institute for Biophysics, Krasnoyarsk, Siberia), with an eye on space travel and colonization (much later similar research was initiated in Europe [27] and Japan [28]). These studies aimed, and still aim, at sustaining human life in space, building so-called closed ecological life support systems (CELSS), which could for example recycle human waste

and CO_2, and hence were less concerned with achieving full closure. In addition, many of these systems are large, with separate compartments for crop production, waste management, etc., and mostly have an undefined species composition. Early research in the US was reviewed by Taub [29]. Later work has been comprehensively reviewed in a recent book [30].

The project in this tradition most relevant to ecology has been the development of Biosphere-2, a large closed model of the earth's biosphere. This project has received widespread attention, and was extensively reported [31-33].

Our interest, however, is primarily in small (1-1,000 mL) aquatic closed ecosystems, which have great potential to increase our understanding of ecosystems, and of the adaptations of organisms to these ecosystems.

A second pertinent line of research started in the late 1950s with the work of H.T. Odum, E.P. Odum, R.J. Beyers and coworkers. They developed the ecosystem concept and constructed some of the first small aquatic microcosms. For example, they addressed the effects of temperature and ionizing radiation on photosynthesis and respiration, invasion by new species, and the effects of eutrophication (all reviewed in [23]). While they advocated the use of closed ecosystems for space exploration [34], their own efforts have remained unpublished ([23], R.J. Beyers, pers. comm.).

By the late 1960s, NASA's interest and support for ecological approaches to life support systems, rather than engineering, had waned. In the late 1970s, a few researchers tried to revive the research on closed ecosystems in the United States, most notably Clair Folsome at the University of Hawaii (see [35] for notes from a workshop conducted in 1982). Folsome's lab has since provided much of the empirical knowledge on small

Figure 2.1: These one liter marine ecosystems were developed by Joe Hanson at JPL/NASA and Clair Folsome, of the University of Hawaii. In 1983, NASA licensed the idea to Engineering & Research Associates Inc., Tucson, AZ . The ecosystems are now produced from "semi-sterile" stock cultures and filtered seawater, with added gravel and *Gorgonia* coral, stated to be non-living, for decoration [6, 7]. According to the EcoSphere website (http://www.eco-sphere.com/care_manual.html), shrimp have survived in such closed ecosystems for up to eight years. Gracing many homes, this may well be the most replicated experiment in ecology.

aquatic closed ecosystems, which will be reviewed in some detail below. His collaboration with Dr. J. Hanson of NASA's Jet Propulsion Laboratory resulted in the commercialization of marine closed ecosystems as EcoSpheres® (Figure 2.1).

How to construct a closed ecosystem?

It is far from clear which ecosystems can persist under closure. For example, what are the constraints on the initial set of species? What interaction patterns are stable? What, if any, is the minimal metabolic diversity required? In general, closed ecosystems can be constructed in two basic ways: in top-down construction, a sample from a natural system is allowed to "self-select" after closure until a remaining set of species persists; in bottom-up construction individual species from stock cultures are combined in a suitable medium.

Most research to date has followed the top-down approach. As a graduate student with H.T. Odum, Beyers showed that samples from a sewage oxidation pond, when left for a few months in 1 L of chemically defined medium, formed a persistent community of eight eukaryotic species and at least 11 bacterial species [9, 36]. This community could subsequently be subcultured in a defined minimal medium with added proteose peptone (hydrolyzed protein) [9, 34, 37]. Kurihara, a former post-doc of Beyers, and Sugiura [38, 39] showed that in open (cotton-plugged) Erlenmeyers, a similar community could be subcultured monthly for 15 years. This community subsequently could persist after closure for at least one year [40] in a 300 mL container with 100 mL atmosphere.

Folsome and coworkers followed a similar approach using samples from Kaneohe Bay, Hawaii, containing sediment and seawater. They initially directly closed samples, but later showed that small samples (<5% of total carbon included) of diluted homogenized sediment/seawater samples could be used to inoculate artificial seawater and obtain 20 mL closed ecosystems which were almost as photosynthetically active as the original closed ecosystems and persisted for over 400 days [25]. The species composition of these ecosystems was not determined.

Few attempts have been made at constructing closed ecosystems bottom-up. Kawabata *et al.* [21] constructed three-species microcosms in a defined medium. They showed that the three species, *E. coli*, *Euglena gracilis*, and *Tetrahymena thermophila*, persisted for over 130 days. They did not explicitly address closure, and it is unclear whether these systems were completely sealed. In Folsome's group, Obenhuber [41, 42] and Wright [43] explored the use of closed ecosystems containing one, two or three species of algae.

Monocultures were shown to persist for over 200 days, but were not studied in detail because biomass did not stabilize and CO_2 turnover was relatively slow.

One can also imagine two hybrid options: isolating individuals from a natural ecosystem, followed by reconstitution in artificial medium (briefly explored by Obenhuber [41]). Similarly, one can take known species, combine them, and wait until a persistent community remains [10, 44].

The role of boundary conditions

The design of a closed ecosystem also entails the choice of boundary conditions: its size, temperature and illumination. The properties of a closed ecosystem presumably scale with its dimensions: its aqueous phase, its atmosphere, its internal surface area. To date, few relevant studies have been published. Maguire [1] examined the persistence of small crustacea in closed ecosystems derived from natural ponds. Sugiura *et al.* [8] changed the ratio of internal atmosphere to water for their closed ecosystem mentioned above, and showed that even after 25 days of closure population densities were markedly dependent on this ratio. Systems with larger atmospheres sustained much higher population densities (results from both Maguire and Sugiura *et al.* are shown in Figure 2.2). The results of Sugiura *et al*. support the notion that open microcosms gradually eutrophy, but presumably depend on which elements are growth-limiting.

Even less is known about the influence of light or temperature on closed ecosystems, and one would have to rely on results from open microcosms [45, 46]. The only result for closed ecosystems comes from the work of Kearns and Folsome [47], who showed that primary productivity in their marine closed ecosystems, for the first 180 days after

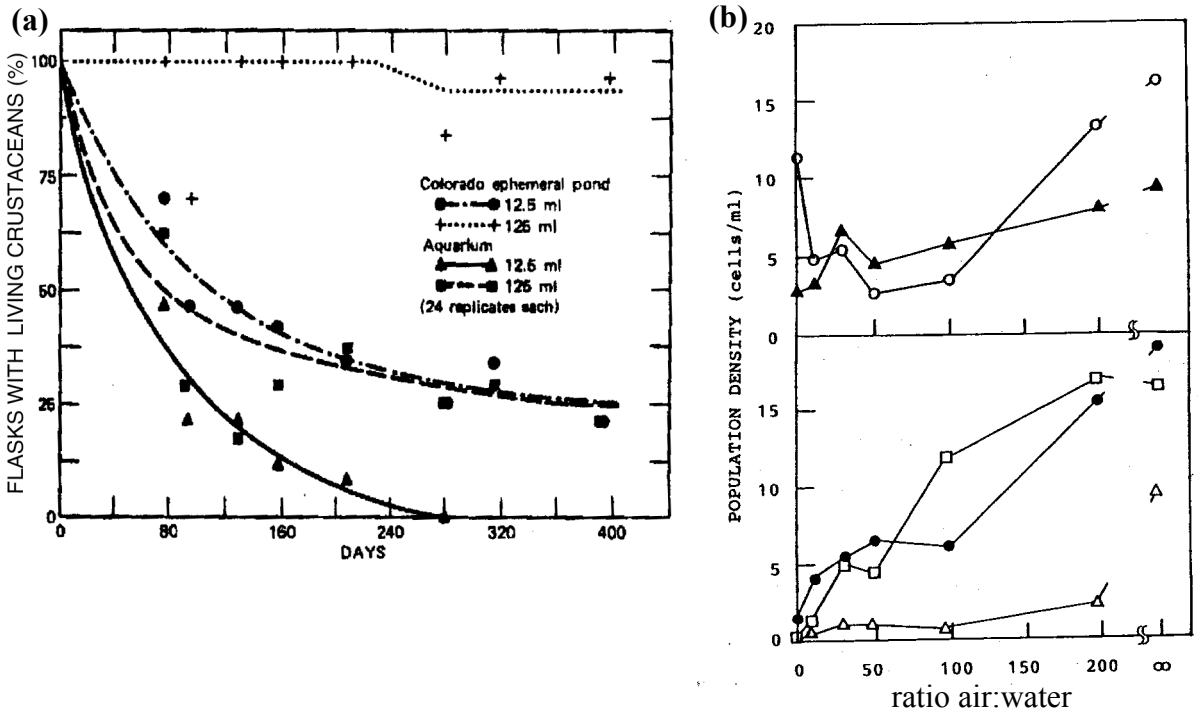


Figure 2.2: Effects of boundary conditions.
(a) Size: Maguire [1] studied the survival of crustacea as a 'focal group' in closed ecosystems of different sizes in flame-sealed glass volumetric flasks. Shown here is survival in ecosystems obtained from a pond community, 125 mL systems (···) and 12.5 mL systems (-·-), and in ecosystems obtained from a lab aquarium which had undergo many additions over the years; 125 mL, dashed line; 12.5 mL solid line. For each set, there were 24 replicates.
(b) Internal proportions: Sugiura *et al*. [8] constructed self-selected 300 mL ecosystems with different proportions of water and atmosphere. Shown are densities for different populations 25 days after closure (open circles: bacteria ($x10^5$), closed triangles: *Chlorella* ($x10^6$), open triangles: *Cyclidium* (x100), closed circles: *Philodina* and *Lepidella*, open squares: *Aelosoma*). Infinity refers to open systems. (Reproduced with permission from the Institute of Space and Astronautical Science).

construction, depended hyperbolically on illumination, reaching half-maximum at about 11 klux.

How long? And, how long is long?

There are anecdotal observations of closed ecosystems surviving for very long times. Folsome's ecosystems reportedly showing signs of persistence for over 18 years [48], Hanson's for over five years [7, 49], and Beyers' closed ecosystems for over thirty years [1, 23, 29], until they were discarded at his retirement (R.J. Beyers, pers. comm.). I am not aware of any published documentation of these claims or extant examples.

The longest documented examples appear to be two-liter closed ecosystems constructed by Kearns and Folsome [50]: Brittain successfully subcultured a persistent community from two of these closed ecosystems after six years of closure [48]. Kearns and Folsome [47] showed ongoing respiration after eight years, comparing total pressure in light and prolonged darkness (see next section).

Ultimately, however, two time scales are more important than chronological age: (1) the time scales of material cycling (see Appendix 2A.1), and (2) the generation times of the species present. Only if a closed ecosystem can efficiently recycle its elements, can it be expected to persist in the long run; and only if species present can coexist for many generations, can we expect to be able to study interactions, their fitness effects and ensuing adaptation.

Material cycling

The study of material cycling has a long tradition in ecology, and studies have been planned for larger terrestrial closed ecosystems [51, 52]. For small closed ecosystems, however, the only results available are simple, but elegant, experiments by Folsome's group, which I will explain in some detail. Obenhuber and Folsome [41] showed that

small (20 mL) closed ecosystems can be formed with a negligible amount of biomass and an otherwise inorganic medium in which elements are supplied in their fully oxidized form[4] (e.g., SO_4^{2-}, NO_3^-, K^+, and crucially HCO_3^- as the only significant carbon source). In such an ecosystem, redox bookkeeping implies that the reducing equivalents (electrons) required for the formation of new biomass are provided by forced oxidation of water in photosynthesis, accompanied by release of molecular oxygen, O_2.[5] Since O_2 is much less soluble in water than CO_2, photosynthesis drives an increase in pressure proportional to the amount of carbon fixated in new biomass[6]. Folsome and Obenhuber indeed found good correspondence between oxygen partial pressure (as measured by gas chromatography) and pressure [41].

Photosynthesis closely links the 'capture' of free energy of illumination to carbon dioxide reduction. Folsome's group made the reasonable assumption that the rate of cycling of other elements is closely linked to carbon cycling [49]. In addition, they assumed that respiration stays approximately constant throughout a light-dark cycle, while photosynthesis ceases in the dark. Pressure changes due to oxygen production and consumption over a light-dark cycle should hence provide information on gross rates of both photosynthesis and respiration. Such pressure changes are illustrated in Figure 2.3 (a) and (b).

Folsome and Obenhuber chose a medium in which no element was present in large excess over biological requirements, and determined turnover times for carbon dioxide/carbonate. Under these conditions a leak in any element (Appendix 2A.1) will

[4] Except for elements more electron-negative than oxygen, such as chlorine (as Cl^-).
[5] Which is not to say, photosynthesis cannot proceed by oxidation of other molecules, such as H_2S at a later stage, but the reducing equivalents for the production of H_2S would first have been provided by H_2O.
[6] Assuming overall biomass oxidation state remains constant.

slow down cycling of all elements once the element becomes growth-limiting. For four complex communities multiple rounds of stable turnover were observed; estimates of mean turnover times were 11, 31, 61 and 75 days. Linear extrapolation of the observed pressure fluctuations (Figure 2.3 (c)) allowed for the projection that these systems would each show positive turnover for at least 1000 days.

Obenhuber's experiments, while addressing carbon dioxide turnover, also allow for estimates of turnover of total biomass (alive and dead combined), ranging from 25-200 days, depending on the initial species composition. This suggests detritus can introduce markedly slow timescales in material cycling.

Following Folsome's death in 1988, work on small aquatic ecosystems largely ceased. Since then, however, the study of elemental fluxes has become much more prominent, and been coined ecological stoichiometry [53]. Elemental fluxes are coupled, and place constraints on population dynamics [54].

Comparison to open ecosystems

While closed ecosystems are arguably perfectly legitimate objects of study *per se*, it is instructive to see how they compare to open microcosms and, when derived from nature, to their "parent ecosystems". Folsome and his students performed three such comparisons.

Brittain [48] opened two six year old closed ecosystems, with a volume of two liters each, containing 300 mL of sediment and 300 mL of seawater found just above it. Using a set of selective growth media, he demonstrated that all metabolic functions found in two fresh samples from the original site, constructed identically, had been retained in

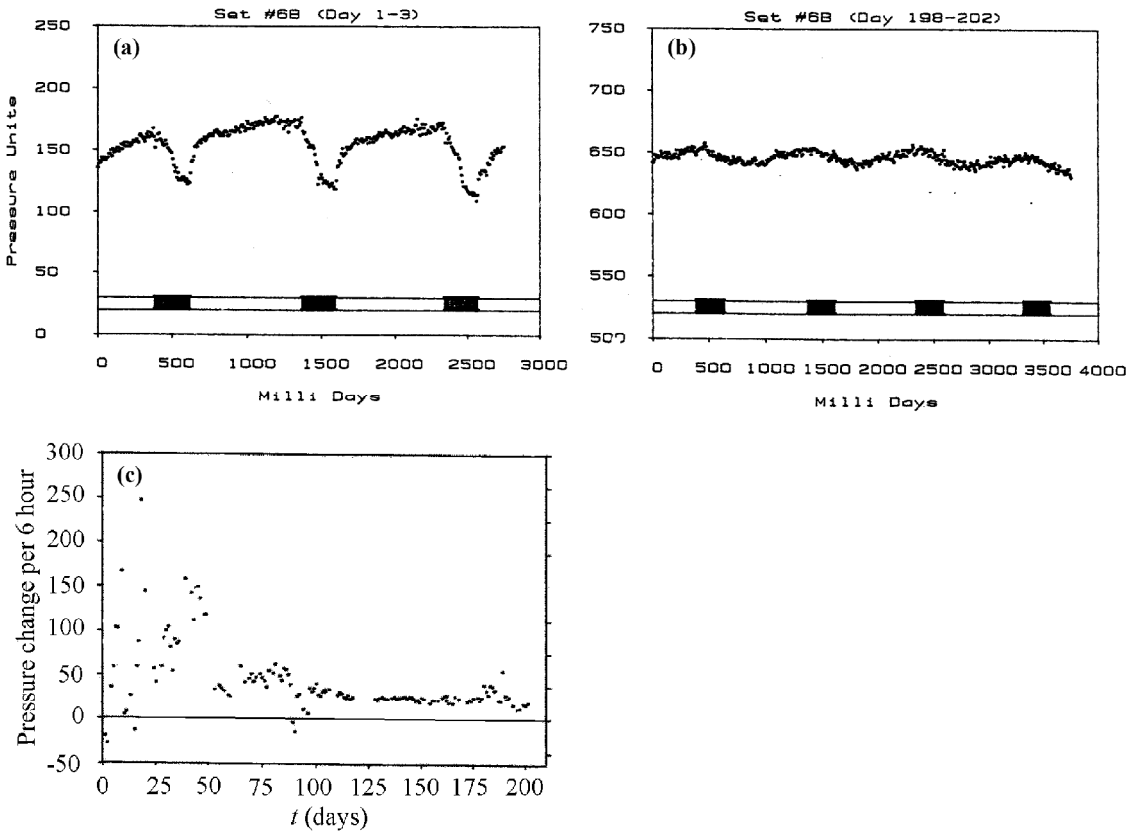


Figure 2.3: Diurnal (daily) changes in pressure (60 PU = 1 kPa), relative to outside pressure at the start of the experiment. Data for a closed ecosystem constructed from a sample from Kaneohe Bay, HI. Shaded bars indicate light and dark periods. (a) shortly after closure, (b) 200 days after closure. Oxygen is produced immediately in photosynthesis and is largely insoluble in water, leading to an increase in total pressure during illumination. At night, photosynthesis ceases, while respiration continues consuming oxygen, leading to a drop in pressure. (c) inferred rate of photosynthesis, proportional to the difference in rate of pressure change during light and dark periods. Note the difference in vertical scale between panels (a) and (b): the marked increase in pressure corresponds to the net production of biomass over the intervening timespan. Reproduced with permission from D.C. Obenhuber's thesis, U. Hawaii [41].

these closed ecosystems, finding only differences in relative numbers of organisms in each metabolic category. Shaffer [25] compared closed and open 20 mL microcosms prepared identically by dilution into artificial seawater of a small inoculum[7] of five-year-old closed ecosystems made by Kearns. After five months, kept under identical

[7] 0.2% of total carbon included was still sufficient.

conditions, there were no major differences in metabolic diversity (using a set of assays similar to Brittain's above) between open and closed ecosystems. The most remarkable difference was in total biomass carbon (alive and dead). As judged by pyrolysis and pressure changes, total biomass carbon stabilized in closed ecosystems after one to four months, depending on the size of the initial inoculum. In open ecosystems, no stabilization was observed, but rather a steady increase, reaching after five months levels five to eight times higher than in closed ecosystems. No parallel increase was seen for nitrogen (as nitrate) or soluble phosphate.

Finally, Shaffer [25] examined the effects of opening similar closed ecosystems after a month of development under closure. Given the previously discussed results it was not surprising that he observed an immediate surge in primary productivity (roughly tripling compared to systems which remained closed), but over time no clear increase in alive biomass was seen, only in total biomass. This suggests that most of the surplus productivity eventually leads to detritus formation. Consistent with this, he observed a 2-3 fold increase in heterotrophic bacterial density.

In conclusion, closure for these marine ecosystems placed strong restrictions on productivity, leading to a much faster stabilization of total biomass than in open ecosystems. At the same time, metabolic diversity did not suffer from closure.

Population dynamics

Up to now, this review has focused on ecosystem properties. Over most of this thesis I will focus on the study of population dynamics in closed ecosystems. Unfortunately, very few data are currently available. Pisman *et al*. [55] observed competitive exclusion

between two algal species, and between a rotifer and a ciliate in an elaborate two-compartment closed ecosystem, but sampling was too sparse for quantification of the interaction. Sugiura [40] obtained time series using destructive sampling over an ensemble of closed ecosystems. Because of the low time resolution (one month) and because of probable divergence between ecosystems, these data did not provide much information on interactions. As argued below, however, closed ecosystems have the potential to provide a framework for the study of interactions between organisms, and for the effects of ecosystem properties on the evolution of individual species. To achieve this, long time series are needed to establish even basic ecological effects, typically at least 50-100 generations long [56].

Despite the absence of good time series data for closed ecosystems, there are some data on the overall rate of division in one of the ecosystems of Folsome's group. Brittain and Karl [57] showed that assays of ^{3}H-thymidine and ^{3}H-adenine incorporation into DNA are problematic for marine ecosystems. Even for thymidine, which is incorporated only into DNA, radioactivity rapidly (<10 min) appeared in protein and RNA. This suggests catabolic pathways act faster on thymidine and adenine than incorporation into DNA. Instead of examining incorporation, Brittain and Karl measured the dilution of the ^{3}H-thymidine and ^{3}H-adenine signals in DNA by cell division after pulse labeling. This method provided an estimate of overall DNA replication time of approximately 1 day, 100 days after construction and closure of the studied ecosystems. As described, Folsome's ecosystems can persist for years. The experiments of Brittain and Karl show that these closed ecosystems can also persist over hundreds or thousands of generations.

The nature of microbial interactions

When studying the interactions within and between species experimentally, we have to grapple with the problem of only observing a few coarse-grained dimensions of the system, while other dimensions can vary at will. This was vividly demonstrated by Matsui *et al*. [58] using the three-species ecosystem developed by Kawabata *et al*. [21] mentioned above. The authors demonstrated the importance of indirect (metabolite) interactions comparing growth curves of species in each other's presence to growth of individual species in medium supplemented with spent medium from single-species cultures. Indirect interactions rivaled direct interactions, such as predation and resource competition, in effect. In another striking example, Warren *et al*. [44] showed that some open-microcosm communities could not persist when assembled by themselves, but could when originally assembled with an extra species which then went extinct. *These examples show that shifts along unobserved chemical or phenotypic dimensions of ecosystems can dramatically affect observed population dynamics.*

Chemostats aim to address this issue by keeping the composition of growth medium as constant (and simple) as possible. If chemostats really kept the chemical composition of the medium constant, the wealth of molecular interactions between organisms could simply not be studied. For adaptation of a single species, such as in Lenski's experiments, chemostats or periodic subculturing seem suitable though, offering the additional advantage of acquiring a detailed "fossile record" [59, 60]. Chemostats were also rapidly applied to the the study of species-species interactions, pioneered by A.G. Fredrickson (for reviews, see [61, 62]). Early examples were the study of T-series bacteriophage and *D. discoideum* on *E. coli.* Such systems can still be interpreted mechanistically as long as

the growth-limiting factors are clear, for *E. coli* the limiting nutrient, and for the predator the prey, *E. coli*. The cases in which a single limiting resource is consistently present, are, however, exceedingly rare. Closure of ecosystems offers a consistent complementary approach: no material fluxes across the system's boundaries. Under such conditions, interactions are not distorted or washed out, but can act on any timescale.

Outlook

This review has focused on the methodological advantages of full material closure, and experiments pertinent to basic questions about closed ecosystems. Closure, however, does pose technical challenges. Any measurement needs to be non-invasive, or if necessary, minimally invasive, using for example an airtight septum and not perturb the system any more than necessary.

The only non-invasive measurements of material cycling discussed so far were the ingeniously interpreted measurements of pressure fluctuations by Folsome's group. Now sensitive spectroscopic measurements are available, such as IR spectroscopy for CO_2, NH_3, and other gasses with (inducible) dipoles. The shown ability to subculture closed ecosystems into chemically defined media also allows for isotopic depletion or enrichment, opening the way for the full repertoire of, for example, ^{13}C and ^{15}N NMR techniques [63]. Likewise, mass spectrometry techniques have become so sensitive that picomoles of substances can be detected [64-66], allowing use of very small samples. Other chemical and biological sensing techniques have been extensively described for oceanography [67], many of which seem readily applicable to closed ecosystems.

Likewise, a range of methods exists for non-invasive monitoring of population dynamics. There are optical methods based on fluorescence or absorbance measurement of various pigments, offering discrimination between phyla in complex communities, or fluorescent proteins [68-70]. Direct imaging, or *in situ* microscopy, has found application in for example oceanography [71] and biotechnology [72] to quantify populations. Acoustic measurements have allowed for the determination of the distribution of algae in seawater [73].

When studying complex communities it may be helpful to choose a "focal species" which can be fluorescently marked or its genetic barcode [74] tracked, for example by sensitive quantitative PCR techniques now available [75, 76]. Specifically, in this thesis I will describe an *in situ* fluorescence microscopy method allowing for the acquisition of long time series of population densities within a closed ecosystem.

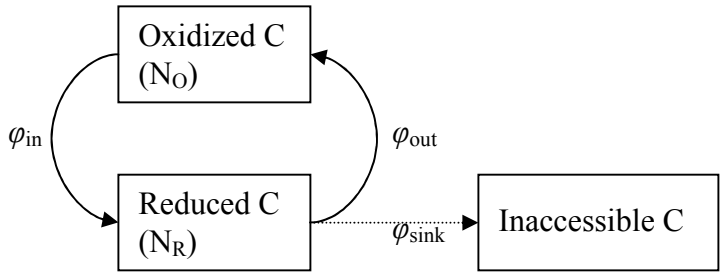
As mentioned in the introduction, any microcosm study involves a certain degree of material closure of an ecosystem from its environment. In this review, I chose to focus on those studies that explicitly examine closure: under which conditions a community of species can persist after closure, its effect on ecosystem properties such as total biomass and elemental fluxes, and on population dynamics.

While progress needs to be made in the application of non-invasive measurement techniques, closed ecosystems have additional advantages. Persistence, besides allowing for the acquisition of long time-series, places a constraint on the community present, making it at once less arbitrary than species-assemblages present in other laboratory microbial ecosystems. Secondly, as seen throughout this chapter, a tight grip on initial and boundary conditions is helpful in learning about material cycling and species-species

interactions, and may one day teach us how the forcing by abiotic factors is transformed by ecosystems into selection pressures on organisms. Finally, closure is conceptually significant: any material exchange of organisms with their environment, whether it be so-called signaling molecules or simple metabolites, impacts the future of the ecosystem, and can contribute to their fitness. Although simplification of ecosystems has taught us a lot, ultimately we need to learn how to address the biological and chemical complexity of ecosystems head on.

Appendix 2A.1 Material cycling

The bewildering chemical diversity of ecosystems can to some extent be simplified by dividing the compounds each element is involved in into a few "compartments". For example, for carbon, C, three compartments can be distinguished: an inorganic compartment, containing CO_2 and (bi)carbonate, an organic, or reduced, compartment, containing C in sugars, fatty acids, etc.; and a compartment for forms of carbon which cannot be transformed within the ecosystem to anything else. Let N_R and N_O be the total amount of C in reduced and oxidized form, respectively:



The **turnover time** of a compartment *i* or *j*, τ_i, is set by the size of the compartment divided by the sum of the outward **fluxes** φ_{ij} (total carbon transformed from one compartment to another), i.e.

$$\tau_i = \frac{N_i}{\sum_j \varphi_{ij}}$$

The **recycling efficiency** of a compartment *i*, η_i, is given by the ratio of fluxes into compartments with outward links $\sum_{j \neq \text{sink}} \varphi_{ij}$ and the sum of outward fluxes to all compartments, $\sum_j \varphi_{ij}$, i.e. in our example, $\eta_R = \varphi_{out}/(\varphi_{out} + \varphi_{sink})$. If there is a "sink", a form of carbon produced, which cannot be transformed to another compartment, one

would expect the ecosystem to run out of reduced carbon (and hence biomass!) on a time scale of $\tau_R/(1- \eta_R)$. In other words, unless every compound can be recycled, or its production negatively regulated, available carbon will end up in dead-end products.

Chapter 3

Choice and Preliminary Characterization of a Model Closed Ecosystem

Abstract

In this chapter I describe the three species closed ecosystem used in this thesis for the development of experimental methods and time series analysis techniques. The system builds on work by Kawabata *et al*. [21] and consists of three single-celled microbes, an **a**lga, *Chlamydomonas reinhardtii*, a Gram-negative **b**acterium, *Escherichia coli*, and a **c**iliate, *Tetrahymena thermophila*. 'ABC' provides a convenient mnemonic. I demonstrate that these three species can persist in a chemically defined medium, examine the properties of its subsystems (e.g., A, AB, BC) and the effects of ecosystem size on persistence. Finally, I illustrate the phenotypic diversity observed in even such a simple ecosystem and discuss its consequences for our ability to make inferences from population density time series alone.

Quote

"It is good thus to try in our imagination to give any [life] form some advantage over another. Probably in no single instance should we know what to do, so as to succeed. It will convince us of our ignorance on the mutual relations of all organic beings; a conviction as necessary, as it seems difficult to acquire." (C. Darwin, The Origin of Species [12])

Introduction

As argued in Chapters 1 and 2, the study of interactions between organisms requires replicable time series of population dynamics over many generations. In this chapter I describe a simple three species closed ecosystem we have used in the development of methods to obtain such time series.

Various methods to construct closed ecosystems were discussed in Chapter 2. I have followed a bottom-up approach, combining a known set of species in a defined medium. This work builds on results by Kawabata *et al*. [1, 2], who developed a three species ecosystem containing *Escherichia coli*, *Euglena gracilis* and *Tetrahymena thermophila*, which proved to be persistent for at least 120 days in 10 mL of medium. They subsequently used this system, typically in volumes of 250-500 mL, to assess the effect of elevated CO_2 levels [77], ionizing radiation and a variety of toxic agents on ecosystem function [78, 79]. In addition, they examined the effects of these species on the competence (DNA uptake) of added *Bacillus subtilis* [80]. As explained below, we chose to replace *E. gracilis* by another alga, *Chlamydomonas reinhardtii*. Properties of all species mentioned are summarized in Table 3.1. Species are shown in Figure 3.1.

As a convenient mnemonic, either system can be abbreviated ABC, short for **A**lgae (*E. gracilis*, *C. reinhardtii*), **B**acteria (*E. coli*) and **C**iliates (*T. thermophila*).

First, I will describe some of the basic properties of the system. The medium used throughout this thesis is not fully chemically defined, but as shown, such a defined chemical medium does exist. Then the properties of subsystems, e.g., AB, containing *C. reinhardtii* and *E. coli* will be examined, and finally, the effect of size on persistence.

Finally, I will describe some phenotypic aspects of the three species. These descriptions are based on the literature and fragmentary observations, and by no means on exhaustive investigation. They merely serve to illustrate the complexity of our model ecosystem, and raise questions about how to interpret results from analysis of population density time series.

'ABC' also highlights how we wish to address this system, without preconceived ecological concepts such as production, consumption and decomposition, predator-prey relationships and limiting resources, however instructive. Abstraction does not imply that even this system is simple in any way. In any biological system only a few degrees of freedom can be observed with adequate temporal and numerical resolution. The goal is to determine what properties can be inferred with such limited knowledge in a system for which at least both initial and boundary conditions can be controlled and replicated.

Development of the system

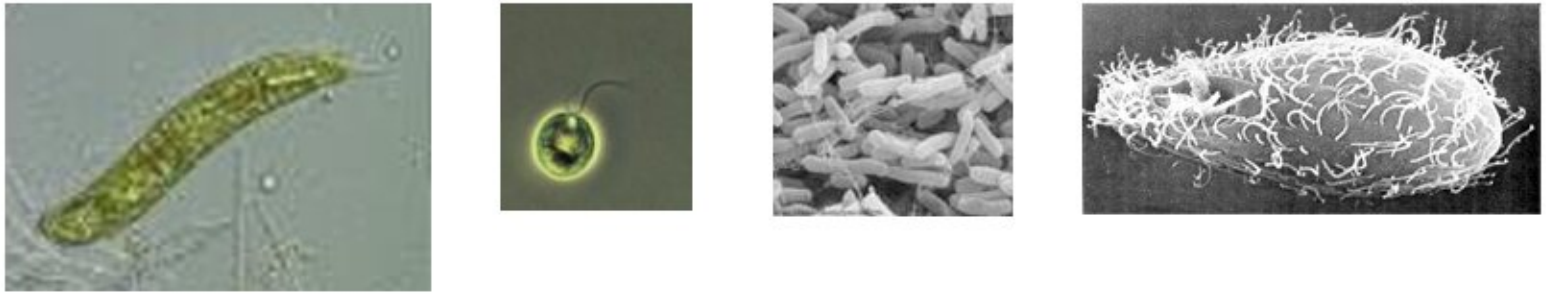
Comparison to the system developed by Kawabata et al.

Kawabata *et al*. [21] showed that *E. gracilis*, *E. coli* and *T. thermophila* can coexist in a medium called ½x Taub #36, with added proteose peptone (see below). Representative data from Kawabata *et al*. [21] are shown in Figure 3.2 for two different concentrations of proteose peptone.

Because few genetic techniques are available for *E. gracilis*, we decided to replace it with another photosynthetic eukaryote, *Chlamydomonas reinhardtii*, which is genetically tractable, can be grown in defined inorganic media [81] and has a faster doubling time (5-

Table 3.1: summary of properties of species discussed in this chapter. Nutritional requirements are those above and beyond a medium containing all bioelements except carbon in accessible form.

Species	**Description**	**Morphology**	**Nutritional requirements**	**Mating types**	**Fluorescence**
Euglena gracilis [82]	Eukaryotic alga, flagellated	Cylindrical, 35-55 x 10 μm	Vitamin B12, CO_2 or reduced C compound	1	Chlorophyll
Chlamydomonas reinhardtii [81]	Eukaryotic alga, flagellated	Round to ovate, 5-10 μm	CO_2 or acetic acid	2	Chlorophyll
Escherichia coli K12 [83]	Gram-negative bacterium, flagellated	Rod-shaped, 1 x 2 μm	A reduced C compound	1	dTomato
Tetrahymena thermophila [84]	Eukaryote, ciliate	Lemon or football-shaped, 30 x 50 μm	Various amino acids, vitamins, nucleosides.	7	H3.2-EYFP



Caption: the four species used in this chapter: (1) *Euglena gracilis*, (2) *Chlamydomonas reinhardtii*, (3) *Escherichia coli*, and (4) *Tetrahymena thermophila*. Objects are roughly to scale, except *E. coli* which is, in proportion, shown five times larger. Illustrations obtained from the internet.

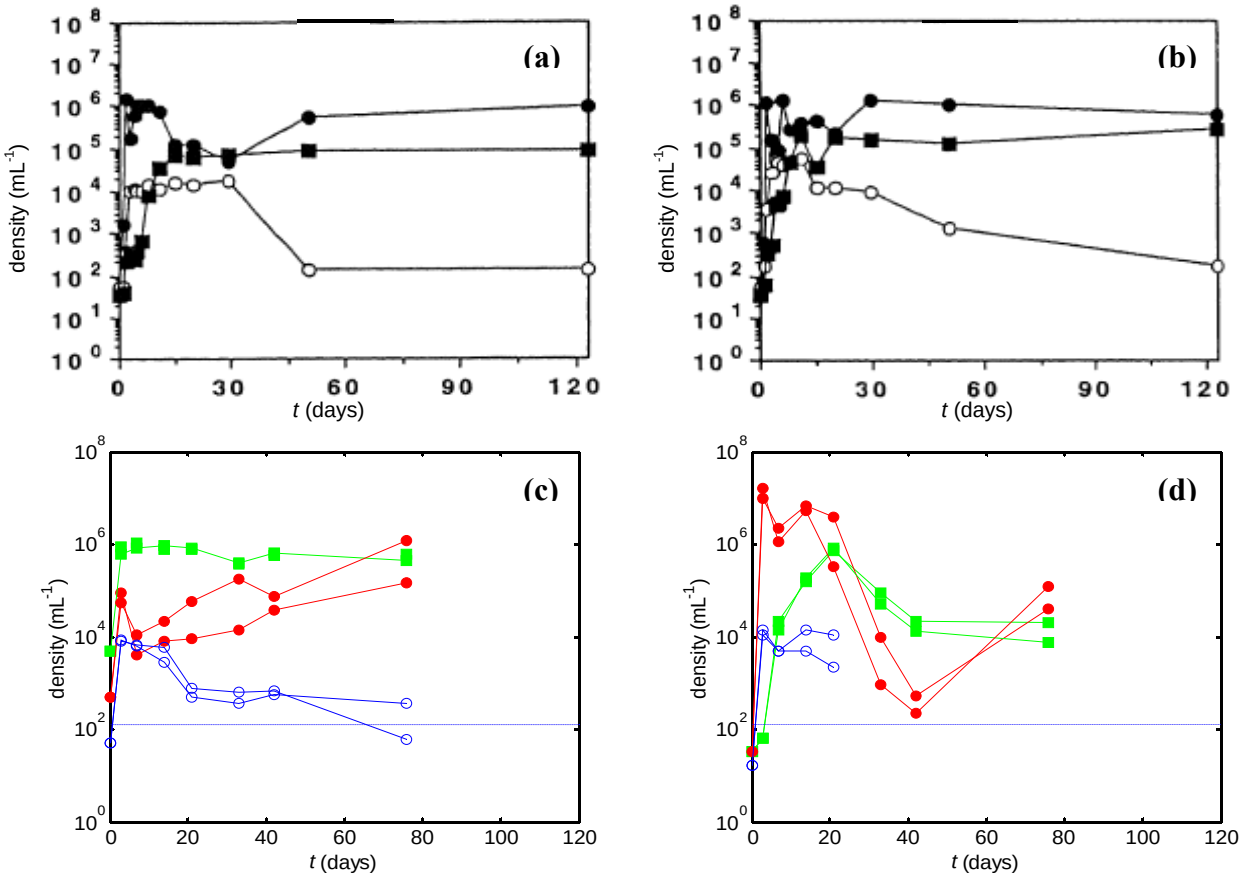


Figure 3.2: comparison of the three species ecosystem of Kawabata *et al*. (J. Protozool. Res., 1995, 5: 23-26) to the one used in this thesis. (a) *E. gracilis* (squares), *E. coli* (filled circles), and *T. thermophila* (open circles) in ½x Taub #36/0.01% proteose peptone (b) idem in 0.05% proteose peptone; (c, d) *C. reinhardtii* (green, squares), *E. coli* (red, filled circles), and *T. thermophila* (blue, open circles) in ½x Taub #36/0.03% proteose peptone starting from two different initial densities. The dotted line indicated the approximate detection limit for haemocytometry (125 mL^{-1}). *T. thermophila* density in (d) drops below this limit.
Note that (a), (b) were obtained by destructive sampling, so each time point is a new sample; (c) and (d) were obtained by repeated sampling of small aliquots (see Materials and Methods).

6 hours under optimal conditions, versus 10-11 h for *E. gracilis*) [81]. The properties of the species discussed in this chapter are summarized in Table 3.1.

The boundary conditions

In nature, temperature and light are dynamic variables and important aspects of a species' environment. Little is known about how fluctuations in these boundary conditions affect density fluctuations within ecosystems (see Chapter 6). Unless otherwise indicated, experiments described in this thesis were conducted at constant light intensity (1200 lux) and temperature (25 ^{o}C). Kawabata *et al*. [21] performed their experiments at 25 ^{o}C, but used a 12/12 hour light/dark cycle at 2500 lux (amounting to about the same time-averaged illumination).

The current system does allow for variation of light (Chapter 7) and temperature. *T. thermophila* [85] has a temperature range of about 10 to 40 ^{o}C, *E. coli* [83] of about 8 to 48 ^{o}C, and *C. reinhardtii* [81] of at least 15 to 35 ^{o}C. Likewise, *C. reinhardtii* is routinely cultured at light intensities up to 20 klux. For comparison, typical office lighting is about 200-400 lux, while bright sunlight can have an intensity of 100 klux. Exposures of 12 or 24 hours of about 7000 lux of fluorescent light with wavelengths ≥ 400 nm was shown to leave *E. coli* cells intact, but strongly inhibited its metabolism and growth [86]. Little other knowledge is available.

The chemistry

The default medium used throughout this thesis is ½x Taub #36 with 0.03% (w/v) proteose peptone no. 3 (described in Chapter 4), which is defined in all its components except proteose peptone[8]. For simplicity, I will call this Taub medium. The medium was originally developed by Taub and Dollar [88] to accommodate *Chlorella sp.* and *Daphnia*

[8] Proteose peptone is bacterial protein hydrolyzed to single amino acids and dipeptides, with additional undefined chemical constituents [87], see www.bd.com/ds/technicalCenter/typicalAnalysis/typ-proteose_peptone3.pdf.

sp. without physiological abnormalities and is supposed to resemble freshwater in its mineral composition[9]. It has since frequently been adapted for freshwater aquatic microcosms, including many of the pertinent closed ecosystem studies mentioned in Chapter 2 [8, 9, 38]. Results with this medium have been reliable and suitable for the development of methodology. In the medium, even in the absence of proteose peptone, a precipitate forms after about three months, which, given the composition of the medium, is likely a calcium phosphate salt.

Given the known nutritional requirements of the three species used (Table 3.1) it should be possible for the initial chemical conditions to be completely defined, with C and N introduced in pure compounds. Figure 3.3 demonstrates the existence of such a medium in which all three species persist for at least ten weeks. Its basis is ½x Taub #36 as used throughout this thesis, but without proteose peptone. Instead it contains 15 mM sodium acetate and 5 mM ammonium acetate. All three ions, sodium, ammonium and acetate, are well soluble in the presence of almost any other ions. The C:N:P ratio of this medium is about 1:8:200, that of the proteose peptone containing medium about 1:3.7:80 (the -fairly homeostatic composition of *E. coli* is about 1:4:45 [83]). In absolute terms the medium contains 40 mM C compared to about 17 mM C for proteose peptone. It is hence fairly rich, and technically outside the range of freshwater in total ion concentration.

Media were also tested with 7.5 mM sodium acetate/2.5 mM ammonium acetate, and 5 mM sodium acetate/1.25 mM ammonium acetate, but at these concentrations, the *E. coli* population crashed after 10 days (data not shown), presumably because the pH reached

[9] Freshwater is defined as having less than 50 mM dissolved salts (1 part per thousand). The medium mostly used for this thesis, ½x Taub #36 with 0.03% PP3 is about 11 mM total ion/molecule concentration (about 0.3 atm osmotic pressure).

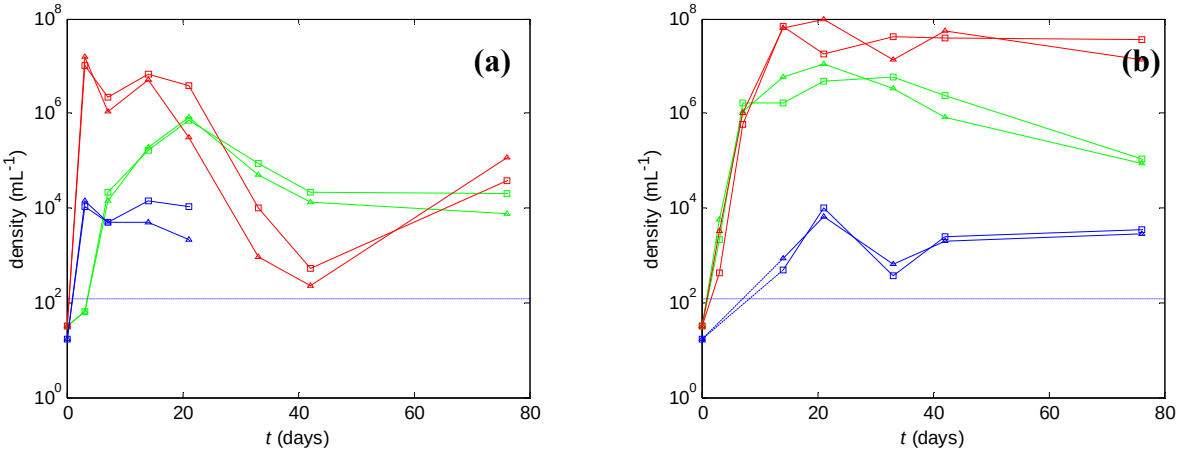


Figure 3.3 Test of a completely defined medium: (a) population dynamics for duplicate three species systems in ½x Taub #36/0.03% proteose peptone, and (b) dynamics in ½x Taub with 15 mM sodium acetate/5mM ammonium acetate as C and N sources.
C. reinhardtii (green), *E. coli* (red), and *T. thermophila* (blue). Duplicate experiments (squares and triangles, respectively) were performed in cuvettes with 3 mL of medium. Dotted horizontal lines indicate the approximate detection limit of haemocytometry. Note that low starting densities were chosen to minimize any medium contributions from the starting cultures.

high values (9 to 9.3). Pending development of an inorganic medium with a more stable pH, however, the remaining experiments in this thesis were performed with ½x Taub #36/0.03% proteose peptone.

Defining the initial chemical conditions is not only conceptually significant. It also opens the door for studies of material fluxes in the ecosystem using for example NMR and mass spectrometry. Equally important is the control of the overall redox state of the ecosystem it allows. As described in Chapter 2, Folsome and his colleagues showed that simple redox bookkeeping allows for powerful inferences about an ecosystem [42]. In particular, assuming the redox states of organisms to be relatively fixed, the redox state of the medium controls the point at which photosynthesis is necessary to form any further biomass. Nitrogen for example can be supplied as ammonium (N^{3-}) or nitrate (N^{5+}), and

carbon can likewise be supplied in a range of oxidation states. Carbon and nitrogen are relatively reduced in our example (acetate/ammonium), requiring little photosynthesis to form biomass. There are no data for our system on the efficiency or rate of material cycling.

Subculturing and cryopreservation

The ability to subculture and cryopreserve samples from an ecosystem are powerful tools. Subculturing allows for the creation of multiple daughter ecosystems and the study of the effects of adaptation. Cryopreservation allows for the isolation of individual clones and chemical analysis of the medium long after an experiment has been concluded.

For our three species ecosystem subculturing seemed to work fine. In the first experiment, described in Chapter 5, 300 μL samples from eleven 200-day-old ecosystems could be successfully subcultured into fresh medium, with all species present showing renewed growth (*E. coli* had gone extinct prior to opening in two of the eleven ecosystems).

Cryopreservation protocols are available for each of the three species [84, 89, 90], and in principle individuals of each species can be picked out of ecosystems, grown up in rich media and frozen down according to standard protocols. Ideally, however ecosystems samples could be frozen in their entirety. For a large three species ecosystem (see Materials and Methods) I explored three cryoprotectants (DMSO, glycerol and methanol) and two different freezing protocols (2 h or 24 h of gradual precooling at-1 °C/min). No protocol allowed recovery of *T. thermophila*. Only one protocol worked for *C. reinhardtii*, the one previously published [90], using 5% methanol as a protectant and a 2

h precooling step, giving at least 1% recovery. All other recoveries were < $1x10^{-4}$. Remarkably, more than 50% of *E. coli* was recovered in any protocol.

To preserve *T. thermophila*, I instead adopted a protocol which included an amplification step: subculture of ecosystems in fresh Taub medium without illumination gave a robust initial bloom of *E. coli* followed by a bloom of *T. thermophila* after 3-4 days, at which point the normal *T. thermophila* protocol was followed [84]. Recovery under this protocol still needs to be assessed.

Subsystems

A natural question about a persistent set of species is whether any of its subsets can also persist. A two-species system is a lot simpler than a three species system, being characterized by at most a single two-species interaction. The ability to study subsets of a larger community also gives us the opportunity to dissect species-species interactions. Matsui *et al*. [58] used subsystems of Kawabata's system [21] to differentiate between direct and indirect species interactions by comparing growth of species in each other's direct presence to growth in supplemented spent medium of the other species. The authors demonstrated that indirect interactions between *E. gracilis*, *E. coli* and *T. thermophila* all increased densities except for strong self-toxicity of spent *E. coli* medium.

To address these questions in our system, duplicate cultures of single species (A, B, C), two-species combinations (AB, AC, BC) and three species (ABC) were formed with the same initial densities (*C. reinhardtii*: 5000 mL^{-1}, *E. coli*: 500 mL^{-1}, *T. thermophila* 50

mL^{-1}) and the same medium, and kept under constant light (1200 lux) and temperature (25 ^{o}C). The results are shown in Figure 3.4.

It appears that *C. reinhardtii* does equally well in any combination, benefitting somewhat from the presence of *T. thermophila*, maybe because of more efficient recycling of debris. *T. thermophila* cannot persist on its own, showing a 'half-life' of about 3-4 days after the first two weeks. In the presence of *C. reinhardtii* or *E. coli*, or both, however, it survives for at least 65 days at densities of about 10^2 mL^{-1}.

E. coli, when grown by itself, decays very slowly in numbers ('half-life' ~11 days), after initial exponential growth and a short period of stasis, consistent with observations by Kolter *et al*. [91]. This persistence is quite remarkable, since *E. coli* does not appear to have any means to exploit the free energy of illumination available to algae. Its densities were, however, strongly suppressed by both *C. reinhardtii* (~100-fold reduction) and *T. thermophila* (~10-fold reduction). After initially being suppressed even stronger in the three species system, *E. coli* eventually appears to overcome this effect. Matsui, Kawabata *et al*. [58] observed qualitatively similar effects, except that *Euglena gracilis*, unlike *C. reinhardtii*, did not appear to have any effect on *E. coli* densities.

Effect of ecosystem size

Smaller ecosystems, at least under restricted immigration, tend to support fewer species than larger ones [92-95]. In smaller systems, the formation of new strains or species is less likely, while extinction due to fluctuations, forced by the environment or internal dynamics is much more likely [96]. However, population sizes in smaller systems are generally easier to measure and smaller systems can be replicated in larger numbers.

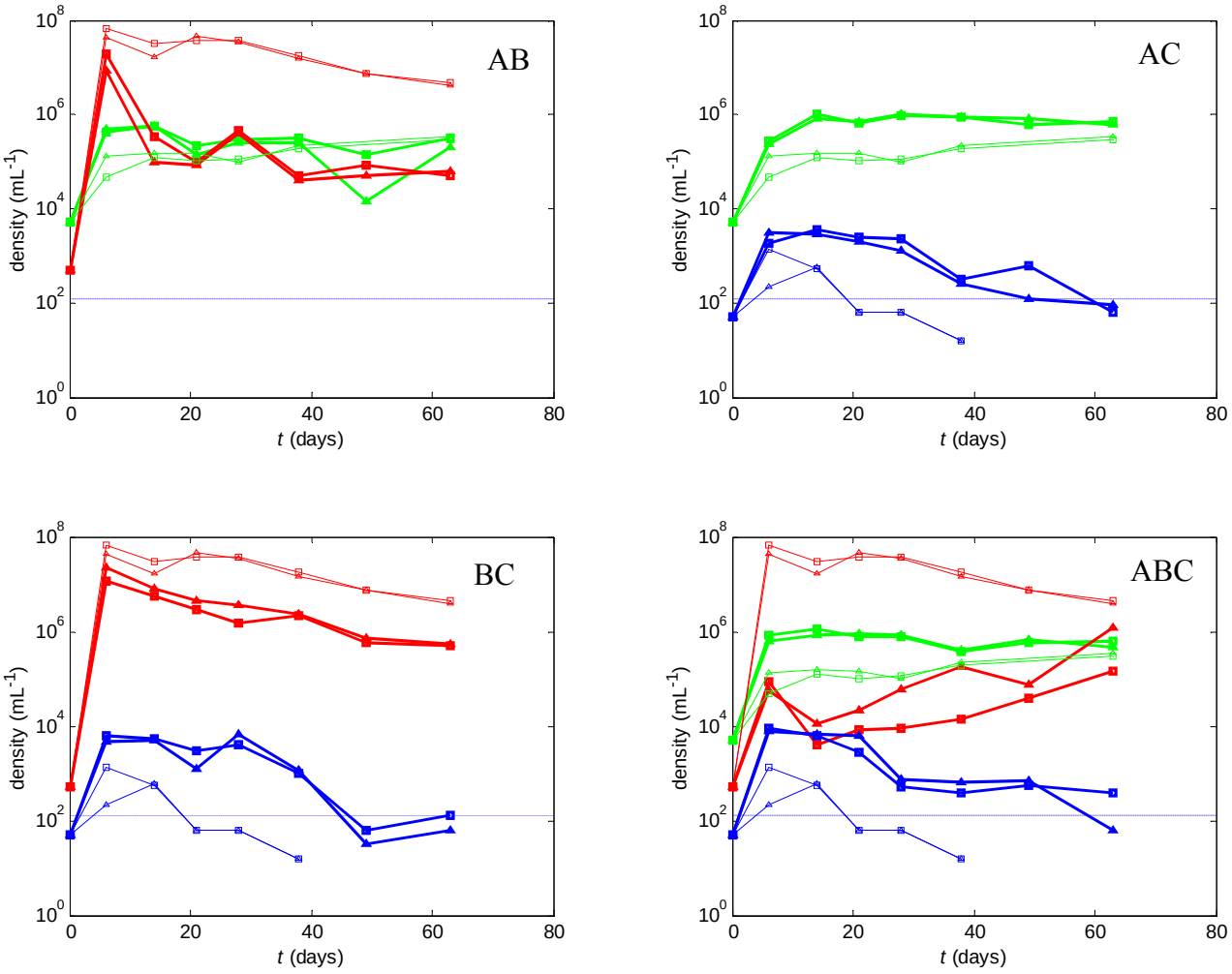


Figure 3.4 Growth of subsystems: two and three species communities (thick lines) compared to pure cultures (thin lines), all in ½x Taub#36 with 0.03% proteose peptone: A, green: *Chlamydomonas reinhardtii*, B, red: *E. coli*, and, C, blue: *Tetrahymena thermophila*. (AB) *C. reinhardtii* and *E. coli*, (AC) *C. reinhardtii* and *T. thermophila*, (BC) *E. coli* and *T. thermophila*, (ABC) all three species.
Duplicate experiments (squares and triangles, respectively) were performed in cuvettes with 3 mL of medium, and sampled manually as described. The dotted line indicates the approximate detection limit for haemocytometry.

For that reason, I sought to find out how far the system could be miniaturized. In a simple experiment, 300 µL wells of a 96-well plate were filled with each of the two (AB, AC, BC) and three species (ABC) ecosystems, with consistent initial densities in total liquid

volumes of 37.5 (n = 8 for each combination), 75 (n = 4), 150 (n = 4) and 300 μL (n = 4). The wells were then assayed for survival of each species about every 11 days over a 66 day period by microscopy.

For the AB subsystem, containing *C. reinhardtii* and *E. coli*, in the 37.5 μL liquid wells, *E. coli* went extinct in 7 out of 8 wells between day 53 and 66. In the eighth well, however, a brighter, larger *E. coli* phenotype became dominant after day 43 with a concomitant disappearance of motile *C. reinhardtii*. A similar displacement was observed in one 75 μL well. No other extinctions were observed.

For the AC subsystems, containing *C. reinhardtii* and *T. thermophila*, *T. thermophila* failed to establish in any of the 37.5 μL liquid wells. In such small volumes, under the initial densities chosen, only 3.8 *T. thermophila* would be present initially on average (chance of empty well at start ~2.2%). It appears initial growth of *C. reinhardtii* was too slow to sustain such small populations of *T. thermophila* (at the same initial density *T. thermophila* did establish itself in the presence of *E. coli*). Two more extinctions of *T. thermophila* were observed in wells with larger liquid volumes.

For the BC subsystem, containing *E. coli* and *T. thermophila*, three extinctions of *T. thermophila* were observed, two in 37.5 μL wells and one in a 150 μL well. It appears such small numbers don't allow for generalizations.

While in all two-species ecosystems (AB, AC, BC), extinction was rare, more extinctions were observed for the three species ABC system, for both *T. thermophila* and *E. coli*. The results are summarized in Figure 3.5. For both species, extinctions are common after two months in 37.5 and 75 μL liquid systems, while by the end of the experiment no extinctions had been observed for either species in 300 μL systems.

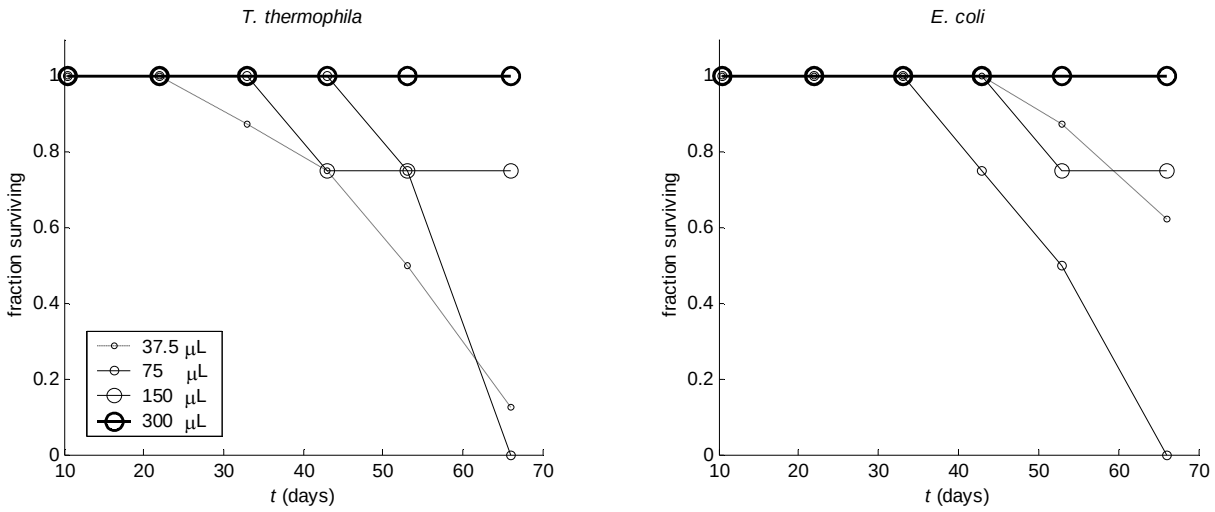


Figure 3.5: Persistence of *T. thermophila* (left panel) and *E. coli* (right panel) in three species (ABC) ecosystems of various sizes on a 96-well plate. The number of replicates is four for volumes of 75 μL and larger, and eight for the smallest volume, 37.5 μL.

In conclusion, in volumes ≥ 150 μL, the three species, or any two-species combination, could persist for at least two months, while extinctions were fairly common in smaller systems. As discussed below, there are significant initial phenotypic changes in these ecosystems. In addition to the intrinsically higher chance for smaller populations to go extinct, their smaller ability to generate fitter phenotypes (which scales linearly with population size) may contribute to their extinction.

Phenotypic diversity

The next chapters will be devoted to the measurement of population densities and the analysis of acquired time series data thereof. Here I want to illustrate the phenotypic complexity that appears to underlie these dynamics. I will start with some considerations

from the literature and then present observations made in the course of experiments described in this and the next chapter, species by species.

As Maharjan *et al*. [97] showed, even in the supposedly constant environment of a glucose-limited chemostat inoculated with a single strain of *E. coli* at a large stationary population size (10^{10} total) significant phenotypic diversity arose within four weeks (about 90 generations). Co-existing strains differed in cell-cell aggregation, *rpoS* genotypes (a σ-factor and global transcription factor active in stationary phase), glucose uptake rate and growth yield, and proportions of metabolites produced when grown on glucose. Another phenomenon in such glucose-limited chemostats inoculated with a single *E. coli* strain was demonstrated by Treves *et al*. [98]. After 100-400 generations in 6 out of 12 replicate chemostats acetate cross-feeding had evolved, one descendant strain specializing in scavenging acetate produced by the dominant other strain in its glucose catabolism.

Batch cultures tend to show even stronger diversification. Probably best-known is the work of Kolter, Finkel and others on changes occurring during prolonged starvation in *E. coli* cultures in rich medium. Mutations in *rpoS* [99] and *lrp* [100], another global transcription factor, were observed providing large fitness advantages in stationary phase. Finkel and Kolter [91] examined colony morphology of *E. coli* samples from batch cultures of 150 to 450 days old. They typically detected three to five morphotypes[10] per sample (with a detection limit around 1% of total population) with substantial dynamics over the observation period. This suggests that these aged cultures are both polymorphic and keep growing and dying (rather than becoming dormant). As a final example, *Pseudomonas fluorescens*, when introduced into an unperturbed batch culture (at $N = 10^9$

[10] Phenotype resulting in a distinct colony morphology.

to 10^{10}), rapidly diversifies, with one morphotype occupying the air-liquid interface, and two the bulk phase [101]. The morphotype occupying the interface did so by over-expressing a cellulosic polymer. In a subsequent publication, Rainey and Rainey showed the emergence of a defector phenotype which had a preference for growth at the interface but did not overexpress the polymer [102].

These examples illustrate that even in 'simple' environments micro-organisms rapidly diversify in visible (aggregation, spatial stratification, etc.) and invisible (altered gene expression and metabolism) ways. Such diversity is remarkable in the light of the principle of competitive exclusion: the statement that no two species or strains can occupy the same niche. This principle was first shown to apply by Gause in simple competition experiments between different yeasts and between paramecia [103], and later formulated by Hardin [104]. There are however many exceptions [105]. First, even simple systems can contain multiple niches, such as the air-liquid interface and bulk of a medium as seen for *Pseudomonas*. A more direct example of frequency-dependent fitness is given by so-called cheater phenotypes, e.g. for the defecting *Pseudomonas* phenotype on the air-water interface. Finally, there can be fitness trade-offs even within a niche so alternative solutions confer nearly identical fitness, as is likely the case for the four-week chemostat experiment of Maharjan and colleagues. Relative fitness differences Δf take $\sim\log(N)/\Delta f$ generations to be "resolved", which can easily be a hundred generations for the fitness differences determined by Maharjan *et al*. [97], comparable to the length of the experiment.

Now I want to describe some known and observed phenotypic properties and polymorphisms of the species used in the experiments described in this thesis.

Chlamydomonas reinhardtii

C. reinhardtii is a eukaryotic green alga with two mating types, designated mt+ and mt-. It can reproduce vegetatively in a haploid state, or undergo complete fusion with the opposite mating type and form a diploid cell. Such diploids normally undergo meiosis, but a small fraction undergoes mitosis, and such diploids can subsequently grow vegetatively [81].

As will be relevant in Chapter 7, where modulation of illumination is explored, *C. reinhardtii* responds in a variety of ways to light. It can display positive or negative phototaxis depending on its cellular state and illumination [81, 106]; it displays photo-kinesis [107], showing increased motility at stronger illumination; and both phototaxis [108] and the cell cycle [109] can be entrained by its circadian clock.

Cell morphology also appears to be variable. I noticed the occurrence of dyads and tetrads of divided cells still sharing the same cell wall. Unhatched tetrads, known as palmelloids in the literature [81], were observed on rare occasion. Commonly 10-20% (on occasion up to 50%) of cells was present as dyads, displaying reduced motility.

A final aspect of *C. reinhardtii* relevant to the experiments in this chapter is some difficulty in assessing the status of cells in light microscopy as dead or alive. While motile cells appeared obviously alive, immotile cells were scored as alive if they appeared structurally intact with respect to cell wall and eyespot. Dead *C. reinhardtii* appears to lose its internal structure and color within one or two weeks.

C. reinhardtii is also capable of surface adhesion and motility, using glycoproteins on its flagella [81]. It could frequently be seen getting stuck to haemocytometer slides, and in

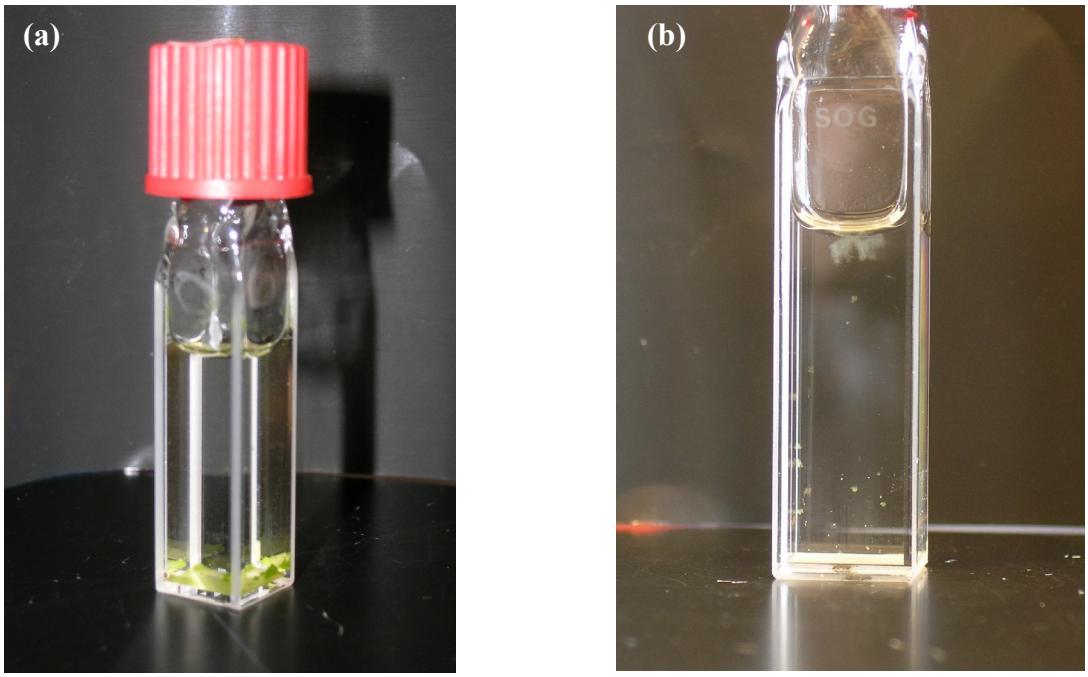


Figure 3.6: two cuvettes from experiment 1 in Chapter 5 after 200 days. (a) much debris and alive material is observed on the bottom; note the green peaks of *C. reinhardtii*. (b) the most extensive interface structure observed so far, extending about 2-3 mm into the liquid; also note specks on the glass walls ≤ 1 mm^2 in area. These patches contain *C. reinhardtii* and *E. coli* growing interspersed on the wall.

some ecosystems, mixed patches of *E. coli* and *C. reinhardtii* were visible on the walls (Figure 3.6).

Escherichia coli

The most conspicuous phenotypic feature of *E. coli* when examined by microscope in BC and ABC systems was its aggregation. The strains used are MG1655 Δ*flu* Δ*fimA hsdR*[11], with either the low-copy number plasmid pZS* 3R dTomato or that same plasmid integrated into the HK022 phage attachment site (for construction details see the

[11] *HsdR* is a mutation in a restriction enzyme, cotransducing at about 30% frequency with *fimA* (distance 41 kb). Carriers of the mutation can be identified by their susceptibility to λ-phage previously grown on an *hsdM*⁻ host (B. Wanner, pers. comm. and DH, data not shown) and it was retained as a convenient marker.

Materials and Methods section of Chapter 4). *Flu* codes for a cell-cell adhesion protein, Ag43, and *fimA* for the main structural unit of fimbriae, a kind of type I pili. Knockouts of fimbrial genes and ag43 do not significantly flocculate[12] in solution [110], and Ag43 knockouts show decreased biofilm formation [111]. While these mutations decreased flocculation, they did not prevent aggregation on surfaces in BC and ABC communities (see Figure 3.7).

Three types of aggregates formed. In the BC communities of the 96-well experiment described above, initially frequent filamentation was observed. Filamentation has been previously observed as a way to escape predation by protozoa [112, 113]. Filamentous cells were eventually displaced in most wells, but one well retained a massive filamentous network for the rest of the experiment, with many cells up to 100 μm in length (Figure 3.7). Since filamentation appears to be caused by septation defects [114], it is natural to expect a competitive disadvantage relative to other aggregation modes. In most other wells with BC communities, initially fairly loose patches of cells dominated, at least on the bottom. These were over a few weeks displaced by either dense microcolonies (with no individual cells distinguishable under fluorescence microscopy), or by elaborate three-dimensional biofilms containing cells spaced quite sparsely in a rigid structure. A multitude of surface structures, such as flagella, curli, and fimbriae have been implicated in biofilm formation in *E. coli*, as well as a host of biopolymers, such as DNA, cellulose, and colanic acid and 'PIA-like polymer', two complex modified polysaccharides [115], each of which may be dispensable by itself [116].

In AB communities, containing *E. coli* and *C. reinhardtii*, no microcolony formation and very little biofilm formation was observed. This suggests there are metabolic

[12] Flocculation: formation of loose aggregates in liquid.

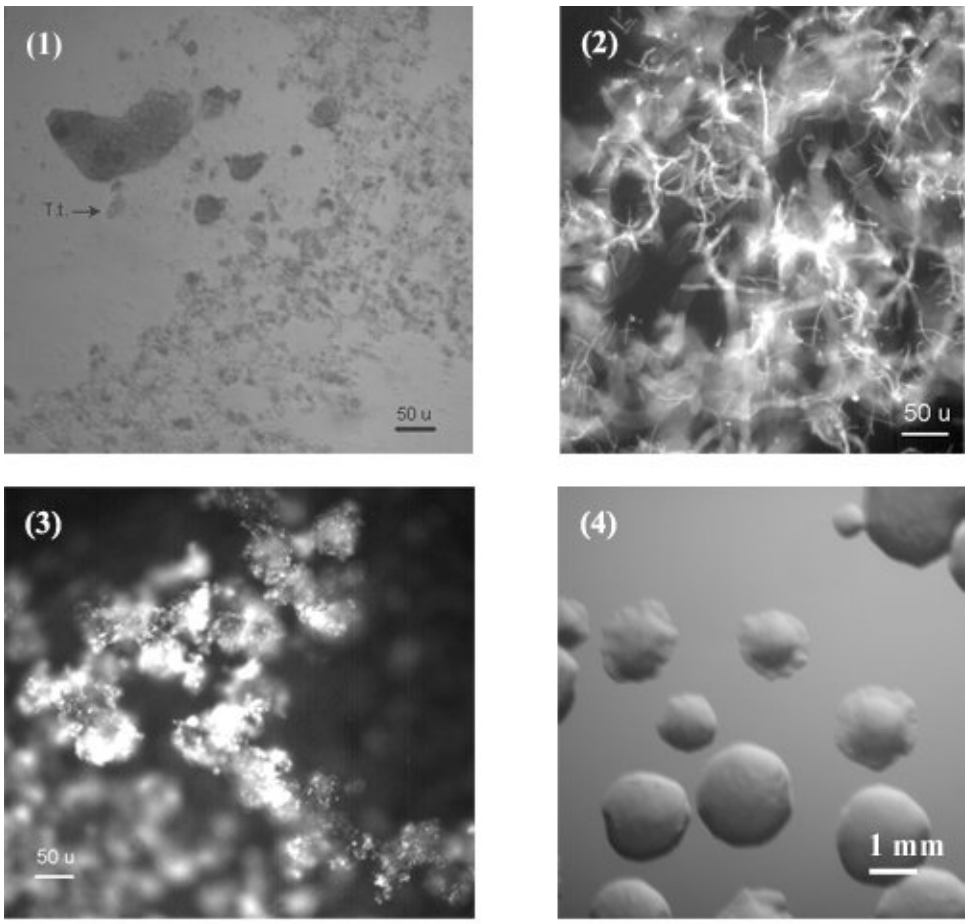


Figure 3.7: Phenotypes encountered for *E. coli*: (1) microcolonies (top) and a fairly loose network of cells (bottom and right) in an ecosystem with *T. thermophila*, one of which is indicated in the figure. (2) a network of filamentous *E. coli* cells in another ecosystem with *T. thermophila*, (3) elaborate three-dimensional *E. coli* matrix extending into the medium, with individual cells visible (slice taken well above bottom). (4) colony polymorphism on YPD agar plates of *E. coli* isolated from an eight month old three species ecosystem.
Scale bars in (1-3) are 50 μm (after calibration using a haemocytometer slide). The scale bar in (4) is approximate. (1) brightfield microscopy, (2), (3) fluorescence microscopy and (4) semi-darkfield conditions on a stereomicroscope.

disadvantages to living in an aggregate, which in the case of BC and ABC systems are outweighed by the benefits of decreased predation. It also suggests an explanation for why initially (Figure 3.4) *E. coli* densities are suppressed more (~10^{-4}x) in ABC

communities than the product of effects in AC (~10^{-2}x) and BC (~10^{-1}x) communities. Remarkably, in the ABC community *E. coli* eventually reaches much higher densities.

Tetrahymena thermophila

The most striking feature of *T. thermophila* is its complicated genetics. It contains a micronucleus, known as the mic, and a macronucleus, the mac. The mic is a diploid nucleus which is transcriptionally inactive. The mac is a polyploid nucleus (N ~ 45 for most DNA [117]), which is transcriptionally active. After meiosis and conjugation (the process of mating in which two cells of different mating types exchange haploid nuclei, without further cytoplasmic fusion) a new diploid mic is formed which gives rise to a new mac, which undergoes extensive genome editing (for a detailed description, see [84]). During vegetative growth, the mic is duplicated by mitosis, while division of the mac is not well understood. Chromosomes in the mac lack a centromere, and while the overall ploidy appears to be preserved, no distinction is made between sister chromosomes and homologous chromosomes, so alleles segregate at random [117].

These genetic properties have important consequences. In particular, unless strains are inbred extensively, the concept of isogenic strains is close to meaningless in ciliates. Sister cells with identical germ line (mic) are likely to differ in their phenotypic details due to stochastic differences in mac composition. Equally important are the effects on mating type. Mating type alleles, while apparently not completely determining mating type [118], are subject to segregation and cells are immature (incapable of mating) for about sixty generations, until the alleles have undergone sufficient segregation. In one of the ecosystems described in Chapter 5, several conjugating pairs were observed at the

conclusion of 200 days of closure, suggesting the initial transformant matured into at least two different mating types, although 'selfing', that is conjugation between cells of the same mating type, is also known to occur on occasion [84].

Two more features should be mentioned. First, *T. thermophila* is nearly always motile. It is known however to have a special 'rapid-swimmer' phenotype [119] in which a large flagellum forms and it can reach speeds up to five times its normal speed (100-200 μm/s, depending on temperature [120, 121]). *T. thermophila* is also known to form patterns in liquid cultures, a process called bioconvection [122][13].

Also, *T. thermophila* can display a significant size polymorphism. This deserves mention because of its apparent implications for our three species system. Large *T. thermophila* cells are capable of eating the algae, *C. reinhardtii*, which at about 5-10 μm diameter is too large for normal-size *T. thermophila* (see Figure 3.8). Significant dependence of size on prey type has been observed for another ciliate, *Blepharisma americanum*, which can double in both length and width when feeding on larger prey [123]. In communities in which the *T. thermophila* size polymorphism is observed, the size distribution appears to be bimodal (Figure 3.8), but this has not been quantified.

Debris

Most life and death in these cuvettes appears to take place on the bottom (Figure 3.6 (a)). When examining the material on the bottom, it appears to consist mostly of algal debris (Figure 3.9), with interspersed alive *C. reinhardtii*, *E. coli*, in various forms, and foraging *T. thermophila*. Especially *C. reinhardtii* cell wall material appears to decay slowly (~1 month), with many empty parts of shells visible, after the other components of the cell

[13] Bioconvection appears to be driven by oxygen gradients and/or gravitaxis, not thermal gradients.

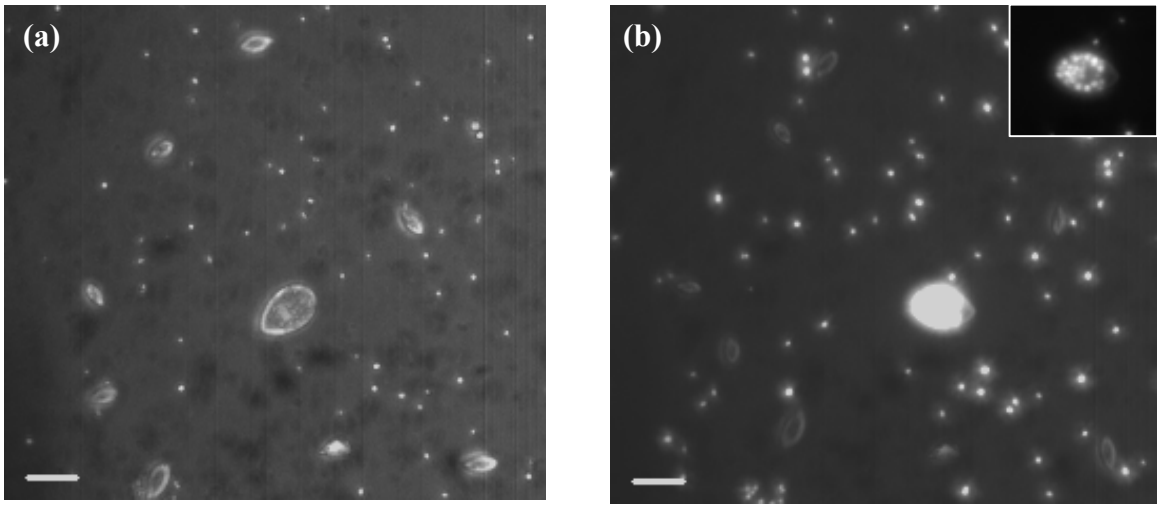


Figure 3.8: Size polymorphism in *T. thermophila*. (a) phase microscopy (b) combined phase and fluorescence image. Images were acquired about 1 second apart, in which the large individual in the center turned its oral region from left/down to facing right. Fluorescence indicates that this individual is filled with chlorophyll. Insert: same individual five times dimmer, fluorescence localizes to endosomes everywhere except the nuclear and oral regions. The other bright objects in (b) are (free) *C. reinhardtii*. Sample from a six-week-old three species ecosystem.
The large individual measures 64 by 40 μm, the other individuals 42 (± 5) by 21 (± 2) μm (standard deviations over 11 cells). The scale bars are 50 μm.

have long disintegrated. This can introduce a slow time scale especially in carbon metabolism (since cell walls consist mainly of polysaccharides).

Another consequence is the introduction of significant spatial structure, potentially leading to new niches. Debris, more pragmatically, also significantly lowers contrast when imaging the bottom on an inverse microscope.

Discussion

What are the consequences of the phenotypic complexity just described? The challenge for the analysis of population density time series is twofold: the observed populations are

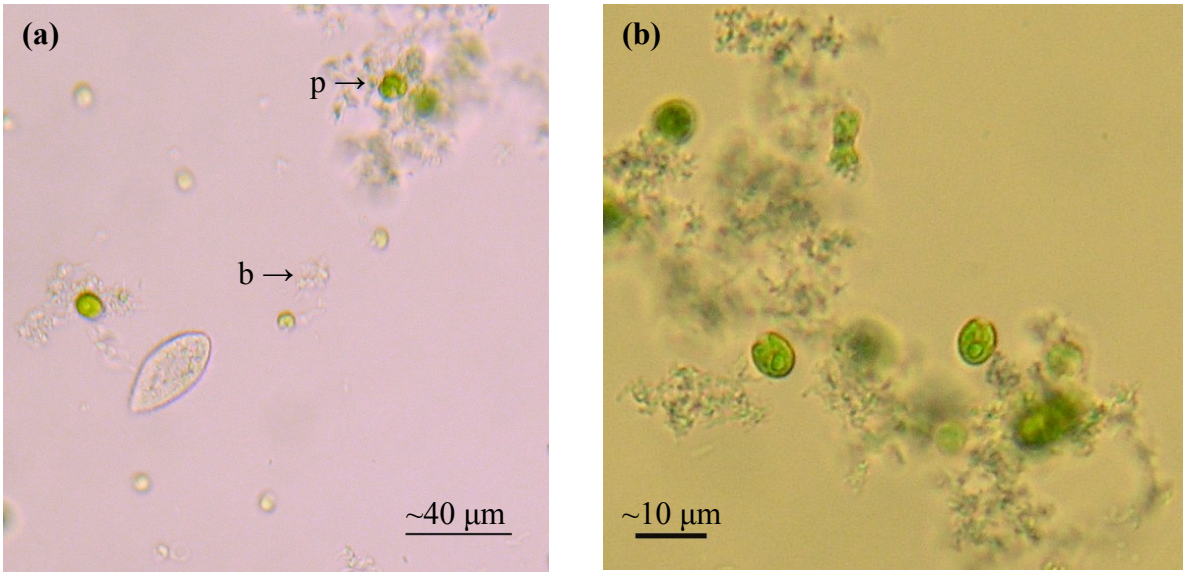


Figure 3.9: Life at the bottom: after most liquid had been pipetted off a 33 day old ABC ecosystem, the remaining liquid was gently mixed by pipetting up and down, placed on a depression slide, and imaged on a Nikon Optiphot microscope with an Olympus C-5060 camera. (a) small groups of bacteria (marked 'b') and a palmelloid cell (see text); (b) close up of algal debris containing dead *C. reinhardtii* in various states of decay, with two alive *C. reinhardtii* in focus. Scale bars are approximate.

themselves heterogeneous and subject to change, and significant populations on surfaces and interfaces are not observed in population density measurements.

Whether population heterogeneity is a problem, depends on the rate at which each phenotype is generated. Do populations get trapped into different phenotypic trajectories, or can, when conditions change, new phenotypes rapidly be generated? Observations for the *E. coli*/*T. thermophila* system in the 96-well plate experiment on system size (see also Figure 3.7 a-c), suggest that initial phenotypic divergence can indeed have long lasting consequences. I will return to this question in Chapter 5.

Regarding the second problem, coverage of as little as 1% of the immersed glass surface of a cuvette by a monolayer of *E. coli* can rival the total bulk population in numbers. I don't know how much exchange there is between surface and bulk populations. Do they have the same dynamics or are they effectively different 'species'? Surface aggregation can be achieved in several ways. Are we then to knock out every single pathway that doesn't fit the model? Should we study systems so large that surfaces become irrelevant? These questions will be considered in the final chapter.

To conclude, an example which illustrates that microbial interactions may be much more complicated than can be imagined in any mechanistic model. Teplitski *et al*. [124] demonstrated that spent *C. reinhardtii* culture medium contains ethyl acetate soluble (i.e., apolar) components which mimic the action of acetylated homoserine lactones (AHL), so-called quorum sensing molecules produced by a variety of Gram-negative bacteria [125]. *E. coli* does not have a full AHL-based quorum sensing system, but, in something of a puzzle, does possess a transcription factor, sdiA, which is homologous to luxR receptors and has been shown to bind a variety of AHLs [126].

Appendix 3A.1 Materials and Methods

Inorganic medium (Figure 3.3)

Inorganic media were designed based on considerations outlined in the text. The recipe follows that of ½x Taub #36 with 0.03% proteose peptone (see Chapter 4), except that appropriate amounts of 0.30 M sodium acetate and 0.10 M ammonium acetate were substituted for proteose peptone.

The experiments were performed with *E. coli* strain MG1655 Δ*flu* Δ*fimA hsdR* HK022::(3R dTomato), *C. reinhardtii* strain UTEX 2244 (mt+, UTEX strain center), and *T. thermophila* H3.2-EYFP. For details on strain construction, please see Chapter 4.

Glassware preparation and purity testing of starting cultures and final cultures were performed as described in Chapter 4. Starting cultures were grown to early stationary phase (*C. reinhardtii* in TAP at 2.0×10^6 mL^{-1}, *T. thermophila* in SPP at 1.0×10^5 mL^{-1}, and *E. coli* in ½x Taub #36 with 0.03% proteose peptone at 8×10^7 mL^{-1}) and densities determined using a Coulter Counter for *C. reinhardtii* and *T. thermophila*, and an OD_{400} calibration curve for *E. coli*. Starting densities were chosen deliberately low at 100 mL^{-1} for *C. reinhardtii* and *E. coli*, and 50 mL^{-1} for *T. thermophila*. Cultures were washed twice in ½x Taub #36 with either 0.03% proteose peptone or 5 mM acetate/1.25 mM ammonium. Clear Fisherbrand screw thread vials with a nominal volume of 4 mL (total internal volume: 4.9 mL) were used with screw caps with a silicon septum lined with teflon on the inside. Duplicate vials were filled with 3 mL of ½x Taub #36 with 0.03% proteose peptone, or, with, respectively 1.25 mM/3.75 mM; 2.5 mM/7.5 mM or 5 mM/15 mM ammonium acetate/sodium acetate, and the three species added at the final densities

mentioned above. Wide-bore pipette tips were used for *T. thermophila* to minimize any cell damage.

All sampling was performed in a laminar flow hood. At each time point, sample septa were cleaned with a cotton tip drenched in 70% isopropanol, left to dry for 5-10 minutes, inverted gently about five times, rapidly passed through a flame, and samples (~100 µL) taken with disposable syringes (gauge 25 or 26) and stored in Eppendorf tubes. *T. thermophila* and *C. reinhardtii* counts were determined in a haemocytometer slide (100 µm deep). To determine *E.* coli densities, the remainder of the sample was vortexed vigorously and diluted appropriately. Vigorous vortexing was repeated after each dilution step. Dilutions were plated on LB plates and counted after a few days.

Subsystems

The experiments were performed with *E. coli* strain MG1655 Δ*flu* Δ*fimA hsdR* HK022::(3R dTomato), *C. reinhardtii* strain UTEX 2244 (mt+), and *T. thermophila* H3.2-EYFP as for the *Inorganic medium* experiment.

Starting cultures were grown to early stationary phase (*C. reinhardtii* in TAP at 9.6×10^6 mL^{-1}, *T. thermophila* in SPP at 2.5×10^4 mL^{-1}, late exponential phase, and *E. coli* in ½x Taub #36 with 0.03% proteose peptone at 3×10^7 mL^{-1}) and densities determined using a haemocytometer for *C. reinhardtii* and *T. thermophila*, and an OD_{400} calibration curve for *E. coli*. Purity tests and glassware preparation were performed as above, and the same vials were used. Starting cultures were washed twice in 1x Taub #36 without added C and N source, and diluted appropriately. Final starting densities were for *C. reinhardtii*: 5000 mL^{-1}, *E. coli* 500 mL^{-1}, and *T. thermophila* 50 mL^{-1}. *T. thermophila* was again handled

with wide-bore 200 μL pipette tips. Sampling was performed as for the *Inorganic medium* experiment.

Size dependence

This experiment was prepared in parallel with the previous experiment, except for *E. coli* strain MG1655 Δ*flu* Δ*fimA hsdR* pZS* 3R dTomato (culture density 2.0×10^8 mL^{-1}) was used, because it is about 5x brighter (flow cytometry observations, data not shown) than the strain with the construct integrated into the chromosome used for the experiments above, a concession necessary for imaging. The plasmid has a loss rate of about 10^{-5} per division (DH, data not shown); see [127] for the basic plasmid design and [128] for the SC101* origin of replication. Initial densities were for *C. reinhardtii* and *E. coli* 500 mL^{-1}, and for *T. thermophila* 100 mL^{-1}. For each species combination (AB, AC, BC, ABC) 4 wells each had final liquid volumes of 300, 150 and 75 μL, and eight a final volume of 37.5 μL.

To insure identical preparation of wells, an eight-channel multipipet was used, and all wells containing a specific species were pipetted from the same solution basin successively. Plates used were clear polypropylene plates from Greiner (#655201) and were autoclaved. Wells were sealed using a silicone Greiner cap (#381070), which was UV-sterilized for 30 minutes. No significant loss of liquid was observed after nine months.

Microscopy was performed on a Leica DM IRBE microscope, using YFP and Texas red filter sets. While both dTomato and chlorophyll fluorescence are transmitted by the latter's emission filter, they were, by eye, clearly distinguishable by color, dTomato fluorescence emission being orange and chlorophyll deep red.

Cryopreservation assay

In a 500 mL Erlenmeyer, 10 µL of an overnight culture of MG1655 Δ*flu* Δ*fimA hsdR* HK022::(3R dTomato) in ½x Taub #36/0.05% proteose peptone, 10 µL of stationary phase culture of *C. reinhardtii* strain UTEX 2244 (mt+) in TAP (Chapter 4), and 100 µL of a 10-day old *T. thermophila* H3.2-EYFP culture in SPP [84], washed once in final medium, were added to 25 mL ½x Taub #36/0.03% proteose peptone. The Erlenmeyer was covered by aluminum foil (not sealed) and placed under fluorescent lighting in a 12/12h light/dark cycle.

After three days 0.4 mL samples from this system were frozen according to 4x2 protocols: using no cryoprotectant, 8% final conc. DMSO, 5% final conc. methanol or 5% final conc. glycerol; and either 2 h or 24 h gradual pre-cooling in a Cryo 1°C "Mr. Frosty" freezing container (Nalgene). Initial and final densities of viable cells were determined by serial dilutions on a 96-well plate followed by incubation in a humidity chamber under fluorescent lighting for a few days. After incubation, wells were scored for presence/absence of *C. reinhardtii* and *T. thermophila* under a dissecting microscope. *E. coli* viability was scored by plating on LB agar plates (1.5%).

Chapter 4
Non-invasive Measurement of Population Dynamics in a Closed Ecosystem

Abstract

In this chapter I describe the development of a method for the non-invasive measurement of population densities in a closed ecosystem consisting of *Chlamydomonas reinhardtii*, *Escherichia coli* and *Tetrahymena thermophila*. The method adapts selective plane illumination fluorescence microscopy to obtain counts of these species in a small observation volume. This method offers a large dynamical range, low classification error and reliable operation over several months.

Introduction

The size of a population has important effects on its future, as well as on the future of other populations it interacts with. For example, the risk of extinction and the rate of generation of new phenotypes depend on size. The growth of populations depends on their own and other species' densities, although such dependence has remained hard to demonstrate. This difficulty has been attributed to the variety of mechanisms, and hence time and length scales, by which such density dependent regulation can take shape. Also, abiotic factors can keep population densities too low for density dependence of growth rates to be significant over many generations [129, 130].

Small closed ecosystems, as argued in Chapter 2, provide the opportunity to acquire long time series of population dynamics under tightly controlled initial conditions (biological and chemical) and boundary conditions (light and temperature), and hence to address the effects of populations on each other and themselves. In principle, any interactions will be included, regardless of mechanism or time scale, and complicating external parameters can be controlled.

I set out to develop a method for the non-invasive, automated measurement of population densities in a model closed ecosystem over several months. Such a method requires (1) a large dynamical range, (2) low classification error between different species, (3) minimal perturbation of the dynamics observed by the measurements, and (4) robust operation over long periods of time.

A host of methods has already been developed to estimate population densities in mixed cultures. Many methods, such as plating on selective media, haemocytometry, flow cytometry, and quantitative PCR rely on invasive sampling and are labor intensive (a

notable exception is the development of automated inline flow cytometry [131]). In addition, any invasive sampling is accompanied by some risk of contamination.

Recently, progress has been made in monitoring population densities non-invasively. For example, when species or strains are naturally pigmented, as are many algae, and the ciliate *Blepharisma*, or can be labeled with a fluorescent protein, bulk fluorescence can be measured and related to population density [69, 70]. Alternatively, attention can be focused on a small observation volume, as is commonly done in fluorescence correlation spectroscopy and light scattering (explored in preliminary experiments).

Direct imaging of individual organisms in a small observation volume is, however, more easily interpreted. Balagaddé *et al*. [132], developed a microfluidic microchemostat in which population density was determined from counts in images obtained by light microscopy. Optical sampling techniques, in which images are taken in observation volumes small compared to the total system volume, have also found application in oceanography [71] and monitoring of fermentation processes in biotechnology [72].

Our model closed ecosystem consists of three species, the green alga *Chlamydomonas reinhardtii*, *Escherichia coli*, and the heterotrophic ciliate *Tetrahymena thermophila*. It is based on work by Kawabata *et al*. [21] who showed a similar system, containing *E. coli* and *T. thermophila* together with another green alga, *Euglena gracilis* could coexist for at least 120 days. I substituted *E. gracilis* with *C. reinhardtii* because of its simpler nutritional requirements, faster growth, and better genetic tractability.

The basis for our method is a technique called thin light sheet microscopy (TLSM) or selective plane illumination microscopy (SPIM), in which a plane of illumination coincides with the object plane of the imaging lens. For fluorescence microscopy this

implies that only fluorophores in the object plane are excited and imaged. Selective plane illumination has a history of several decades in fluid dynamics (see e.g. [133]) and was introduced into oceanography by Fuchs, Jaffe *et al*. [134] and developmental biology by Huisken *et al*. [22].

Technical development

A method was developed to measure population density time series for a three-species closed ecosystem consisting of *C. reinhardtii*, *E. coli* and *T. thermophila* in an aquatic environment with minimal perturbation. These species normally occur as single cells, allowing for straightforward counting of individuals in an observation volume.

The apparatus is shown in Figure 4.1. At the heart of the SPIM method lies the coincidence of the illumination plane and the focal plane of the imaging lens [135], which implies orthogonality of the illumination and observation axes. The illumination beam is initially expanded, its central part selected using an iris or slit, and the resulting, fairly uniform beam focused in one dimension by a cylindrical lens, effectively forming a sheet of light over the field of view of both objectives. This illumination sheet excites the fluorophores that are imaged in each emission channel. Important aspects of the setup will be motivated here; a full technical description can be found in Appendix 4A.1.

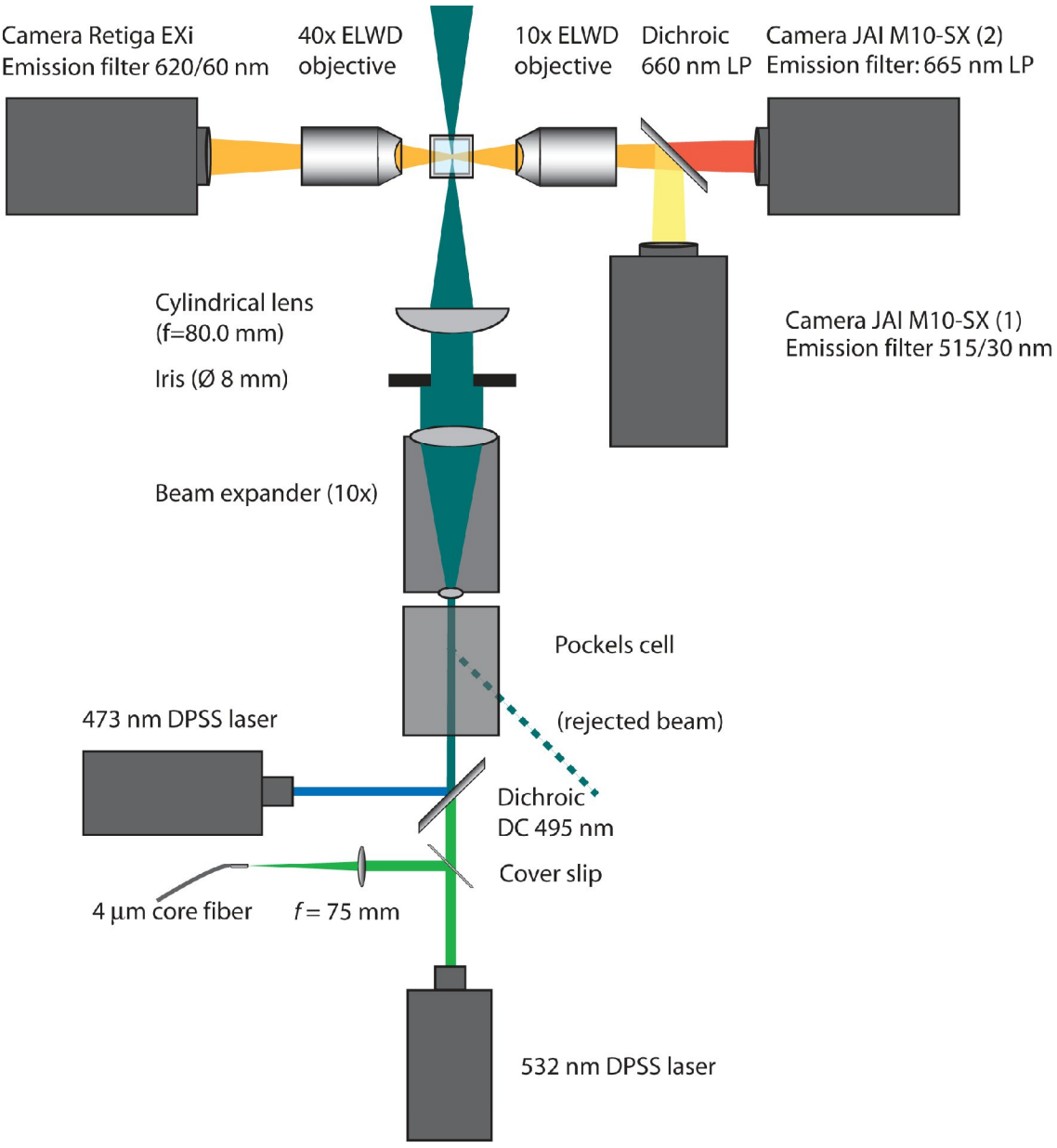


Figure 4.1: Setup for in situ fluorescence microscopy of a three-species closed ecosystem. Top view. The excitation axis runs vertically (blue, green), the observation axis horizontally (emission shown in yellow, orange, red). The objectives can be micropositioned along the observation axis. The cameras can be translated in X,Y and Z.

Strains

C. reinhardtii, like all green algae, contains abundant chlorophyll which is naturally fluorescent with absorption bands from about 440-490 nm and 660-690 nm and emits in a sharp band around 689 nm. Besides chlorophyll it possesses a number of auxiliary pigments [136], which cause it to have weak fluorescence over most of the visible spectrum. The construction of fluorescent strains of *E. coli* and *T. thermophila* is detailed in the Materials and Methods section. Briefly, an *E. coli* MG1655 derivative was constructed with genes for fimA (structural protein for fimbriae) and ag43, a cell-cell aggregation protein [110, 111, 137] knocked out. Such strains have been reported to show markedly reduced flocculation in liquid cultures. In addition, dTomato, a red-fluorescent protein ([138], excitation and emission maxima at 554 and 581 nm, respectively) was integrated into the chromosome under the control of the bacteriophage λ R promoter.

For *T. thermophila* a yellow-fluorescent strain (EYFP, excitation and emission maxima at 514 and 527 nm, respectively) was constructed by replacing the native open reading frame for histone 3.2 by the open reading frame for a histone 3.2-EYFP fusion protein[14]. Expression of most fluorescent proteins in *T. thermophila* has been difficult, most likely because of its distinct codon usage compared to that of organisms like the mouse and *E. coli*[15,16]. *T. thermophila* histones appear to have highly optimized codon usage, with only the most common codons for each amino acid used, suggesting that efficient expression is important. Gerami-Nejad *et al*. [143] optimized several fluorescent protein DNA sequences for expression in *Candida albicans* which has codon usage very similar to *T.*

[14] GFP = green fluorescent protein; EYFP = enhanced yellow fluorescent protein.
[15] See http://www.kazusa.or.jp/codon/ for extensive codon usage data.
[16] However, good expression has been obtained for the original GFP construct isolated from *Aequorea victoria* (Y.F. Liu, pers. comm.), Enhanced GFP has also been successfully expressed in *T. thermophila*. [139] [140] [141] [142]

thermophila. Using their EYFP sequence indeed cells far brighter than with ones expressing the original GFP were obtained.

Illumination

After initial experiments with an argon laser, diode-pumped solid state lasers were chosen because of their longer life-times and lower cost. Specifically, used were a 20 mW 473 nm DPSS laser (Lasever, Inc.) to excite chlorophyll and EYFP, and a 50 mW 532 nm DPSS laser (AOTK, Inc.) to excite dTomato and, to some degree, chlorophyll. These lasers can be operated continuously for months, but have one significant shortcoming. Limited thermal management can lead to pointing instability[17]. The nominal pointing stability was 50 μrad (r.m.s.[18]) for the 532 nm laser, and 20 μrad for the 473 nm laser. The corresponding transverse displacement of the light sheet is determined by the focal length of the cylindrical lens (here 80 mm), resulting in nominal r.m.s. displacements of 4 μm and 1.6 μm respectively. To monitor these fluctuations, some of the beam was deflected using a cover slip, and focused the light onto a 4 μm core fiber (FSSN-3224, Thorlabs, inc.) using a 75.6 mm spherical lens. These conditions mimic the focal length of the cylindrical lens and the typical depth of field of a microscope objective under our imaging conditions. Pointing instability was observed to be in large measure driven by environmental temperature fluctuations and hence the entire setup was placed in an environmental room with long-term temperature fluctuations of ±0.1 °C. In

[17] Attributed to buckling of cavity mirrors (http://www.point-source.com/).
[18] R.m.s.: root mean square.

addition, fans were placed over each laser to improve thermal equilibration. These measures effectively eliminated any pointing instability[19].

Sheet properties

Ecosystems were maintained in 4.6 cm^3 glass cuvettes (special optical glass, Starna Cells, Inc.) with 10x10 mm optical path lengths, filled with 3 mL of medium. The focal length (f = 80 mm) and iris aperture (ø = 8 mm) were chosen to obtain a reasonably uniform and thin sheet over the field of view.

The sheet converges and diverges at a full angle α = 0.075 rad ($\arctan\left(f/\text{ø}\cdot n_w\right)$, with n_w = 1.33 the refractive index of water). *C. reinhardtii* and *T. thermophila* are imaged at a magnification of 5.2x (see below) and a CCD width of 4.75 mm. This results in a field of view width w of 900 μm and, assuming proper centering, a sheet thickness at the edges of the field of view of $\frac{1}{2} w \tan\alpha$ = 35 μm. For the *E. coli* imaging conditions, this is 16 μm.

In the center of the sheet, the thickness is limited by diffraction, yielding lower bounds of $1.69\frac{\lambda_0 f_0}{\text{ø}}$ for a sheet formed from a uniform beam[134], and $2.54\frac{\lambda_0 f_0}{\text{ø}}$ for a sheet formed from a beam with a Gaussian profile[20] [133], giving 4-σ thickness limits of 8.5 and 12.5 μm respectively for Gaussian fits to the transverse intensity profile[21]. The thickness in air of a sheet formed by an argon laser was measured by scanning with a 4 μm core fiber

[19] Alternatively, fiber coupled lasers can be used, where the fiber end becomes a new, stable approximate point source [22].

[20] Subscript 0 refers to values in vacuum or air.

[21] Cross-over between the diffraction limit and geometrical considerations occurs when $\lambda \cdot z \approx \pi\delta_0^2$ with z the distance along the illumination axis to the center of the sheet, and δ_0, one standard deviation of the Gaussian fit to the center thickness of the sheet. This yields a cross-over at about 40 μm from the center [133].

mounted on a precision XYZ-stage, and obtained an estimated 4-σ thickness of 14 μm after compensation for the fiber core size.

Imaging conditions

Our main concern in choosing objective lenses was to maximize light-gathering power, while retaining working distances long enough to work with deep samples (>1 cm across), and image quality. Brightness in our setup scales as NA^2/M^2, with NA the numerical aperture of the objective, and M its magnification.

Long-working distance (LWD) objectives are designed to meet these challenges, offering large working distances and large NA. In comparison with conventional microscopy, we eliminated the use of a tube lens and directly imaged the "intermediate" image formed by the objective lens. For the JAI cameras 10x LWD objective were used, while a 40x LWD objective was used to image *E. coli*. In order to maximize brightness, the distance between the objectives and cameras was reduced to close to the point where geometric distortions became noticeable[22]. This decreased magnification from 10x to 5.2x and 40x to 16x, and hence increased brightness about 3.6 and 6.2 fold, respectively[23]. Under these imaging conditions, pixel size is a little larger than a diffraction limited point source image.

[22] These distortions are due to deviations from paraxial approximations.

[23] A standard 5x Mitutoyo objective in the same series has twice smaller NA. NA is fairly insensitive to tube length, so we gain approximately a factor 4 in brightness by using a 10x objective at shorter tube length instead. For the 40x objective, the brightness gain over using a 20x ELWD objective from the same series is about 1.5x. Total excitation needed under these circumstances appeared approximately the same as when imaging on a conventional microscope.

Measurements

The setup has two states. In the measurement state a sample is exposed to selective plane illumination, and otherwise dark. In the rest state the sample is illuminated by white LEDs, allowing for photosynthesis (for a typical spectrum, see Appendix 4A.1, Figure 4A.3).

The exposure time during measurements is chosen to collect sufficient emission from individuals of each species. I chose an exposure time of 35 ms. The dimmest species is *E. coli* and its movement is dominated by convection in the system. Convection speeds vary from close to 0 to about 100 μm/s (convection within the system is nearly inevitable[24]). Since at 100 μm/s objects move 3.5 μm over a 35 ms exposure, longer exposure would only smear emission out over more pixels rather than increase the signal per pixel.

Exposure of the sample is controlled by a Pockels cell, an electro-optical modulator rotating the polarization of light when an electric field is applied. Since both DPSS lasers are linearly polarized, and the Pockels cell has an integral polarization filter, application of an electric field leads, in the proper orientation, to rejection of the beam. In the absence of a field, the beam is transmitted. Use of a Pockels cell allows for long term reliable operation, quick on and off times (0.1 and 0.5 ms, respectively), and high extinction ratios (about 200-fold for the combined beam).

Flow and motility also influenced the choice of measurement frequency. As shown in the next chapter, *E. coli* and *C. reinhardtii* stay in the field of view for about 3 and 5 seconds, respectively, while the dwell time is about 400 ms for *T. thermophila*. Measurements were made at a frequency of about 1.2 Hz, which means most *E. coli* and *C. reinhardtii*

[24] The Rayleigh number under these conditions is $3.9 \times 10^5 \cdot \Delta T$, with ΔT the vertical temperature difference across the ecosystem in Kelvin, while a Rayleigh number of about 2×10^3 is sufficient to induce convection.

were observed multiple times, while faster sampling could still somewhat improve *T. thermophila* density estimates.

Image analysis and classification errors

Because of the large number of images obtained (about 250,000 per day during continuous operation), it was necessary to develop automated image analysis methods. The three species were primarily distinguished by their different fluorescence emission characteristics (see above). The Retiga EXi camera (Figure 4.1) has a 620/60 nm emission filter, suitable for capturing dTomato emission. The two less sensitive JAI cameras have, respectively, a 665 nm LP emission filter for chlorophyll and a 515/30 nm emission filter for EYFP. These filters by themselves, however, offer insufficient discrimination under most circumstances.

E. coli identification is complicated by *C. reinhardtii* autofluorescence, which occurs over nearly the entire visible spectrum. *E. coli* cells are however brighter and smaller, and this distinction can be exploited to yield additional discrimination. After subtraction of a background image and smoothing with a Gaussian filter (see Appendix 4A.1), the maximum object intensity of *C. reinhardtii* is markedly lower than for *E. coli*. This allows for the application of additional thresholds depending on area and maximum brightness as shown in Figure 4.2 for a 20-day-old three-species ecosystem[25].

A similar improvement can be made for the distinction of *T. thermophila* and *C. reinhardtii*. Possibly in part because of scattered excitation (the EYFP filter has a cut-off at 530 nm, but still transmits about 0.3% of the 532 nm excitation), as much as 50% of *C. reinhardtii* leak through the EYFP filter. Both species are detected through the same

[25] This set of thresholds works equally well in discrimating *E. coli* from *T. thermophila*.

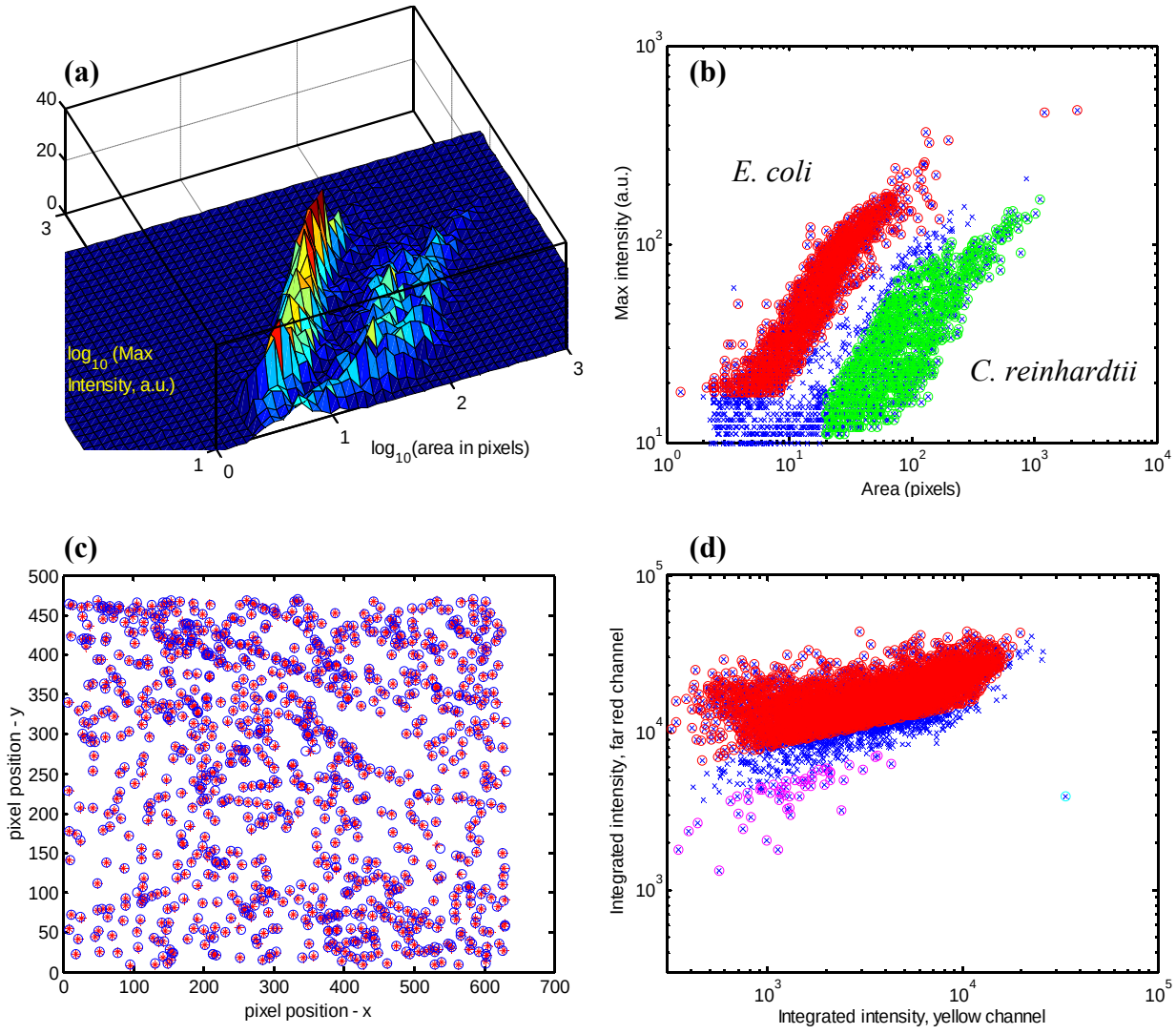


Figure 4.2: Additional thresholds aid in discrimination.
(a) While *C. reinhardtii* leaks through the *E. coli* emission filter, in an area versus maximum brightness plot, they form two distinct clusters, or ridges, as shown in this two-dimensional histogram (counts on *z*-axis). (b) Thresholded data: red, *E. coli*; green: *C. reinhardtii*. Blue crosses: all observations.
(c) Alignment of objects seen in both JAI emission channels. The best linear transformation described in the text aligns objects to within one pixel on average. The curvature of the sheet can be seen as a slightly sparser horizontal band. (d) Thresholding of objects observed in both channels. The vast majority of these objects are *C. reinhardtii* (red), with one *T. thermophila* (cyan) and a group of dead or dying algae shown in magenta (confirmed by flow cytometry with vital stains, data not shown).

objective, and the cameras can be aligned with multifluorescent beads to have the same field of view within a few pixels. Further alignment, to within one pixel on average, is performed a posteriori, by finding the linear transformation $\mathbf{y} = \mathbf{A} \cdot \mathbf{x} + \mathbf{B}$ (where **x**, **y** are coordinates of a list of paired objects in both emission channels[26]) that minimizes the sum of square distances of paired objects. **A** and **B** can be expressed in closed form, and were found to be nearly invariant over experiments lasting weeks.

After this alignment, objects appearing in only one channel were simply thresholded based on total intensity and area. These thresholds were chosen based on extensive manual inspection of acquired time series of pure and mixed cultures and designed to separate in-focus from out-of-focus objects.

For objects appearing in both channels, the excess intensity above background was integrated over the object in both channels, as shown in Figure 4.2 (d), and based, again, on extensive manual inspection, a set of thresholds was determined.

In pure cultures of *T. thermophila*, individuals almost never leak through the chlorophyll filter (~1%). In two- or three-species ecosystems, however, a small fraction of the population becomes sufficiently large that they can ingest *C. reinhardtii* (see Chapter 3 and Figure 5.2). Species densities often differ by several orders of magnitude, so it is important to determine classification errors. Classification errors were determined for pure cultures, and the results shown in Table 4.1.

[26] This list was obtained from time points at which only one object was seen in both emission channels, which almost always was the same individual.

Table 4.1: classification errors on a count/count basis, based on calibration runs with pure cultures. Number of observations given in parentheses. Actual classification errors in mixed cultures might be larger.

Inferred \ True	*C. reinhardtii*	*E. coli*	*T. thermophila*
C. reinhardtii	**>99 %**	**$<10^{-5}$** ($0/2.1\text{x}10^5$)	**$<10^{-3}$** **($0/2.4\text{x}10^3$)**
E. coli	**$1.1 \text{ x } 10^{-3}$** ($49/4.5\text{x}10^4$)	**>95 %**	**$2 \text{ x } 10^{-3}$** ($5/2.4\text{x}10^3$)
T. thermophila	**$3.8 \text{ x } 10^{-4}$** ($17/4.5\text{x}10^4$)	**$1 \text{ x } 10^{-5}$** ($2/2.1\text{x}10^5$)	**> 95%**

False positives, from electronic or photon noise were completely absent for *T. thermophila* and *C. reinhardtii*. The Retiga EXi used for imaging *E. coli* suffered from occasional pixel shot noise, but this could largely be eliminated by imposing an upper threshold in the maximum intensity versus area plane, as in Figure 4.2, reducing such errors to below 10^{-4} per frame.

Calibration

Counts in a small observation volume should be directly proportional to the (local) density, with a proportionality constant given by the observation volume. Results from calibration runs at different densities are shown in Figure 4.3. Error bars in the figure are based on a model of counting noise developed in the next chapter. These errors appear appropriate for *T. thermophila* and *E. coli*. For *C. reinhardtii*, the error bars are much too small, suggesting that counting noise was not the dominant source of error in the

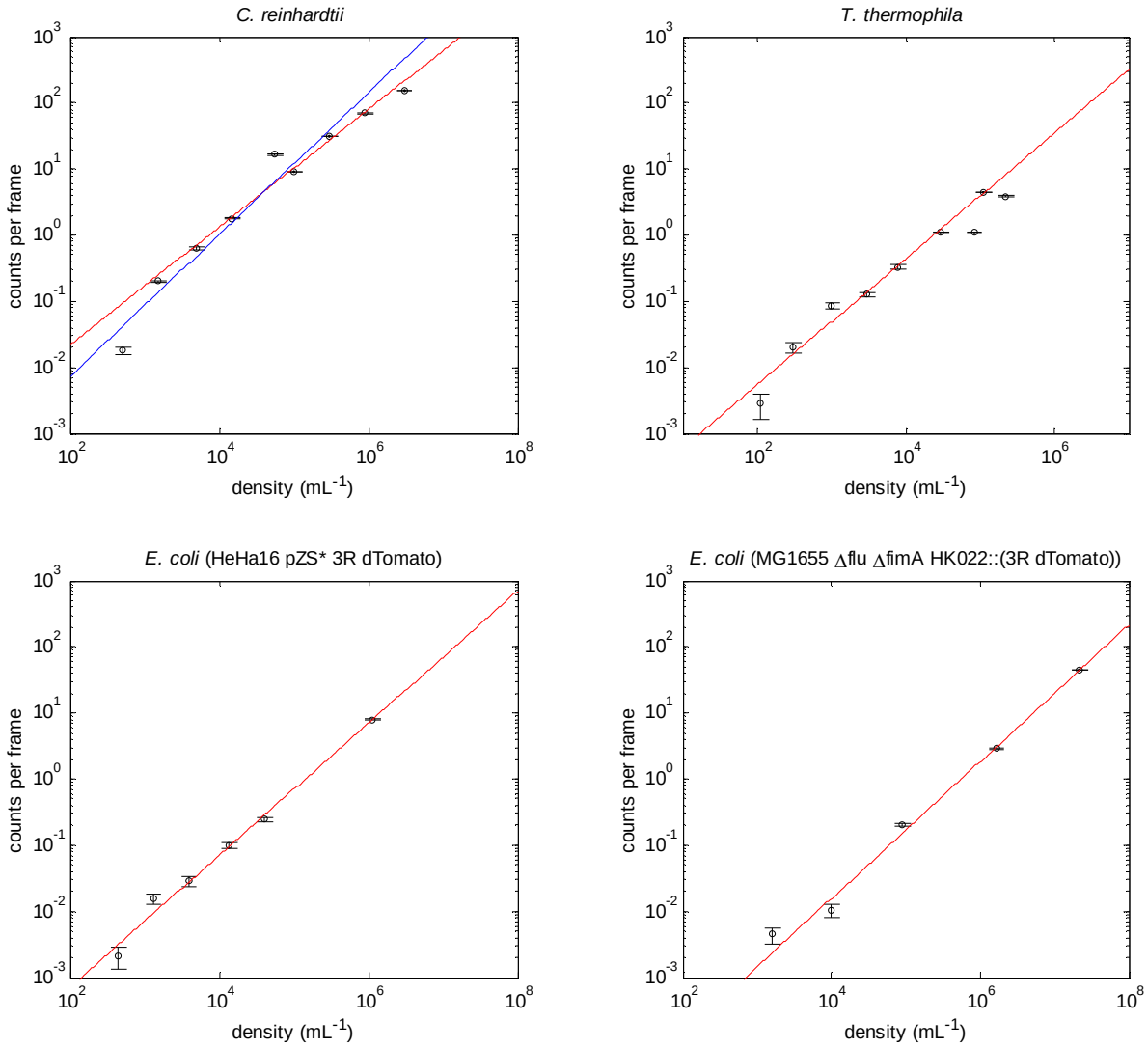


Figure 4.3: Calibration data: shown are the average number of counts per frame versus density for *C. reinhardtii* (top left), *T. thermophila* (top right) and *E. coli* (bearing dTomato on a plasmid, bottom left, or integrated in the chromosome, bottom right; detailed genotype in graph title).
Error bars are based on Poisson statistics correcting for serial correlations as described in chapter 5. These error bars appear inadequate for *C. reinhardtii*. Shown in red: weighted linear regression, in blue unweighted linear regresion. Slopes are 0.89 (weighted) or 1.07 (unweighted) for *C. reinhardtii*, 0.95 for *T. thermophila*; 1.00 and 1.03 for the two *E. coli* strains.

experiment[27]. Consistent with this, an unweighted linear regression gives a fit more in line with our expectations than the weighted regression line (see caption for details).

The lateral dimensions of the observation volume are set by the chip size and the magnification. The effective thickness of the observation volume, however, is more complicated, and is limited by the sheet thickness (δ) and depth of focus (*d*) of the respective objectives[28], as well as the size, *s*, of the objects imaged (any part can be in the imaged plane), and the transverse speed, $v_\perp$, of the object (if bright enough, it can spend any part of the exposure time in the sheet and be imaged), so, approximately,

$$\text{effective thickness} \approx \sqrt{\frac{\delta^2 d^2}{\delta^2 + d^2}} + s + v_\perp \cdot \tau_{\text{exp}} \qquad (4.1)$$

Approximate values of each of these parameters are given in Table 4.2. It is worth noting that for all three species, the three terms in equation 4.1 are of similar magnitude, with

Table 4.2: Parameters contributing to total observation volume as described in the text. *V* is the inferred observation volume based on the parameters in the table. V_{obs} is the observation volume based on the proportionality constant of the regression line in figure 4.3. All values are in μm, except for the observation volumes. τ_{exp} = 35 ms; *E. coli* motion was assumed to be dominated by convection up to 100 μm/s.

	d (2σ)	δ (2σ)	$\sqrt{\frac{d^2\delta^2}{d^2+\delta^2}}$	s	$v_\perp \cdot \tau_{\text{exp}}$	A_{CCD}/M^2	V (nL)	V_{obs} (nL)
C. reinhardtii	16	10	8.5	~5	2-4	1220 x 910	**20**	**54**
E. coli	3.8	7.5	3.4	1-2	0-2	550 x 420	**1.5**	**1.1**
T. thermophila	16	10	8.5	~10	4-8	1220 x 910	**32**	**68**

[27] I intend to repeat *C. reinhardtii* calibration with a phototactic mutant, hopefully allowing for more reliable density determinations.

[28] Both *d* and δ taken as two standard deviations in a Gaussian fit. Perfect alignment of focal and illumination plane is assumed. Depth of field calculation as in ref. [133].

the movement term being the smallest, making the observation volume fairly robust against changes in any of them. If anything, the discrepancies between calculated and measured observation volumes for *T. thermophila* and *C. reinhardtii* suggest our approximation for the effective sheet thickness is an underestimate. Over most of the extent of the sheet, under the imaging conditions for these two species, thickness is not limited by diffraction (see footnote 9 in this chapter) and its transverse intensity profile uniform rather than Gaussian as assumed in the calculation.

At high optical density (> 0.1), as for the highest density point in both the *C. reinhardtii* and the *T. thermophila* calibration curves, image quality starts to suffer under scattering, and counting becomes harder. Such high densities are, it seems, only observed in pure cultures in rich media. The lower limit of the dynamical range is set by either the total measurement time, or by false positives resulting from the presence of other species.

Two point setup

In order to provide some information on spatial density correlations, a second apparatus was built which can measure densities at two points in space. An outline of this apparatus is given in Figure 4.4. As can be seen, the construction is nearly identical to that of the apparatus described above. A smaller magnification objective was used for *E. coli* because the camera has smaller pixels (4.65 x 4.65 versus 6.45 x 6.45 μm^2). In this embodiment, *C. reinhardtii* densities are measured at two points, while *T. thermophila* and *E. coli* densities are measured at one point each. Measurement of two species at two points instead is straightforward. Details can, again, be found in Appendix 4A.1.

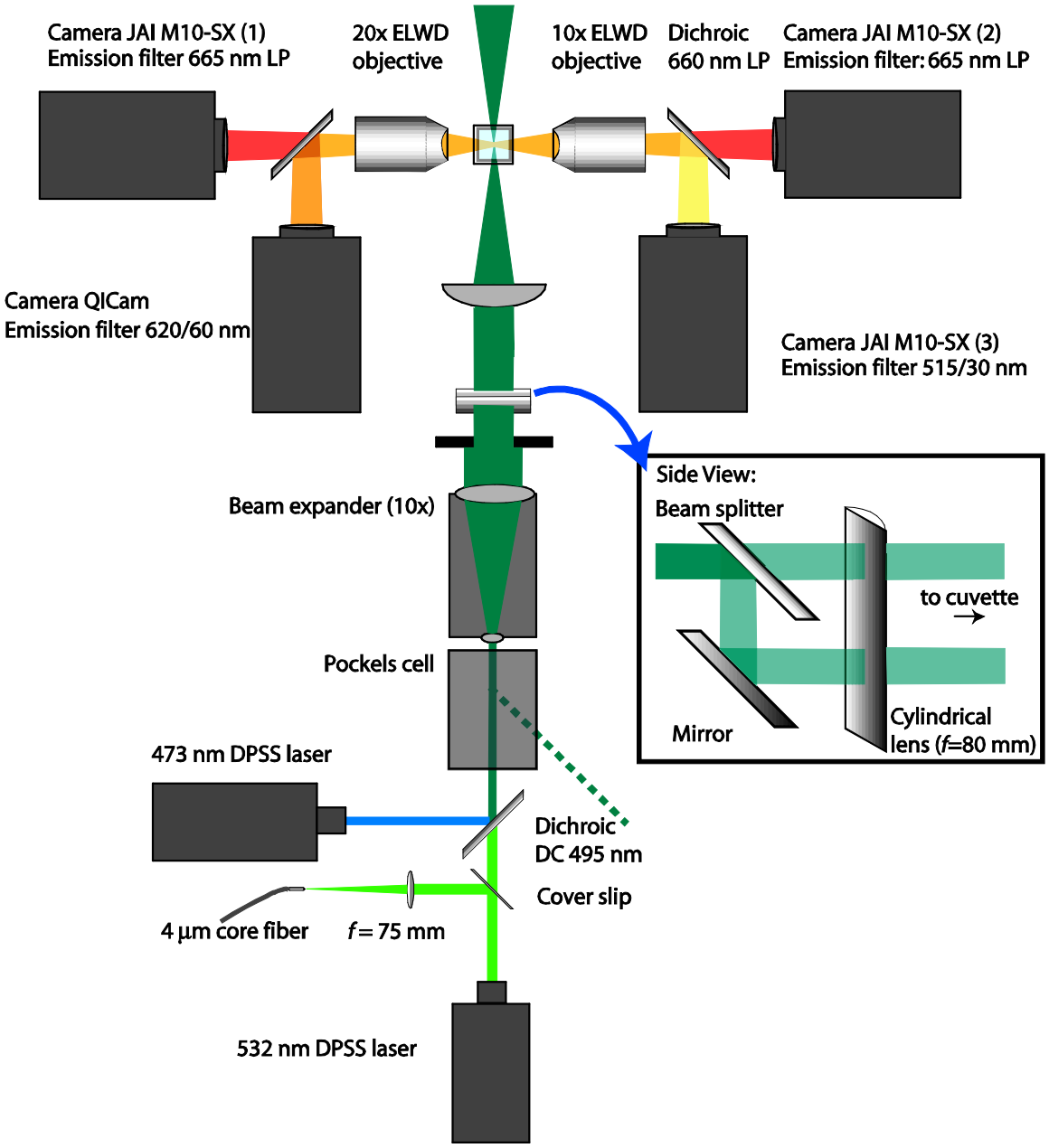


Figure 4.4: Two-point measurement apparatus. The design is completely analogous to the design in figure 4.1. The excitation axis runs vertically (blue, green) and the observation axis horizontally (yellow, orange, red). The insert shows a side view of the action of the beam splitter, splitting the beam into two equal power beams which each excite one observation volume.

Conclusions

Selective-plane illumination fluorescence microscopy was shown to be a suitable technique to measure population densities for all three species in a model closed ecosystem. The method provides a large dynamical range for all three species with low classification error. The lower limit of detection is limited by either classification error, if densities of one species are low, while others are much higher, or total measurement time. The method can be used either to take measurements of a number of ecosystems, for example once per day (Chapter 5), or continuous measurements can be taken from a single ecosystem (Chapter 6 and 7).

To conclude, a few comments on improvements and extensions of the setup that can be made. The current protocol has two sources of perturbation: sample handling and illumination. Sample handling can in principle be made reproducible by some kind of automatic handling, with samples placed for example on a movable vertical rail. In the current experiments swarms could often be seen being displaced from the meniscus during handling.

Maximum illumination in the current setup is approximately the same as that of a conventional microscope using similar objectives[29]. In principle, a rectangular slit can reduce the fraction of the ecosystem currently exposed to excitation (now ~1%) five to tenfold over the current iris. Also, more sensitive cameras are available with larger pixels, per pixel amplification, and higher quantum efficiency, allowing for reduction in

[29] The total exposure of the sample was 10 lumens from fluorescence excitation (time average: 0.4 lumen), and 0.4 lumen from the white LEDs, so excitation is still large relative to the LEDs.

exposure time by as much as 20-fold[30] (this would reduce excitation times to ~0.2% of total time).

While the method was developed for a particular ecosystem, it could be extended to more and different species. Other green algae can be substituted for *C. reinhardtii*, as can other bacteria for *E. coli*, as long as they can be fluorescently labeled.

The trick used to increase separation between *C. reinhardtii* and *T. thermophila*, alignment of fields of view of two cameras can be extended to more than two using the basic idea of infinity-corrected optics. Any number of optical elements (e.g. dichroic filters) can be inserted after an infinity-corrected objective, in effect separating the image into emission channels much like is done in flow cytometry[31]. This way more fluorophores can be separated and even multiple fluorophores can be used per species, providing for example simultaneous information on gene expression.

Finally, population densities are observed in only a small part of the ecosystem. As such, the bulk of the ecosystem is probably fairly well-mixed, although *C. reinhardtii* and *T. thermophila* can swim faster than typical flow seen in the system.

As shown in the previous chapter, these ecosystems become spatially and phenotypically heterogeneous over time. This means the densities observed may not accurately reflect total population size. On the contrary, surface populations can rival the bulk population in size (see Chapter 3). It remains to be seen how much we can say (a) despite and (b) about this underlying complexity.

[30] Experiments were performed imaging *E. coli* side by side with a Retiga EXi and a demonstration Princeton Instruments Cascade camera.
[31] Each camera would require its own additional lens.

Appendix 4A.1 Technical Details

Optics

Experiments were performed in a Honeywell HotPack environmental room, kept at 22.8 °C, with long-term stability of ±0.1 °C. DPSS lasers were purchased from Lasever, Inc. (LSR473-U-20mW) and AOTK, Inc. (MGL-Green series 532 nm, 50 mW), and cooled by external convective cooling to within 0.1 °C over a three week period.

Pointing stability of the 532 nm laser was monitored by splitting the beam using a cover slip and subsequent focusing using a spherical lens (f = 75.6 mm) onto a 4 μm core single-mode optical fiber (FSSN 3224, Thorlabs, Inc.). The output of this fiber was collimated using an aspheric collimator (Thorlabs) and directed at a light sensor (Vernier, Inc., Beaverton, OR). Output of the sensor was collected by a LabPro unit (Vernier), and recorded via serial cable by a custom Visual Basic program. In general, root mean square variation between and during measurement days was below 10%. On the time scales of months of operation, however, there was a slow decrease in laser power.

Laser beams were combined into a single beam using a 495 nm long-pass dichroic (Chroma Technology, Rockingham, VT), and collimation verified over a two-meter span, as well as by use of multifluorescent beads in the setup (see below).

Electronic shuttering was performed using a Pockels cell (350-50 -01 E-O Modulator, ConOptics, Inc., Danbury, CT) which contains a KD_2PO_4 crystal and an integral Glan type polarizer. A custom driver circuit for the Pockels cell was coupled to a circuit controlling the rest of the setup (see Figure 4A.2, and text below). Charge and discharge

times for the Pockels cell were 0.5 and 0.1 ms, respectively. Orientation of the Pockels cell and voltage were optimized for the combined beam using a light sensor.

The combined laser beam was expanded using a 10x beam expander (unknown model), constricted using an adjustable iris (diameter 8 mm), and focused using a cylindrical lens (f = 80 mm, LJ1105L2-A, Thorlabs, Inc.) with an antireflective coating for the visible range.

Ecosystems were maintained in fluorimetry cuvettes (Special Optical Glass, Starna Cells, Inc., Atascadero, CA) with a nominal volume of 3.5 mL (inner dimensions 10x10x35 mm; total volume: 4.6 mL), filled with 3 mL medium (see below). Cuvettes were closed with screw caps. Caps for the first experiment described had either a silicone septum or were closed and contained a silicone insert lined with Teflon®. Caps for the second experiment were all closed with Teflon-lined insert. Measurements were done 5 mm below the lowest part of the meniscus.

Fluorescence emission was collected using a 10x CF Plan Mitutoyo objective (nominal NA = 0.30, working distance = 16.5 mm, infinity-corrected) and a 40x ELWD Nikon objective (nominal NA = 0.5, working distance = 11 mm, 210 mm tube length). Objective position was controlled using Mitutoyo linear stages. Light from the 10x objective was split by a 45° 660 nm long-pass dichroic (Chroma Technology). The distance from the base of the 40x objective to the CCD was about 75 mm. This distance was about 90 mm for the 10x objective.

Red fluorescence emission (dTomato, *E. coli*) was collected on an uncooled monochrome Retiga EXi camera (QImaging Corp., Surrey, BC, Canada) with a D620/60 emission

filter. To increase image brightness, pixels were binned 2x2 (four pixels forming one effective pixel). Native pixel size is 6.45 x 6.45 μm.

Yellow fluorescence (EYFP, *T. thermophila*) was collected on a CV-M10 SX monochrome progressive scan camera (JAI, Inc., San Jose, CA, pixel size 9.9 x 9.9 μm) fitted with a HQ515/30 filter. Far-red fluorescence (chlorophyll, *C. reinhardtii*) was collected by a second CV-M10 SX camera, fitted with a custom 665 LP filter. All filters were produced by Chroma Technology. The two CV-M10 SX cameras were controlled by a Flashbus MX capture card (Integral technologies, now Pelco). A software development kit (SDK) is available for the capture card.

Alignment

Alignment of the two laser beams is described above. All alignment was performed using polystyrene multifluorescent beads (XPR-1251, ø = 3 μm, Duke Scientific, Palo Alto, CA). Cameras were centered visually through the objectives (the center of the sheet can be identified when the cuvette is slightly moved transversely, causing a distinct twinkle as beads move in and out of the sheet faster in the center).

Relative positioning of the two JAI cameras, mounted on XYZ stages (Edmund Optics, Barrington, NJ), was performed by simultaneous imaging in both channels using a custom C++ program and ImagePro Plus version 4.5.1.29 (Media Cybernetics, Bethesda, MD).

After placement of a sample in the setup, due to cuvette to cuvette variability, focus often had to be adjusted by up to about 10 μm using a Mitutoyo linear stage. This was done by

inspection of live video in all three channels. These shifts in focal plane position between cuvettes were largely reproducible from day to day.

Integration

As described in the text, the setup has two states. In the measurement state, the sample is exposed to fluorescence excitation and all cameras acquire simultaneously. In the rest state, excitation is off, and the sample is illuminated by LEDs. Operation of the setup is controlled by a custom C++ program which coordinates exposure and image acquisition and controls the intensity level of the LED illumination. It is linked to libraries provided by the capture card manufacturer, Integral Technology, and QImaging. Most timing, however, is done electronically. The Retiga EXi provides an exposure mask, a TTL signal that is high (5 V) as long as the CCD is exposed, and 0 V otherwise. The Retiga itself is controlled by the custom program through a FireWire connection.

The TTL signal from the Retiga EXi is used to control the action of the Pockels cell, the LEDs and the other two cameras by a circuit shown in Figures 4A.1 and 4A.2. The two JAI CV-M10 SX cameras have a so-called pulse-width controlled exposure mode, in which the exposure timing can be controlled by an external TTL signal (0/5V).

Sample illumination

When in the rest state, about 96% of the time, the sample in the setup is illuminated by four LEDs (BW03, Philips Lumileds Lighting Co., San Jose, CA), a representative spectrum for which is shown in Figure 4A.3. The panel holding the LEDs is mounted orthogonal to the incoming beam (Figure 4A.4). To decrease the anisotropy of

illumination, a mirror is placed parallel to the LED panel across the cuvette, and the space spanned by the mirror and the panel is covered with white paper. The entire setup was enclosed in a black cardboard box. Illumination intensity was 1200 lux, measured at the position of the observation volume with a light sensor facing the LEDs. Overall light intensity was probably somewhat higher.

LED intensity could also be varied, as done for the experiments in Chapter 6. To this end a tuning curve was constructed of the measured light intensity versus a control voltage sent from the capture card (see Figure 4A.1) using a random sequence of control voltages. This randomization served to check for hysteresis effects, which were not observed.

Computer hardware

The experiment was controlled, and data collected, on a custom-built PC, with an ASUS A8N-SLI Premium motherboard and an AMD Athlon 64 processor. The Lian Li PC-6077B ATX Mid tower computer case provides housing for four Kingwin KF-812 or 813 trays for removable hard drives (on some occasions this required a restart). The computer ran Windows XP Pro with SP2.

SP2 causes Windows XP to run out of free page file table entries after handling about 700,000 images on this system, causing system failure. The underlying NTFS error was corrected for using a hotfix available from Microsoft upon request (see KB Article Number: 918338).

Images from the Retiga EXi were compressed losslessly using Lempel-Ziv-Welch compression (libtiff version 3.8.1, www.libtiff.org). JAI images were, due to capture card

SDK limitations, saved as bitmap files. The available compression option, run length encoding, did not decrease image size appreciably, and images were hence saved uncompressed. With few exceptions, images were stored on Seagate Barracuda 7200.10 320 GB hard drives.

Two-point Setup

The outline of the two-point setup is shown in Figure 4.4 of the text. Its construction was nearly identical to that of the initial measurement apparatus ('one point setup'). Two more powerful lasers were used, a 50 mW 473 nm (blue) DPSS laser (LSR473-U-50mW, Lasever, Inc.), and a 100 mW 532 nm (green) DPSS laser (MGL series, CNI lasers, Inc.[32]). The beams were combined as above, and split into two equal power mixed beams using a 50/50% beam splitter (Thorlabs, Inc.). The two observation volumes were approximately 5 and 25 mm above the bottom of the cuvette (total liquid height 30 mm). The objectives used were a Mitutoyo 10x/0.28 NA M Plan Apo (for infinite tube length) objective with 16.5 mm working distance and a Nikon 20x/.5 NA M Plan ELWD objective (for 210 mm tube length, all nominal values). Images were acquired on JAI CV-M10 SX cameras as before, and a QImaging QICam camera with 4.65 μm square pixels. All software written for the Retiga EXi camera above could be used to control the QICam without modification. The JAI cameras were controlled from two capture cards (Flashbus MX, Integral Technologies). For this, an additional software library is available from Integral Technologies, entirely analogous to the one used for the one-point setup.

[32] It appears that CNI took over this line of lasers from AOTK, Inc. from which the previous green laser was purchased.

In comparison with the one-point setup, control of the boundary conditions was improved. A fan and holes in the enclosure, still without significant light leakage, provided convective air flow from and to the environmental room. Fluctuations in temperature due to illumination ranged from ±0.04 to 0.08 °C, comparable to fluctuations due to room temperature (in)stability. Illumination was made much more isotropic by placing four LEDs (BW03, LumiLEDs, as before) symmetrically around the cuvette, and diffusing white plastic in between. All remaining electronic and software integration was identical to that of the one-point setup discussed above.

Sample stands

Sample stands consisted of two parts (see Figure 4A.4). A grid of 2x2 (for experiments in Chapter 4) or 4x4 (for experiments in Chapter 3) LEDs (BW03 as before) with a center to center spacing of 32 mm with reflective film in between was covered with two layers of diffuse white film (Lee #216, Barbizon, New York, NY) at distances of 3 and 5 cm above the grid. The surrounding box was made of matte-silver cardboard (Barbizon).

At 2 cm above this illumination box, a clear acrylic sheet (McMaster-Carr, Princeton, NJ) 2 or 3 mm thick supported the cuvettes which were held into position by a second acrylic sheet through a grid of holes of appropriate size. To improve isotropy of the illumination, cuvettes were surrounded by a matte-silver box of cardboard and matte aluminum foil.

Light intensity of the LEDs was controlled by constant current source circuits like the one included in Figure 4A.1. The light intensity as measured by a Vernier light sensor at meniscus level facing downwards was 1200 lux ± 10%. Temperature of the cuvettes was maintained by forced air flow between the illumination box and the cuvettes at 23.1 °C,

or about 0.3 °C above the temperature in the room. The analogously made sample stand for the 96-well experiment in Chapter 3 was a bit warmer at 23.6 °C.

Measurement protocol

On sample days, each cuvette was taken from the sample stand and placed in the setup. After allowing at least three minutes for flow to subside, which was generally adequate, the position of objectives was fine-tuned to compensate for cuvette to cuvette variability (see above). Images were taken at a frequency of about 1.2 Hz (with an attainable maximum of about 3 Hz) and an exposure time of 35 ms, as explained in the text.

The temperature in the setup box was about 25.5 °C, a little higher than in the rest of the room due to heating by the LEDs and cameras. A typical measurement consisted of 3000-5000 images (30-60 minutes). This length was adjusted depending on previously observed densities.

Image analysis

Image analysis proceeded in two rounds. In the first round, objects were identified in images and a list of properties of each object written to file. In the second round, additional thresholds were applied, as described in the text, to improve species discrimination. Analysis was not performed in real-time (during acquisition) in order to improve system stability. Also ImagePro Plus version 4.5.1.29 (Media Cybernetics) experiences a resource leak in Windows XP when XP is not run in "Classic" mode.

Images from the JAI camera on channel 1 of the capture card (see above) occasionally for long stretches (~10^5 images) would be misaligned, with the top part of the image

appearing at the bottom. All the information in the images was retained. This defect could be corrected automatically in ImagePro Plus after manual identification of the break line.

The first round, object identification, was performed in ImagePro Plus. To JAI images a mask was applied identifying pixels above an intensity threshold. Groups of connected bright pixels were then subjected to an area threshold and the properties, such as area, aspect ratio and intensity-weighted center coordinates, of large enough objects written to a text file. For Retiga EXi images (intended to count *E. coli*) a background image was first subtracted to account for any heterogeneity in base pixel brightness and the image filtered with a low-pass Gaussian filter (kernel size 4 pixels, 5 passes). Afterwards size and intensity thresholds were applied as for the JAI images. All thresholds were extensively validated by manual inspection of threshold performance on images acquired from pure and mixed cultures.

In the second round, additional thresholds were applied in Matlab (Mathworks, Inc.) as described in the text. At high optical densities (for example in sample (5) of the first experiment, see Figure 4.4) it was necessary to scale all thresholds for maximum intensity in Retiga images by a factor ranging from 0.5 to 1, as judged by shifts in cluster position.

Finally, counts per frame were converted to density estimates using the calibration curves described in this chapter. Where applicable, for the purpose of analysis, 0-counts were replaced by ½ counts to prevent the occurrence of infinities when considering log n.

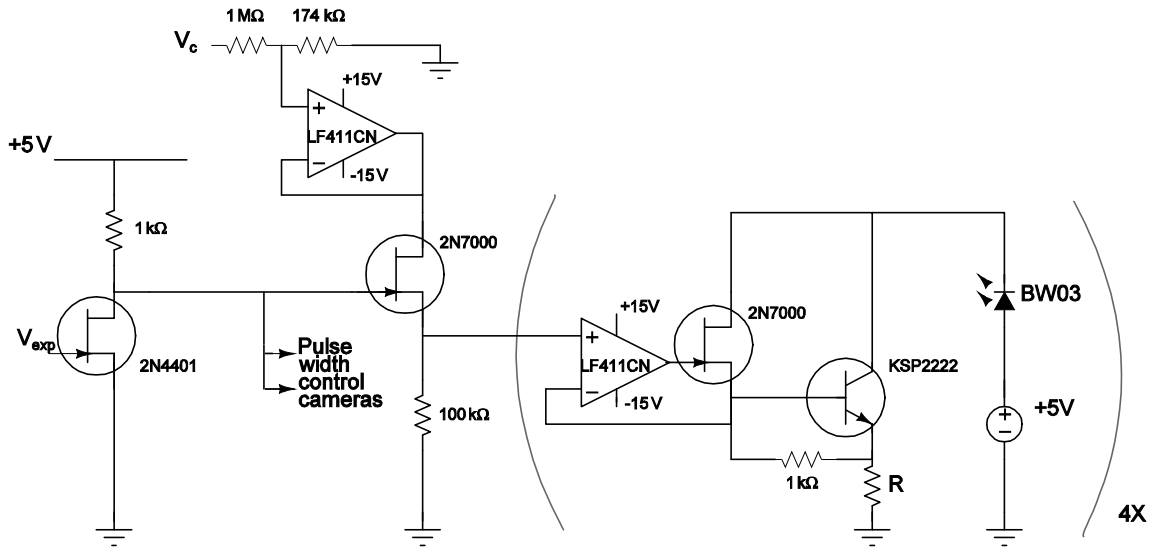


Figure 4A.1: Circuit diagram for the control of LEDs and JAI cameras.
LED intensity is determined by a control voltage (V_c) coming from the capture card DAC pin, which ranges from 0 to 10 V in 256 steps. LEDs and pulse-width controlled exposure are turned on and off by a TTL signal (V_{exp}) coming from a QImaging Retiga EXi camera which indicates its exposure. The circuit within parentheses is repeated for each LED, and based on a constant current source proposed in The Art of Electronics, Chapter 4 [144]. The current through the LED is approximately $\gamma V_c/R$, with R = 2.2 Ω, and γ determined by the voltage divider on top of the diagram (here 0.15).

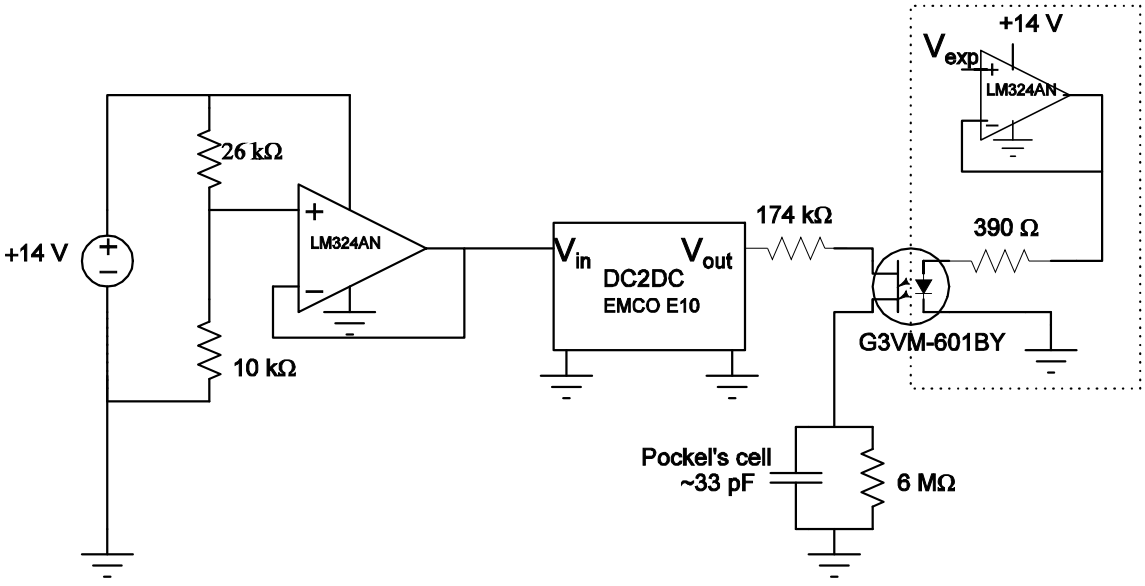


Figure 4A.2: Circuit diagram for control of the Pockels cell.
The circuit consists of two optically-coupled parts. A voltage divider with a voltage follower set a precise input voltage for a DC to DC voltage converter which converts to high voltages (EMCO E10). The Pockels cell rotates the polarization of light dependent on the voltage across it (about 500V for a 90° rotation). As such it effectively acts as a small capacitor (about 33 pF). Shown in the bounded box is a circuit that places the TTL exposure mask of a QImaging Retiga camera across a LED inside a MOS solid state relay, which acts as a high voltage transistor. The circuit has charge and discharge times of the voltage across the Pockels cell of about 0.5 and 0.1 ms, respectively, and were limited by the capacity of the relay (G3VM-601BY) rather than that of the Pockels cell.

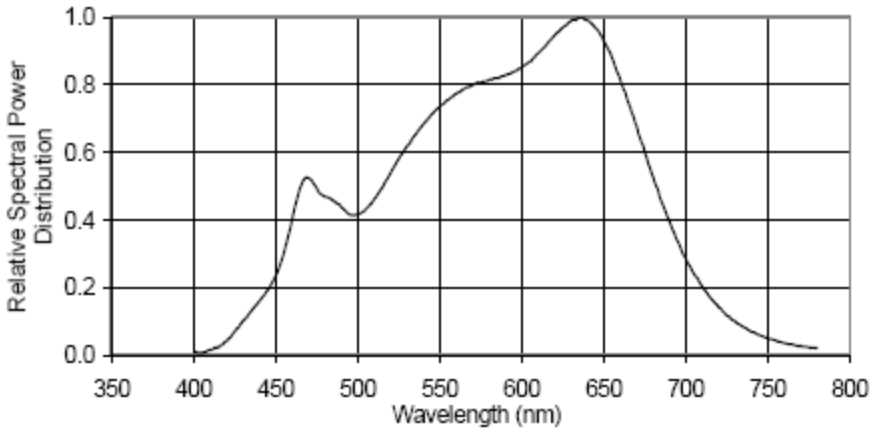


Figure 4A.3: power spectrum of a typical warm white Luxeon LED, as described by the manufacturer (Technical Data Sheet DS47, Lumileds Lighting, U.S., LLC, San Jose, CA).

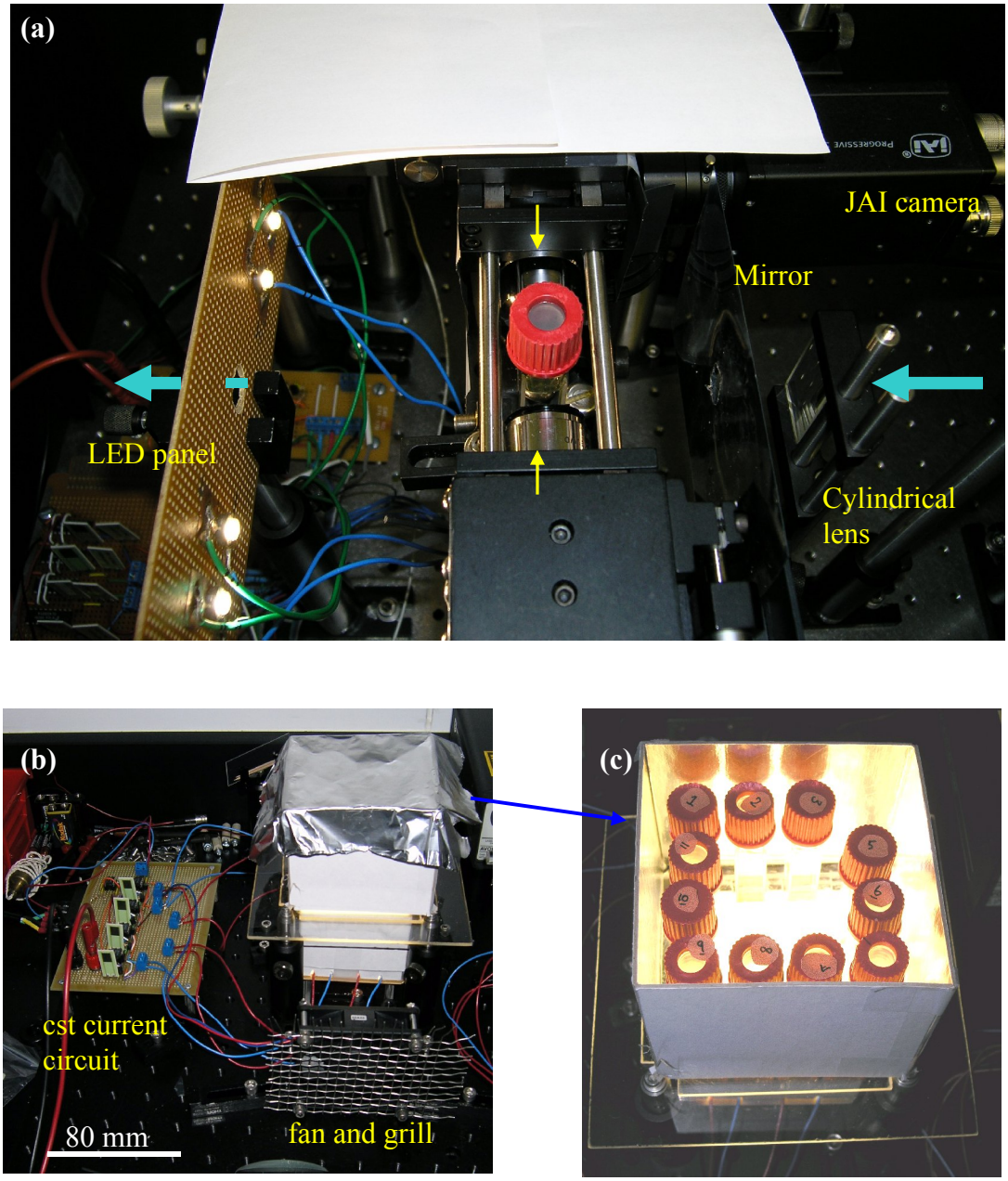


Figure 4A.4: Illustration of the setup and sample stands. (a) The setup, top view. To provide a sense of scale, the outer width of the cuvettes is about 12 mm. Yellow arrows indicate the objectives; cyan arrows the excitation laser beam. In the rest state in between image acquisition, the four white LEDs provide illumination, and mirrors and a white paper cover are intended to decrease overall illumination anisotropy. (b) and (c) sample stand in which samples were kept when not measured. Samples were illuminated from below and surrounded by matte silver cardboard (sides) and matte aluminum foil (cover). A fan maintained airflow between the illumination box and the samples.

Appendix 4A.2 Experimental Protocols

Glassware

Bottles and beakers used for preparation of Taub medium, and stock solutions for it, as well as cuvettes for the first experiment were cleaned (1) in 10% HNO_3 at least overnight (50% HNO_3 for cuvettes), (2) extensive rinsing in ddH_2O (from a Millipore Milli-Q water purification system, Quantum EX cartridge), (3) three rinses with isopropanol (Fisher Scientific, >99.9%), with 5-10 minute drains in between, and (4) three rinses with ddH_2O with draining in between.

For the cuvettes in the second experiment, as well as the subsystem experiment in Chapter 3, step (3) in the above protocol was replaced by placing the cuvettes inverted in a stainless steel steamer over boiling and refluxing 100% ethanol (absolute, 200-proof, AAPER, Shelbyville, KY) for at least 30 minutes. All cuvettes were autoclaved for 30 minutes well in advance of use. Glass pipettes used in the preparation of individual ecosystems were used as provided by the Rockefeller University glass washing facility.

Purity testing

Cultures of *C. reinhardtii* and *T. thermophila* were inspected directly under a light microscope. All cultures and media used in the construction of ecosystems were also tested for purity by plating 100 μL on each of Luria-Bertani agar (Teknova, Inc., Hollister, CA), Nutrient Broth agar (Difco/BD, Sparks, MD), Thioglycollate Broth agar (Difco) and Yeast extract Peptone Dextrose (YPD, Teknova) agar, all at 1.5 % agar (Bacto agar, Difco). Purity testing of ecosystems from experiment 1 in this chapter used

duplicate plates for each with 50 μL culture per plate. All cultures and stocks were maintained without any added antibiotics.

Marker stability

The stability of *T. thermophila* fluorescent labeling was assessed by serial subculture. For 16 months, about once every three or four weeks, each of fourteen transformants was subcultured 1/20x into 1.5 or 2 mL of fresh medium (SPP [84]) in a 24-well plate. No marker loss was observed over this period.

Ecosystem divergence: experiments 1 and 2

The experiments were performed with *E. coli* strain MG1655 Δ*flu* Δ*fimA hsdR* HK022::(3R dTomato), *C. reinhardtii* strain UTEX 2244, and *T. thermophila* H3.2-EYFP. For details on strains, please see Appendix 4A.3. For the second experiment, *T. thermophila* H3.2-EYFP sustained in a long-term (5 month) soy bean culture was used after cryopreserved *T. thermophila* failed to recover (see [84] on such cultures).

Starting cultures were grown to: for *C. reinhardtii* in TAP: (1) 2.9×10^5 mL^{-1} and (2) 2.8×10^6 mL^{-1}, *T. thermophila* in SPP (1) 5.2×10^4 mL^{-1} and (2) 2.7×10^5 mL^{-1}, and *E. coli* in ½x Taub #36 with 0.03% proteose peptone (1) 6.6×10^7 mL^{-1} and (2) 6.0×10^7 mL^{-1}. Densities were determined using a Coulter Counter for *C. reinhardtii* and *T. thermophila*, and an OD_{400} calibration curve for *E. coli*.

Cultures were washed twice in ½x Taub #36 with 0.03% proteose peptone (300 *g*, 5 minutes for C. reinhardtii and T. thermophila; >10,000 *g*, 1 minute for *E. coli*). Ecosystems were contained in fluorimetry cuvettes (Special Optical Glass, Starna Cells,

Inc., Atascadero, CA) with a nominal volume of 3.5 cm^3 and a total volume of 4.6 cm^3. A master mix of ½x Taub #36 with 0.03% proteose peptone with *C. reinhardtii* at 5000 mL^{-1}, *E. coli* at 500 mL^{-1} and *T. thermophila* at 50 mL^{-1}. From this master mix, 3 mL was pipetted into each cuvette, and the cuvette sealed by tightening the screw cap.

Cuvettes with detectable leakage at the end of the experiment (100-200 μL over a 200 day period) were a posteriori excluded from all analysis. This concerned 0/9 cuvettes for experiment 1, and 1/10, 1/10, 2/10, 1/10, 1/10 for the respective measurement frequencies in experiment 2. In the second experiment, wide-bore pipette tips were used for *T. thermophila* to minimize any cell damage.

Two-point experiment (Chapter 7)

Glassware, washing, medium and purity testing followed experiments 1 and 2 above. The experiment was performed with *E. coli* strain MG1655 Δ*flu* Δ*fimA hsdR* pZS* 3R dTomato, *C. reinhardtii* strain UTEX 2244, and *T. thermophila* H3.2-EYFP. For details on strains, please see Appendix 4A.3. All strains were obtained from frozen stocks. Starting cultures were grown to: for *C. reinhardtii* in TAP: $7.6x10^6$ mL^{-1}, *T. thermophila* in SPP $2.1x10^5$ mL^{-1} and *E. coli* in ½x Taub #36 with 0.05% proteose peptone: $1.5x10^7$ mL^{-1}. Densities were determined using a haemocytometer for *C. reinhardtii* and *T. thermophila*, and an OD_{400} calibration curve for *E. coli*. Densities were the same as in the preceding experiments, with *C. reinhardtii* at 5000 mL^{-1}, *E. coli* at 500 mL^{-1} and *T. thermophila* at 50 mL^{-1}.

Calibration experiments

Magnifications of the objects were determined by imaging a 100 gauge mesh (SPI supplies, West Chester, PA). Classification error and calibration curves were determined from the same data sets.

C. reinhardtii was grown in TAP, *T. thermophila* in SPP and *E. coli* in ½x Taub #36 with 0.05% proteose peptone to late log-phase, as for the experiments described above. Serial dilutions were made for each culture into ½x Taub #36 with 0.05% proteose peptone, and placed in a fluorimetry cuvette. Measurements were performed as for the above experiments. After each measurement, a 100 μL sample was taken at approximately the position of the sheet, and densities determined using a Coulter Counter (ZM, 200 μm aperture, Beckman Coulter) for *C. reinhardtii and T. thermophila*, and plate counts for *E. coli*.

Appendix 4A.3: Strain Details

MG1655 Δflu ΔfimA HK022 att::(cat $P_{\lambda R}$ dTomato) hsdR

A plasmid containing the ORF (open reading frame) for tdTomato (tandem dTomato, [138]), pRSET-B tdTomato, was kindly provided by the Tsien lab. TdTomato consists of two copies of dTomato transcriptionally fused to each other. In wild-type *E. coli* strains, this construct appeared to be highly unstable. Instead, using primers containing KpnI and HindIII restriction sites, dTomato was obtained from pRSET-B tdTomato, gel purified, and cloned into a pZA 3R vector (present in the lab, see ref. [127] for a general map of these plasmids)[33].

By sequencing, it was determined that the final construct had one mutation, H173R, but this did not affect fluorescence, and it was hence retained.

Next, the 3R dTomato fragment, containing chloramphenicol acetyl transferase (cat), which confers chloramphenicol resistance, and dTomato under the strong constitutive λ_R promoter, was isolated by restriction with Sac I and Hind III, and ligated into pZS* and pZS* HKatt backgrounds. pZS* plasmids, like the pZA plasmids before, have been described by Lutz and Bujard [127]. HKatt, the attachment site for lambdoid bacteriophage HK022, had been previously inserted in the pZS* background by PCR by John Chuang in our lab.

The remainder of the chromosomal integration follows the protocol of Haldimann and Wanner [145]. The pZS* 3R dTomato (HKatt) plasmid was digested with Avr II and Spe I, the unique restriction sites of which flank the origin of replication. The cohesive ends

[33] Forward primer: GCATACGGTACCATGGTGAGCAAGGGCGAGG (Kpn I)
Reverse primer: CACTCCAAGCTTACTTGTACAGCTCGTCCATGCC (Hind III)

after digestion are compatible and were religated. The hybrid restriction site cannot be recut with either restriction enzyme. Recutting with Avr II was used to linearize any remaining plasmid which had retained the ori. MG1655 was transformed with a helper plasmid, pAH69, which is only maintained at 30 °C and expresses the HK022 integrase. In a second round of transformation the ori-less religated '3R dTomato HKatt' circular fragment was introduced by electroporation, with outgrowth at 42 °C and subsequent plating on LB cam agar[34] at 37 °C. Individual colonies were screened for integration by colony PCR using primer pairs as described [145].

The effects of the knock out of *flu* and *fimA* were described by Hasman *et al*. [110]. Briefly, fimbriae (for which fimA is the main structural unit) and ag43 (the protein encoded by *flu*, implicated in cell-cell adhesion) are involved in flocculation of liquid cultures. The knockouts were done by two rounds of P1-transduction [89] from KEIO collection strains [146]. In each KEIO collection strain, an identified ORF has been replaced by a neomycin resistance gene flanked by two FRT sequences. This casette can subsequently be removed by expression of FLP from a helper plasmid, leaving a 102 nt (34 aa) in-frame scar as described [146]. After the first transduction (*flu*), the neomycin resistance casette was removed to facilitate marker selection in the second round (*fimA*). After the second round, neomycin resistance was retained as a convenient marker. Both deletions were confirmed by chromosomal sequencing.

The KEIO collection donor strains have a known set of differences from MG1655 (being derivatives of BW25113 [146]), of which only one had a reasonable chance (~30%) of being cotransduced with the resistance marker. The mutation, hsdR results in loss of ability to restrict DNA not methylated by hsdM. I tested transductants by sensitivity to λ

[34] Luria Bertani broth with 12.5 μg/mL chloramphenicol.

phage (advised by Barry Wanner, pers. comm.) grown on MG1655 versus methylation deficient strains WA803 and ED8767 (CGSC, Yale University). For the experiments described, a clone that *does* carry the hsdR mutation was used to allow for additional strain verification.

Chlamydomonas reinhardtii

Chlamydomonas reinhardtii strain UTEX 2244 (mating type: +) was obtained from the University of Texas at Austin Culture Collection[35]. It is also known as CC-125 and reportedly nitrate reductase deficient [81].

Tetrahymena thermophila H3.2-EYFP

This strain was created with much help and advice from Dr. Yifan Liu in the Allis lab (now at the Department of Pathology, University of Michigan Medical School). He performed the transformations. The protocol follows the description in [147], with the exception that the somatic rescue plasmid is different. It contains the wild-type HTT2/HHF2 locus[36] with a 6-amino acid linker encoded by GAT CCA CCT GTC GCC ACC (DPPVAT) and the EYFP-coding sequence inserted into the HHT2 gene right before the stop codon. The fusion ORF was constructed by fusion PCR and confirmed by sequencing.

[35] http://www.utex.org/
[36] HTT2 is the ORF for histone 3.2, HHF2 is the ORF for histone 4.2.

As explained in the text, the EYFP sequence was derived from a plasmid, pYFP-URA3 with sequence optimized for expression in *C. albicans*, kindly provided by Mrs. M. gerami-Nejad, Berman lab, University of Minnesota[37]. Transformation is easily verified by fluorescence microscopy. Mating types of the transformants were not determined.

[37] Forward primer: GAT CCA CCG GTC GCC ACC TCT AAA GGT GAA GAA TTA TTC ACT
Reverse primer: AGA CTA GTT GTT ATA TTA TGC TCA TTT GTA CAA TTC ATC CAT ACC

4A.4 Culture Media

TAP

Tris-Acetate-Phosphate (recipe obtained from the Volvocales Information Project, http://www.unbf.ca/vip/), used for the culturing of *C. reinhardtii*. I replaced $CoCl_2.6H_2O$ by $Co(NO_3)_2.6H_2O$. *C. reinhardtii* was grown at room temperature under fluorescent lighting.

For 1 L of medium:

2X Filner's Beijernicks Solution	25 mL
1M Potassium Phosphate	1.0 mL
Trace mineral solution	5.0 mL
Tris-Base	2.42 g
Glacial Acetic Acid	1.0 mL

Adjust with ddH_2O to 1000 mL and adjust pH to 7.2, and autoclave. Store at room temperature.

Stock solutions:

2X Filner's Beijernicks Solution (500 ml)

NH_4Cl	8.0 g
$CaCl_2.2H_2O$	1.0 g
$MgSO_4.7H_2O$	2.0 g

Add deionized/distilled water to 500ml and autoclave. Store at 4°C.

Trace Mineral Solution (500 ml)

5 g disodium EDTA – dissolve in 400 ml water by heating and stirring

Neutralize to pH 6.5 with 5N NaOH

Add each of the following in order. Allow each to dissolve completely before adding the next.

$FeSO_4.7H_2O$	0.5 g
$ZnSO_4.7H_2O$	2.2 g
H_3BO_3	1.14 g
$MnCl_2.4H_2O$	0.51 g
$CuSO_4.5H_2O$	16 mg
$Na_2MoO_4.2H_2O$	73 mg
$Co(NO_3)_2.6H_2O$	19.6 mg

Bring volume up to 500 ml and autoclave. The solution should be pale gree, turning deep orange to purple upon storage.

1M Potassium Phosphate Stock (50 ml)

KH_2PO_4	6.8 g
K_2HPO_4	8.7 g

Bring volume to 50 mL with ddH_2O and autoclave.

Taub #36 medium

For 100 mL ½x Taub #36 with 0.03% proteose peptone, add 47 mL ddH2O and 3 mL 1.00 % (i.e. 1 g/100mL) Bacto™ Proteose Peptone No. 3 (Becton, Dickinson and Company, formerly Difco) to 50 mL 1x Taub #36 (see ref. [88]), and filter-sterilize.
For 1 L 1x Taub #36:

Part A (final solution)

Add 175.3 mg NaCl, 2.0 mL part B, 2.0 mL part C and 1.0 mL partG. Bring volume up to 998 mL with ddH_2O. While for long-term experiments fresh medium was prepared, this solution seemed stable without addition of parts D and F, and was used in the preparation of some simpler experiments.
For final medium, add 6.67 mL part D and 125 μL part F. Filter sterilize and store at room temperature for up to a month. Afterwards a precipitate forms.

Part B

Dissolve 12.35 g $MgSO_4.7H_2O$ in 500 mL ddH_2O and autoclave.

Part C

Dissolve 6.80 g KH_2PO_4 and 1.60 g NaOH (careful!) into 500 mL ddH_2O, adjust the pH to 7.5 and autoclave.

Part D

Dissolve 3.33 g $CaCl_2$ (anhydrous, hygroscopic) into 100 mL ddH_2O and autoclave.

Part E: no part E.

Part F

Dissolve 26.1 g EDTA into 268 mL 1M NaOH.

Add 24.9 g $FeSO_4.7H_2O$ and bring volume to 1 L with ddH_2O.

Aerate overnight in chemical hood.

Filter sterilize, and store at room temperature shielded from light.

In the dark the solution appears stable for at least a year.

Part G

Dissolve sequentially in 1 L ddH_2O:

H_3BO_3	1.854	g	
$ZnSO_4.7H_2O$	0.287	g	
$MnCl_2.4H_2O$	1.36	g	(or 1.98 g $MnCl_2$.9H2O)
$Na_2MoO4.2H_2O$	0.242	g	
$CuSO_4.5H_2O$	0.0499	g	
$Co(NO_3)_2.6H_2O$	0.291	g	

Filter sterilize and store at room temperature, shielded from light.

Chapter 5 Divergence of Population Dynamics in a Closed Ecosystem.

Abstract

The dynamics of replicate closed ecosystems under constant boundary conditions (light and temperature) are studied. Population dynamics diverge for all three species, *C. reinhardtii*, *E. coli* and *T. thermophila*, in a reproducible way, with characteristic divergence times of about 20 days for *T. thermophila* and about 40 days for the other two species. For the first 60 days, an approximately linear increase in system divergence is observed. Afterwards, there is a marked decrease in the overall speed of system divergence. It is shown that deviations after three to four weeks have long-lasting effects on these ecosystems. The results show that historical effects, as defined in Chapter 1, are important for the dynamics of such ecosystems.

Introduction

In Chapter 1 historical effects were described as long-lasting effect of random changes, be they genetic, chemical or numerical, such as by random mutation or the random timing of birth and death processes. In this chapter, I examine whether such effects leave their footprint in population density time series. Specifically, the pattern of divergence between replicate ecosystems kept under constant boundary conditions is studied. Under constant light and temperature, exogenous factors can be excluded as significant drivers of observed divergence, but a number of potential causes of divergence still remains, such as the exponential amplification of small differences in initial conditions often observed in nonlinear systems [148, 149], the occurrence of random phenotypic change (e.g., by mutation), and demographic noise, possibly amplified by species-species interactions.

Divergence between ecosystems also has important consequences for the design of experiments. Faster diverging ecosystems offer a shorter time window, or alternatively require larger numbers of replicates, for obtaining reproducible results. As discussed in Chapter 1, since selective and historical effects can't be predicted reliably, the ability to perform reproducible measurements is a necessity.

To be precise, divergence is understood as the variation between replicate ecosystems within the same experiment, prepared from the same 'master mix' and kept under the same boundary conditions. Reproducibility, on the other hand, is understood as the ability to reproduce the same results in independent experiments; with results in the present context meaning the statistical properties of a replicate set of ecosystems.

Below I describe two experiments on the dynamics of a set of replicate ecosystems started from the same initial conditions and kept under the same boundary conditions (1200 lux, 23.1 °C, see the Appendix of Chapter 4 for details). Before that, however, it is necessary to make some introductory remarks about the data analysis methods.

Analytical methods

The analysis of population density time series is limited by the lack of a proper stochastic model for our ecosystem. This ignorance influences the choice of a divergence measure, which I will address first. Then I will consider the consequences for the assignment of confidence intervals to estimated statistical quantities [148, 150].

Traditionally [150], divergence over time in dynamical systems is measured by looking at the change in distance between initially similar state vectors, here population densities. This leads to the definition of a Lyapunov exponent [148, 149], Λ,

$$\Lambda = \lim_{t\to\infty} \left\langle \log\left(\left| n_i(t) - n_j(t) \right| \right) \right\rangle , \tag{5.1}$$

with $n_i(t)$ the population density of a species in system *i* at time *t*, $n_j(t)$ its density in a system *j*, and the average performed over all pairs of trajectories considered. In practice, estimation is done in one of two ways (see [150] for a review): if the dynamics display limit cycle behavior, approximate recurrence of system states will be observed. For pairs of nearby, non-sequential observations, the rate of divergence can then be calculated directly from observations and averaged over the limit cycle. The alternative approach is to fit a sufficiently flexible phenomenological model to the observations (e.g., a neural

network), and determine the Lyapunov exponent from its Jacobian [151]. In principle, these methods can be applied to either single systems or a set of replicates.

Such Lyapunov exponents appear inadequate for two reasons. First, neither way of estimation appears appropriate for our data. Recurrences are rare, and there is no clear sign of a single fixed point, limit cycle or strange attractor to suggest divergence is driven by low-dimensional (non)linear dynamics. More importantly, the definition of the Lyapunov exponent appears inappropriate when population densities vary over orders of magnitude: we are more interested in relative density differences between ecosystems.

Instead I chose to look at $\sigma(\log n)$, the standard deviation of the logarithm of the population density for individual species, n, at each time point and $\Sigma(\log \mathbf{n})$, the covariance matrix for the set of all three species $\mathbf{n} = (n_A, n_B, n_C)$.

The metrics $\sigma(\log n)$ and $\Sigma(\log \mathbf{n})$ have important benefits[38]. First, they are invariant under proportional scaling of densities. In a simple thought experiment imagine a set of ecosystems, and dilute each of them twofold. It is a desirable feature of a divergence metric to be invariant under this operation, since the relative differences between ecosystems have not changed. The metric $\log(|n_i(t) - n_j(t)|)$ in the definition of the Lyapunov exponent would change by log(2), and $\sigma(n)$ would be affected linearly. $\sigma(\log n)$ on the other hand would not be affected. More pragmatically, the distribution of population densities n at any particular time point appears strongly positively skewed, making log(n) an appropriate variance-stabilizing transformation for statistical analysis

[38] $\sigma(\log n)$ is akin to another measure of variability, the coefficient of variation (CV), i.e. $\sigma(n)/\mu(n)$. They are equivalent when $\sigma(n) << \mu(n)$.

[130, 150]. Finally, dynamical models for population growth are quite naturally expressed in terms of log n:

$$\frac{dn}{dt} = r \cdot n \cdot f(n) \Leftrightarrow \frac{d \log n}{dt} = r \cdot g(\log n) \tag{5.2}$$

where r is a growth rate at low densities, and f and g describe density-dependent effects ($f = 1$ describes exponential growth, $f = (K - n)$ logistic growth).

Ignorance of an underlying probability distribution has consequences as well for the determination of confidence intervals on inferred statistical quantities like means and standard deviations. Absent such assumptions, I will assign confidence intervals using bootstrapping [152]. Conceptually, in bootstrapping the empirical cumulative density function (c.d.f.) is assumed to be the best available knowledge about the true c.d.f. This c.d.f. can then be computationally resampled a large number of times, providing histograms for the parameter of interest under the empirical c.d.f.. Methods like jackknifing and cross-validation schemes can be shown to be approximately equivalent to bootstrapping [152]. Unfortunately, no such technique can properly deal with properties strongly dependent on tail behavior.

Results

In the first experiment, densities in 9 ecosystems were measured on four to seven days per week for a ten week period, for 30-60 minutes each day. Over the next 130 days they were measured once every two or three weeks. The results are shown in Figure 5.1. The measurement protocol appears largely appropriate: on only three occasions not a single *T.*

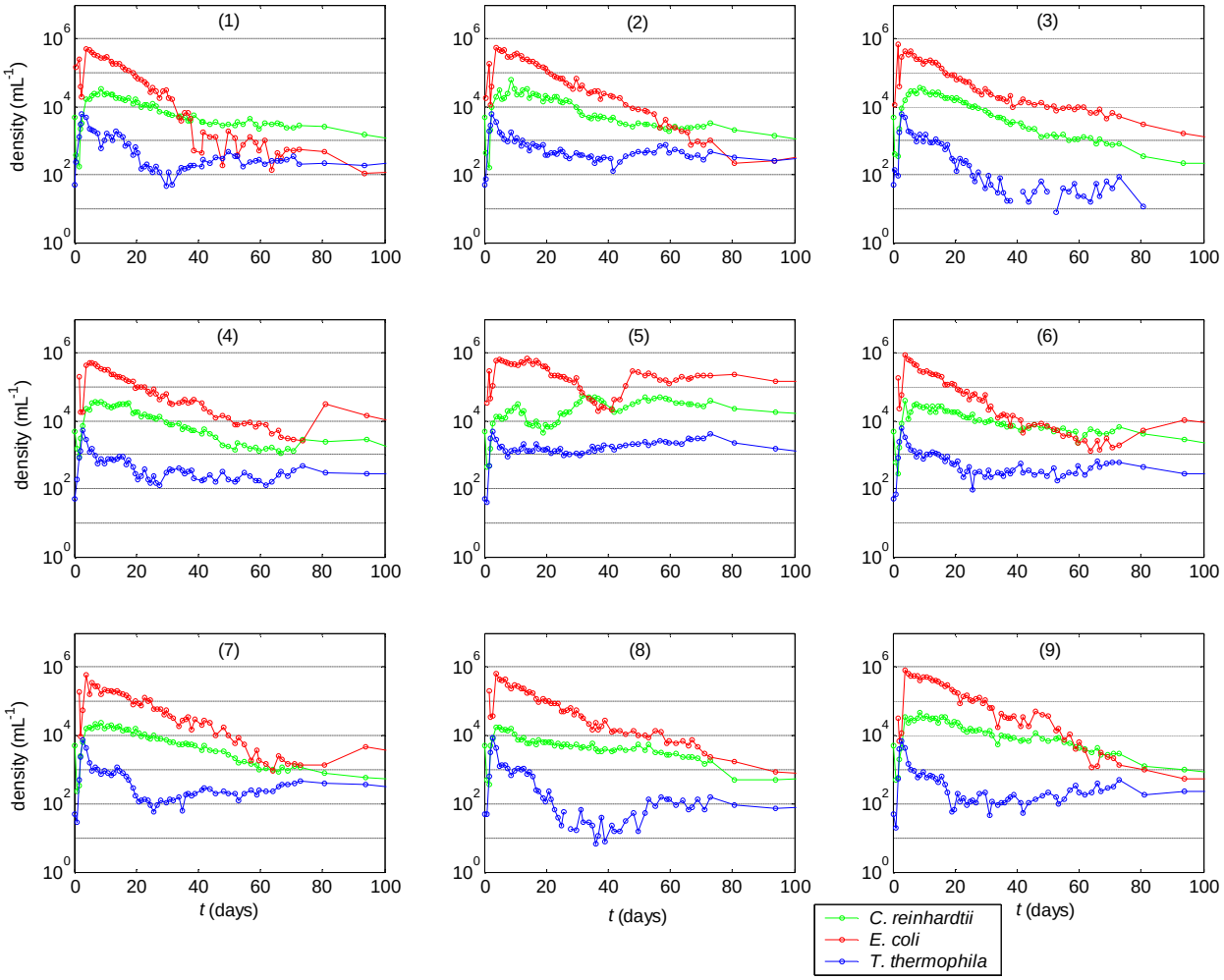


Figure 5.1: Population density trajectories for nine three-species ecosystems in experiment 1. Irregular behaviour at low densities is due to limited sampling. For *T. thermophila*, in three instants missing data points indicate no *T. thermophila* were observed during the 30-60 minute measurement for that day.

thermophila was observed[39], and measurement noise became significant at densities below 10^2 mL^{-1} for *T. thermophila* and 10^3 mL^{-1} for *C. reinhardtii* and *E. coli*[40], as can be seen by wild wiggles in estimated density at low density.

Despite gradual divergence, the dynamics in eight of nine ecosystems appeared similar. In the other ecosystem (number 5 in Figure 5.1) dynamics were markedly different. I

[39] Such counts were replaced by ½ counts when taking logarithms.
[40] About 10 counts.

tentatively attribute this to the early appearance of a *T. thermophila* clone capable of ingesting *C. reinhardtii* (see Chapter 3, Figure 3.8 for an illustration). As shown in panel 5.2 (b), a significant number of *T. thermophila* has a strong chlorophyll fluorescence signal, unlike *T. thermophila* grown by itself, which rarely even leaks through the chlorophyll emission filter. Panel (c) of Figure 5.2 shows that, unlike in the other eight ecosystems, in system 5 a considerable fraction of cells has a strong chlorophyll signal, that is, eats *C. reinhardtii*, after day 9. Paradoxically, after about 25 days this system sustains higher than average densities of *C. reinhardtii*, maybe because *T. thermophila* causes more efficient recycling of algal material.

As an initial exploration of total ecosystem divergence, examine $\Sigma(\log \mathbf{n})$, the covariance matrix of the logarithm of the three species densities over ecosystems. As discussed, this measure is invariant under dilution and should be more robust than measures based on **n**. A measure linear in the differences between ecosystems can be obtained by examining $|\Sigma(\log \mathbf{n})|^{1/6}$, which, if the species A, B and C varied independently, would reduce to $\sqrt[3]{\sigma(\log n_A)\cdot\sigma(\log n_B)\cdot\sigma(\log n_C)}$ and is the cube root of the volume element spanned by the covariance matrix in ($\log n_A$, $\log n_B$, $\log n_C$) space.

Results for this metric are shown in Figure 5.3. First, there is an initial apparent drop in the divergence between systems. Initial exponential growth is so fast, however, that the dynamics are undersampled (see also Figure 5.5). Also, over the first few days clear spatial structure in *E. coli* density is visible, forming drape-like density variations, potentially affecting density estimates from the observation volume significantly. Because of a decrease in optical density, it was not established whether this spatial structure remains.

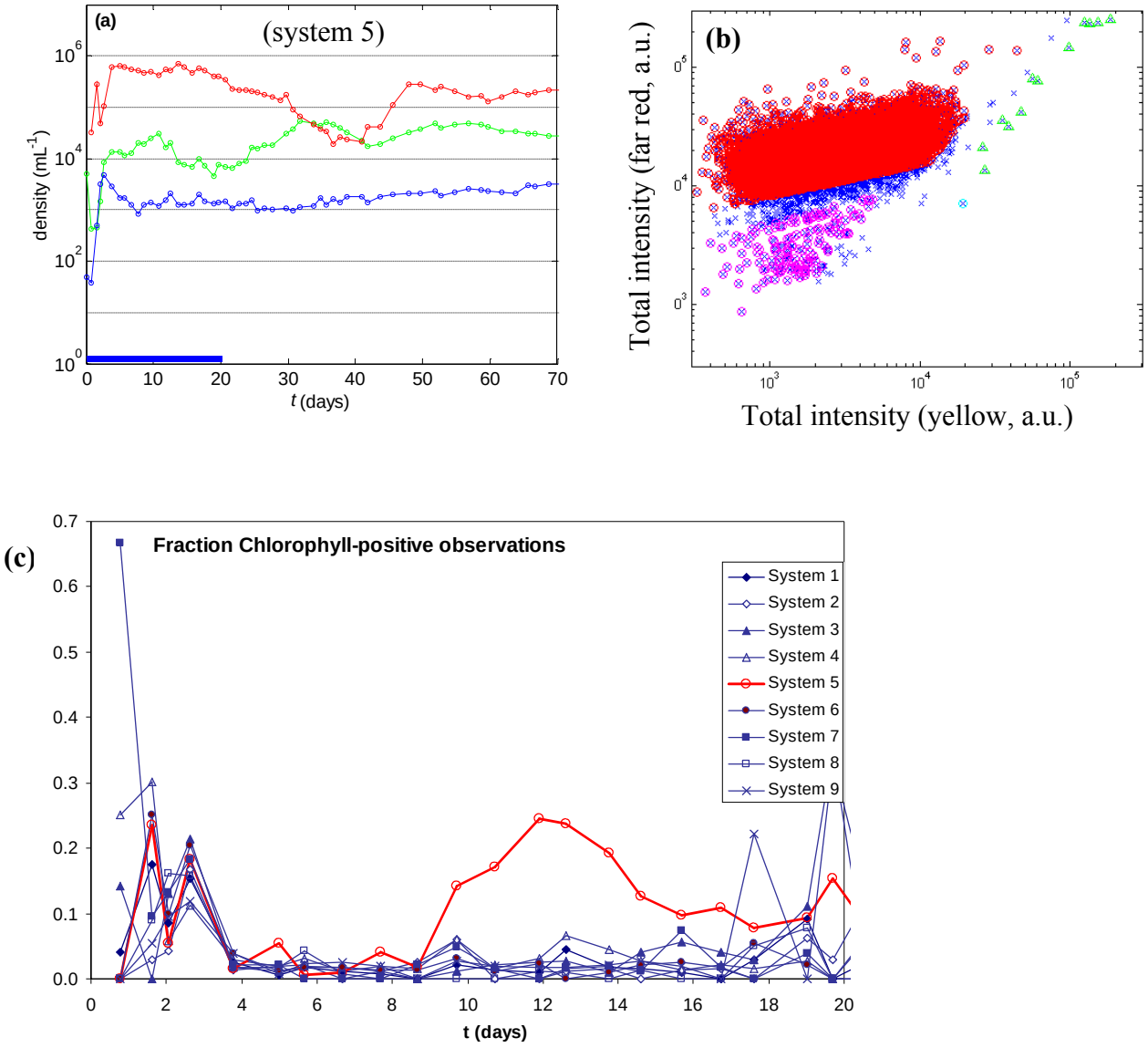


Figure 5.2: Putative explanation of different population dynamics in system (5) of experiment 1. Since both JAI cameras can be aligned, the fraction of *T. thermophila* which do have a significant chlorophyll signal (see chapter 3) can be estimated.
(a) the observed population dynamics (green: *C. reinhardtii*, red: *E. coli* and blue: *T. thermophila*; The blue bar indicates the time window shown in panel (c). (b) comparison of objects in both channels (day 35), shown in red are *C. reinhardtii*, magenta: dead or dying *C. reinhardtii*, green: inferred *T, thermophila*. Blue: unclassified objects. (c) fraction of *T. thermophila* with a significant chlorophyll fluorescence signal.

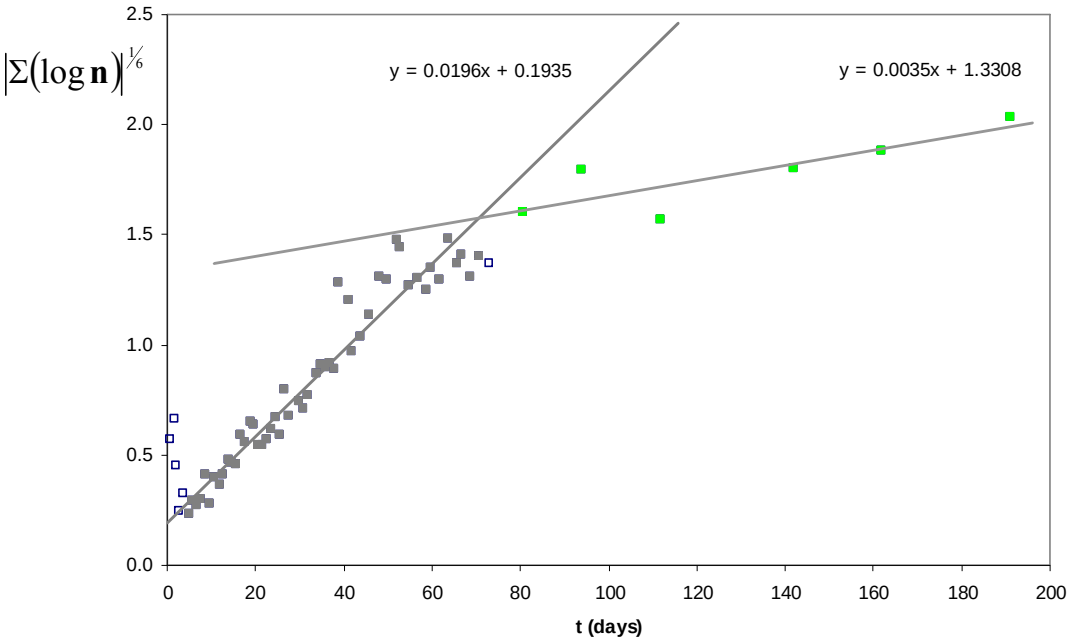


Figure 5.3: Divergence of replicate ecosystems in the first experiment as measured by $|\Sigma(\log \mathbf{n})|^{1/6}$, an aggregate measure linear in differences in log **n**. The two trend lines shown are fits over day 5-70 (grey squares) and 80-200 (green squares), respectively, and are based on unweighted linear regression. The slope of the second trend line is 5.6x smaller than for the first one. After 70 days one of the systems ((4) in Figure 5.1) was measured nearly continuously (see chapter 6 and 7), while the other systems were measured once every two or three weeks. System 4 was excluded from the analysis in this plot, without affecting the results significantly.

Over day 5 to 55 (gray squares, Figure 5.3), the divergence of the system shows a fairly linear increase, followed by much slower increase over day 60 to 200 (green squares, Figure 5.3).

A second experiment

These results raise a number of questions which were addressed in Experiment 2 (Figure 5.4). First of all, are there any fundamental reasons to expect a linear increase of total divergence? Was the decrease in the rate of divergence after day 60 really due to a

decrease in measurement frequency? In other words, do the measurements perturb the systems so much, and so unequally, that ecosystem divergence is directly driven by the measurements? And, along these lines, would measurements affect ecosystem dynamics themselves (i.e., the mean trajectories observed)? Ideally, measurements have no significant impact on the systems being measured.

In the second experiment, fifty replicate ecosystems were divided into five groups of ten[41], and each group subjected to a different measurement frequency, at 2, 1, 0.5, 0.25 and 0.125 week^{-1} for 90 to 100 days. For comparison, in the first experiment densities were measured in each ecosystem four to seven times per week. The boundary and initial conditions were the same, within experimental precision, as in the first experiment. Examples of observed dynamics are given in the Figure 5.4.

To start, the mean dynamics of the ecosystems in the first experiment were reproduced almost completely in Experiment 2 over 100 days (Figure 5.5). In this figure, the mean logarithmic density, log n, of each species is shown for different measurement frequencies, with 90% confidence intervals of the mean of log n estimated by bootstrapping (see Appendix 5A for the bootstrap method used). The only exceptions to complete reproducibility were for *C. reinhardtii* in experiment 1 over day 10 to 25, with mean densities lower than for any of the measurement frequencies in experiment 2; and for *E. coli* measured at ¼ week^{-1}, which had significantly higher densities than any of the other sets of ecosystems (red circles in the second panel).

Secondly, the general patterns of divergence for each species were the same in each treatment (Figure 5.6, note the large uncertainty in these estimates). *T. thermophila* typically showed a large increase in divergence between systems around day 20, roughly

[41] One or two cuvettes per group were excluded over the time of the experiment because of leakage.

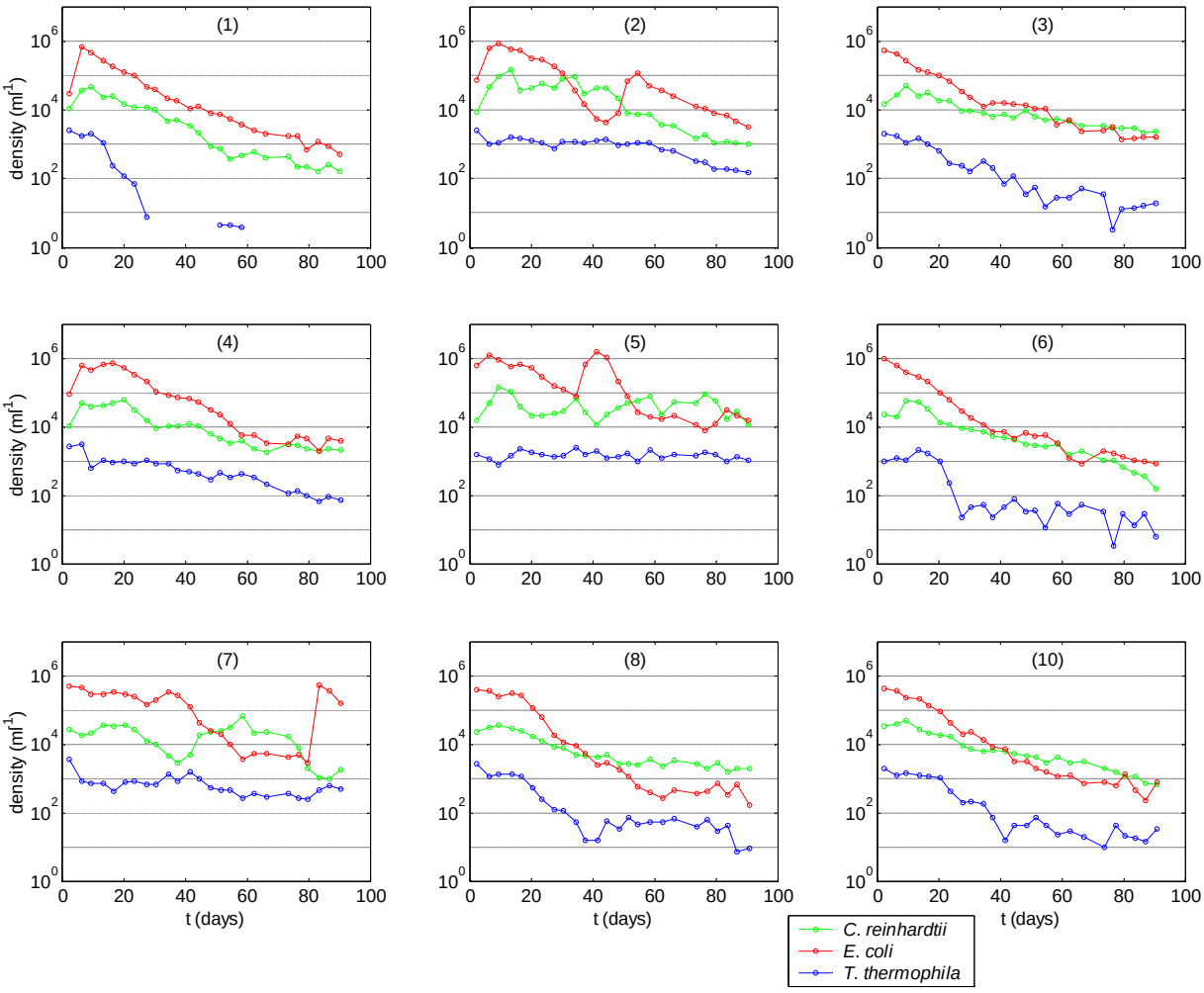


Figure 5.4: Population dynamics in Experiment 2, for nine ecosystems measured twice a week. Note for exampe the remarkable increase in *E. coli* density in sample 7 after about 80 days.

coincident with a one order of magnitude drop in (geometric) mean density, suggesting differential ability of populations at that point to survive under more difficult conditions. This increase in divergence is followed by slower increases in *E. coli* and *C. reinhardtii*. The only exception from the general pattern appears to be that *E. coli* densities diverge less in the two least-sampled sets of ecosystems (panels 5 and 6 of Figure 5.6), but given

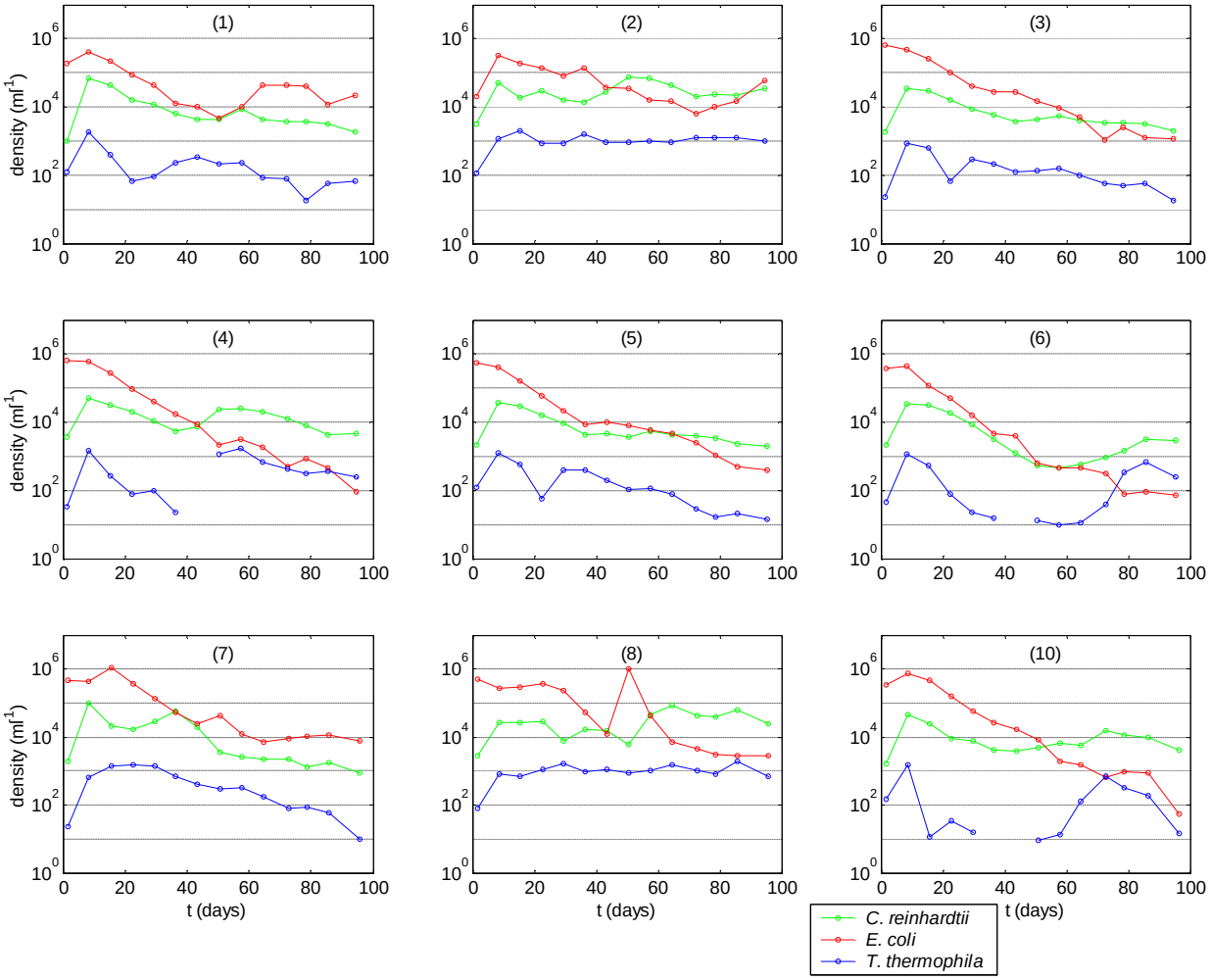


Figure 5.4 (continued): Population dynamics in Experiment 2: nine ecosystems measured once a week.

the large and asymmetric bootstrap intervals of the other data sets, it is possible that these two sets lacked more atypical ecosystems by chance.

Total system divergence for the measurement protocols in both experiments are compared in Figure 5.7. Again the patterns are largely the same for measurement frequencies down to ½ week^{-1}.

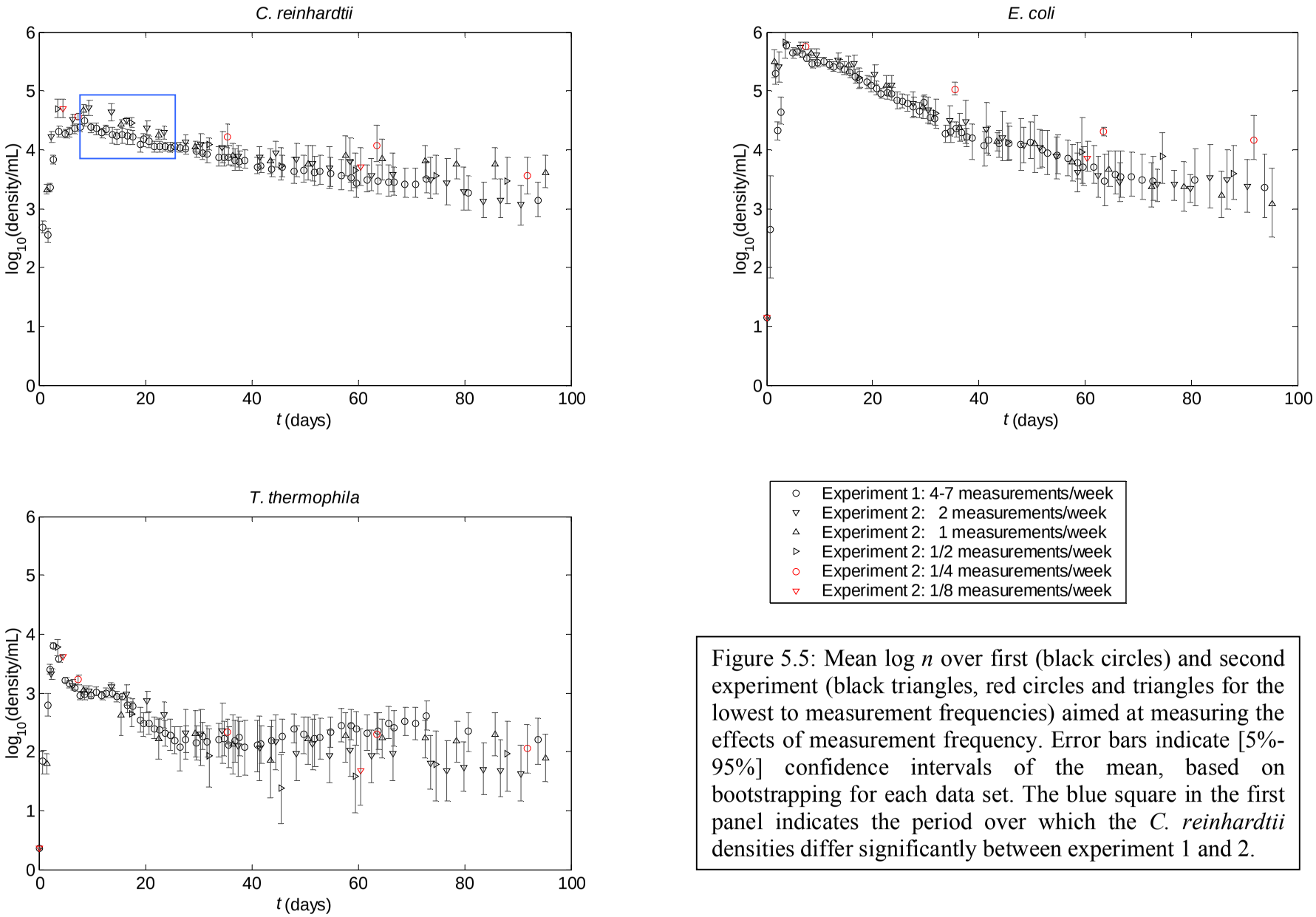


Figure 5.5: Mean log n over first (black circles) and second experiment (black triangles, red circles and triangles for the lowest to measurement frequencies) aimed at measuring the effects of measurement frequency. Error bars indicate [5%-95%] confidence intervals of the mean, based on bootstrapping for each data set. The blue square in the first panel indicates the period over which the *C. reinhardtii* densities differ significantly between experiment 1 and 2.

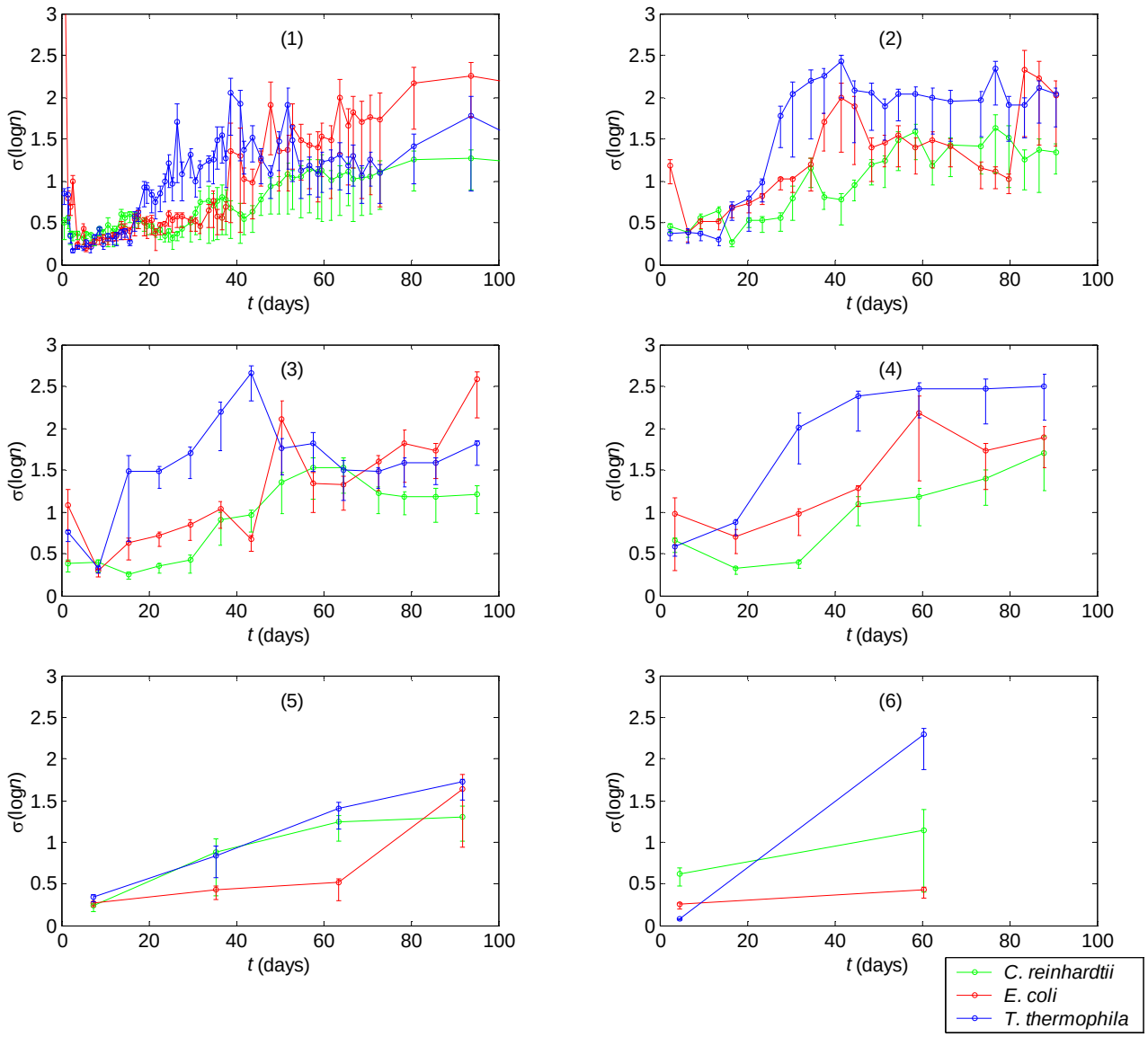


Figure 5.6: Divergence between populations of individual species in sets of replicate ecosystems measured at a different frequencies (8 or 9 replicates per set). Experiment 1: panel (1), 4-7 week^{-1}. Experiment 2: panel (2-6): (2) 2 week^{-1}, (3) 1 week^{-1}, (4) ½ week^{-1}, (5) ¼ week^{-1}, (6) ⅛ week^{-1}. Species: *C. reinhardtii*, green; *E. coli*, red; *T. thermophila*, blue. Error bars indicate 50% confidence intervals based on bootstrapping for clarity (larger confidence intervals would take up most of the range of each graph).

I now consider the questions raised above after the first experiment. First, it remains hard to distinguish between a linear and sigmoidal increase over day 5 to 55 (red data points, Figure 5.7). Yet, it appears that the different patterns of increase of divergence for individual species lead to a more linear increase of total ecosystem divergence. To distinguish this, though, a larger number of systems would need to be measured. Secondly, a marked slowdown of the rate of divergence, *also* occurs when measurement continue (Figure 5.7, panels 2 to 4, after about 60 days), and total divergence appears to reach similar levels (~1.4) under all measurement frequencies (for the once per eight weeks group I would have had to measure longer).

Next, the rate of divergence for each species is shown in Figure 5.8. For *C. reinhardtii* and *T. thermophila*, measurement frequency does not seem to affect the rate or pattern of divergence significantly, that is, a horizontal fit would be consistent with the data within the error bars. For *E. coli* there may be an effect, but chance cannot be excluded given the large asymmetric confidence intervals seen at higher measurement frequencies[42]. This asymmetry indicates that total divergence estimates for each measurement frequency depend fairly strongly on the presence of one or two atypical ecosystems.

Persistence of early differences between ecosystems

Finally, I address the whether differences between ecosystems tend to be preserved. Specifically, do individual ecosystems tend to relax back to the mean behavior, or once apart do they stay apart? In an analogy, consider the case of diffusing particles. Within a harmonic potential well, trajectories of particles starting at the same point will diverge,

[42] In addition, the error bars at the lowest two measurement frequencies may be underestimated, because the effect of having few data points on the error in the divergence rate is not included.

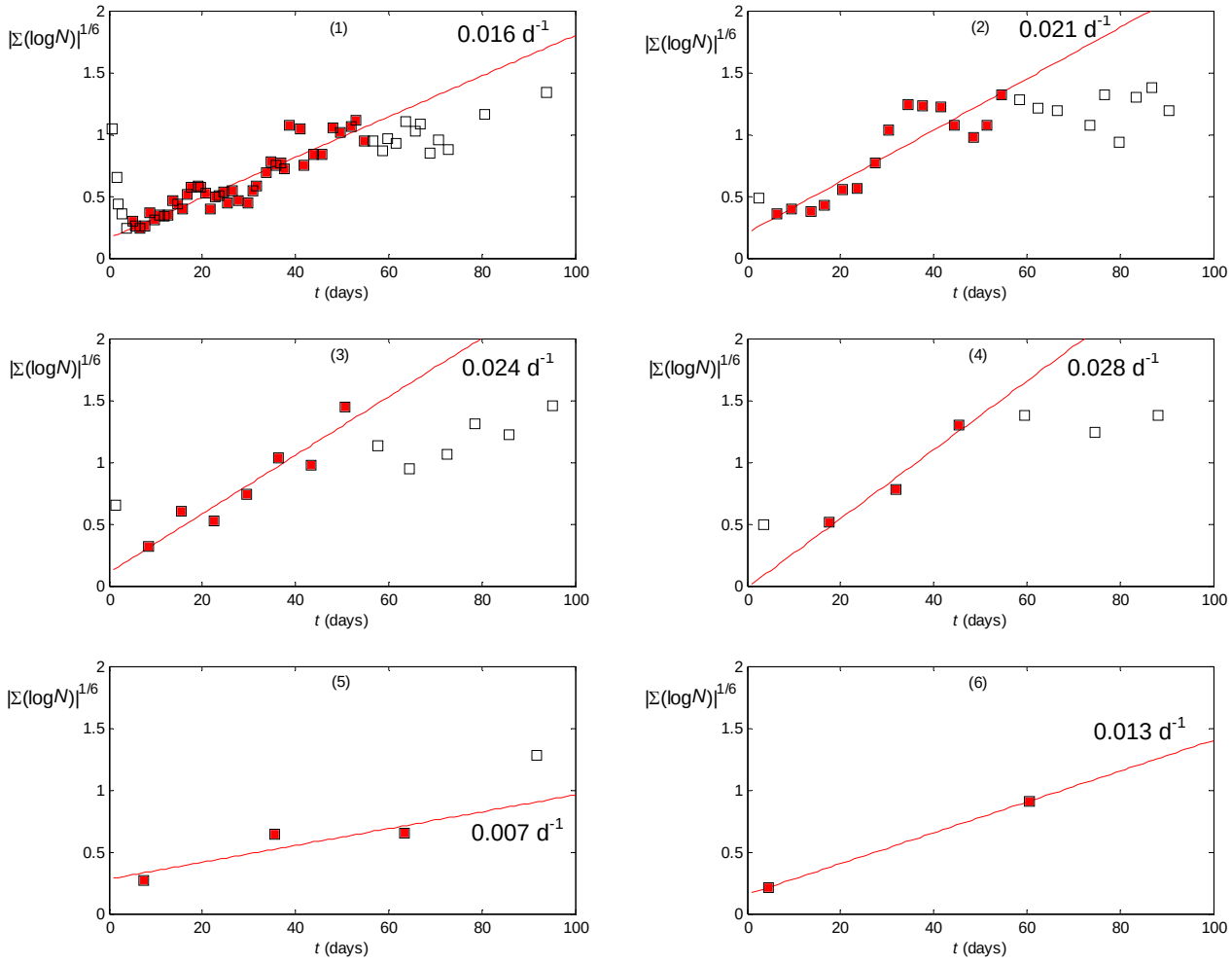


Figure 5.7: Linear dimension of the state space element spanned by the covariance matrix, Σ, of log **n** = log(n_A, n_B n_C): it is the root (variance to standard deviation) of the one-third power (for a three-dimensional state space) of the determinant of the covariance matrix. Filled red squares indicate the time points included in the estimate of divergence rate: over day 5-55, except for the last two panels, where the first data point after day 55 is included. Red lines are linear regression fits to the data points indicated in red. The slope of each regression line is shown next to it.
Measurement frequencies: (1) experiment 1, 4-7 week^{-1}, experiment 2: (2) 2 week^{-1}, (3) 1 week^{-1}, (4) ½ week^{-1}, (5) ¼ week^{-1}, (6) ⅛ week^{-1}.

but individual particles tend to revisit the mean position many times. However, in this analogy, particles diffusing on a flat or rugged landscape do not necessarily tend back to

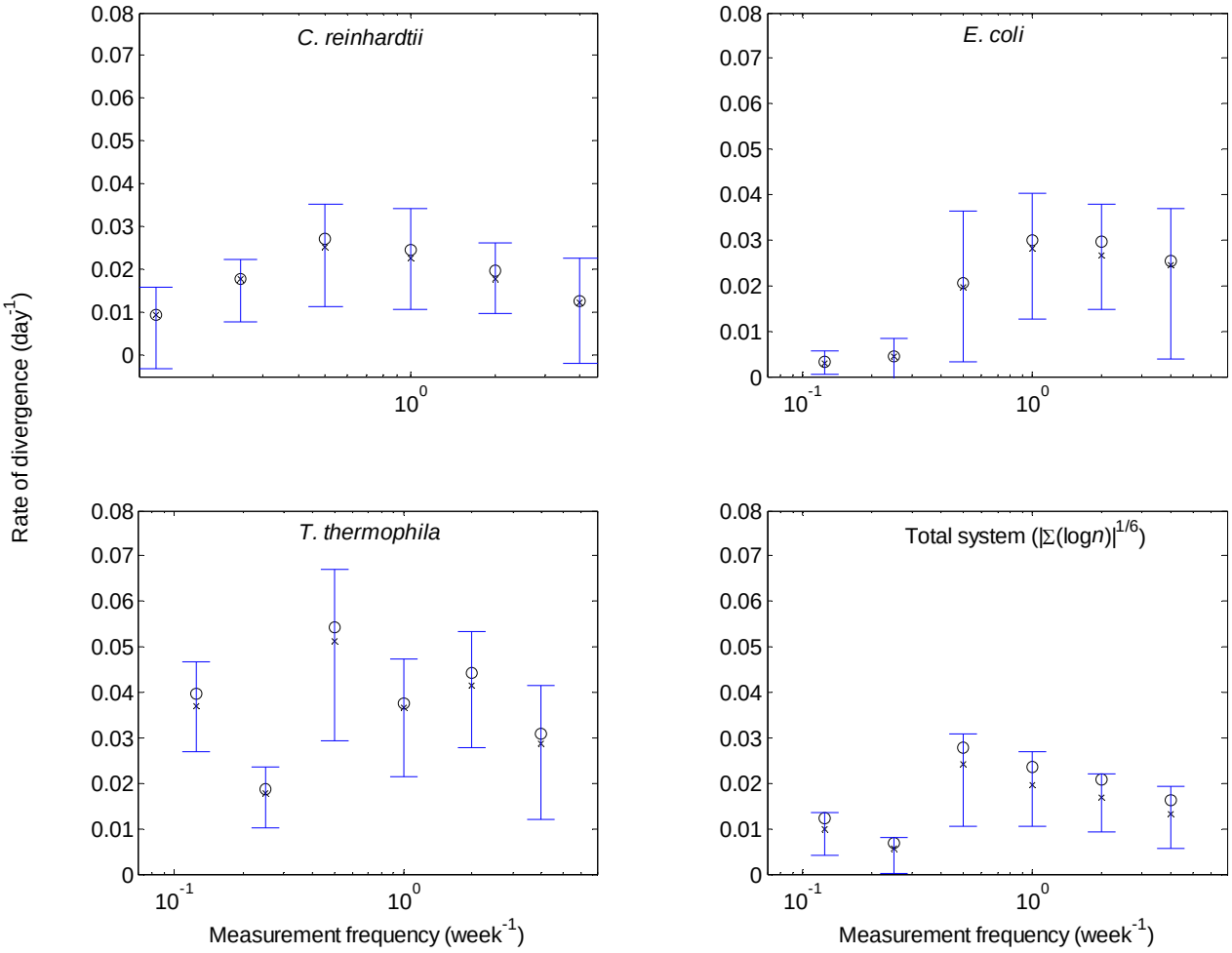


Figure 5.8: dependence of divergence rate on measurement frequency for each species individually, and for all three species combined. Divergence rate is measured as the slope of increase of $\sigma(\log n)$ for individual species and $|\Sigma(\log \mathbf{n})|^{1/6}$ as an aggregate measure over day 5-55, or day 5-65 for the lowest two measurement frequencies (red squares in Figure 5.7). Error bars indicate 90% confidence intervals based on bootstrapping. The uncertainty of the regression slopes themselves was not included in the error bars (possibly increasing the error at low sampling frequencies). Circles are the slopes fit to the full data sets; crosses are median bootstrap estimates.

the mean after a deviation. Instead, the effects of individual fluctuations can be frozen. Either scenario can be consistent with the increasing divergence observed.

The mutual information *I* can be considered a robust analog of the correlation function (while sacrificing information on the sign of the correlation). Specifically, I examine the mutual information between species densities at different points in time along population density trajectories [153]. The ensemble will be the set of 27 ecosystems from experiment 1 and 2 that were sampled at least once a week. From the sets measured at higher frequencies 1 week^{-1} time series are obtained by subsampling (with at most one day error). There are advantages to using the mutual information rather than autocorrelation functions (ACFs) and crosscorrelation functions (CCFs). First, the mutual information is invariant under any monotonic transformation applied to both its arguments: $I(n_a(t), n_b(t+\tau)) = I(f(n_a(t)), f(n_b(t+\tau)))$ for any monotonic transformation *f*, with *a* and *b* species indices[43]. Secondly, the mutual information is much less sensitive to outliers than the ACF. When *f* is such that $f(n_a(t))$ is normally distributed, the mutual information is closely related to the ACF or CCF for $f(n_a(t))$.[44]

As shown in Figure 5.9, the densities in the ecosystems three weeks after closure are predictive of the trajectories until at least 7 weeks later. A corresponding graph using the correlation coefficient rather than mutual information is shown in Figure 5.10. To highlight this, slices forecasting from three weeks and four weeks after closure are shown in Figure 5.11. Deviations from the mean appear to have long persistence for *T. thermophila* and *E. coli* (>6 weeks in the left panel of Figure 5.11, ~5 weeks in the right

[43] So $I(n_a(t), n_b(t+\tau)) = I(\log n_a(t), \log n_b(t+\tau))$.

[44] If *f* is such that *f*(*n*) is normally distributed, $I(f(n_a(t)), f(n_b(t+\tau))) = -\frac{1}{2}\log(1-\rho^2_{t,t+\tau}) \approx \frac{1}{2}\rho^2_{t,t+\tau}$ for $\rho << 1$, with $\rho_{t,t+\tau} = \langle f(n_a(t)) \cdot f(n_b(t+\tau)) \rangle - \langle f(n_a) \rangle \langle f(n_b) \rangle$ the traditional ACF (*a* = *b*) or CCF (*a* ≠ *b*).

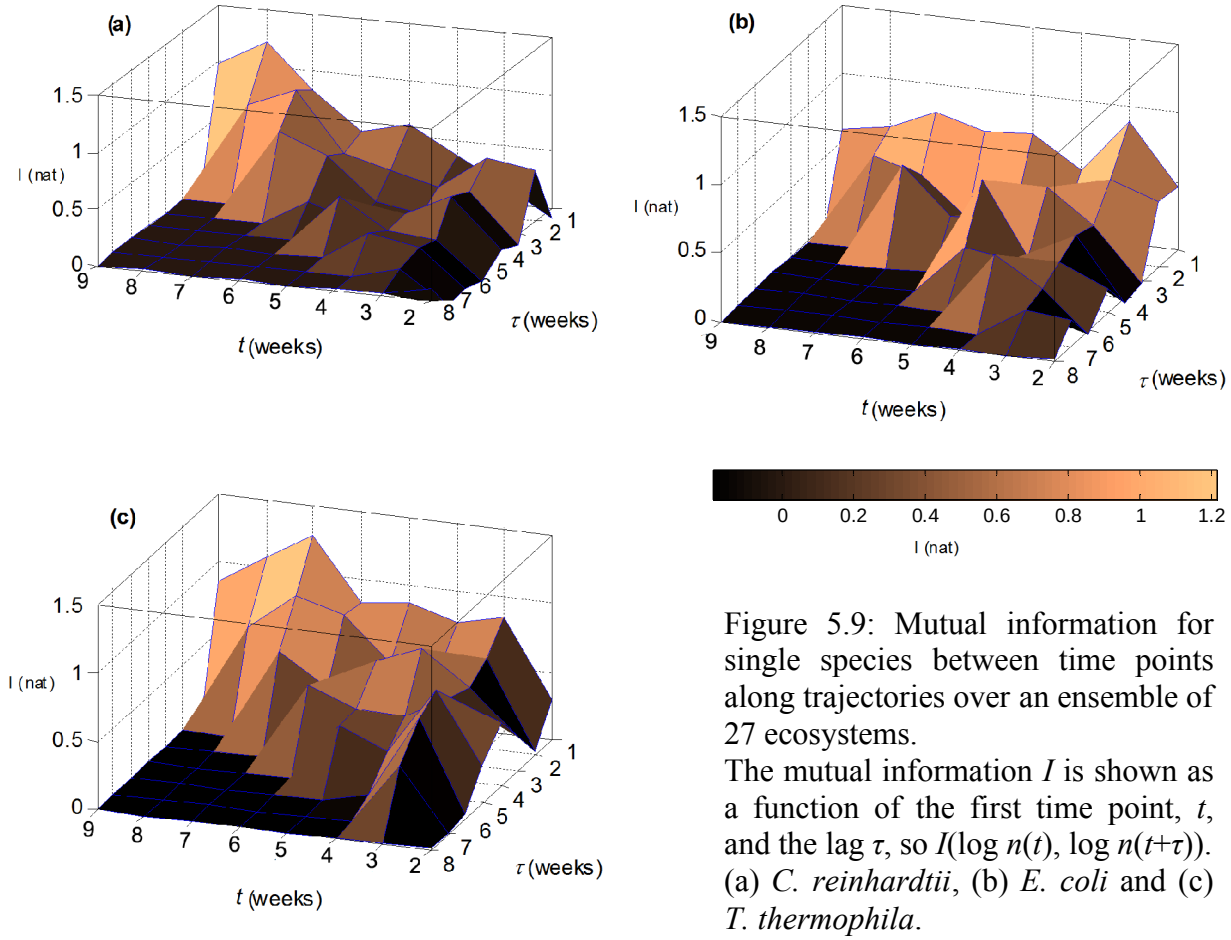


Figure 5.9: Mutual information for single species between time points along trajectories over an ensemble of 27 ecosystems.
The mutual information I is shown as a function of the first time point, t, and the lag τ, so $I(\log n(t), \log n(t+\tau))$. (a) *C. reinhardtii*, (b) *E. coli* and (c) *T. thermophila*.

panel), but shorter for *C. reinhardtii* (2 to 4 weeks). Weaker persistent correlation is also seen for species pairs (Figure 5.12), suggesting slightly different interspecies interaction patterns are established in different replicates.

The time scales on which these deviations from the mean relax are much longer than one would expect for the relaxation of demographic fluctuations (approximately the time the inverse gross growth rate (or birth rate), thought to be on the order of one day).

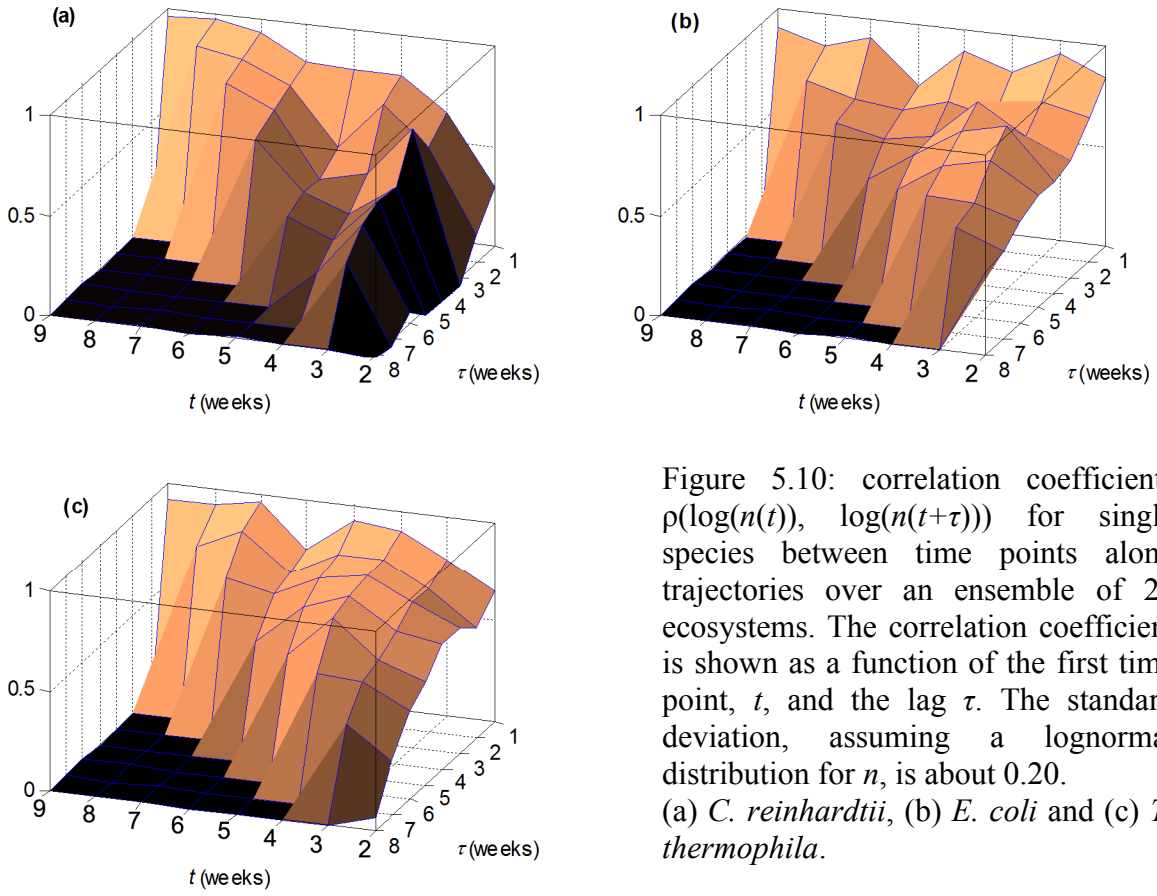


Figure 5.10: correlation coefficients $\rho(\log(n(t)), \log(n(t+\tau)))$ for single species between time points along trajectories over an ensemble of 27 ecosystems. The correlation coefficient is shown as a function of the first time point, t, and the lag τ. The standard deviation, assuming a lognormal distribution for n, is about 0.20.
(a) *C. reinhardtii*, (b) *E. coli* and (c) *T. thermophila*.

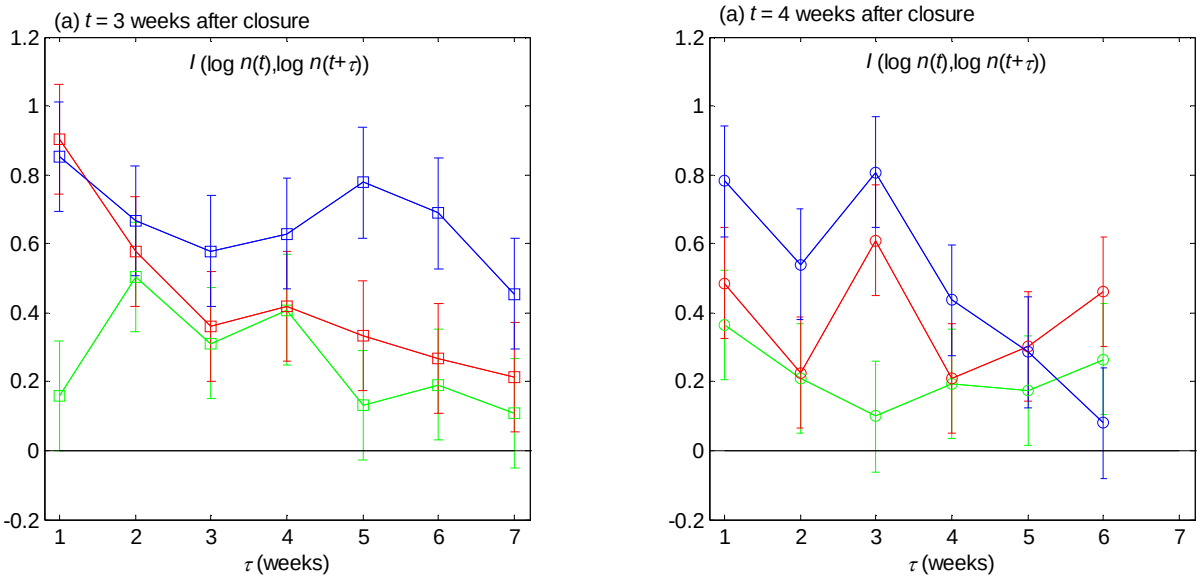


Figure 5.11: Mutual information between different time points. At each initial time point t and lag τ, we estimated $I(\log n(t), \log n(t+\tau))$ over a set of 27 ecosystems which were measured at least once per week (algorithm described in ref. [4]). (a) t = 21 days after closure, (b) t = 28 days after closure. Green: *C. reinhardtii*, red: *E. coli*, blue: *T. thermophila*. The error bars are based on data shuffled with respect to replicate index and are consistent with estimates given in [4].

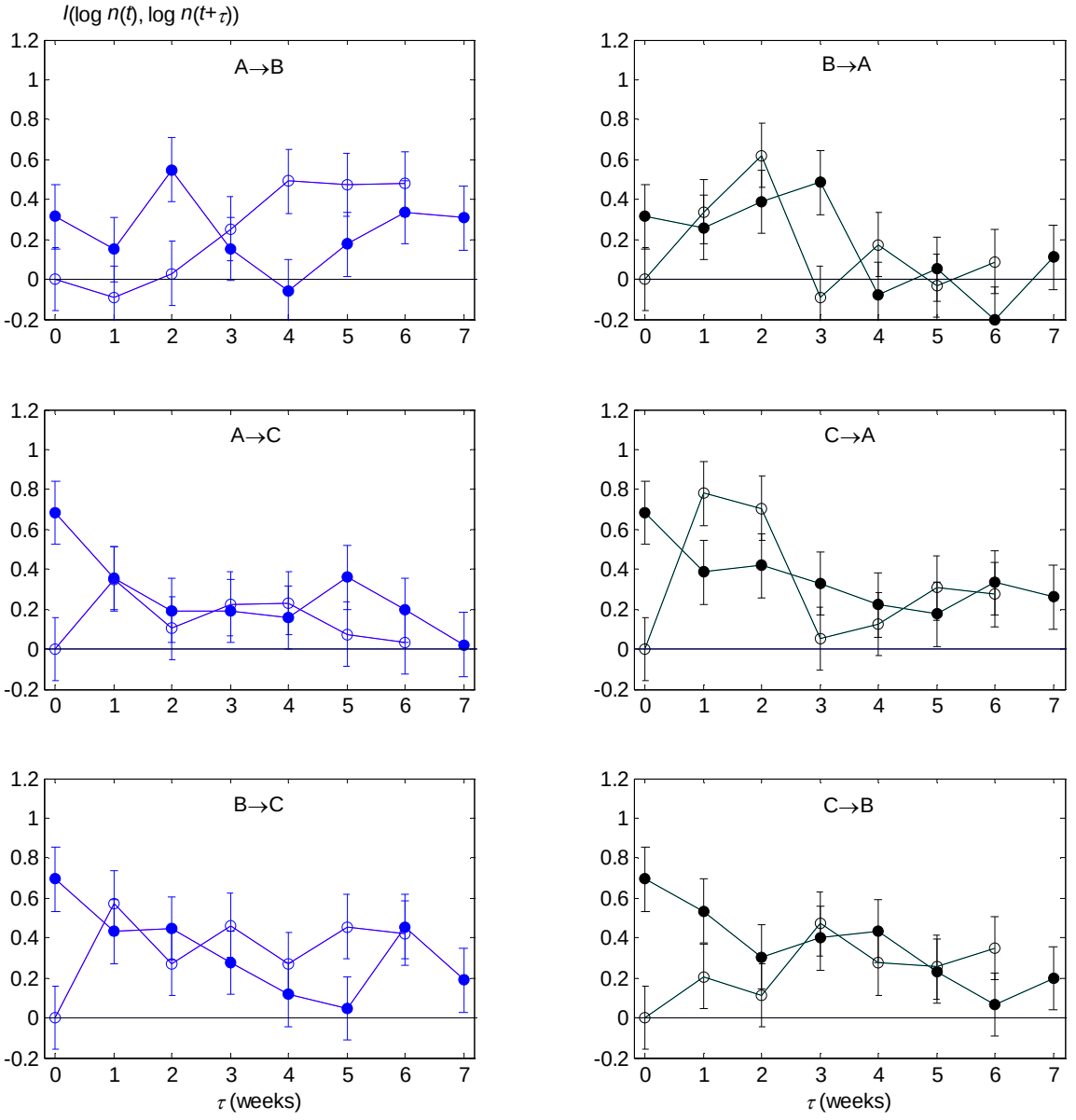


Figure 5.12: Mutual information across species between different time points. At each initial time point t and lag τ, $I\left(\log n_a(t), \log n_b(t+\tau)\right)$ was estimated as for Figure 5.11 [4], except $a \neq b$, that is, for species pairs. Open circles: t = 21 days after closure, filled circles t = 28 days after closure.
Notation in each graph is as follows: A→B means: $I\left(\log n_A(t), \log n_B(t+\tau)\right)$, that is we consider the correlation of early deviations from the mean in A (*C. reinhardtii*) with later deviations in B (*E. coli*). C: *T. thermophila*. The error bars are again based on shuffled data

Conclusions

First of all, these experiments show the utility of both closed ecosystems and the measurement method developed in the previous chapter. The ecosystems described in this chapter allow for long time series to be acquired. In fact, the ecosystems of experiment 1 were opened after 200 days (including two controls which were not measured), and all three species had survived in nine out of eleven systems (82%). The only two extinctions were of *E. coli*. The apparatus developed in the previous chapter was operated without significant problems. In addition, comparison between the two experiments in this chapter shows that the tight control of boundary and initial conditions allows for nearly complete reproducibility of mean population densities over at least two months.

Population densities varied over orders of magnitude, over time and over replicates, with a few ecosystems showing patterns strongly deviating from the dominant trends. To address this, a robust metric for divergence was introduced, based on log *n* rather than *n* (*n* being the density). The patterns of divergence of individual species could also be reproduced, within the uncertainty of the results.

Remarkably, the total system divergence, as measured by a metric linear in fluctuations, $\left|\Sigma\left(\log \mathbf{n}\right)\right|^{1/6}$, increased fairly linearly over time up to about 55 days. This linear increase may prove to be a more general feature of ecosystems. After this period, the increase in divergence slows down markedly. The observed divergence was shown to be largely independent of measurement frequency, which is important since the measurements were probably the most heterogeneous exogenous factor present. In order to extend these results to other ecosystems, however, two kinds of information would be needed: about the gross growth rates of the species present (see Chapter 8) and about the scaling of

divergence with ecosystem size. For example, a decrease in gross growth rates after two months might explain the observed decrease in divergence between replicates ecosystems at that point.

Finally, I examined the divergence between ecosystems in more detail and demonstrated that differences between ecosystems as early as 20 days after closure provided information about differences between ecosystems up to 6 weeks later (Figures 5.9 and 5.10), with the strongest memory seen for *E. coli* and *T. thermophila* (Figure 5.11). Results for the correlation between different species over periods of weeks were much weaker (Figure 5.12). Together with observations in Chapter 3 on the phenotypic divergence within replicate subsystems (A, AB, …), these results strongly suggest that initial phenotypic change can push ecosystems into different futures.

Discussion

While in this chapter I characterized divergence, it is still unclear how much different factors contribute to it. One approach would be to exaggerate possible causes in a controlled manner and extrapolate back to zero variability. For example, there may still have been some variability in boundary conditions between ecosystems. While it may not be possible to reduce this variability much more, boundary conditions can easily be made more variable in a controlled way (for example, by varying light or temperature randomly over time) and the scaling of divergence with this variability can be measured in the same way the dependence of divergence on measurement frequency was measured in this chapter. Likewise, additional genetic noise can be introduced by additional illumination

with UV (to increase mutation rate) and the initial numbers of each species can be controlled accurately with FACS[45] (S. Mazel, pers. comm.).

Tighter control of boundary and initial conditions does not imply less divergence, however. Imhof and Schlötterer [154] for example showed that *E. coli* microcosms with a more diverse inoculum (of previously adapted strains) had a more predictable final composition. Likewise, Massin and Gonzalez showed that occasional shaking of *Pseudomonas fluorescens* microcosms actually decreased the rate of phenotypic divergence as judged by colony morphology [155].

Finally, the results on ecosystem divergence and reproducibility also inform our design of experiments. We saw that about 10-20% of ecosystems deviated substantially from the dominant trends ('atypical' population density trajectories), suggesting at least ~30 ecosystems are required to capture some of the 'long tail' of population density trajectories (expect 3-6 atypical trajectories). Secondly, the longer the measurement period over which reproducible results are needed, the more replicates are needed. Alternatively, the period over which ecosystems can be considered to be the same is given by Figure 5.7. The period for which relative fluctuations are less than twofold ($\left|\Sigma\left(\log \mathbf{n}\right)\right|^{1/6} < \log 2$) is about 20-25 days for *T. thermophila*, and 40 days for *C. reinhardtii* and *E. coli* (Figure 5.6). In the next chapter I will examine continuous measurements under constant boundary conditions. This allows us to see whether any significant components of ecosystem dynamics are overlooked by performing measurements at most once per day.

[45] Fluorescence-assisted cell sorting.

Appendix 5A.1: Bootstrapping

The principle of bootstrapping was explained in the text, and has been described in the literature [152]. Briefly, confidence intervals on the mean were determined by resampling the set of observed densities for a species on a single measurement day (8 or 9 samples for each set) 1000 times with replacement. Each bootstrap sample consisted of the same number of measurements as the original set (i.e. 8 or 9). This generated a set of 1000 inferred means given the empirical cumulative density distribution (c.d.f.), providing a confidence interval under the empirical c.d.f..

Confidence intervals on $\sigma(\log n_a(t))$ for a given species *a* and sampling day *t* were determined analogously. In the determination of confidence intervals for $\Sigma(\log \mathbf{n}(t))$, for the set of three species densities $\mathbf{n} = (n_a, n_b, n_c)$ resampling was over ecosystems. In other words, the joint c.d.f. was used rather than considering each species independently. Likewise, for the determination of confidence intervals on divergence rates, entire ecosystems were considered as replicates from which was sampled, rather than sampling independently from different species or sampling days.

All operations were performed in Matlab 6.1 (Mathworks, Inc.) using custom scripts, rather than the included Matlab bootstrap functions to increase transparency of the applied methods.

Chapter 6

Population Dynamics in a Closed Ecosystem under Constant Boundary Conditions

Abstract

A species' population density has important consequences for long-term survival, as it affects the ability to generate new phenotypic variants, risk of extinction and interactions with other species. Population density time series were obtained by continuous measurement of closed ecosystems under constant illumination and temperature. First a model is developed to describe fluctuations due to measurement noise alone. The model is shown to compare well to data.

Two means of analysis of these data sets are compared. Classical Fourier power spectra show that the data are nonstationary, with similar scaling of power with frequency observed for all three species. As shown, correlation functions are hard to estimate for such nonstationary data. As an alternative, wavelet analysis is explored. Wavelets are localized probe functions which are trend-insensitive. It is shown that fluctuations in the different species densities are correlated. The results raise questions about the underlying causes of the observed fluctuations. It also appears likely that there is no clean separation of ecological and evolutionary timescales, posing a challenge for the determination of selective constraints.

Introduction

Species densities are important components of the fluctuating environment any strain or species faces, both those of its own and those of others species, and likely result in additional fluctuations in chemical and physical variables.

The expected growth rate and risk of extinction of a strain are evaluated over a probability distribution of the environments a population or lineage of organisms may face. Such fitness calculations play a role for example in the prediction of optimal phenotypes, as in the theory of foraging. In our model system, the optimal foraging strategy for *T. thermophila* feeding on *E. coli* and/or *C. reinhardtii* likely depends on the spatiotemporal pattern of their population densities. In another example, Kussell and Leibler [19] compared induction[46] and random switching[47] as means of changing phenotype in fluctuating environments. They showed that if the environment fluctuates sufficiently slowly random switching can provide a larger population growth rate than induction, and that the optimal random switching rates match the corresponding rates of transitions between different environmental conditions. In other words, the pattern of environmental fluctuations can have a significant effect on which adaptations convey the most fitness. Few data have been recorded, however, on such fluctuations. In this chapter, then, fluctuations in species densities are studied under constant boundary conditions. In the next chapter, the effect of an oscillating boundary condition is examined.

Dynamics in two ecosystems described in the previous chapter were measured continuously for several weeks, providing time series with high temporal resolution.

[46] A direct change in phenotype (< generation time) when sensing a change in environment.
[47] A heritable change in phenotype, regardless of the underlying molecular mechanism, not caused by a change in environment.

First, I will introduce a null model for observed fluctuations in our ecosystem. That is, a model describing the fluctuations observed when population densities remain constant. Individual organisms can remain in the observation volume for various lengths of time, increasing observed fluctuations over uncorrelated counting noise. I then proceed to study the fluctuations in population densities under constant light and temperature.

Data sets

The continuous measurements described in this chapter were performed on two ecosystems from experiments described in Chapter 5. After both of the experiments in that chapter a single cuvette was left in the setup and measured at a frequency of 1.1 Hz (experiment 1) or 1.25 Hz (experiment 2). The measurement history of these ecosystems is shown in the appendix (Figure 6A.3 & 6A.4). The segments analyzed in this chapter are shown in Figures 6.1 and 6.2, respectively, and were acquired starting at day 142 and day 118 after closure, respectively.

Although not considered in the analysis below, I have also included the time course of the number of chlorophyll containing *T. thermophila* in Figure 6.2. This time course shows distinct features which appear to correlate with *C. reinhardtii* densities, suggesting that predation by *T. thermophila* on *C. reinhardtii* can be a significant factor. Ultimately, we would like to be able to infer such fluctuations in phenotypic composition of populations (as in panel (b)) from observed densities, as in panel (a) of Figure 6.2.

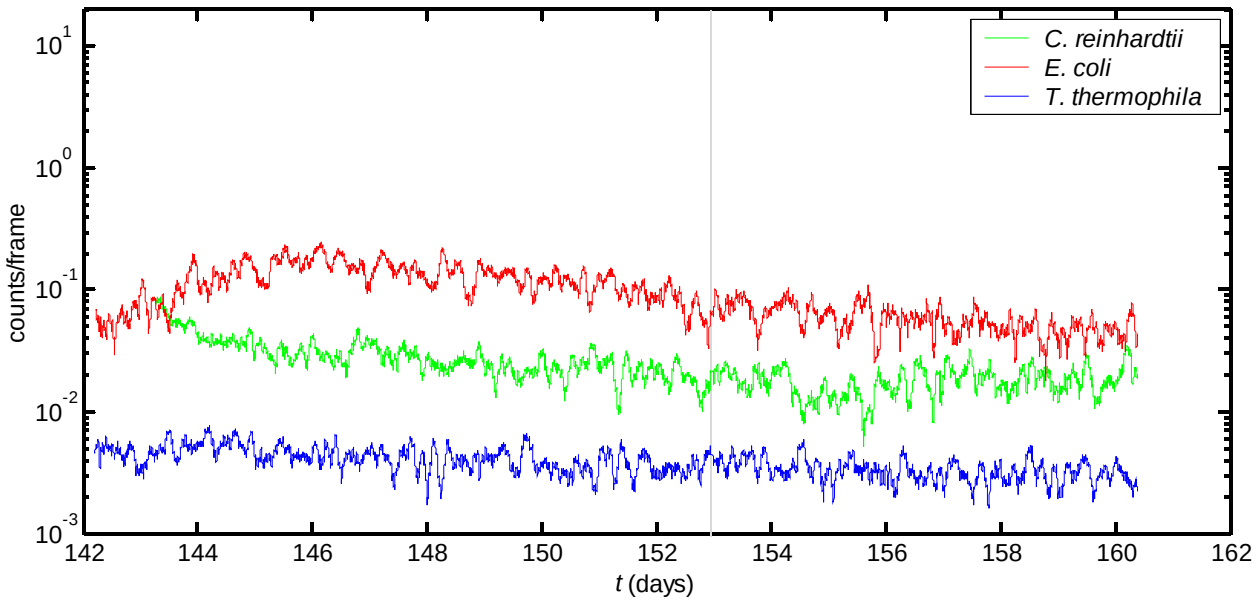


Figure 6.1 Data set 1: Population density time series for ecosystem 4 of experiment 1 (see Figure 5.1). After conclusion of ten weeks of daily measurements, nearly continuous measurements were made for another six months (see Figure 6A.5). Shown is the data segment analyzed in this chapter. Counts for *E. coli* were multiplied by 5 to improve image clarity. The gray line at t = 153 days after closure represents a brief interruption in data acquisition.

A null model for observed fluctuations

In order to evaluate the significance of any fluctuations observed, their probabilities under only measurement noise need to be known. The quantification of this measurement noise is complicated by the serial correlation of observations: individual organisms can dwell in the observation volume for various amounts of time. The probability of organisms staying in the observation volume for various lengths of time can be evaluated from time series of observed counts instead of from the tracking of individuals entering and leaving the observation volume. That is, while the continued presence in the volume of individual organisms is never certain, the observed transition rates between counts, Q, can be used:

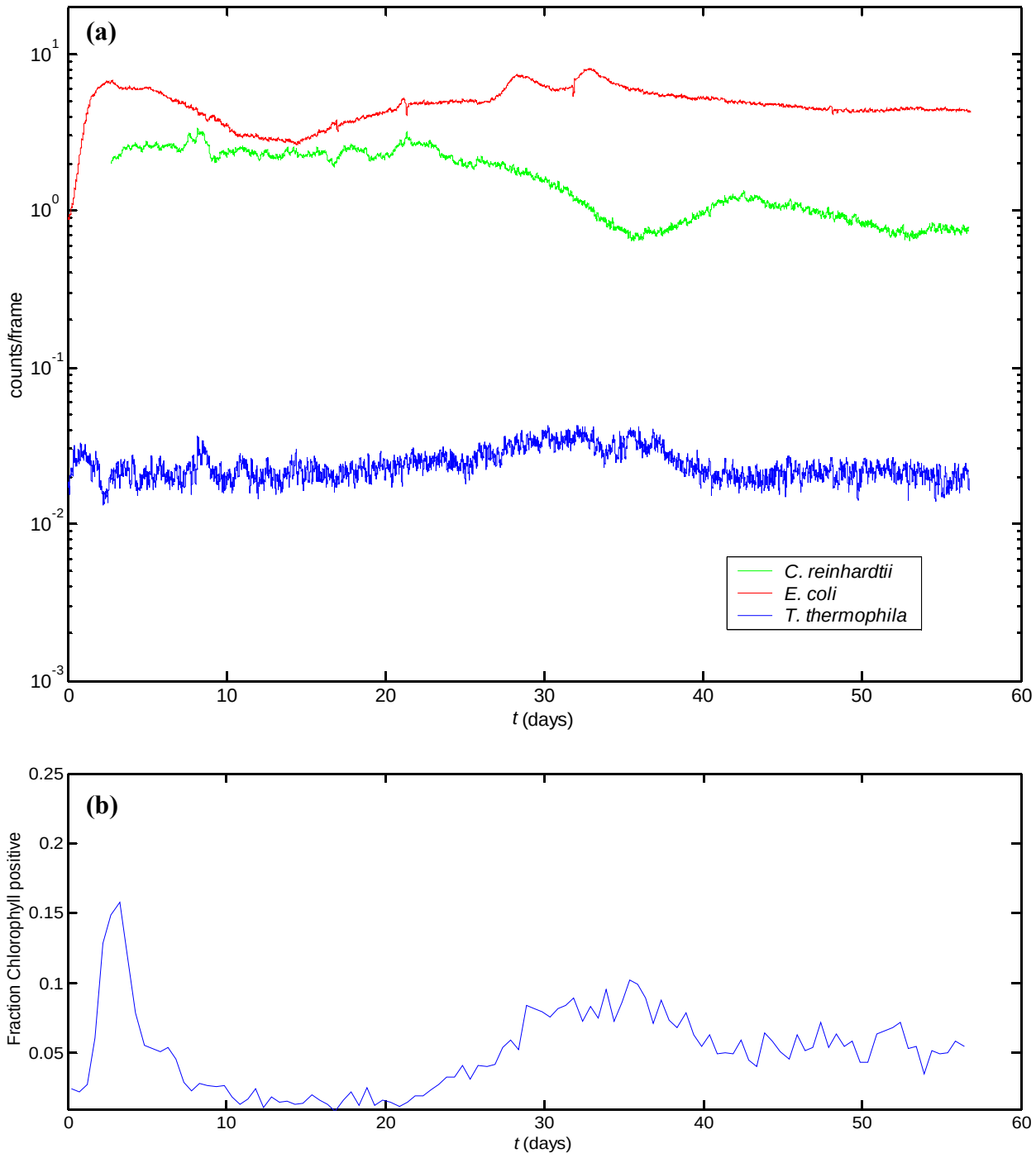


Figure 6.2: Second experiment under constant boundary conditions. An ecosystem from the second experiment in chapter 4 (measured at 2 week^{-1}) was placed in the setup for eight weeks and measured continuously at 1.25 Hz. (a) 2.2 h moving average of observed counts (without any multiplication); (b) fraction of *T. thermophila* observations with a significant chlorophyll signal (suggesting ingestion of alive or dead *C. reinhardtii*). Coarse-grained, with each data point a 12 h average. Data suggest an increase in algae-eating *T. thermophila* preceded a decrease in *C. reinhardtii* density.

$$Q_{1\to 0}(l) = \text{Prob}(\text{1 present leaves after } l \text{ lags}) \cdot \text{Prob}(\text{any entering leave as well})$$
$$Q_{0\to 0}(l) = \text{Prob}(\text{any entering leave as well})$$

Where l is the lag between frames at which the transition probabilities are estimated and the subscripts indicate the relevant transitions. So,

$$F(l) = 1 - \frac{Q_{1\to 0}(l)}{Q_{0\to 0}(l)} \tag{6.1}$$

is the probability that an individual stays up to l frames in the observation volume. In Figure 6.3 the probability is shown that given the observation of an organism at time 0, it is still observed after 1, 2, 3, … frames later (logarithmic plots are shown in Figure 6A.5).

The definition of an effective 'escape rate' of organisms, that is, the rate at which organisms escape from the observation volume, can simplify the description. Such an escape rate assumes the presence of one characteristic time scale, most simply of a Markov process [156]. Two such approximations are shown in Figure 6.3. The simplest approach is to a use the escape rate after one frame, i.e., $p_e = -\frac{1}{\Delta t}\log F(l=1)$, with p_e the escape rate and Δt the time between images, and extrapolate the result to larger time scales (red line in Figure 6.3). A second approach is to calculate a mean escape rate from $F(l)$ over all lags l (at least up to l = 20). Such a mean escape rate can be calculated in a number of ways. Most robust results were obtained by making the *prior* assumption that $F(l)$ is exponential. The mean lag can then be calculated as:

$$\langle l \rangle = \frac{\sum_{l \geq 1} l \cdot F(l)}{\frac{1}{2} + \sum_{l \geq 1} F(l)} \tag{6.2}$$

where the extra ½ follows from the trapezoid approximation of the area under the curve of $F(l)$. A fit of $F(l)$ using $p_e = 1/\langle l \rangle$ is shown in Figure 6.3 in blue. While the blue fit may seem less appropriate than the (red) one-step fit, it provides a much better estimate of tail behavior (see Figure 6.5 for a logarithmic vertical scale). Estimates of p_e are about 0.24 s^{-1} for *C. reinhardtii*, 0.40 s^{-1} for *E. coli* and 1.16 s^{-1} for *T. thermophila*[48]. As shown in the Appendix 6A.4, escape rates were stable over time (fig 6A.3), but varied somewhat between cuvettes (fig 6A.4). These rates suggest that there is little to gain from an increase in the frequency of image acquisition (about 1.2 s^{-1} for the experiments in this thesis), especially for *E. coli* and *C. reinhardtii*, since mostly the same individuals would be observed a few more times.

The description of count correlations using an effective escape rate is not necessarily an accurate description of the dynamics in the observation volume[49], but has the advantage of condensing the description of measurement error.

The next question is how these correlations affect inferred population densities. Such density estimates were obtained by applying a moving average to the observed counts. In what follows I will assume this moving average to be simply an unweighted moving average. The full details of the procedure are given in the appendix. Here I will provide a sketch. The moments of the distribution of the total count over a window of duration T,

[48] These values are for data set 1. For data set 2, the escape rates estimated were slightly less for *E. coli* and *T. thermophila* at 0.30 s^{-1} and 0.80 s^{-1}, respectively.

[49] In fact, given isotropic ballistic motion through an infinitely extending observation plane, the dwell time distribution scales as t^{-2}, with finite size effects truncating the tail.

$N_T = \sum_{t=1,\ldots,T} N(t)$, with $N(t)$ counts from individual images, can be calculated using a moment generating function formalism. The construction of an equivalent Markov process characterized by a single escape rate for each species allows for factorization and efficient evaluation of this moment generating function[50].

An approximate full distribution for N_T can then be obtained by fitting the coefficients $\{a_0,\ldots,a_4\}$ of a simple exponential form, $\exp(a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4)$, under the constraints given by its first four moments $\langle N_T{}^k \rangle$, k = 1, 2, 3, 4, and proper normalization (k = 0). Surrogate population density trajectories can be rapidly drawn from this distribution given the escape rate, the mean density and window length T. Such surrogate data are used in this and the next chapter to determine the significance of observed fluctuations.

[50] Extension of this method to weighted moving averages as well as finite-memory processes is conceptually straightforward, but comes at the cost of additional computation and, for finite memory, additional parameters.

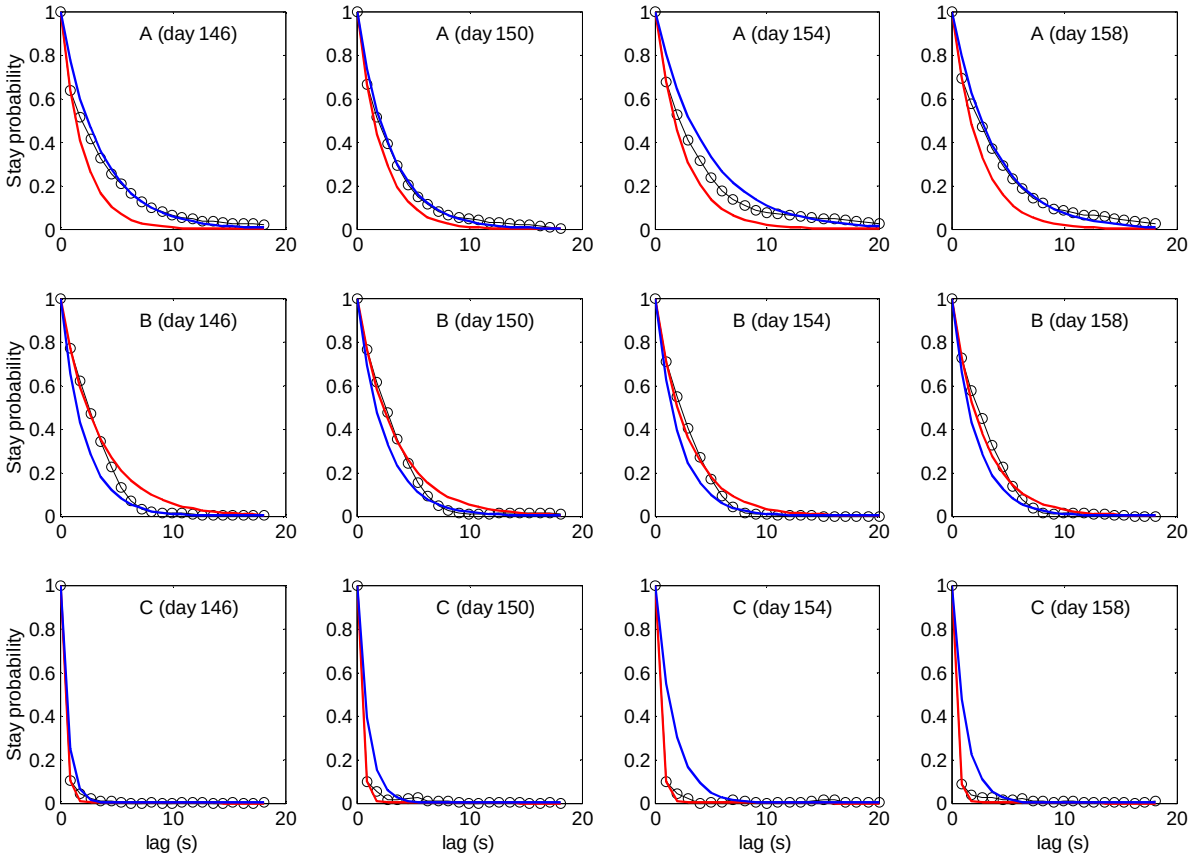


Figure 6.3: Probability for individuals of the different species to stay in the field of view in subsequent frames given that they were observed at time 0. Each curve was calculated over a 1 day window as indicated (black circles). Data based on Figure 6.1. Red: expectation for a memory-less process with decay time given by the decay over the first lag. Blue: expectation based on memory-less process with decay time given by the entire curve. While the red curve may appear a more appropriate fit to the data, the blue curve provides a better desscription of the behavior at large lags (see also Figure 6A.5).

In Figure 6.4, the null model is compared to data from set 1 and shown to be in good agreement with observed fluctuations for all three species on time scales shorter than 1 day. Deviations between the null model probability distribution and the data are consistent with expectations under the null model ($p > 0.1$ for 11 out of 12 segments shown, see Appendix 6A.2 for details). It should be mentioned, though, that these observed fluctuations are in fact consistent with a roughly twofold range for each escape rate[51].

Analysis of Observed fluctuations

The analysis of fluctuations in observed densities is complicated by several factors. Most importantly, the densities are not stationary, even after 142 days (set 1) or 118 days (set 2). This may be in part due to the difference in conditions (23.1 versus 25.5 °C, and illumination from the side versus from the bottom) between sample stand and measurement apparatus (see Chapter 4), causing transient dynamics after transfer. But nonstationarity appears to continue for at least 8 weeks (set 2), suggesting the dynamics are intrinsically nonstationary.

Time series analysis methods for stochastic nonstationary signals are still much less developed than for stationary [2] or deterministic signals [157]. Moreover, the dynamics in our system don't appear to be driven by low-dimensional (non)linear dynamics, although this needs to be explored in more detail.

[51] The wavelet-based power spectra place more stringent constraints on the escape rate.

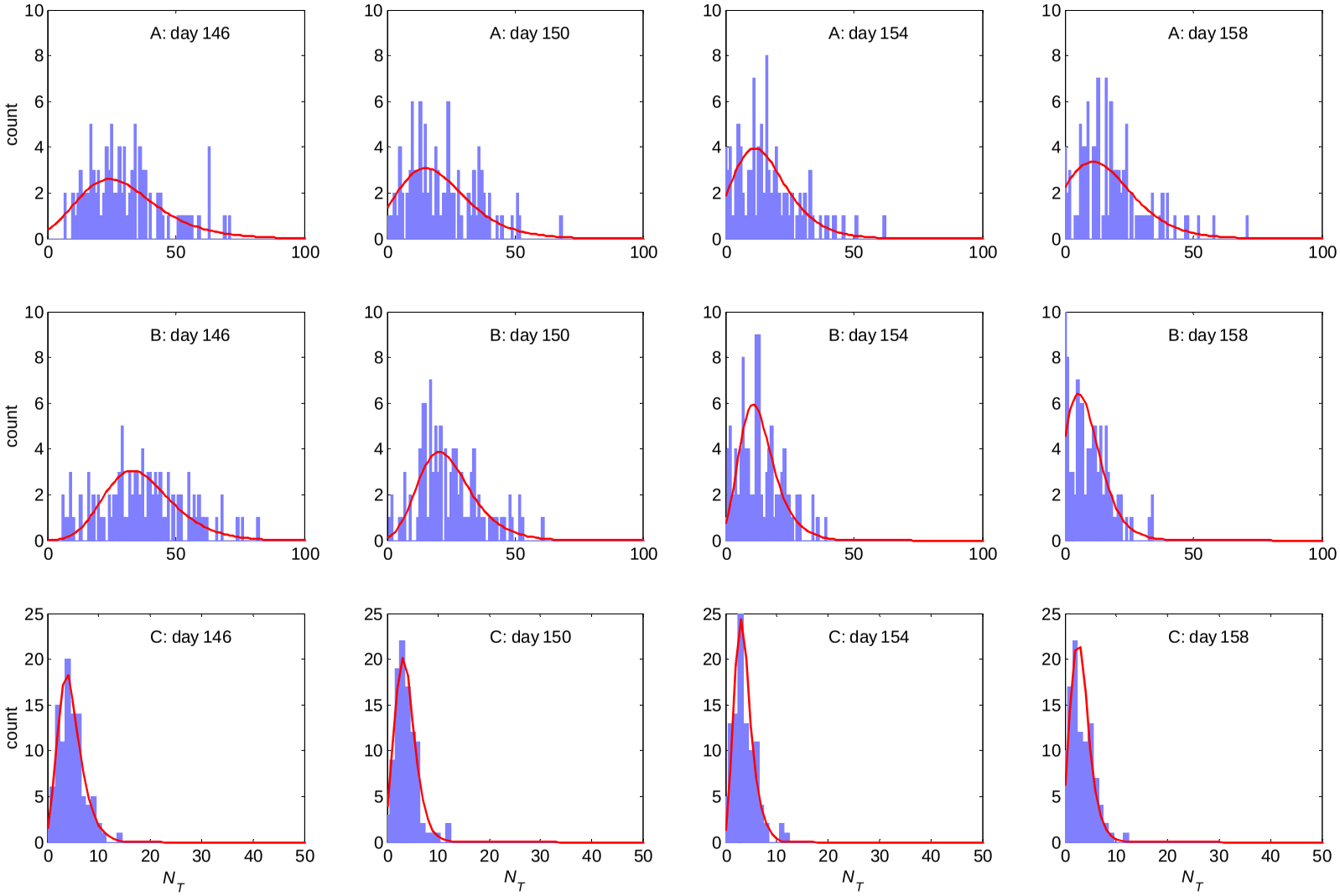


Figure 6.4: comparison of the total number of counts in non-overlapping windows of duration T, N_T (N_T in the text), compared to the null model for four one day periods, with window size $T = 1{,}000$ (that is 15 minutes). All p-values under the null model are >0.10 except for *T. thermophila* on day 154. However, the escape rate estimate used for *E. coli* was 10% lower than inferred from equation 6.2.

For that reason, typical nonlinear dynamics approaches like embedding or the construction of recurrence maps [149, 158], do not seem promising.

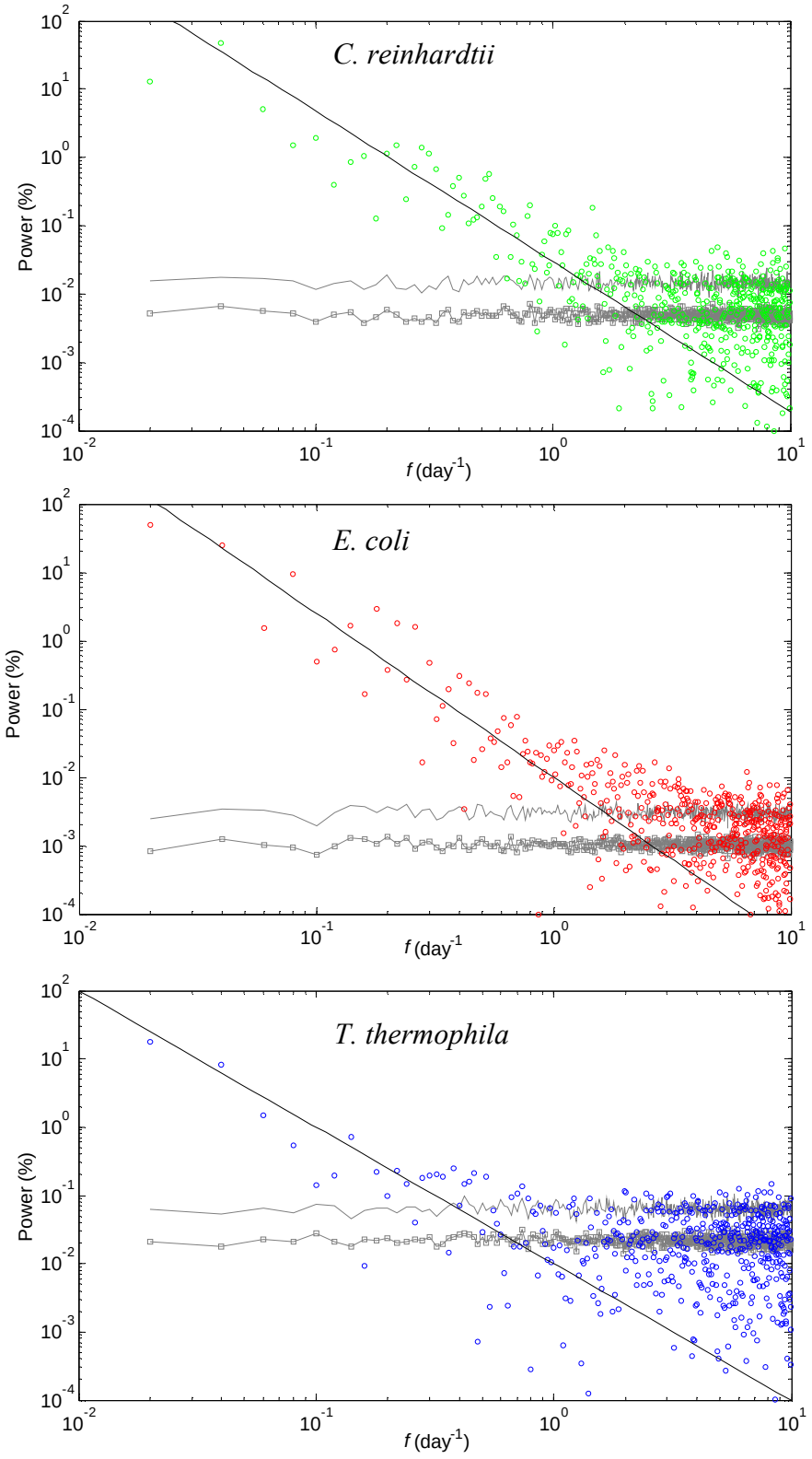
In what follows I will focus on the results for data set 2, since its higher densities and longer duration allow for stronger inferences. First, the data will be analyzed by Fourier analysis and correlation functions. Then I introduce an alternative technique, the use of wavelets, which addresses some of the problems raised by the Fourier analysis.

Fourier power spectra and classical correlation functions

Figure 6.5 shows the power spectra for data set 2 on a log-log scale. The fitted lines (see caption) suggest that the power S at a frequency f scales as $S(f) \propto f^{-\alpha}$, α = 2.0-2.4, for the different species. While this is not to imply that S follows a power law, it does mean that most of the power is present in the lowest frequencies. Importantly, it suggests that for all three species the longest relevant time scale, if it exists at all, would be longer than ~50 days. Another implication is that the measurement frequencies of 2 or 4 week^{-1}, as in the preceding chapter, are largely appropriate, because fluctuations on time scales shorter than 2 to 3 days are weak compared to fluctuations on longer time scales.

Figure 6.5 (next page): power spectra for data set 2. Power is expressed as a fraction of the total power (in %) and compared to the null model developed in this chapter. Shown are the spectra for data (*C. reinhardtii*, green; *E. coli*, red; *T. thermophila*, blue, respectively), and the mean (gray squares) and mean + 1 standard deviation (gray line) for 50 surrogate trajectories. Comparison with the high frequency components of the observed time series suggests that the null model is appropriate.

The black lines indicate fits by eye of the power spectrum at low frequencies. The slopes (on a log-log scale) are: *C. reinhardtii*: -2.2; *E. coli*: -2.4; *T. thermophila*: -2.0.


C. reinhardtii
E. coli
T. thermophila
Power (%)
f (day-1)

As a consequence of nonstationarity, auto- and crosscorrelation functions cannot be determined reliably. Specifically, $C_{ab}(\tau) = \langle (n_a(t) - \bar{n}_a) \cdot (n_b(t+\tau) - \bar{n}_b) \rangle$ is the autocorrelation function (ACF) of the population density $n_a(t)$ for $a = b$, and the crosscorrelation function (CCF) between densities of species a and b for $a \neq b$; $\langle ... \rangle$ denotes averaging over time. The mean densities ($\bar{n}_a, \bar{n}_b$) are not well-defined for a nonstationary process[52]. To demonstrate the difficulties in using ACFs and CCFs for a nonstationary process, ACFs are shown for data set 2 in Figure 6.6, and the corresponding crosscorrelation functions in Figure 6.7. The curves are entirely dominated by the longest time scale, leaving no room for information on shorter time scales. An *ad hoc* approach is to detrend data over shorter time windows, which is equivalent to approximating $\bar{n}_a, \bar{n}_b$ by polynomial functions of time. An example of this approach is shown in Figures 6.8 (ACFs) and 6.9 (CCFs): ACFs and CCFs were determined for 5 subsequent 10-day windows after linear detrending. Now most of the remaining correlation is seen at time scales < 1 day, yet the correlation functions still show wild variation. In the next chapter, I will use 10 day windows in combination with polynomial detrending. Such an approach appears adequate for analysis of frequencies shorter than ~1 day, especially since the observed effects under modulation of illumination are much stronger than the intrinsic fluctuations on those time scales. In general, however, the choice of a time window appears arbitrary and probably affects the observed dominant time scale. In addition, the current estimates of the ACF and CCF don't use any of the opportunities for averaging

[52] This can be seen from the power spectrum: the mean is related to the Fourier component at frequency 0. Since $S(f) \propto f^{-\alpha}$, $S(0) \to \infty$, and the mean cannot be determined.

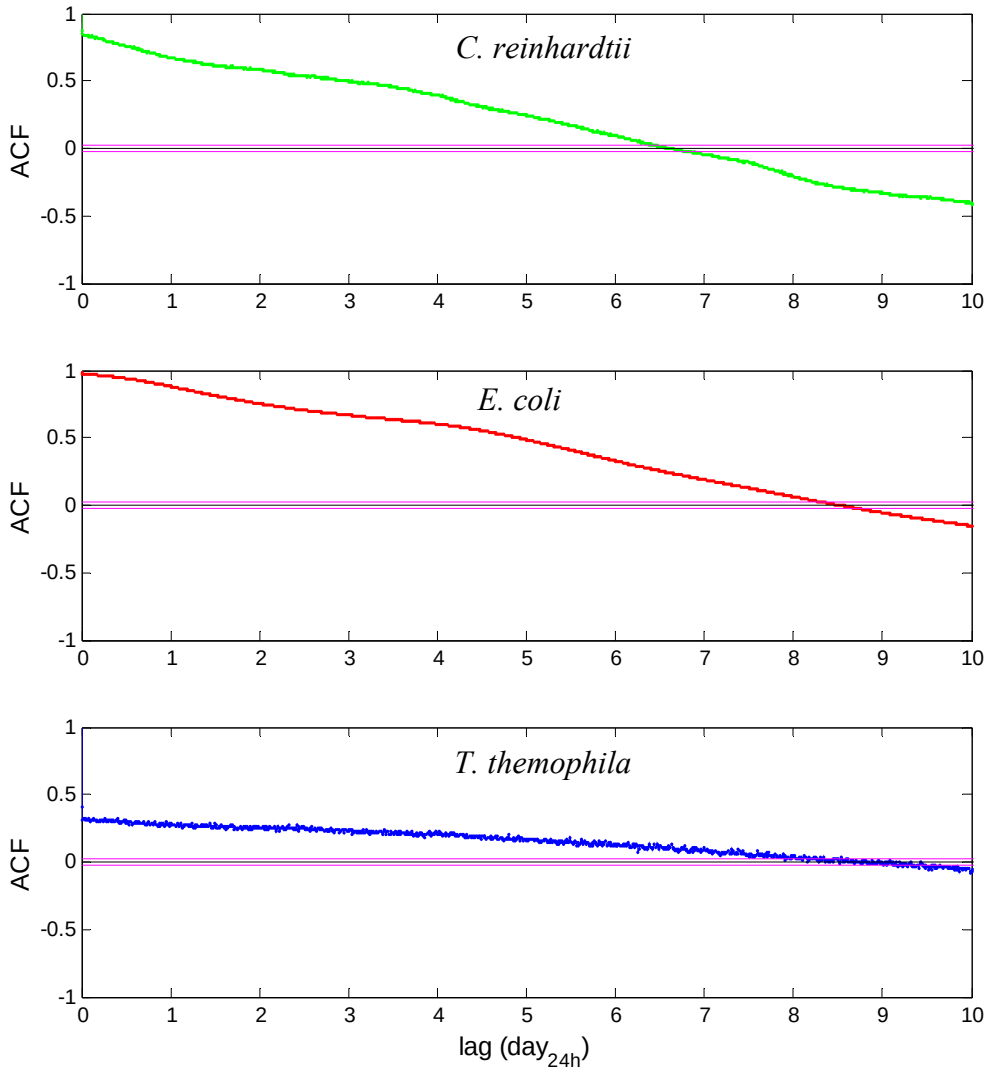


Figure 6.6: autocorrelation functions for *C. reinhardtii*, *E. coli* and *T. thermophila* under constant temperature and illumination for data set 2. Data were linearly detrended. The magenta lines indicate the ±2 σ interval for uncorrelated data ([2], on this time scale, the null model gives nearly identical bounds).

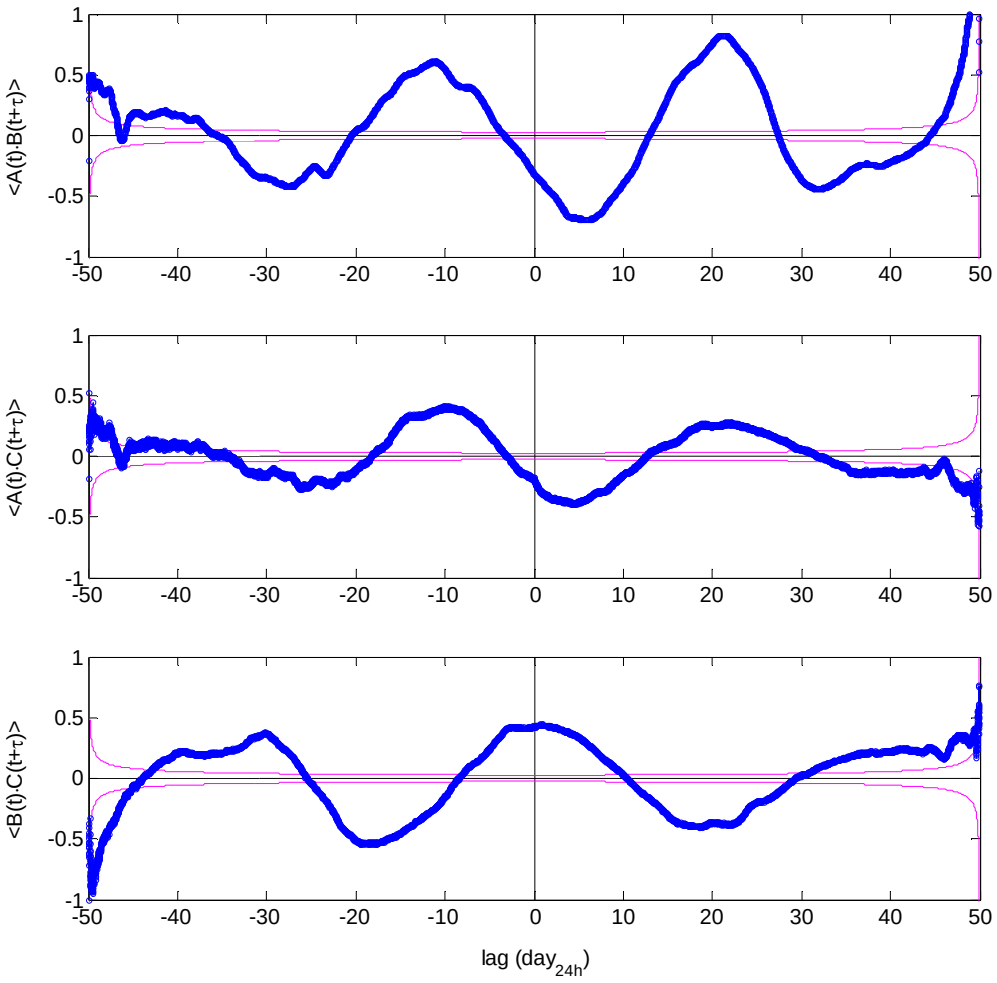


Figure 6.7: Crosscorrelation functions for a linearly detrended 50 day window of data set 2. The species pairs are indicated on the vertical axis, with A = *C. reinhardtii*, B = *E. coli*, and C = *T. thermophila*. Due to nonstationarity the autocorrelation function only yields an signal on the longest time scales while there is no clear interpretation for correlations on shorter time scales. The magenta lines, as in Figure 6.6, indicate the ±2 σ interval for uncorrelated data.

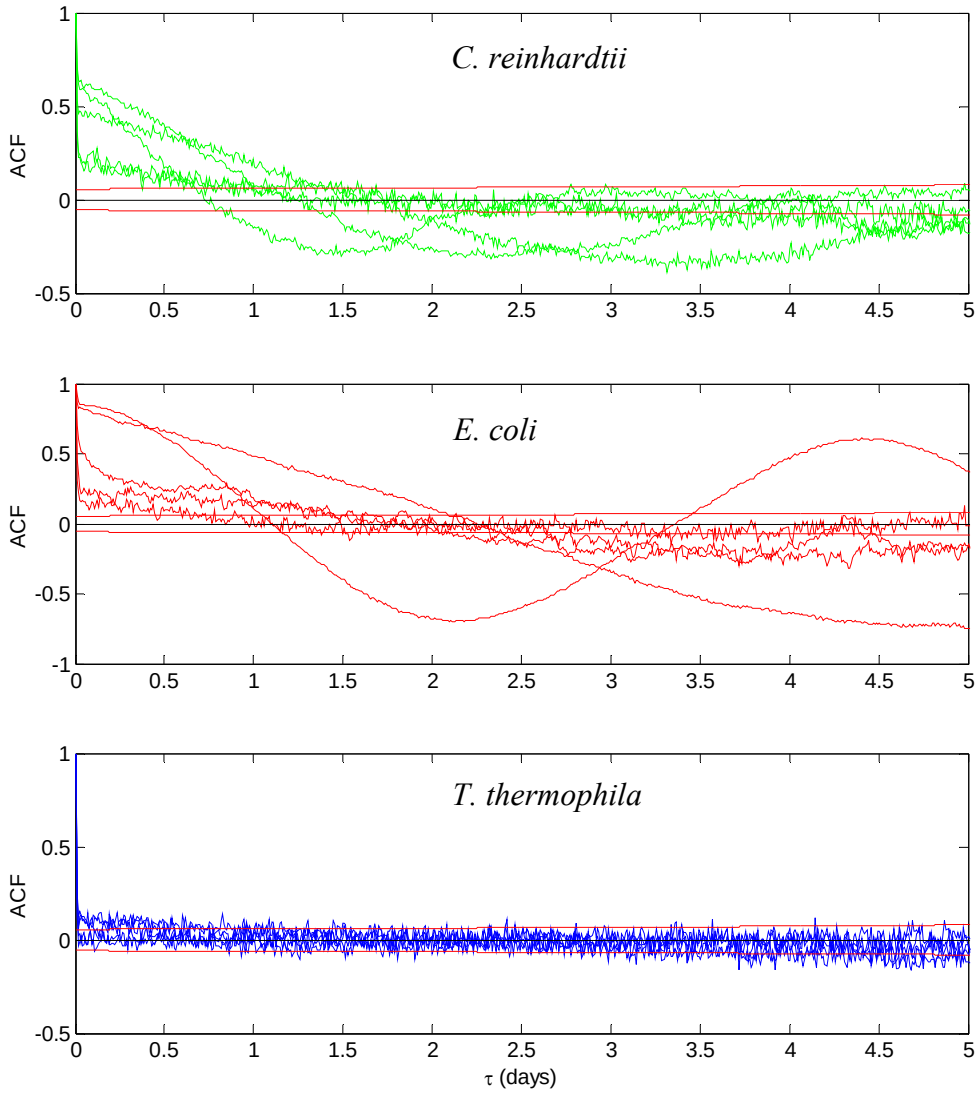


Figure 6.8: Autocorrelation functions for five subsequent 10-day windows of data set 2 (Figure 6.2). Densities in each window were linearly detrended prior to calculation of the correlation functions. The red lines, as in Figure 6.6, indicate the ±2 σ interval for uncorrelated data.

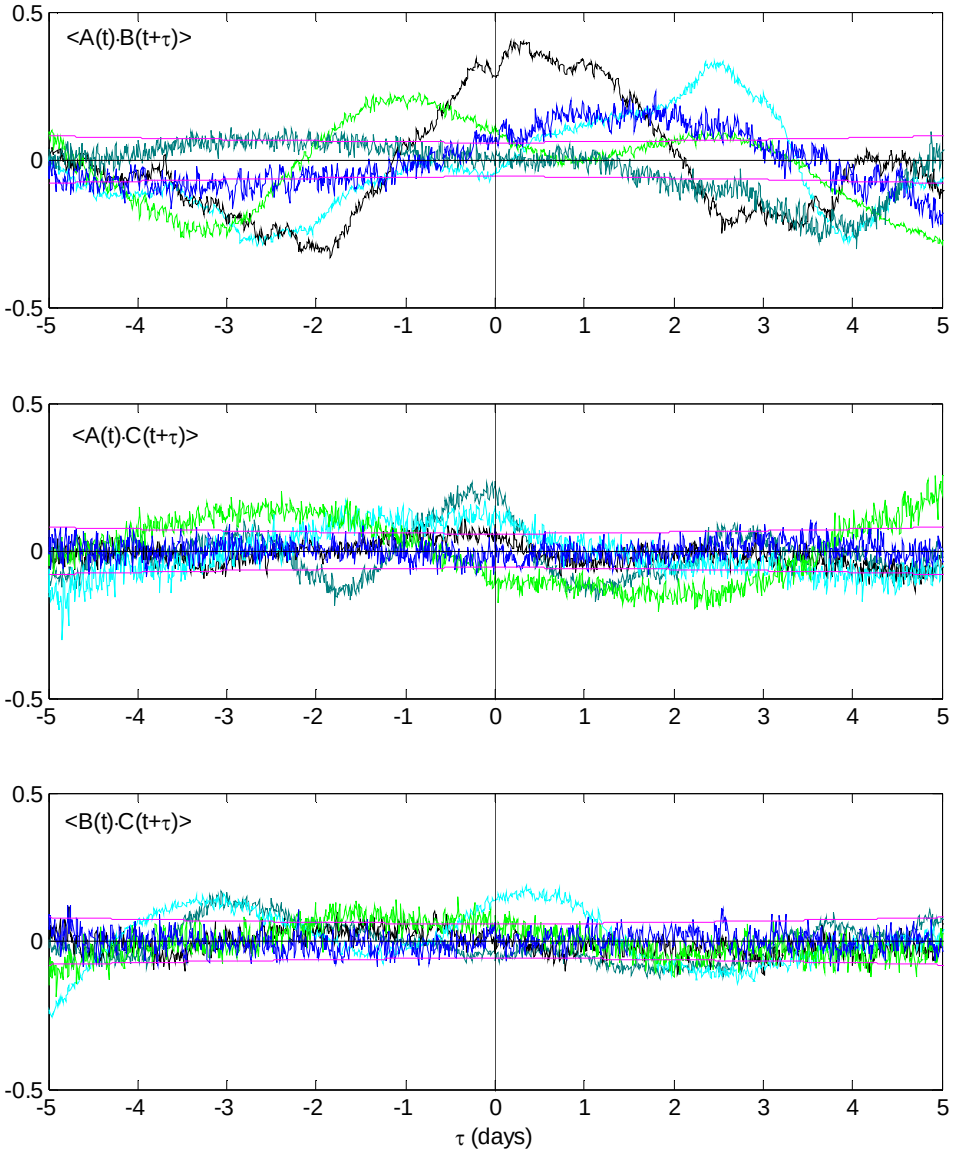


Figure 6.9: Crosscorrelation functions for five subsequent 10-day windows of data set 2 (Figure 6.2). The species pairs are indicated on the vertical axis, with A = *C. reinhardtii*, B = *E. coli*, and C = *T. thermophila*. Densities in each window were linearly detrended prior to calculation of the correlation functions. The magenta lines, as in Figure 6.6, indicate the ±2 σ interval for uncorrelated data.

offered by larger time scales. One could progressively smoothen data and consider larger time windows, but it seems smoothing would introduce yet another arbitrary time scale. Wavelet analysis may provide a more consistent approach to data with power present over a range of time scales.

Wavelet transforms

Over the past two decades, wavelet analysis has seen rapid development as a method to analyze nonstationary time series, such as in the study of weather and climate [5] and EEG signals [159]. Wavelets are localized functions used to probe a signal for particular features such as fluctuations (that is, deviations from the mean or a trend), discontinuities and intermittent periodicities (see [10-12] and Appendix 6A.3 for details). They are characterized by a single scale parameter τ and a localization parameter t. That is, let ψ be the 'mother wavelet', then:

$$\psi = \psi\left(\frac{t-t'}{\tau}\right) \tag{6.3}$$

Localization means that they only have a limited support set over which they differ significantly from 0. Wavelets are especially suitable because they locally detrend the signal. Roughly, they locally subtract the mean, addressing the problem of an ill-defined mean posed above. Specifically,

$$\int dt\, \psi\left(\frac{t-t'}{\tau}\right) = 0 \tag{6.4}$$

For example, the db-3 wavelet transform[53] used in this chapter does not depend on (up to) third-order polynomials added to or subtracted from the signal. To be precise,

$$W_s(t',\tau) = \int dt \left[s(t) + a_0 + a_1 \cdot t + a_2 \cdot t^2 + a_3 \cdot t^3 \right] \cdot \frac{1}{\sqrt{\tau}} \psi\left(\frac{t-t'}{\tau} \right) = \int dt\, s(t) \cdot \frac{1}{\sqrt{\tau}} \psi\left(\frac{t-t'}{\tau} \right) \tag{6.5}$$

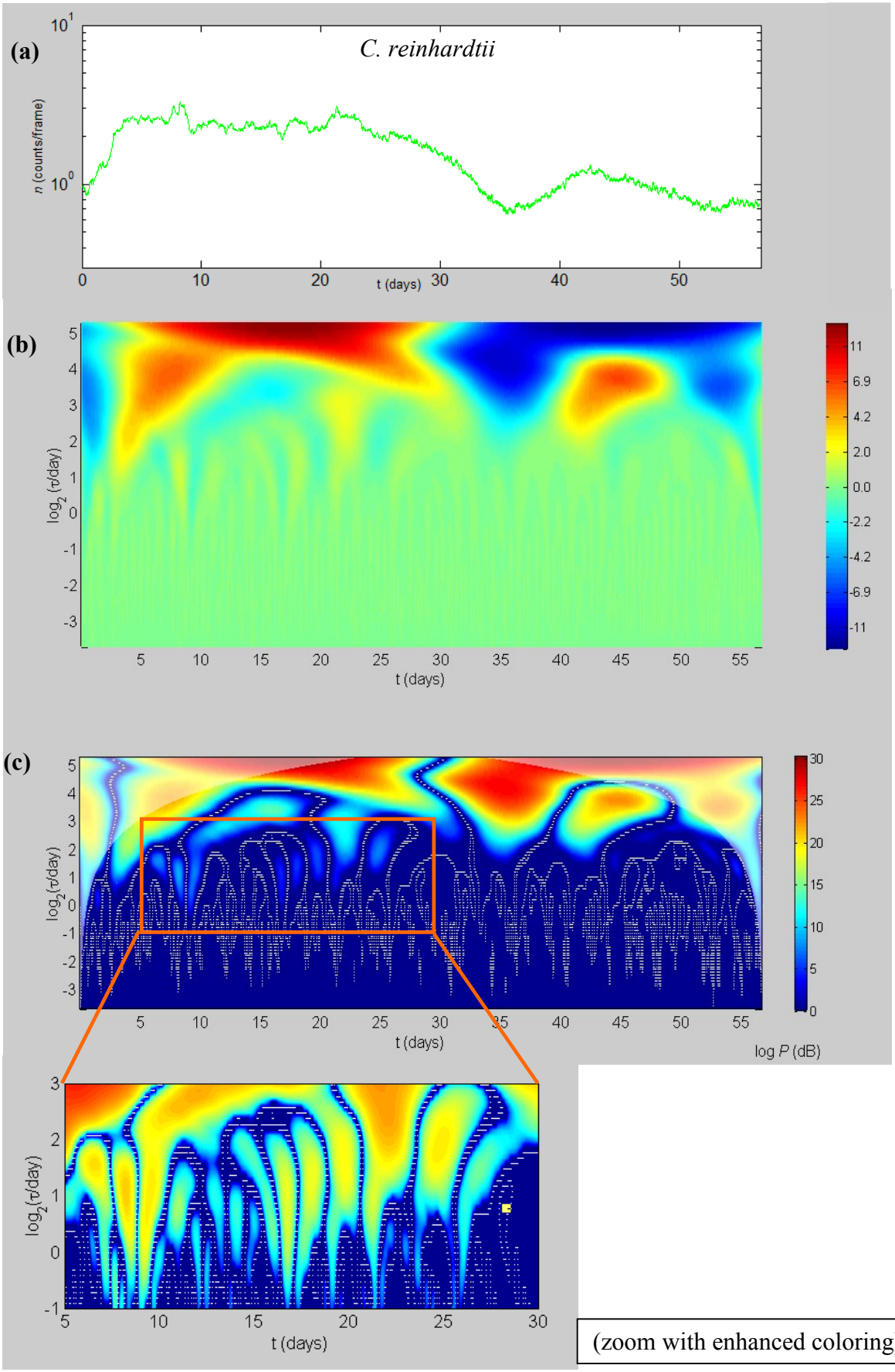
With $W_s(t',\tau)$ the wavelet transform of the signal $s(t)$ at time scale τ and time t', and $a_0, a_1, \ldots$ coefficients characterizing a polynomial trend added to the signal. As such, wavelet transforms can isolate fluctuations in the signal on the time scale of the fluctuations (τ) and in time (t'). The factor $1/\sqrt{\tau}$ is explained by normalization. The wavelets on each time scale, w, are normalized versions of the mother wavelet ψ, with scale-invariant (unit) power:

$$\int dt \left| w(t,\tau) \right|^2 = 1 \;\Rightarrow\; w(t,\tau) = \frac{1}{\sqrt{\tau}} \psi\left(\frac{t-t'}{\tau} \right) \tag{6.6}$$

The results of the continuous wavelet transform are easily visually interpreted, and I will use it to illustrate the ability of wavelet transforms to isolate fluctuations within a signal in time and characteristic timescale. In Figure 6.10, the application of the db3 wavelet transform to data set 2 is shown (results for set 1, being more noisy and relatively featureless, are not shown). As can be seen, features in the signal are mapped to their appropriate timescales.

[53] Named after Ingrid Daubechies, see appendix 6A.3 for the shape of this wavelet.

Figure 6.10 (next three pages): wavelet transforms and time-dependent wavelet power spectra for set 2 (log transformed densities).
Panels **a**, **d**, **g**: time series for *C. reinhardtii*, *E. coli* and *T. thermophila* respectively. Density estimates were obtained by application of 3 h (**a** and **d**) and 12 h (**g**) moving averages.
Panels **b**, **e**, **h**: corresponding continuous wavelet transforms using a Daubechies-3 wavelet (see Appendix 6A.2).
Panels **c**, **f**, **i**: power present in the fluctuations in the preceding panels in dB.
At the boundaries continuation of the initial and final value of the time series, respectively, was assumed. This introduces boundary effects within the so-called cone of influence [5], which was estimated in Matlab to be 0.6x the scale of the wavelet. The cone of influence is shown in panels **c**, **f** and **i** as two shaded regions. The wavelet transforms were slightly smoothened in both the time and scale domain using a Gaussian kernel with σ_t = 55 minutes, and σ_τ = 15% (**b**, **c**) or 7% (**e**, **f** and **h**, **i**).
To test statistical significance, 200 trajectories were generated according to the null model with the same escape rates, mean, and coarse-graining (each data point used for the wavelet transform was the average count of a non-overlapping 512 frame window).
The same wavelet transform and smoothing were applied to these surrogate trajectories. Values of these transformed surrogate data were ranked and 98% percentiles (95% for *T. thermophila*) obtained using a one-pass algorithm that does not require storage of complete histograms, the P^2-algorithm [11]. Outlines of the regions with power larger than the 98% percentile of surrogate data are demarcated with white dots.
The color scheme for time-dependent power spectra was truncated at 0 dB (arbitrary reference) for display clarity. This appeared to yield a good separation between signal and background.


(a)
C. reinhardtii
n (counts/frame)
t (days)
(b)
log2 (τ/day)
t (days)
(c)
log2 (τ/day)
t (days)
log P (dB)
(zoom with enhanced coloring)

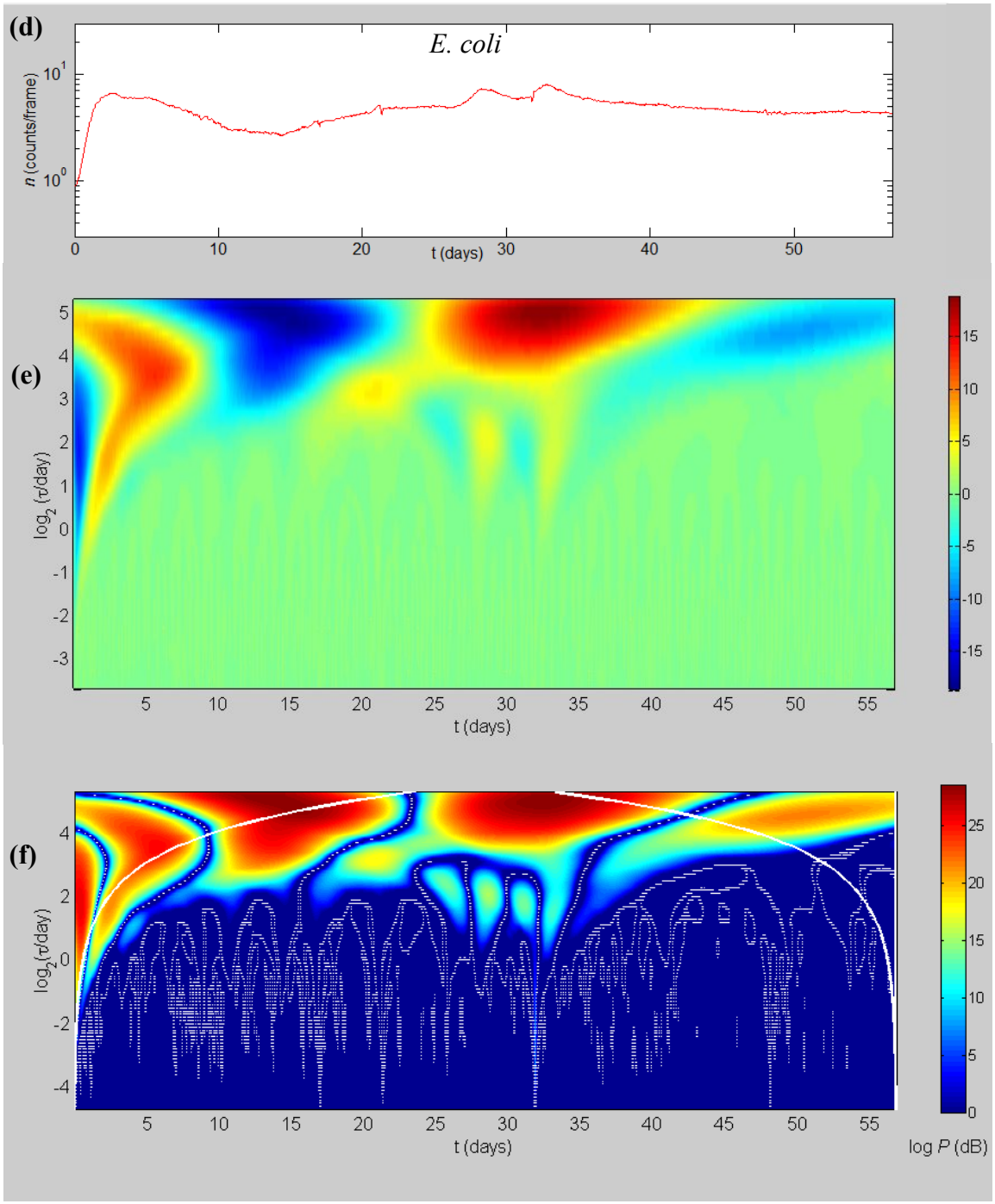


Figure 6.10 (d – f)

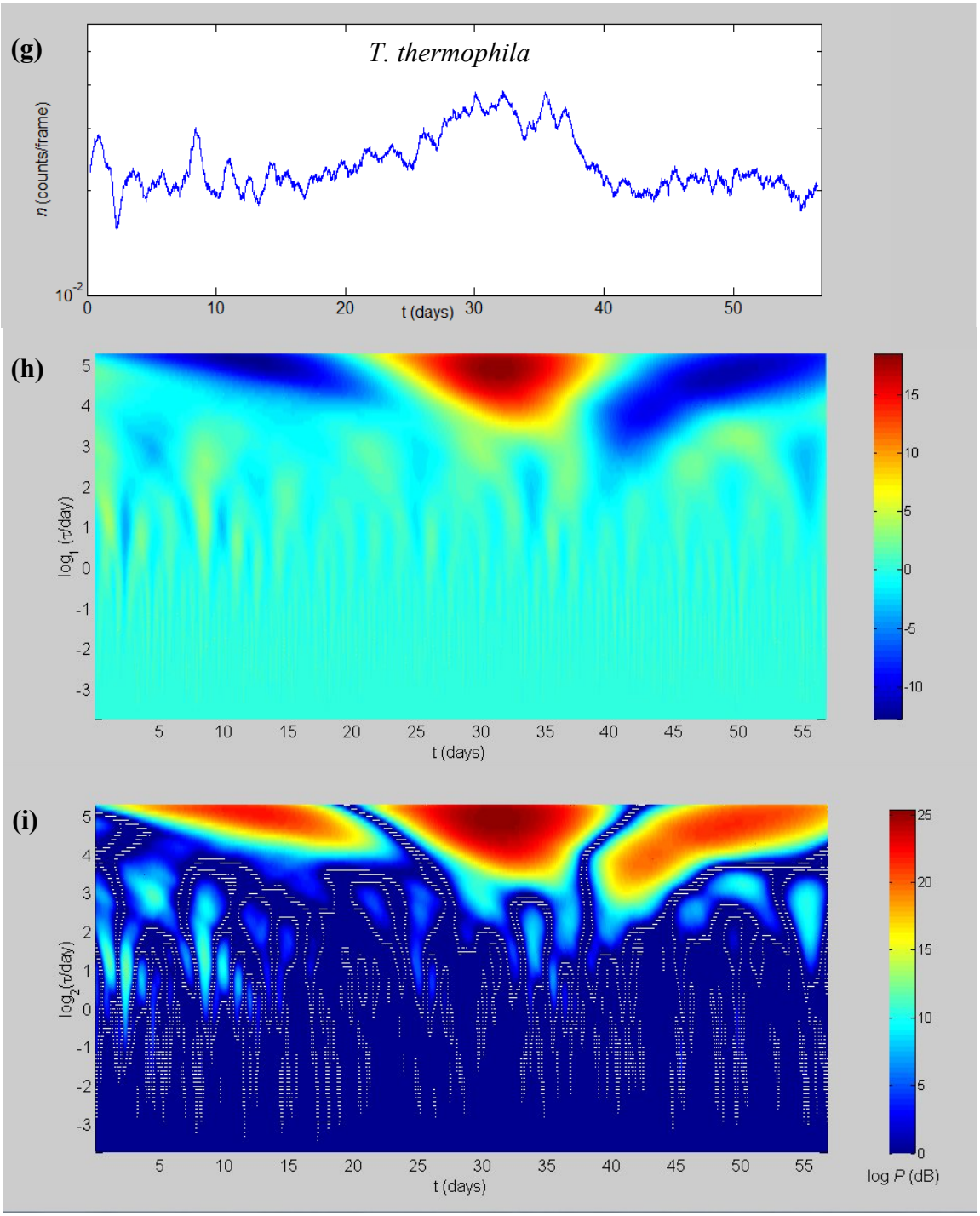


Figure 6.10 (g - i)

A few remarks are in place here. First, wavelet transforms shown are for log n, with n the population density. Hence, as in the previous chapter, the spectrum of relative fluctuations is captured (spectra are very similar for n itself). The continuous wavelet transform yields a highly redundant representation of the data, and is best suited for visualization. Testing of statistical significance of the spectra was done as outlined in the caption of Figure 6.10, and based on the null model described above[54].

\For *C. reinhardtii* (Figure 6.10 (c), zoom insert) fluctuations are observed at timescales of 1-3 days at high densities (~10^5 mL^{-1}, day 3-25 [55]). These relative fluctuations disappear at lower densities. Similar fluctuations are seen for *T. thermophila*. For *E. coli*, the time series is relatively smooth, but the high densities (>1 count/frame) allow for detection of even small fluctuations. Most striking are two large[56] fluctuations over day 27 – 35, during which there is a marked decline in *C. reinhardtii* density and an increase in *T. thermophila* density.

To quantify the correlation between fluctuations in the different species on different time scales, the crosscorrelation function (C_{ab}) can be defined between the continuous wavelet transforms of densities of species a and b for different timescales (that is, horizontal slice from Figure 6.10 c,f and i) [5]. In this way, wavelets provide a 'wavelet zoom' of the data [161]:

$$C_{AB}(\tau|\tau_0) = \frac{\langle W_A(t,\tau_0)\cdot W_B(t+\tau,\tau_0)\rangle - \langle W_A\rangle\cdot\langle W_B\rangle}{\sigma(W_A)\cdot\sigma(W_B)} \qquad (6.3)$$

with

[54] Because of the redundancy of the continuous wavelet transform, false positives (apparently significant fluctuations due to measurement noise) also appear in patches. Areawise significance test are actively being developed to address this [160].

[55] Day 121-143 after closure.

[56] Amplitude about a factor 2 in density.

$$W_A(t,\tau_0) = \int dt' \frac{1}{\sqrt{\tau_0}} \psi\left(\frac{t-t'}{\tau_0}\right) \cdot \log n_A(t') \qquad (6.4)$$

the continuous wavelet transform shown in Figure 6.10, with τ_0 the wavelet timescale considered and τ the lag of the correlation function. The resulting crosscorrelation functions are shown in Figure 6.11-6.13 for different slices from the continuous wavelet transforms. At the timescales considered, except for *T. thermophila* at 7 h, the errors in the wavelet transform coefficients are small compared to their variation over time. Since we are testing the significance of correlations, rather than of the transforms themselves, the obtained crosscorrelation function are compared to crosscorrelations of shuffled transforms[57]. All three species pairs exhibit significant crosscorrelation[58] on scales of 15-120 h. In the clearest example, *E. coli* fluctuations appear to track those of *C. reinhardtii* with a slight lag (0.1-0.3 τ, marked by black arrows in Figure 5.11[59]). Caution is warranted, however: the estimates of the crosscorrelation function are themselves correlated with a timescale τ_0; specifically, within this timescale the estimates tend to reflect the wavelet shape. Any feature over the timescale τ_0 of the wavelet can in principle reinforce this shape. For example, either for 10 day fluctuations *T. thermophila* is negatively correlated with *C. reinhardtii* densities three days earlier, or positively with *C. reinhardtii* densities five days later, or both (Figure 5.12, last panel).

The detailed interpretation of these wavelet correlation functions requires further study.

---

[57] The error in the cross correlation is approximately $\sqrt{\tau_0/L}$, with L the length of the time series.

[58] More than two standard deviations from the mean.

[59] I am not yet sure about the error in these lag estimates.

Time-independent power spectra

As in Fourier analysis, wavelets can provides an orthonormal basis for signal decomposition, known, for wavelets, as multiresolution analysis [162]. As a consequence, the power of the signal ($\sum_t \left(\log n(t) - \langle \log n \rangle\right)^2$ can be decomposed into the contributions of individual time scales [5, 159]. This is completely analogous to the decomposition of a signal in Fourier analysis followed by the calculation of a power spectrum (see equation 6A.13). Such wavelet-based time-independent power spectra are shown in Figure 6.14 and 6.15. Just like in Fourier analysis, a mismatch between the analyzing wave (e.g. a sine wave) and the analyzed signal (e.g. a square wave) can lead to some bleeding of power into higher frequencies[60]. This suggests some wavelets may prove more appropriate for population density time series than others[61].

[60] This is a matter of taste. For audio signals, for example, it is appropriate to say that a square wave actually *contains* higher frequencies, while this is rather silly for many electronic applications.
[61] So far I have only tried one other wavelet, the Meyer wavelet, and obtained very similar results.

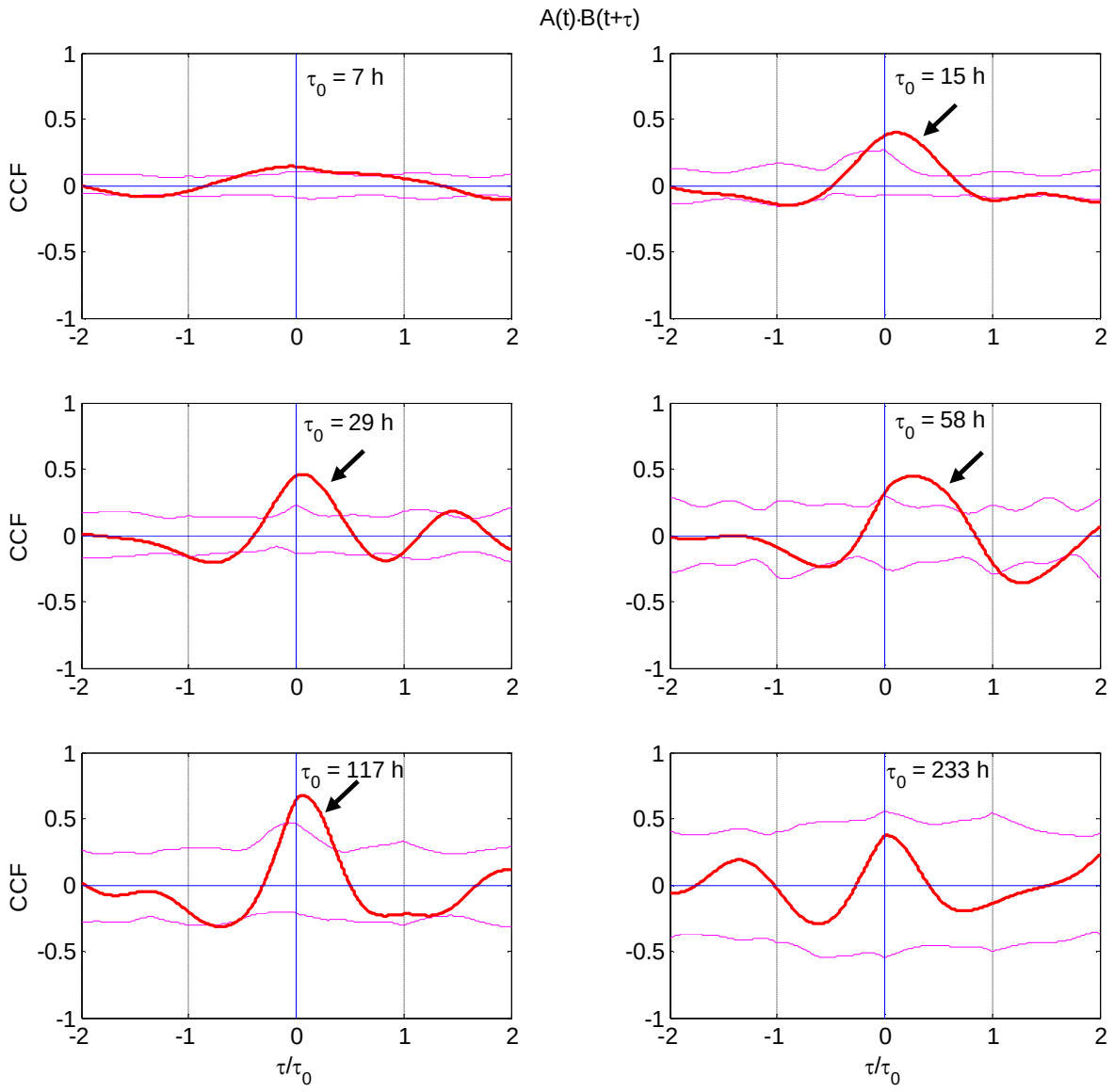


Figure 6.11: Crosscorrelation function between *C. reinhardtii* and *E. coli* continuous wavelet transform coefficients (corresponding to different horizontal slices of the transforms in Figure 6.10). For each time scale of the transform, the lags in the cross correlation function are scaled by that time scale. Magenta lines indicate two standard deviations from the mean for shuffled data (shuffled in blocks of the same size as the time scale examined). The grid lines emphasize the fact that within one lag, the data are not independent but tend to assume the shape of the wavelet function.
The arrows point to significant correlations between *E. coli* and *C. reinhardtii* mentioned in the text.

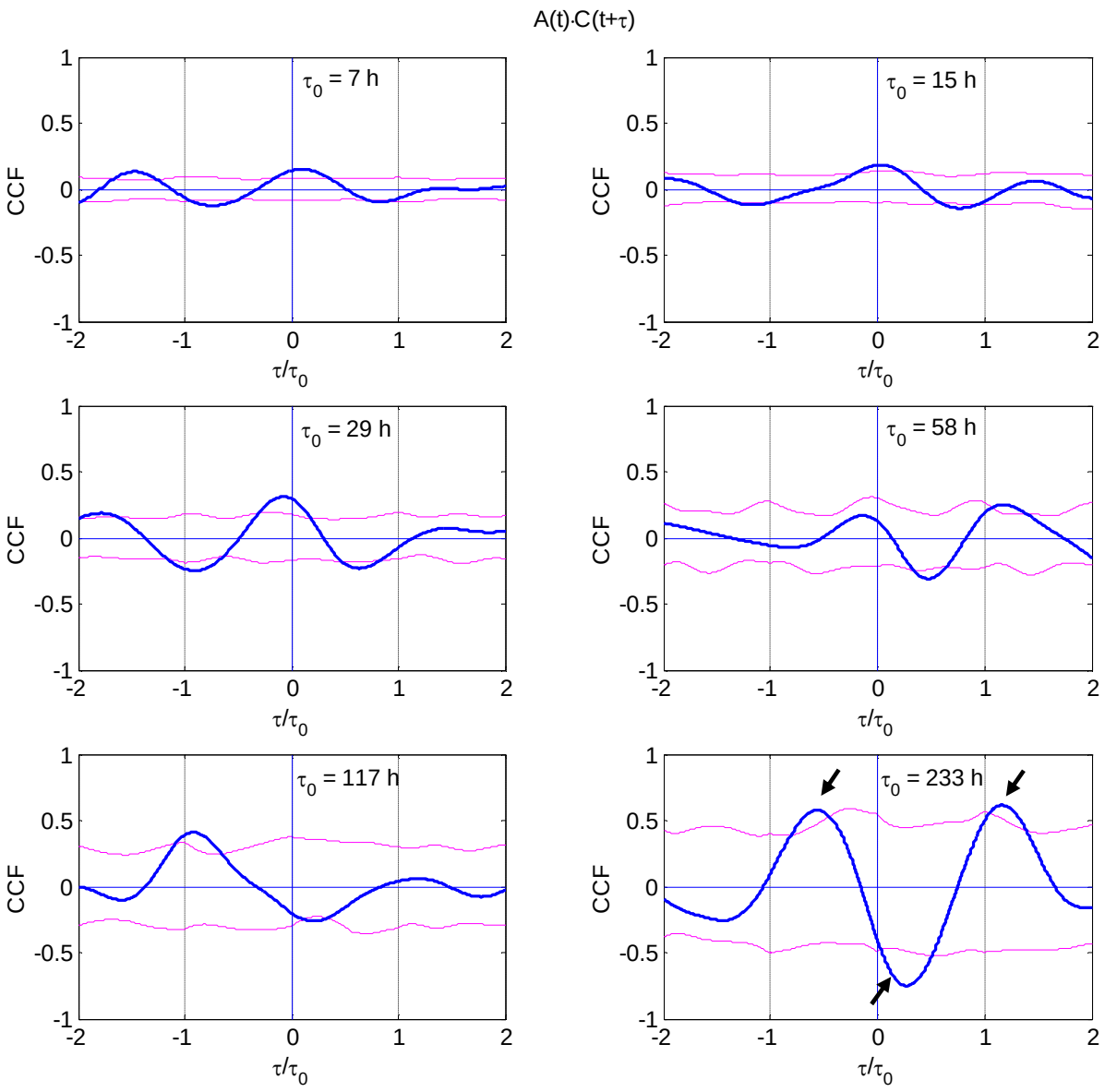
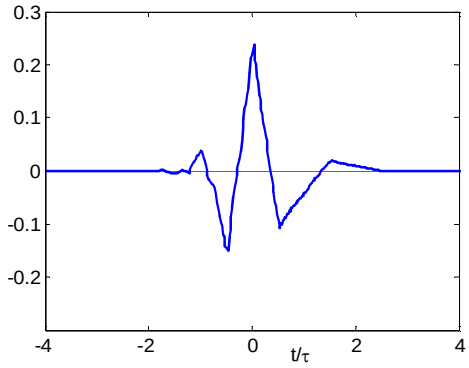


Figure 6.12: Crosscorrelation function between *C. reinhardtii* and *T. thermophila* continuous wavelet transform coefficients as in Figure 6.10. Arrows in the last panel highlight significant correlations between *C. reinhardtii* and *T. thermophila* mentioned in the text. These correlations are not independent, but reflect the shape of the wavelet (below), complicating their interpretation.



Analyzing wavelet, db3, described in appendix 6A.3.

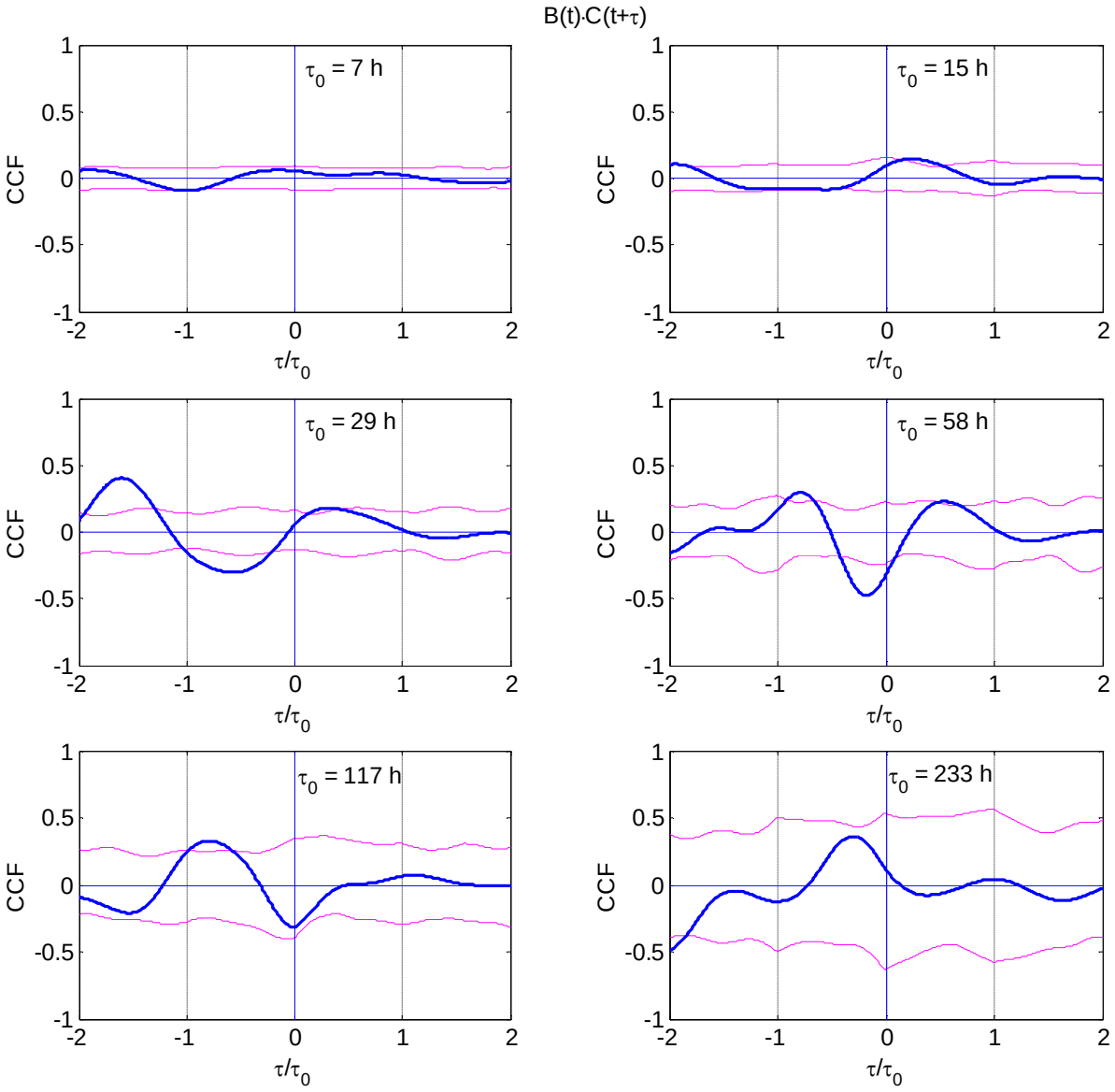


Figure 6.13: Crosscorrelation function between *E. coli* and *T. thermophila* continuous wavelet transform coefficients as in Figure 6.10.

Fluctuations in obtained population density time series are significantly larger than measurement noise for *C. reinhardtii* on time scales larger than 0.1 day (set 1) or 0.3 day (set 2) timescales. For *E. coli*, density significant fluctuations on timescales larger than 1 day (set 1) or 1 hour (set 2) can be distinguished. Note that this difference can be largely due to the larger statistical power for the second data set offered by the higher densities and longer measurement duration. Also, the power in these fluctuations is not obvious from the Fourier power spectrum (Figure 6.5) at 1-5 hour time scales. For *T. thermophila*, significant fluctuations were detected only on timescales larger than ~5 days (set 1) and 1 day (set 2). The power on timescales less than one day is generally small, however, suggesting that a daily measurement protocol, as in the previous chapter, can be largely appropriate.

Excess power over measurement noise is shown in the panels on the right in Figures 6.14 and 6.15. As seen before in the Fourier power spectra, the power keeps increasing as far as our time horizon allows us to see. This increase, seen for all species in both sets, confirms that the observed population dynamics are nonstationary.

Figure 6.14 (next page): Power (left panels) and excess power over a fit to surrogate data (right panels).

Left panels: squares: power spectrum based on Daubechies-3 wavelet as a function of time scale (τ_{min} = 4 min, some of the longer time scales are indicated below the *x*-axis of the first panel). Boxplots are over 40 surrogate time series, with whiskers indicating 1.5x interquartile distance and red plusses indicate outliers. The straight black line is a model fit under uncorrelated noise (power $\sim 2^{-\tau}$). The panels on the right show the excess power in fluctuations on each time scale obtained by subtracting the uncorrelated-noise fit from the observed fluctuations. Error bars are based on the standard deviation for simulated population density time series. Escape rates used were 0.237 s^{-1} for *C. reinhardtii*, 0.30 s^{-1} for *E. coli* and 0.90 s^{-1} for *T. thermophila*. These values are somewhat different for *E. coli* and *T. thermophila* from those estimate from Figure 6.3, and were optimized to fit the observed data.

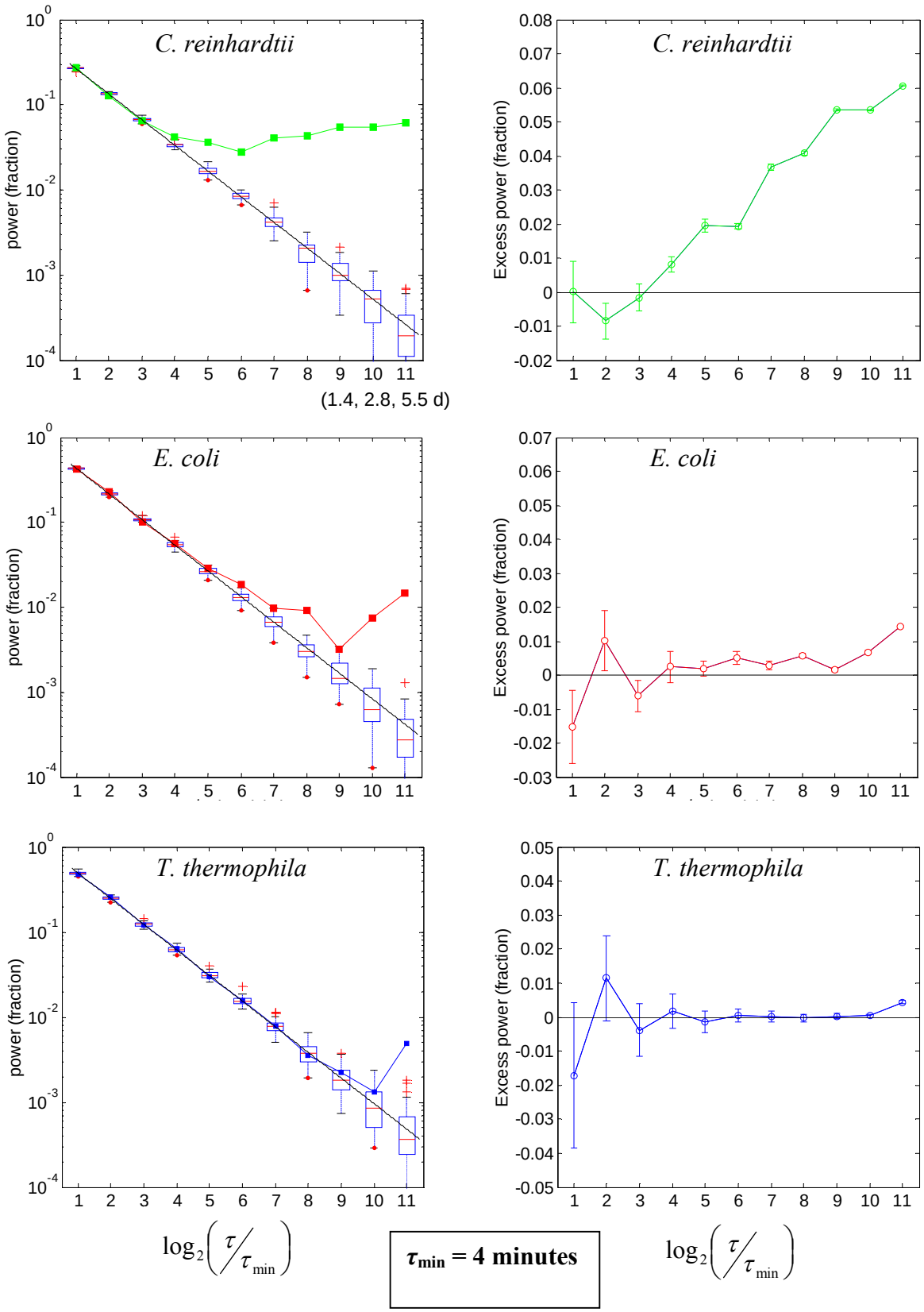
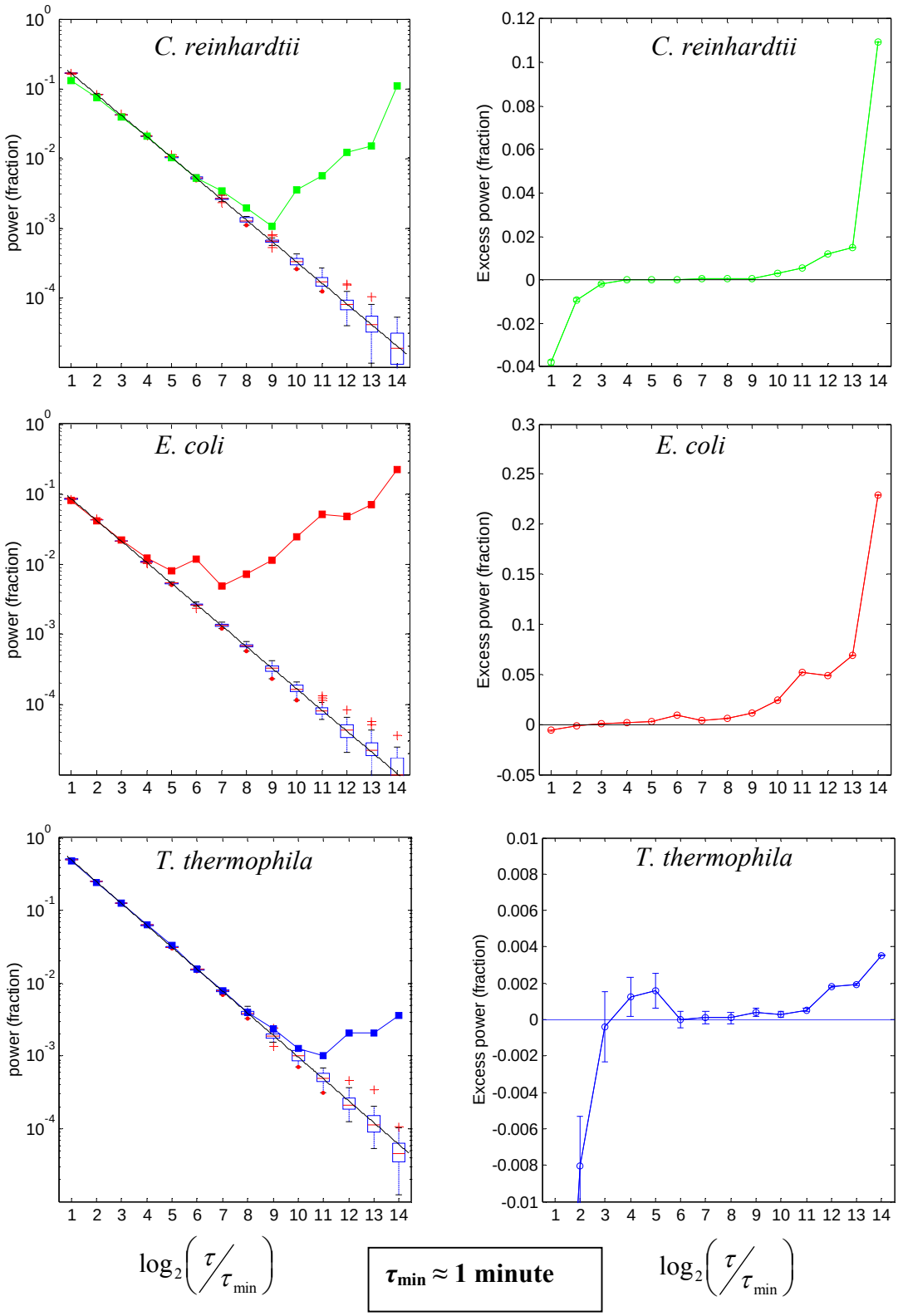

C. reinhardtii
power (fraction)
(1.4, 2.8, 5.5 d)
C. reinhardtii
Excess power (fraction)
E. coli
power (fraction)
E. coli
Excess power (fraction)
T. thermophila
power (fraction)
T. thermophila
Excess power (fraction)
log2(τ/τmin)
τmin = 4 minutes
log2(τ/τmin)

Figure 6.15 (next page): as Figure 6.9, for data set 2. The longer extent of the time series and higher densities allowed for extension to longer and shorter timescales than in Figure 6.9. τ_0 = 51 sec (64 frames). For longer time scales: j = 11: 29 hours, 12: ~2.5 days, 13: ~5 days, 14: ~10 days.

Escape rate estimates used were 0.23 s^{-1} for *C. reinhardtii*, 0.40 s^{-1} for *E. coli* and 0.81 s^{-1} for *T. thermophila*. j = 14 corresponds to 10 days (j = 13: 5 days, 12: 2.5 days, etc.).


C. reinhardtii
power (fraction)
C. reinhardtii
Excess power (fraction)
E. coli
power (fraction)
E. coli
Excess power (fraction)
T. thermophila
power (fraction)
T. thermophila
Excess power (fraction)
log2(τ/τmin)
τmin ≈ 1 minute
log2(τ/τmin)

Conclusions

In order to analyze the population density time series under constant boundary conditions (light and temperature), I first introduced a null model for fluctuations in estimated densities due solely to correlated count statistics. The effects of these count statistics can be adequately described by an effective escape rate (Figures 6.4, 6.14 and 6.15) on timescales longer than ~1 minute.

In the subsequent analysis, two analysis methods were compared. First classical correlation functions and their Fourier transform, the (Fourier) power spectrum, were examined. This analysis showed that all three species densities were nonstationary, with similar scaling behavior of the power per frequency. This nonstationarity complicates, if not invalidates, the use of correlation functions. Their appearance depends on the arbitrary application of time windows, detrending, and smoothing.

Wavelet analysis appears to provide a natural alternative, since wavelet transforms are local and trend-insensitive. Wavelet transforms show proper localization of species density fluctuations in time and map them to the appropriate time-scale. In addition, time-averaged power spectra are comparable to Fourier power spectra. Specifically, there was little detectable power in fluctuations in excess of measurement noise on timescales of minutes to several hours according to either method. After that, the power in timescales increases with timescale. In addition, there was significant correlation between the continuous wavelet transforms of densities of the three species on several timescales.

Which approach is more natural depends on whether interaction delays scale with the timescale of the corresponding fluctuation in density (in which case wavelet analysis as

in Figure 6.11 to 6.13 is appropriate) or not (in which case we should resort to classical crosscorrelation functions, Figure 6.6 to 6.9).

In essence, information is needed on what the likely causes are of observed fluctuations. There are a few candidates. First deterministic systems can exhibit fluctuations as seen in examples ranging from simple oscillations in two-dimensional linear systems to chaotic dynamics in three-dimensional nonlinear models or one-dimensional discrete-time models [149]. In our system, processes like the decay of debris, precipitation (Chapter 3) and deterministic phenotypic change can introduce slow timescales and rapid, but delayed, change in conditions.

For stochastic causes there are two main candidates: stochastic phenotypic change, such as by mutations, and, demographic noise: the random timing of individual divisions and deaths, potentially amplified by interactions between species (see, for example Figure 6.2: the density of *T. thermophila* containing chlorophyll ranges from 100 mL^{-1} to less than 10 mL^{-1}). Comparison with the previous chapter shows that stochastic causes are significant. The long timescales of fluctuations appear more consistent with occasional phenotypic change than with the relaxation of demographic fluctuations. It remains to be seen whether a rigorous connection can be established between the fluctuations of densities in individual ecosystems, as in this chapter, and the divergence observed in densities between ecosystems, seen in Chapter 5.

Appendix 6A.1 Calculation of the distribution of coarse-grained counts

In our experiments the raw observations for each species were series of counts per image $\{N_t\}$, with $t = 1, ..., T$ [62]. It is natural to coarse-grain these observed counts using a moving average to obtain population density estimates over time. Since these counts are serially correlated, a null model is necessary describing fluctuations in $N_T = \sum_{t=1}^{T} N_t$ under counting noise, with T the width of an unweighted moving average. I will proceed in two steps:

(1) calculation of the moment generating function for N_T and the moments of N_T.

(2) approximation of $P(N_T)$ by an exponential form given the first four moments of N_T.

To simplify the problem, I will make two additional assumption about the dynamics of objects in the observation volume.

The first assumption is that the dynamics are Markovian: $P(N_t|N_{t-1}, N_{t-2}, ...) = P(N_t|N_{t-1})$. This is an approximation, but it allows for convenient factorization in the calculation of the moment generating function below. Extension to finite memory is, however, straightforward. Secondly, organisms are assumed to move in and out of the observation volume independent of each other, i.e., the process is characterized by only two

[62] To keep notation simple, a subscript t will be used for time in this appendix. All results apply to individual species, so no species index is required.

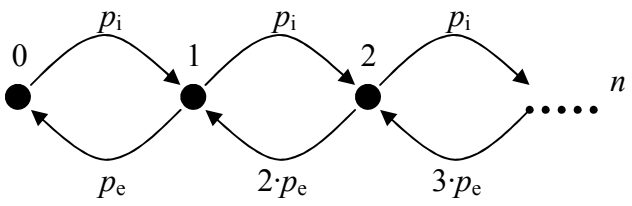


Figure 6A.1: model of count statistics: the process is assumed memory-less and objects are assumed to move in and out of the field of view independently. Arrows are marked with the corresponding transition rates.

---

parameters, p_i (entrance probability) and p_e (escape rate). Figure 6A.1 illustrates this. Its stationary distribution is a Poisson distribution with mean, $\rho = p_i / p_e$.

Moment generating function formalism

Let
$$G(x) = \sum_{\{N_t\}} P(\{N_t\}) \cdot x^{\sum_t N_t}, \tag{6A.1}$$

be the generating function for $N_T = \sum_{t=1}^{T} N_t$. We can describe $P(\{N_t\})$ using a transition matrix M, such that $M(i,j)$ is the probability of observing j counts one image after observing i counts, with $i, j = 0, 1, 2, \ldots$. That is,

$$\begin{aligned} G(x) &= \sum_{\{N_t\}} \left(P(N_1) \prod_{t=1}^{T-1} M(N_t, N_{t+1}) \right) x^{\sum N_t} \\ &= \sum_{\{N_t\}} \left(P(N_1) x^{N_1} \right) \cdot \prod_{t=1}^{T-1} \left(M(N_t, N_{t+1}) x^{N_{t+1}} \right) \\ &= \sum_{\{N_t\}} \left(P(N_1) x^{N_1} \right) \cdot \prod_{t=1}^{T-1} \left(\widetilde{M}(N_t, N_{t+1}) \right) \\ &= \sum_{N_1} \left[\left(P(N_1) x^{N_1} \right) \cdot \sum_{N_T} \widetilde{M}^{T-1}(N_1, N_T) \right] \end{aligned} \tag{6A.2}$$

With $\widetilde{M}(N_t, N_{t+1}) = M(N_t, N_{t+1}) x^{N_{t+1}}$. Under the assumptions described above,

$$M = \exp(Q \cdot \Delta t) \tag{6A.3}$$

with Q the instantaneous rate matrix of the process and Δt the time step between frames. To be precise [156],

$$Q = \begin{pmatrix} -p_i & p_e & 0 & 0 & \ldots \\ p_i & -(p_i + p_e) & 2p_e & 0 & \ldots \\ 0 & p_i & -(p_i + 2p_e) & 3p_e & \ldots \\ 0 & 0 & p_i & -(p_i + 3p_e) & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \end{pmatrix} \tag{6A.4}$$

We already know that $P(\{N_t\}) = P(N_1)\prod_{t=1}^{T-1} M(N_t, N_{t-1})$ and $P(N_1) = \frac{\rho^{N_1}}{N_1!} e^{-\rho}$, so $G(x)$ can be evaluated for each x. Then,

$$\left[\frac{d^n}{dx^n} \log G(x)\right]_{x=1} = \left\langle N_T^{\,n} - \left\langle N_T \right\rangle^n \right\rangle, \tag{6A.5}$$

that is, the n^{th} central moment, or, alternatively,

$$\left[\frac{d^n G(x)}{dx^n}\right]_{x=1} = \left\langle N_T^{\,n} \right\rangle, \tag{6A.6}$$

the n^{th} raw moment.

In practice, these moments are evaluated as follows: $G(x)$ is evaluated over the interval [0.97, 1.03] and a sixth-order polynomial fitted over this range. Differentiation of this polynomial then yields the desired moments. This works well for both equations (6A.5) and (6A.6). The full distribution $P(N_T)$ is then approximated by an exponential form $\exp(a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4)$, with the parameters a_0, a_1, …, determined by the

moments $\langle N_T{}^n \rangle$, found by a modified Newton-Rhapson algorithm ([163], G. Iyengar, pers. comm.). $P(N_T)$ is now determined by only three parameters: the mean frequency of observation, ρ, the escape rate from the field of view, p_e, and the length of the time window, T.

6A.2 Statistical evaluation of the null model

The aim is to compare the model for fluctuations, $P(N_T)$ to the observed histogram of total counts in time windows of length T, $\hat{f}(N_T)$. However, the observed histogram $\hat{f}(N_T)$ is itself an estimate of the true underlying distribution of fluctuations, $f(N_T)$. This was addressed by classical hypothesis testing using a metric provided by information theory for the distance between two distributions, the relative entropy or Kullback-Leibler distance [153]. For two discrete distributions, p and q, that is:

$$D(p\|q) = \sum_i p_i \log \frac{p_i}{q_i} \tag{6A.7}$$

Histograms of the total number of counts per window, N_T, for 100,000 frame periods (≈ 1 day) divided into 100 non-overlapping windows of 1000 frames each were determined (see Figure 6.4, main text). Hypothesis testing then proceeds in two steps:

(1) Given the null hypothesis, that $P(N_T)$ accurately describes the fluctuations, 1000 sets of simulated data are generated with the same mean number of observations as the real data (per day).

(2) For each of these simulated data sets the empirical histogram of $D(f\|P)$ under the null hypothesis is then calculated. This histogram provides a *p*-value for the observed $D(\hat{f}\|P)$.

Relative entropies are used in information theory to distinguish whether observed messages are typical for a source [153]. For several notions of typicality, and convergence properties of sample estimates of the relative entropy, please see S.-W. Ho and R. W. Yeung, "On information divergence measures and a unified typicality" (http://iest2.ie.cuhk.edu.hk/~whyeung/publications/preprint.html).

6A.3 Further notes on wavelets

Wavelets were introduced in the text, some more details are described in this section. Wavelets satisfy an admissibility condition necessary for invertibility of the transform. Let $\psi(t)$ be a mother-wavelet[63], and $\hat{\psi}(\omega)$ its Fourier transform:

$$\int_{-\infty}^{+\infty} d\omega \, \frac{|\hat{\psi}(\omega)|}{\omega} < \infty \tag{6A.8}$$

Which implies:

$$\int dt' \, \psi(t') = 0 \tag{6A.9}$$

A wavelet $w(t|t',\tau)$ on a particular time scale τ, and localized at a time t is a scaled version of the mother-wavelet, satisfying an additional normalization constraint:

$$\int dt \, |w(t|t',\tau)|^2 = 1 \; \forall \tau \quad \Rightarrow w(t|t',\tau) = \frac{1}{\sqrt{\tau}} \psi\left(\frac{t-t'}{\tau}\right) \tag{6A.10}$$

The continuous wavelet transform of a signal $s(t)$ is given by:

$$W_s(t',\tau) = \int dt \, s(t) \cdot \frac{1}{\sqrt{\tau}} \psi\left(\frac{t-t'}{\tau}\right) \tag{6A.11}$$

In addition, some wavelets form an orthonormal basis, allowing for non-redundant decomposition of a signal. Such a basis is shown for the simplest of wavelets, the Haar

[63] That is, the basic shape of the wavelet. I will assume the wavelet to be real for simplicity, but the results directly extend to complex wavelets.

wavelet (Figure 6A.1 (a) and (b)). For this basis, $\tau_j = 2^j$ and $t'_{j,k} = \tau_j \cdot k$, with $j,k \in Z$. As for any orthonormal basis, a variant of Parseval's theorem holds. Specifically, let

$$C_{jk} = \int dt\, s(t) \cdot \frac{1}{2^{j/2}} \psi\left(\frac{t - 2^j k}{2^j}\right), \tag{6A.12}$$

then,

$$\sum_{j,k} \left|C_{jk}\right|^2 = \sum_t \left|s(t) - \bar{s}\right|^2 \tag{6A.13}$$

with $\bar{s}$ the mean of the signal. This decomposition also allows us to determine the power per time scale [5], $\sum_k \left|C_{jk}\right|^2$. An example of an orthogonal wavelet basis is shown in Figure 6A.2.

The wavelet used throughout this chapter is the Daubechies-3 wavelet (db-3, see Figure 6A.2(c)). In the text I mentioned that it is insensitive to addition of a third-order polynomial to the signal. This property is known as the number of vanishing moments, i.e. 3 for the db-3 wavelet. The special property of the class of Daubechies wavelets is that they are the most compact orthogonal wavelets, that is they have smallest non-zero support, for each corresponding number of vanishing moments [161]. This tends to make their shape irregular. The db-3 wavelet is the first one in the series, by ascending number of vanishing moments, to be once-differentiable. In fact, the Haar wavelet is the Daubechies wavelets with one vanishing moment.

Wavelets suitable for the analysis of signals with periodic components are also available and are very similar to traditional approaches like the Wigner-Ville transform and

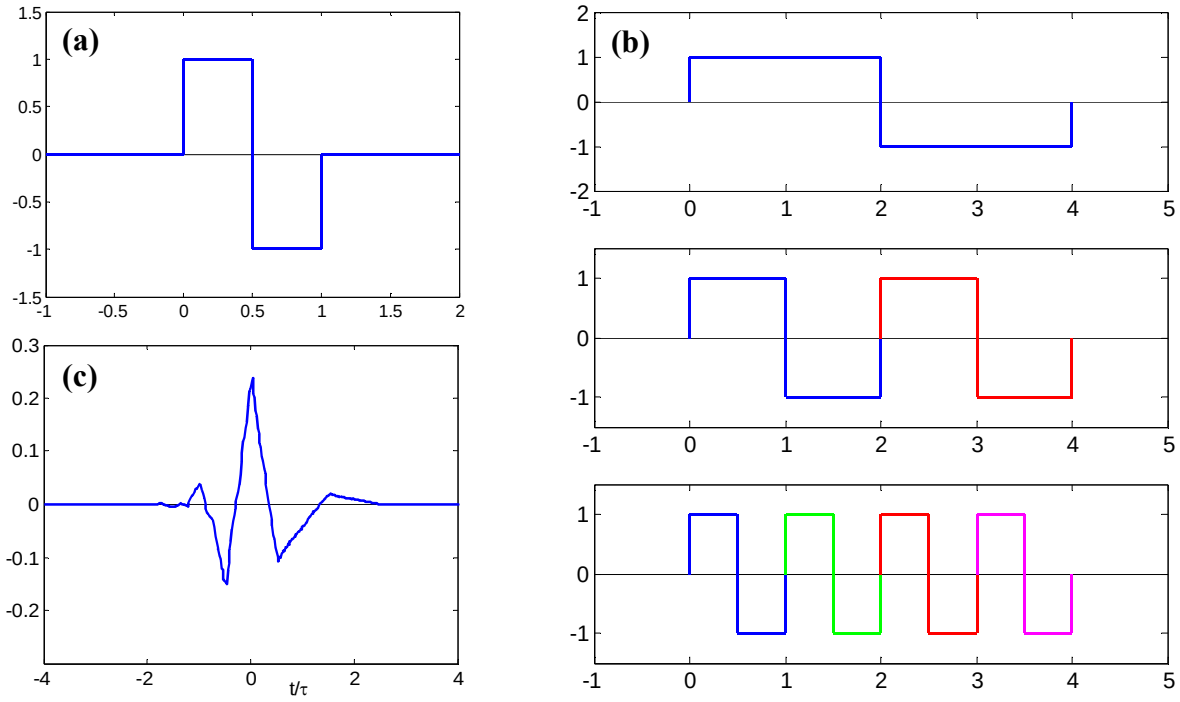


Figure 6A.2: (a) the most basic wavelet, the Haar wavelet, (b) discrete wavelet transform demonstrated for the Haar wavelet, using unnormalized wavelets (to emphasize these are individual wavelets, not block waves, different colors are used). As can be seen, the wavelets are orthogonal and can form a complete, non-redundant base for signal decomposition. (c) the Daubechies-3 wavelet used in the text.

windowed Fourier transforms (see [161] for an extensive description of wavelets and their relation to traditional transforms).

Appendix 6A.4 Additional Figures

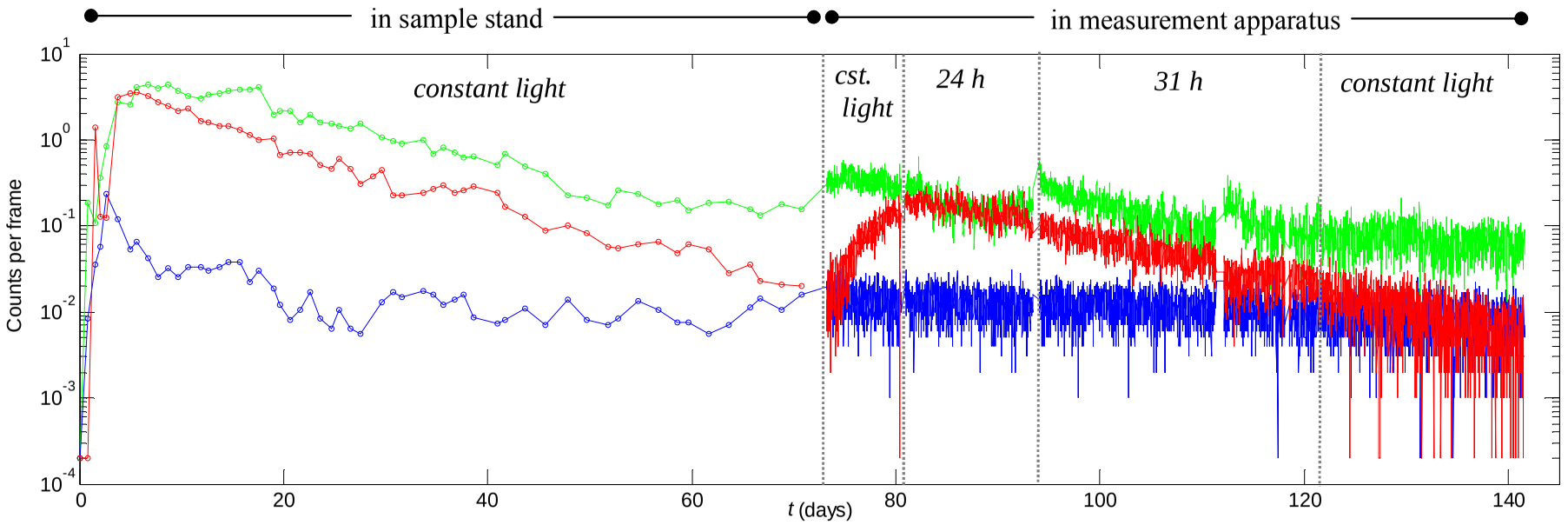


Figure 6A.3: Overview of experiment 1. The measurements over the first 70 days were described in chapter 5 (where the ecosystem was number 4 in experiment 1). After day 70, measurements were performed almost continuously. Most of the constant light section after day 122 is described in this chapter in data set 1 (the other part was not included because of an intervening measurement of the other 8 ecosystems of experiment 1, chapter 5, although the gap in the data is too small to be seen on this scale). The data on 24 and 31 h period modulation of illumination are described in chapter 7.

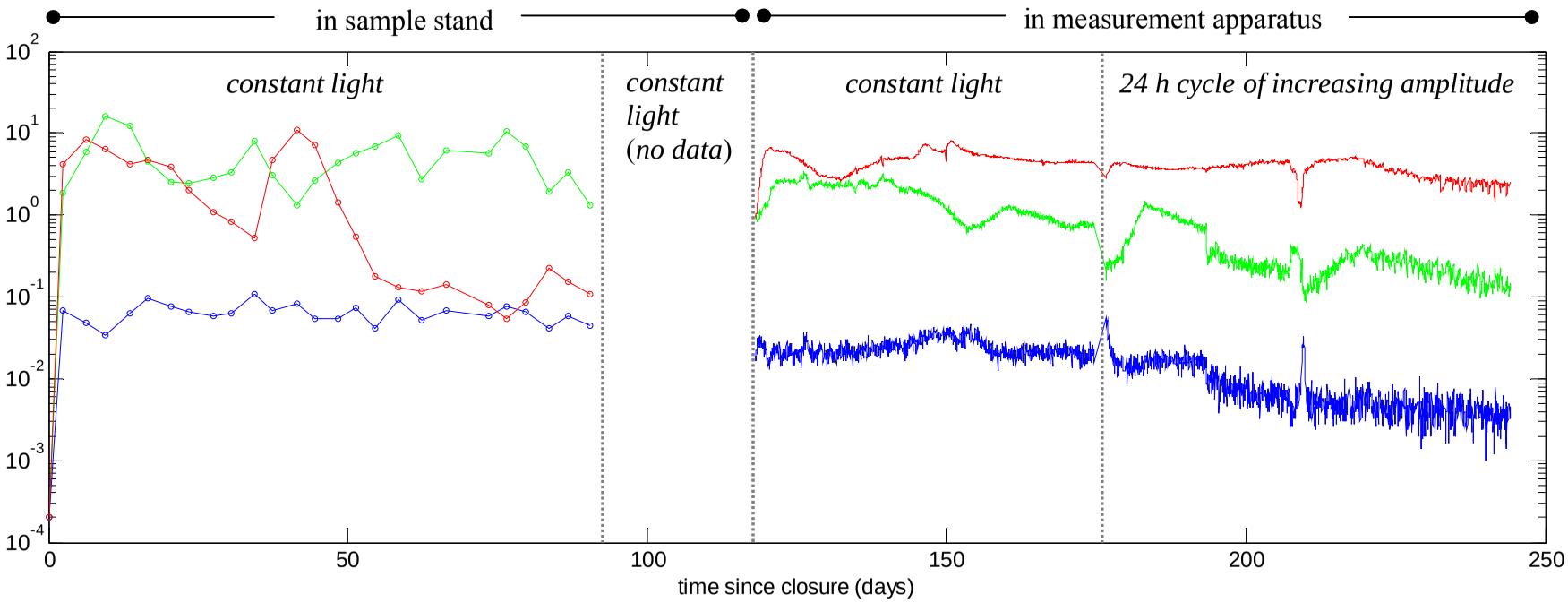


Figure 6A.4: Overview of experiment 2. The measurements over the first 90 days were described in chapter 5 (where the ecosystem was number 4 in experiment 1). After day 118, measurements were performed almost continuously. The constant light section after day 118 is described in this chapter in data set 2. The data on 24 h period modulation of illumination is described in chapter 7.

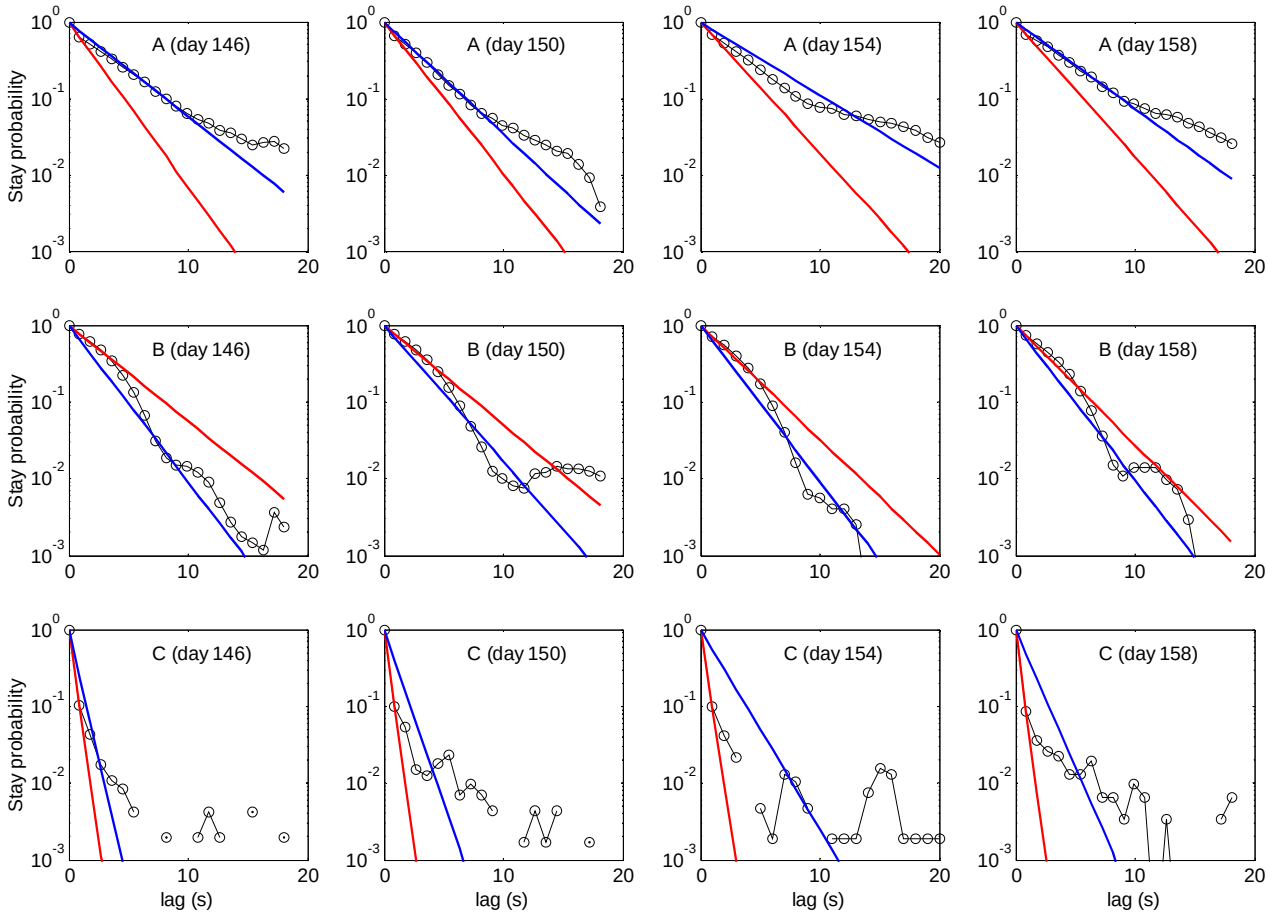


Figure 6.A.5: Logarithmic version of Figure 6.3. Probability for individuals of the different species to stay in the field of view in subsequent frames given that they were observed at time 0. Each curve was calculated over a 1 day window as indicated (black circles). Data based on Figure 6.1. Red: expectation for a memory-less process with decay time given by the decay over the first lag. Blue: expectation based on memory-less process with decay time given by the entire curve.

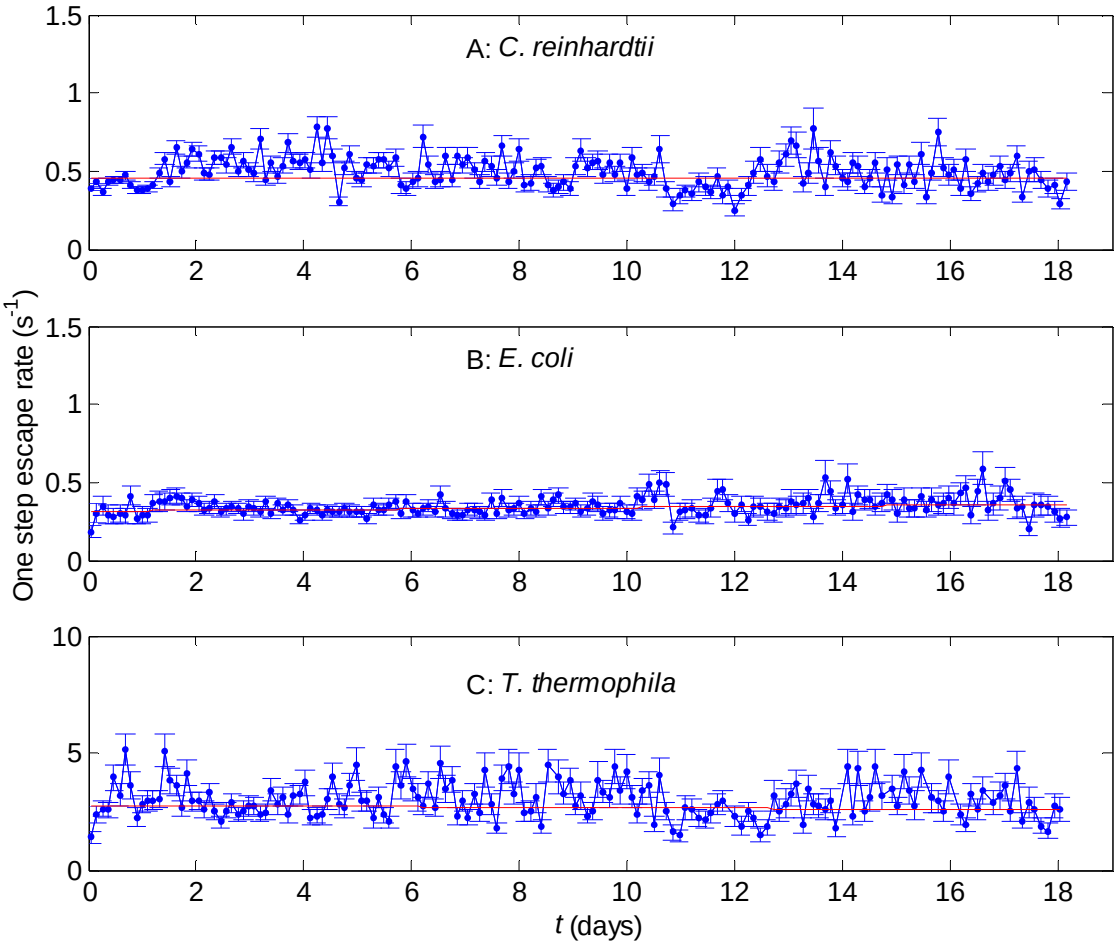


Figure 6A.6: Escape rate variation over a period of 18 days (experiment 1). The one-step estimate of escape rate (see text) was used because more accurate estimates can be obtained over small time windows. In this graph each estimate was over a 10,000 frame window (or 2.5 h). Slopes in A and C were not significantly different from 0 ($p > 0.1$). For B, there was a 0.7% increase per day (or 0.5% after second-order bias correction), which was significant ($p \approx 10^{-4}$). Error bars were estimated from the number of observed escapes from the field of view.

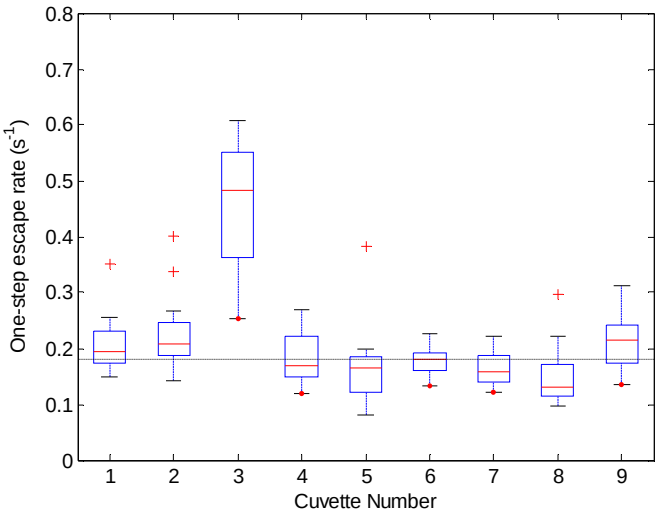


Figure 6A.7: variation in escape rate (estimated from one lag) for the set of cuvettes in Experiment 1, Chapter 4.
For each cuvette the one-step escape rate was estimated for 16 different measurement days over the first six weeks of the experiment. Each box contains the 25-75% quartiles with the median indicated in red.
The dashed black line is a guide to the eye.

Chapter 7 Population Dynamics in a Closed Ecosystem under Modulated Illumination

Abstract

This chapter describes initial experiments on the effects of the perturbation of a boundary condition, illumination. The perturbations used are sinusoidal modulations of light intensity with periods of 24 and 31 hour. Both *C. reinhardtii* and *E. coli* densities respond strongly to modest perturbations (relative amplitude ≥ 10%). *E. coli* shows positive autocorrelation for both half and whole period lags. The interpretation of these results is complicated by a number of factors, however. For example, both migration within the ecosystem and growth and death can contribute to the observed response in density. The circadian clock of *C. reinhardtii* almost surely affects its response to modulation of the illumination. Finally, the observed response is itself variable over time. I describe experiments aiming to address these issues and propose additional experiments.

Introduction

The study of the effects of external forcing, that is, of the perturbation of boundary conditions, on ecosystems is important for two reasons. First, natural ecosystems do not have constant boundary conditions. Instead, even if they experience little material transport or migration, they are subject to variations of abiotic conditions such as light and temperature. For natural ecosystems, the multitude of dimensions of the boundary conditions makes it hard to discern the effects of any one of them. Closed ecosystems allow for the elimination of material fluxes across the system boundaries and tight control of the remaining boundary conditions, light and temperature (chapter 2). In the previous chapter, I described intrinsic fluctuations in population densities under constant temperature and illumination. It is an important question how these intrinsic fluctuations are affected by external forcing, if any of the results in this thesis are to be extended to natural ecosystems.

The perturbation of boundary conditions is important for a second, related reason. Imagine for a moment that a separation of ecological and evolutionary time scales is possible for our closed ecosystem, such that on short time scales the dynamics are (demographically driven) fluctuations around a fixed point. This can, for example, be the case in the Lotka-Volterra equations introduced in chapter 1. On longer time scales, in this scenario, the parameters in the model would be subject to the vagaries of evolution. In such a scenario, studying the dynamics on long time scales does not provide much information on the parameters of the model, but only on their changes. Instead, if the system could be deliberately shifted from its fixed point, the ensuing relaxation

dynamics, for small deviations, would provide information on (combinations of) parameters of the model.

There are many examples of the effects of forcing on ecosystems in the literature. I will discuss a few representative examples[64]. Ollason [164, 165] subjected an undefined set of species in a continuous flow reactor to illumination with a period of 4 days and a relative amplitude of about 30%. The numbers of individuals in different taxonomic groups were determined once every twelve hours and significant fluctuations were observed for the algae at the illumination frequency. Unfortunately the results were not subjected to any statistical analysis and are probably hard to reproduce.

More recently, Fontaine and Gonzalez [166] explored the effects of resource fluctuations on the population density of a rotifer. The resource in this case was an alga, *Chlorella vulgaris*, and its densities were controlled by replacing the entire medium every 2 days. The constructed resource fluctuations consisted of sums of two sines with periods of 5 and 9 days, respectively (the approximate generation time and total life span of the rotifer). Replacement of the entire medium seems, however, like a drastic perturbation in its own right, and the dynamics are probably undersampled. Unfortunately, the authors did not show any power spectra, and little can be concluded from their observations.

In a final example, Grover and colleagues [167] studied the effect of thermal forcing on the density of *Daphnia sp.* in freshwater microcosms (species largely undefined). Different measurements were made every 2, 4 and 8 days, with the measurements at 2 and 8 day periods requiring invasive sampling and replacement with fresh medium, which itself was drawn from a pond during various seasons over two years. As in the other examples, the three replicates were open to air. *Daphnia* densities were measured

[64] From reputable authors published in reputable journals.

every four days, while the period of forcing was 6.8 days. The observed power spectra showed dominant peaks at about 30-50 days, roughly the generation time of *Daphnia* in the experiments. Unfortunately, the power spectra did not include (!) the thermal forcing frequency. In summary, in many current studies, even in the laboratory, the measurement frequency is comparable to the forcing frequency, additional frequencies are introduced by the maintenance protocols, and the systems are ill-defined.

In this chapter, I examine the effects of perturbations in the external variable most easily controlled, illumination intensity. The experiments described are exploratory in nature. The starting point is one of the simplest conceivable perturbations: a 24 hour sine wave superimposed on the prior constant illumination intensity. The first experiment I will describe immediately showed both the promise and the pitfalls of the approach. In particular, the experiment raised questions about spatial effects of the perturbation and the role of the circadian clock in the observed response. I will describe additional experiments aimed at addressing these complications and suggest further ones.

Data sets

Data set 1 contains data from measurements on the same ecosystem used in data set 1 of chapter 6 and experiment 1 in chapter 5 (see Figure 6A.3 for its complete history). After experiment 1 of chapter 5, the ecosystem was placed in the measurement apparatus on day 73[65], and was allowed to adjust to the apparatus until day 82 when the perturbation experiments started. First, the illumination intensity was modulated sinusoidally with a 24 hour period for 11.5 days (figure 7.1), then with a 31 h period for 16 days (figure 7.16). The amplitude of the perturbation in both cases was ±17 %.

[65] That is, after 10 weeks of nearly daily measurements.

Data set 2 was acquired from the same ecosystem as data set 2 in the previous chapter and was part of experiment 2 in chapter 5 (see Figure 6A.4). An overview of the data set is given in Figure 7.6, including data on temperature and laser stability. These data are shown because of two disturbances during the experiment. The first was the introduction of two fans in the measurement enclosure to reduce temperature swings due to variation in illumination on day 193 after construction and closure of the ecosystem. This decreased the measured temperature next to the cuvette by 0.6 °C, causing a rapid shift in *C. reinhardtii* density (twofold drop over 2.5 h). It did not, however, affect the amplitude of the measured temperature response to LED intensity variation (±0.1 °C at ±17% amplitude), suggesting heat flow between the enclosure and the environmental room, not flow within the enclosure was limiting. The second disturbance was due to a refrigerant leak of the environmental room, causing a spike in temperature on days 208 and 209 after closure up to 35 °C, or 10 °C higher than normal. The 10 days after this spike were not used for data analysis (the data segments used are highlighted by black bars in Figure 7.6 a, and shown in Figure 7.7).

The third data set was acquired in an additional measurement apparatus designed to measure densities at two points in space (see chapter 4). Densities in a new ecosystem were measured directly after its construction and closure (Figure 7.12). Its construction proceeded as for the other ecosystems, and the ecosystem had the same initial conditions (see chapter 4, appendix). Because of adjustments to the imaging optics over the first five days, these data were not included for the two time series affected. The illumination pattern in this setup was different, as described in chapter 4 and summarized below in the discussion on '*migration versus growth*'.

Results

Response to a 24h sinusoidal perturbation

As an initial exploration of the effects of modulation in light intensity, one ecosystem from experiment 1 in chapter 5 was subjected to illumination with the same prior mean intensity, say μ, with a 17% sine wave superimposed, so $I(t) = \mu \cdot \left(1 + A \cdot \sin\left(\frac{2\pi t}{T}\right)\right)$, with μ = 1200 lux, T = 24 h and A = 0.17. The amplitude, A, was, arbitrarily, chosen to be a modest perturbation. This perturbation also resulted in temperature fluctuations with an amplitude of 0.1 °C, comparable to the long-term room temperature stability. The resulting population dynamics are shown in Figure 7.1. The densities of both *C. reinhardtii* and *E. coli* show an obvious response to the perturbation. This is shown in more detail by their autocorrelation functions (over 10 days, Figure 7.2). Surprisingly, though, *E. coli* density shows positive autocorrelation for both whole- and half-day lags. For *T. thermophila*, no response was evident in its autocorrelation function.

These responses are even clearer in the power spectra (Figure 7.3). Power spectra make better use of the periodicity of the perturbation[66]. For example, for *T. thermophila,* a small but significant peak is now visible at 1 day^{-1}. The signal for *C. reinhardtii* yielded a dominant frequency of 1 day^{-1}, as anticipated based on the autocorrelation function, and a small peak at 2 day^{-1}. The corresponding power spectrum for *E. coli* is more complex. First of all, the localization of peaks is less precise than for *C. reinhardtii*[67], but localization of dominant peaks at 1 and 2 day^{-1} is obvious. Significant peaks at low

[66] Since the power spectrum is, by the Wiener-Khinchin theorem (e.g., [168]) equivalent to the Fourier transform of the correlation function, it can be seen that information of the entire correlation function, not just at a lag of 1 day, contributes to the signal at 1 day^{-1}.

[67] Periodograms are notoriously hard to estimate, see e.g. [2, 169] and the discussion under *Reproducibility and scaling.*

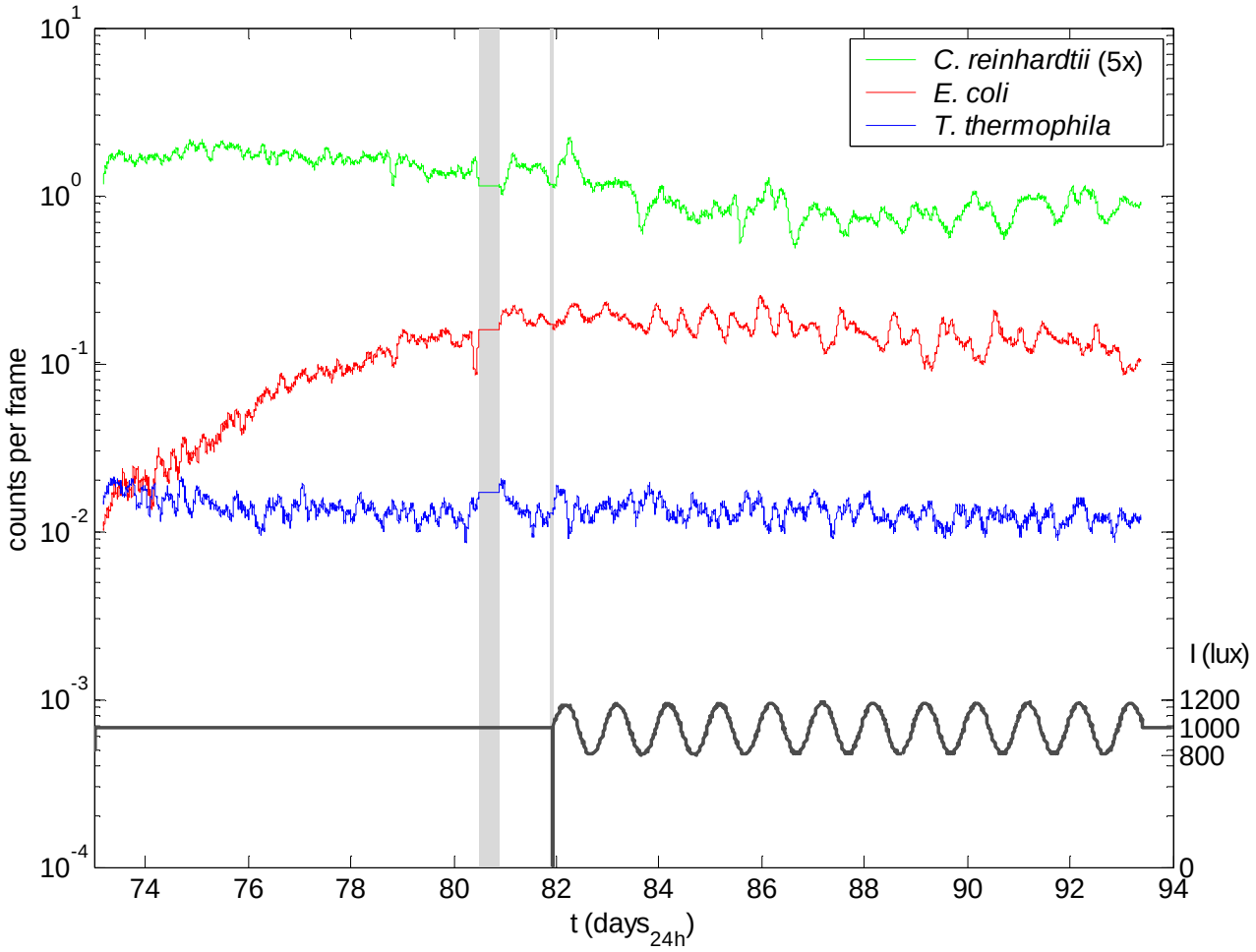


Figure 7.1: Data set 1, 24 hour perturbation. After an initial period of adaptation to the conditions in the measurement apparatus, the illumination of the sample was modulated (gray line). For clarity, the *C. reinhardtii* counts were multiplied by 5. Density estimates were obtained by smoothing the observed counts with a 2.5 h moving average. The illumination intensity is shown in gray.

frequencies ($\leq$ 0.5 day^{-1}) suggest removal of nonstationarity effects by subtraction of a third-order polynomial fit was incomplete, but largely effective.

The *E. coli* autocorrelation function appears qualitatively different from the autocorrelation function of *C. reinhardtii*. To make this more rigorous, the sums of 'integer' cosine and sine components, i.e. 1, 2, …, 6 day^{-1}, are plotted against each other (Figure

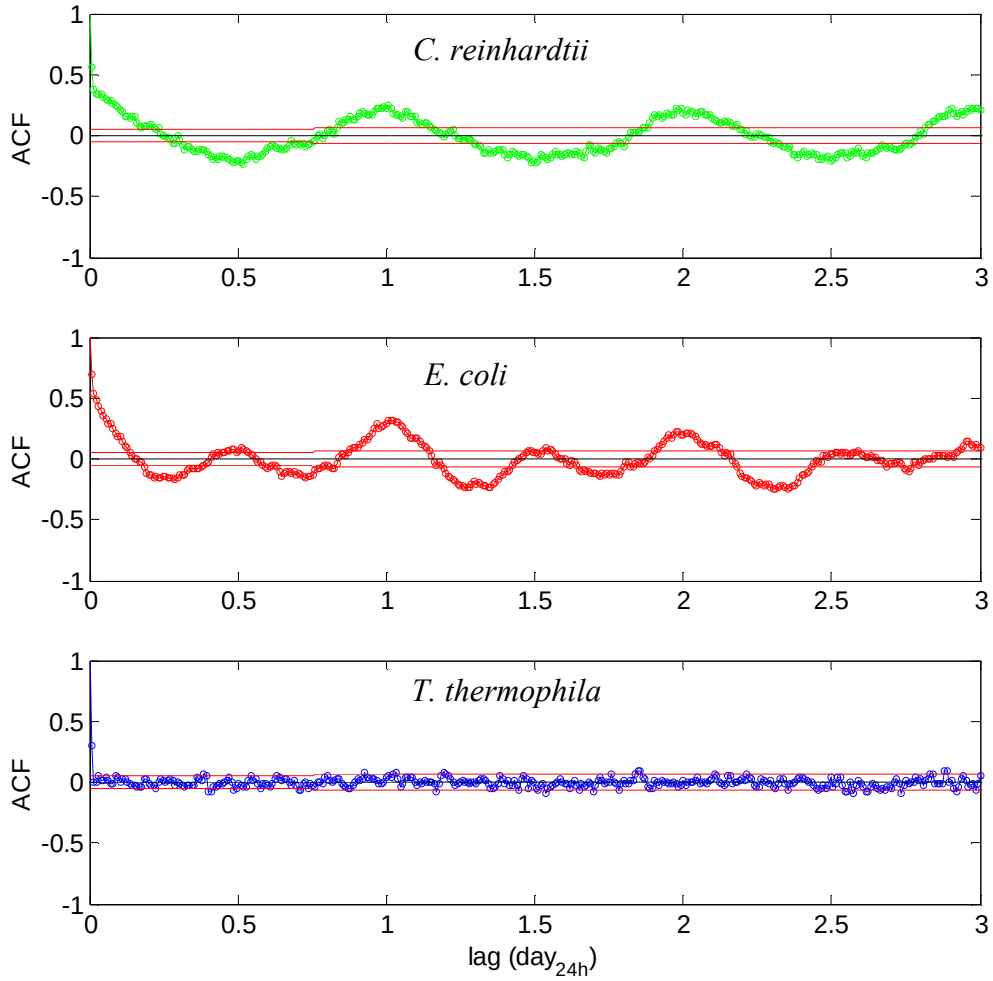


Figure 7.2: Autocorrelation functions for detrended population density time series over a 10-day window of data set 1 (day 83-93).
Data were coarse-grained using a 1000 frame moving average (15 minutes) and interpolated to an approximately 1 minute grid, so Δt = 1 minute. Data were then detrended by subtraction of a third-order polynomial fit. Application of the moving average leads to some decrease in spectral power for frequencies > 100 day^{-1}.
Red lines indicate 2 standard deviations from the mean for uncorrelated data ([2], the null model from chapter 6 yields nearly identical results).

7.4). Taking only integer modes into account is equivalent to performing Fourier transforms for one-day windows and averaging them.

An alternative interpretation of these 'phase plots' is as follows. The sine and cosine components of a particular Fourier component, A_j and B_j, respectively can be considered to define a complex number:

$$z_j(t) = A_j \cos(2\pi\nu_j t / T) + i \cdot B_j \sin(2\pi\nu_j t / T), \tag{7.1}$$

with $\nu_j = 1, 2, \ldots$ day^{-1} and $T = 1$ day. For each number, a phase $\varphi_j = \arctan(B_j / A_j)$ can be defined[68]. Likewise, a phase can be defined for a sum of complex numbers, that is[69]:

$$\varphi(t) = \arg\left(\sum_j z_j(t)\right), \text{ with } \arg(\cdot) = \arctan\left(\frac{\mathrm{Im}(\cdot)}{\mathrm{Re}(\cdot)}\right) \tag{7.2}$$

A winding number quantifies how often a curve winds around a particular point (e.g. [149]). For the phase plots in Figure 7.4, this offers a qualitative distinction. To be precise, the winding number for a closed curve, C, around the origin is:

$$WN = \frac{1}{2\pi}\int_C dt\, \varphi(t) \tag{7.3}$$

For data without measurement noise, the phase of the signal is defined anywhere except at the origin. For data in the presence of measurement noise, the phase can be approximately determined anywhere except within a region around the origin, the radius of which depends on the amount of measurement noise (indicated by red and magenta

[68] With the signs of A and B determining the domain of the transformation.
[69] These are conventional aspects of complex numbers and can be found in any text book (e.g. [170]).

circles in Figure 7.4). As shown in that figure, the phase of the density time series can be inferred for *C. reinhardtii* and *E. coli*, and their respective winding numbers are 1 and 2 day^{-1}, respectively. For *T. thermophila*, no unambiguous phase and winding number determinations are possible.

There are statistically significant features in the crosscorrelation functions of all three pairs of species (Figure 7.5), but unlike for dynamics in unforced systems, such correlations do not necessarily imply interactions. The data, however, strongly suggest that the fluctuations in *E. coli* are caused by its interaction with *C. reinhardtii*, since *E. coli* is commonly thought not to have a circadian clock (because of unpublished negative results, http://circadiana.blogspot.com/, [171]).

Before conclusions can be drawn about the existence and nature of such an interaction, however, some questions need to be addressed. First of all, can the results be reproduced? Secondly, how do the observed periodic density fluctuations scale with the amplitude of the perturbation? For example, if the amplitude of the response were proportional to the amplitude of the perturbation, it could simplify analysis. Thirdly, are the observed density fluctuations the result of migration, or do they result from actual growth and death of organisms? Fourth, what, if any, is the role of the circadian clock in the response of these organisms? Fifth, are the observed patterns stable over many cycles of the illumination? Sinusoidal perturbations allow for making efficient use of obtained time series and the detection of small effects [172], but these effects need to be stable themselves. Finally, can the existence of an interaction be proven in the presence of continuous forcing? I will consider these questions in turn.

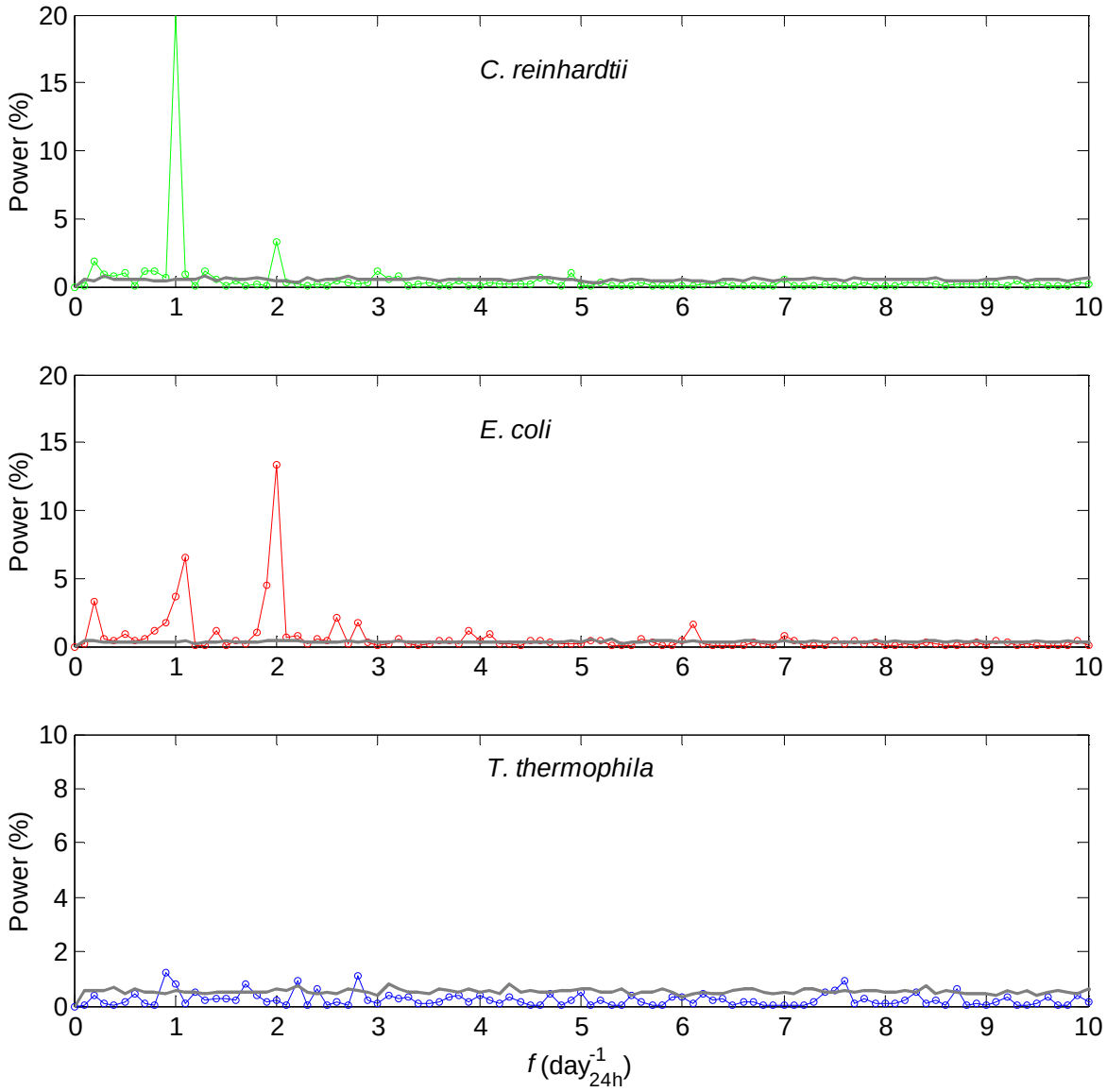


Figure 7.3: Fourier power spectra (periodograms) for data set 1 under 24 h forcing with ±17% amplitude. The power spectrum was calculated over day 83-93 as for figure 7.2. Power is expressed as a percentage of the total power in the time series.
The gray line indicates the mean + 2 standard deviations of 50 surrogate data sets generated from the null model described in chapter 6.
The Matlab Fast Fourier Transform distributes power equally between each frequency ω and its aliasing partner frequency $\omega' = \omega_N - \omega$, with $\omega_N = \pi/\Delta t$ the Nyquist frequency. (In essence, for a regularly spaced time series, $\cos(\omega t)$ cannot be distinguished from $\cos((\omega_N - \omega)t) = \cos(-\omega t)$.) For the power spectra in this chapter, all of this power was allocated to the lower frequency, also for surrogate data obtained from the null model (hence signal to noise ratio estimates are conservative).

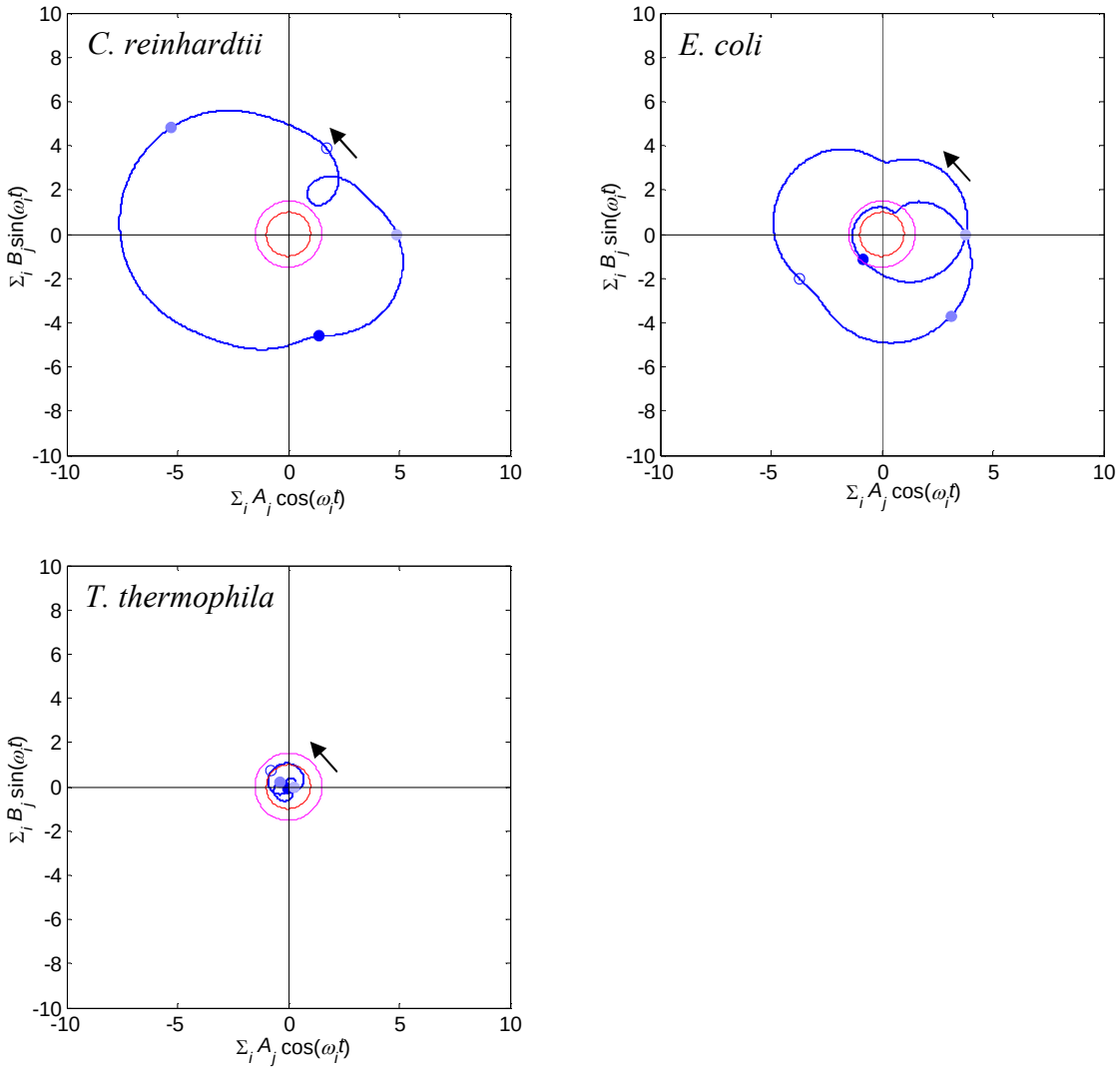


Figure 7.4: Phase plots for all three species in data set 1 under 24 h forcing, based on the first six integer modes (blue lines). Arrows indicate the direction of progression through time, dark dots the moment of least illumination, light dots of maximum illumination, and pale blue dots the illumination midpoints. All curves were rotated to give the upward midpoint of illumination zero phase, that is lying on the positive *x*-axis ("6 AM" if the moment of least illumination is set to 12 AM). The red circle indicates the mean values for surrogate data, and the magenta line their mean + 1 standard deviation. The values of the sums on both axes were normalized by division by the corresponding mean absolute value for surrogate data.

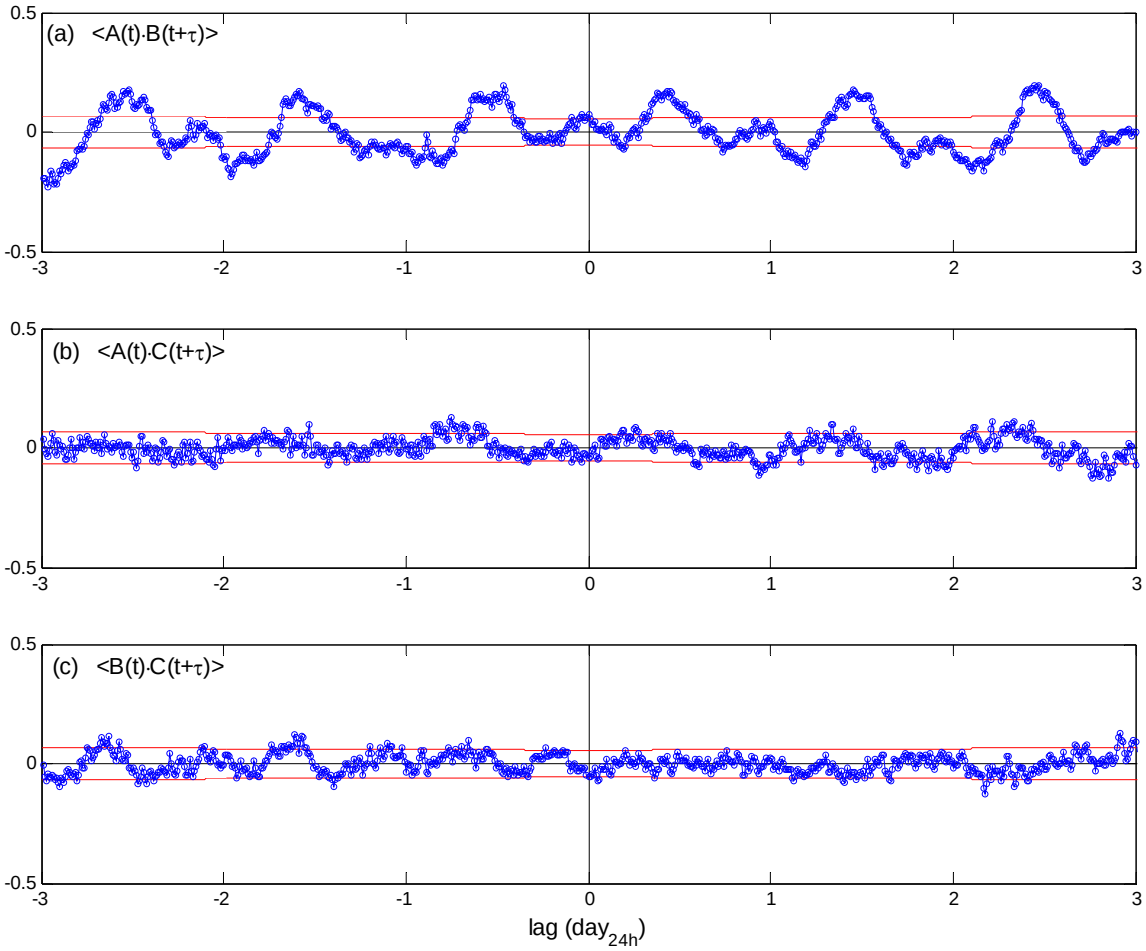


Figure 7.5: Crosscorrelation functions for data set 1 during illumination modulated sinusoidally with a 24 h period (see caption figure 7.2). Crosscorrelation function for (a) *C. reinhardtii* and *E. coli* (b) *C. reinhardtii* and *T. thermophila* and (c) *E. coli* and *T. thermophila*. The red lines indicate two standard deviations from the mean for uncorrelated data. Data treatment is described in the caption of figure 7.2.

Reproducibility and scaling

Experiment 2 aimed to address the first two questions. Namely, can the features of the power spectrum of the first experiment be reproduced? And, how does the power in the different integer modes scale with the amplitude of the perturbation? In experiment 2, an ecosystem (176 days old at the start of the experiment) was subjected to sinusoidally modulated illumination at four increasing amplitudes, 3.2, 5.6, 10 and 17% (logarithmically spaced). As mentioned above, there were two disturbances during the experiment, yet four windows of 10 days could be isolated which appeared to be free of the effects of these disturbances. The experiment is summarized in Figure 7.6, and the four 10-day windows are shown separately in Figure 7.7. An inspection of the power spectra in Figure 7.8 suggests two things. First, the same qualitative features are seen as in experiment 1 for the 17% amplitude. The *C. reinhardtii* response is dominated by the 1 day^{-1} mode. For *E. coli* there is significant power in the spectrum around both 1 and 2 day^{-1}. In contrast to the first experiment, however, the reconstructed phase, using only integer modes, has a winding number of 1 day^{-1} for both *C. reinhardtii* and, with considerable ambiguity, *E. coli* (Figure 7.9, the phase curve enters the region of indeterminate phase indicated by the red circle, around time of maximum illumination).

Figure 7.6 (next page): Overview of the second experiment: (a) counts per frame after application of a 66 minute moving average for *C. reinhardtii* (green, x5 for display), *E. coli* (red) and *T. thermophila* (blue, x50 for display). The black bars indicate the four 10-day windows which were analyzed in more detail in figures 7.7 to 7.11. (b) temperature of one of the lasers. (c) temperature next to cuvette (d) light level next to the cuvette. The dips in the signal are due to interference between the light measurement frequency and the imaging frequency (during imaging the LEDs are off; the sensor measurement frequency was adjusted to reduce such interference at about day 210). (e) light output from the optical fiber controlling for pointing instability.

At about day 207 after closure, a refrigerant leak in the environmental room caused a dramatic rise in temperature. This region is approximately marked between two dotted lines. At day 194 an additional fan was introduced to improve thermal equilibration within the setup enclosure. Despite this, modulation of the LED intensities resulted in temperature fluctuations, suggesting equilibration between the measurement enclosure and the environmental room, not equilibration within the enclosure, was limiting.

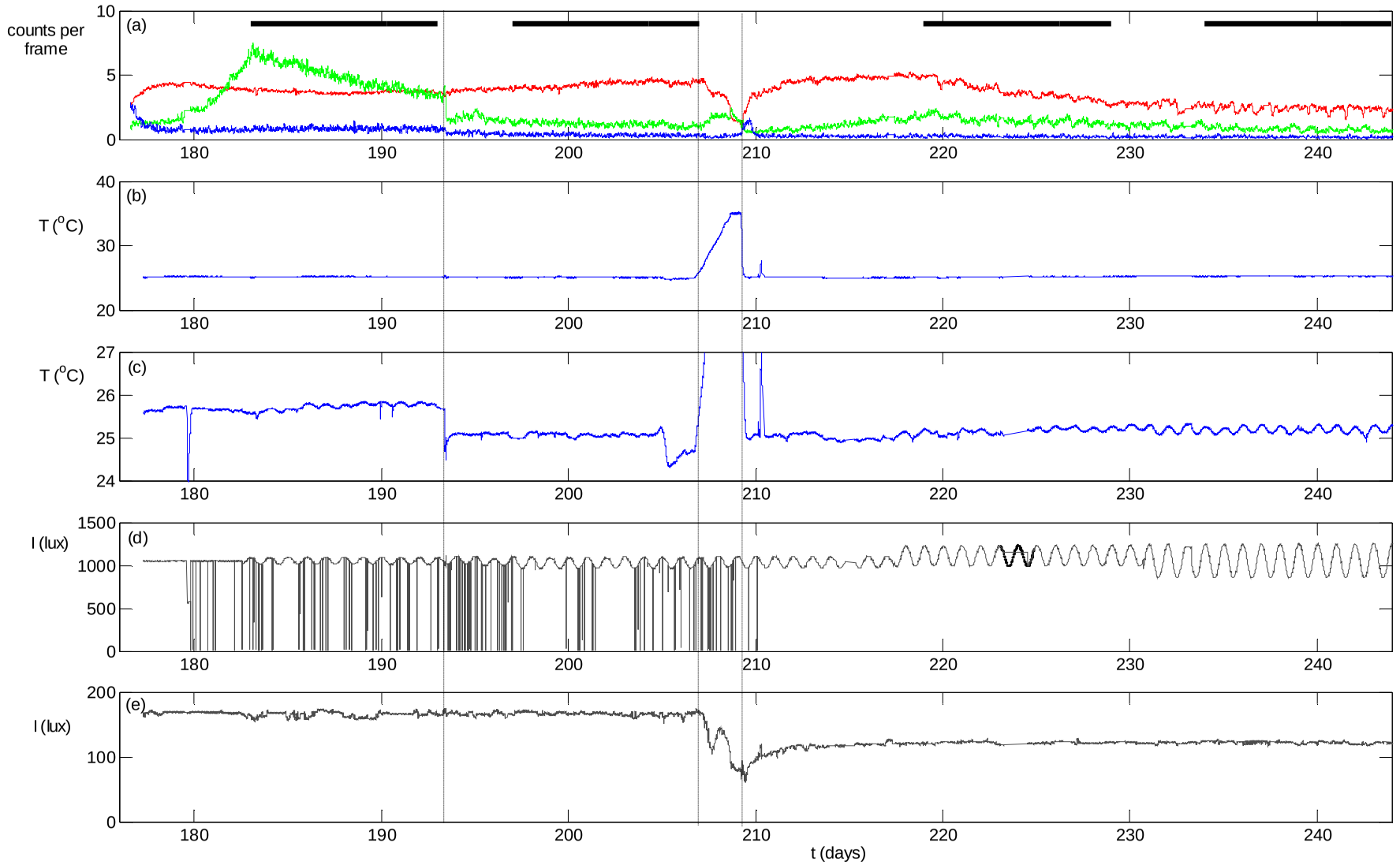

counts per frame
(a)
10
5
0
180
190
200
210
220
230
240
T (°C)
(b)
40
30
20
T (°C)
(c)
27
26
25
24
I (lux)
(d)
1500
1000
500
0
I (lux)
(e)
200
100
0
t (days)

Figure 7.6

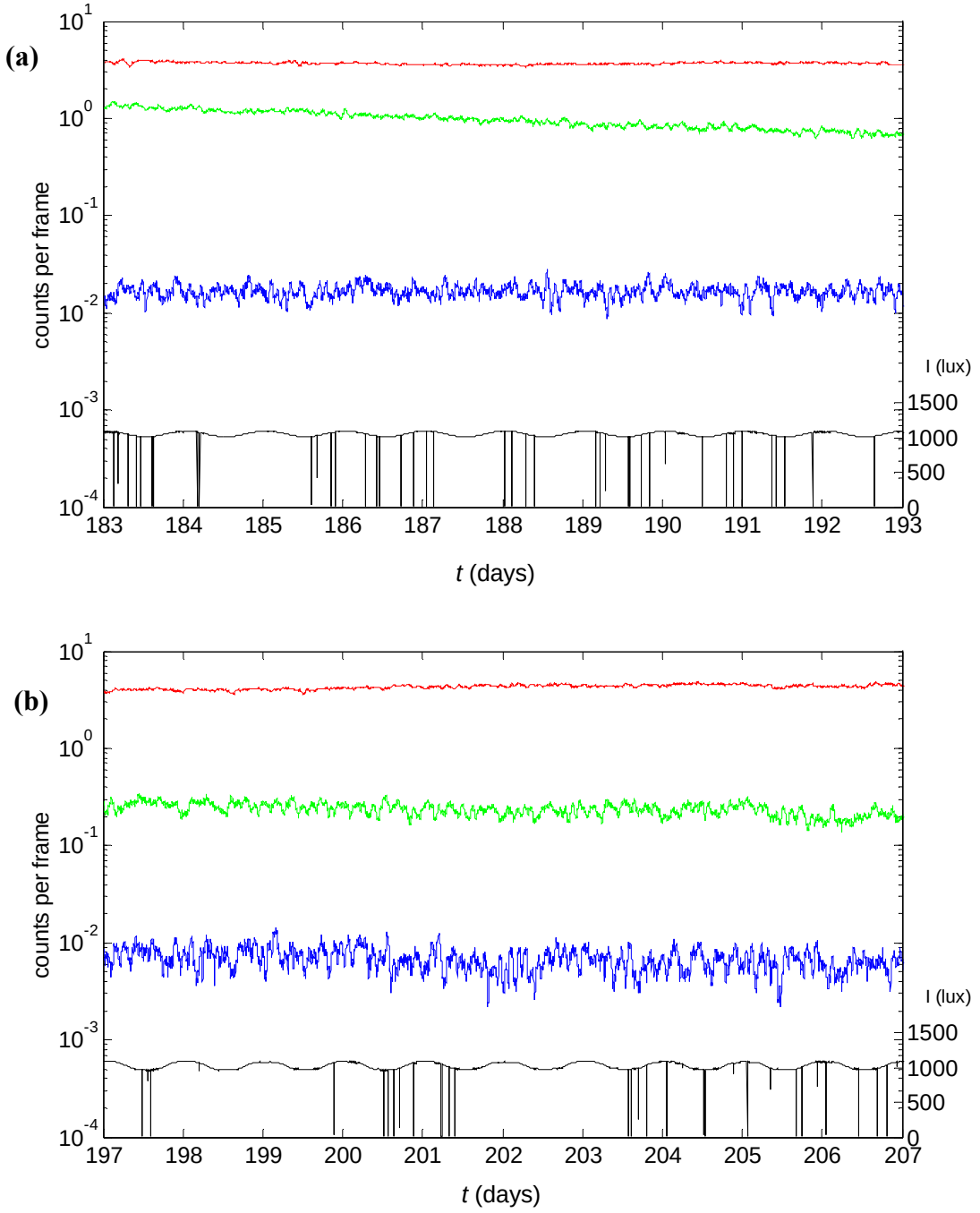


Figure 7.7: Close ups of the analyzed segments of figure 7.6. *C. reinhardtii*, green; *E. coli*, red; *T. thermophila*, blue; illumination intensity, black. Amplitude of illumination: (a) 3.2%, (b) 5.6%, (c) 10%, (d) 17%. A short segment of measurements of illumination intensity is missing due to an acquisition defect. The intended intensity is shown by a dotted line around day 224, dips are due to coincidence of photodiode measurements with exposure (LEDs off).

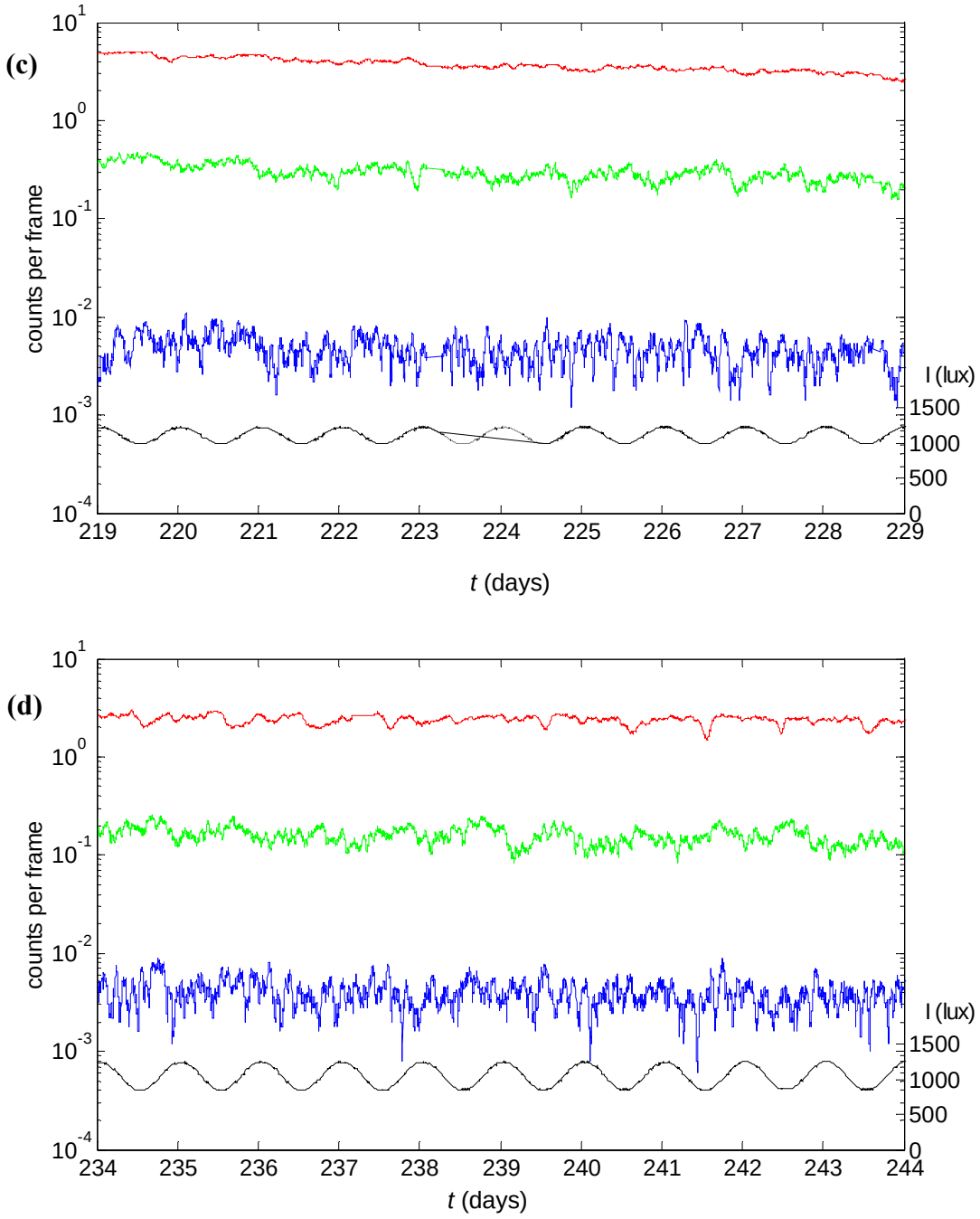


Figure 7.7 (continued)

Figure 7.8 (next two pages): Power spectra for data set 2 at four amplitudes of the illumination modulation, corresponding to segments (a)-(d) in figure 7.7, also indicated in the upper right corner of each panel. *C. reinhardtii*, green; *E. coli*, red; *T. thermophila*, blue. Data treatment followed the description in figure 7.2 and 7.3. The gray line indicates the mean + 2 standard deviations for surrogate data.

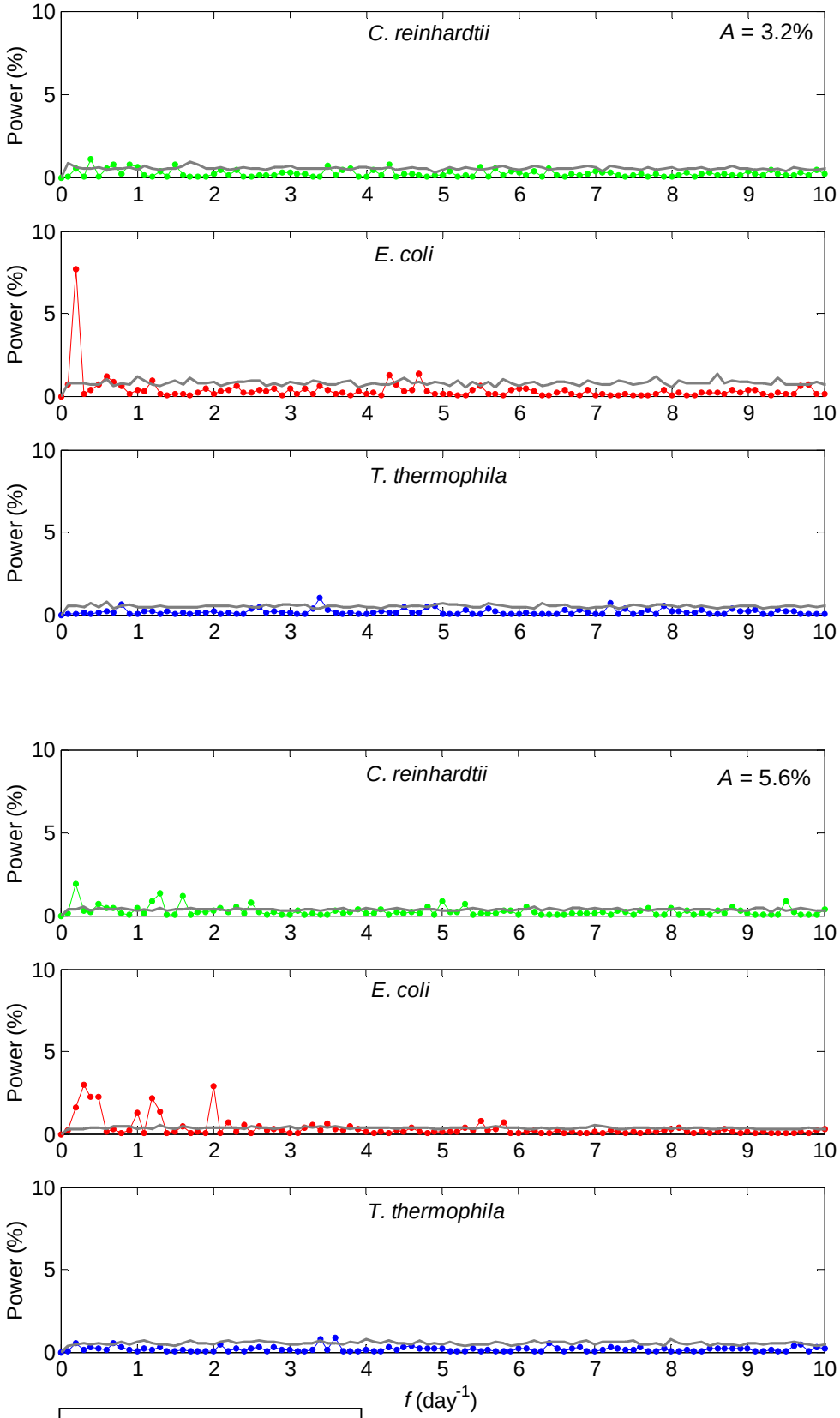


Figure 7.8 (continued)

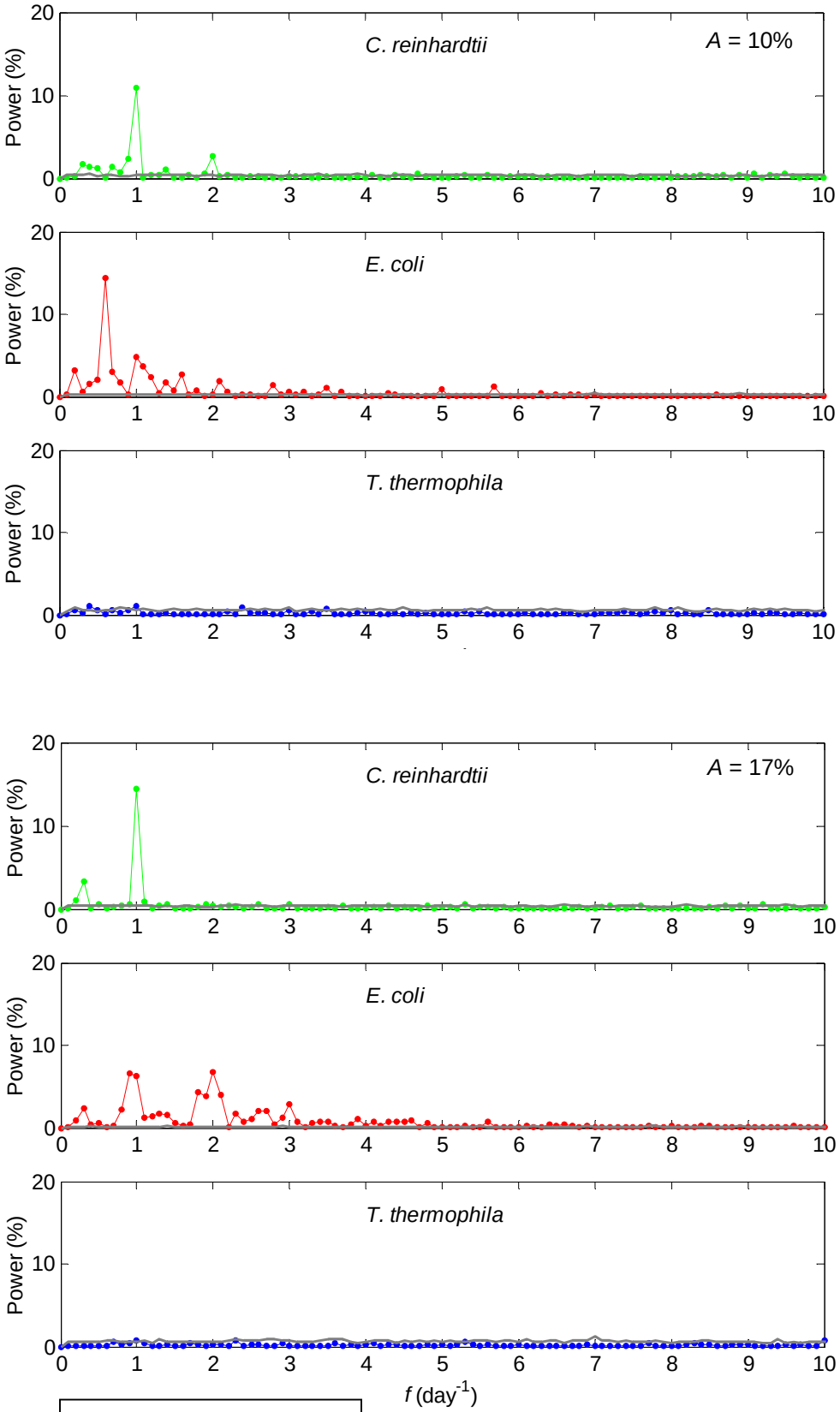


Figure 7.8 (continued)

To address the second question, Figure 7.10 shows that the amplitude of the response increases with the amplitude of the perturbation. *C. reinhardtii* and *T. thermophila* show a strong response only at 1 day^{-1} at perturbation amplitudes of 10% and 17%. For *E. coli*, the scaling of the response appears both more linear, and occurs at all three frequencies considered (1, 2 and 3 day^{-1}). The estimation of power spectra is a hard problem, however, with an especially large chance of observing spectral densities much lower than the true value ("false negatives" are likely, that is, the distribution of the true value, *S*, given a periodogram estimate, $\hat{S}$ has a long tail scaling as $\left(S/\hat{S}\right)^{-2(+\varepsilon)}$). Given these wide confidence intervals, no stronger inferences can be made (much longer time series would be required). In conclusion, the power spectral features of the first experiment were reproduced, but a clear distinction between the dynamics of *E. coli* and *C. reinhardtii* could no longer be made. In addition, the precise scaling of the response in density with the perturbation amplitude could not be determined.

A role for the Circadian Clock?

A rhythm is defined as circadian if it meets three phenomenological criteria [171]. First, a circadian rhythm persists with a period of approximately 24 h under constant conditions (the clocks can lose their synchronization over time in a population under constant conditions). Secondly, a circadian rhythm is temperature compensated, that is, a circadian clock tends to keep running at the same speed at different temperatures, unlike, for

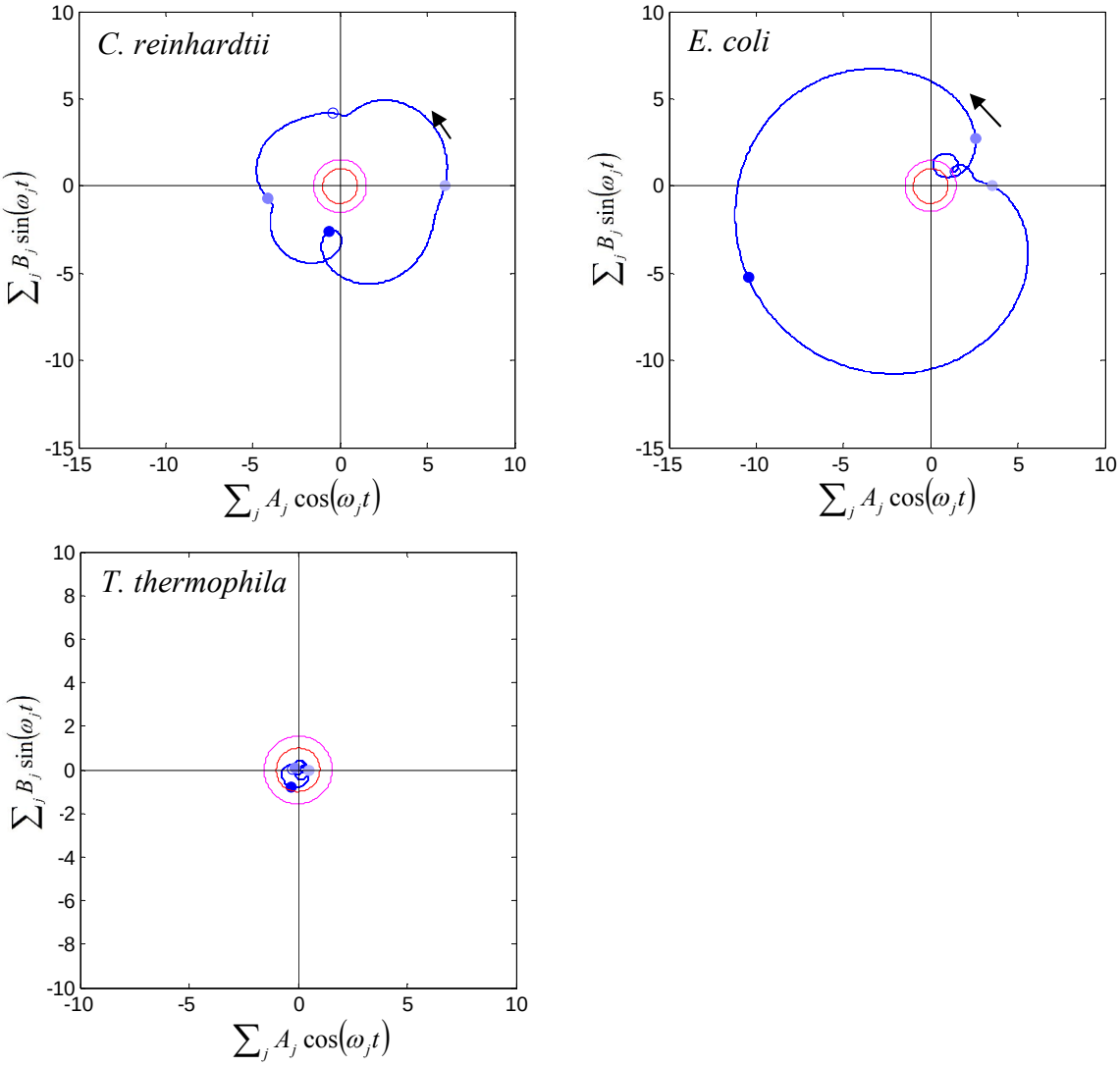


Figure 7.9: Phase plots for all three species in data set 2 under 24 h forcing (17% amplitude), based on the first six integer modes (blue lines). Arrows indicate the direction of progression through time, dark dots the moment of least illumination, light dots of maximum illumination, and pale blue dots the illumination midpoints. All curves were rotated to give the upward midpoint of illumination zero phase ("6 AM" if the moment of least illumination is set to 12 AM).
The red circle indicates the mean values for surrogate data, and the magenta line their mean + 1 standard deviation. The values of the sums on both axes were normalized by division by the corresponding mean absolute value for surrogate data.

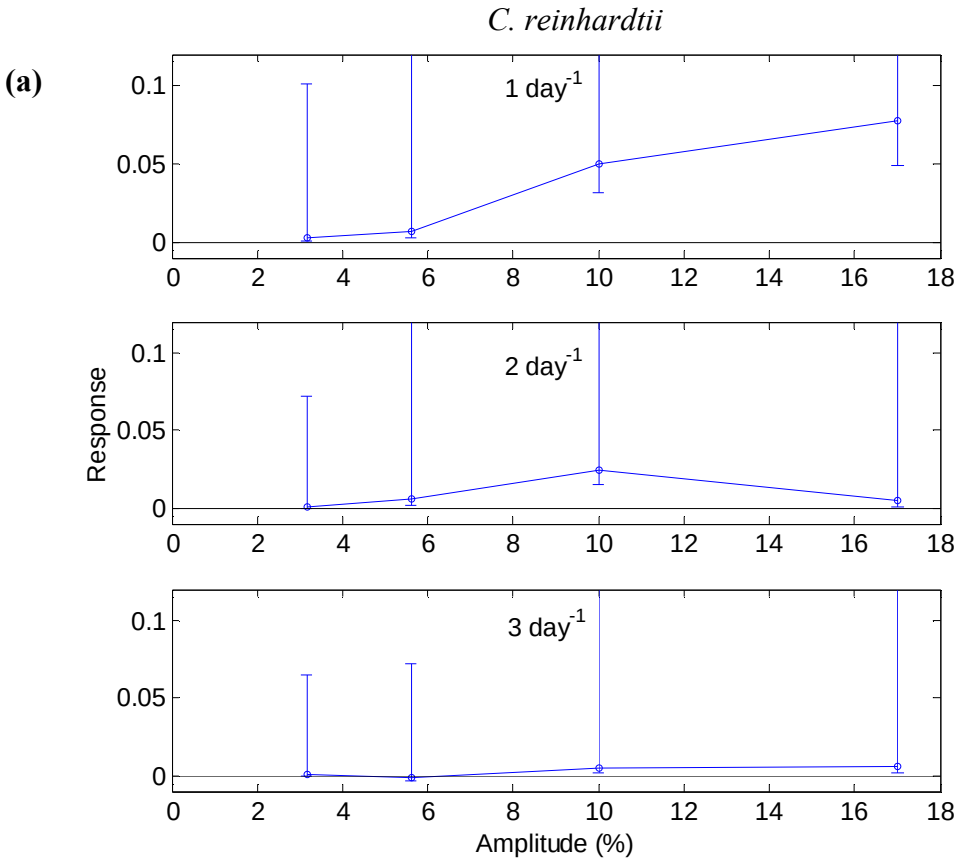


Figure 7.10 (this page and next two pages): response amplitude of (a) *C. reinhardtii*, (b) *E. coli* and (c) *T. thermophila* to sinusoidal modulation of various amplitudes. Shown is the response as $\sqrt{|S(\omega)|} - \langle\sqrt{|S(\omega)|}\rangle_{\text{null model}}$, with $S(\omega)$ the power spectrum of logarithmic densities, that is, *a measure of the mean amplitude of relative fluctuations due to the perturbation*, at the frequencies indicated at the top of each panel.
Data from figure 7.7, with 90% confidence intervals according to [3] (reaching negative values when allowing for a small error in the estimate of the mean of the null model with 250 surrogate sets). The large asymmetry of the confidence intervals is a property of periodogram estimates.

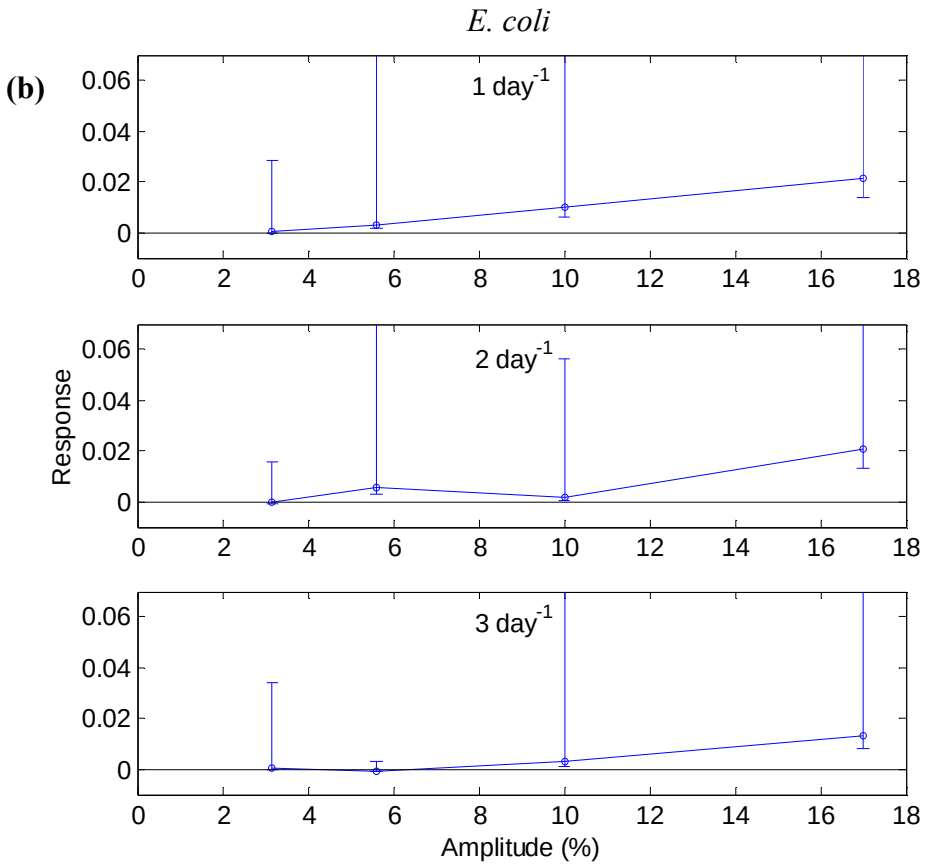


Figure 7.10 (continued)

instance, most enzymatic reactions. Thirdly, these rhythms can be entrained by external stimuli, also known as *Zeitgebers*, over a certain *range of entrainment* around the free-running period. That is, if the external stimuli occur with some period T_{ext}, the circadian clock will run at that pace.

C. reinhardtii has a circadian clock, but its effects can be masked by direct effects of variation in illumination, such as a direct increase of motility with illumination [173, 174]. Judging by data for other eukaryotic algae, its range of entrainment is large (e.g.,

(c)

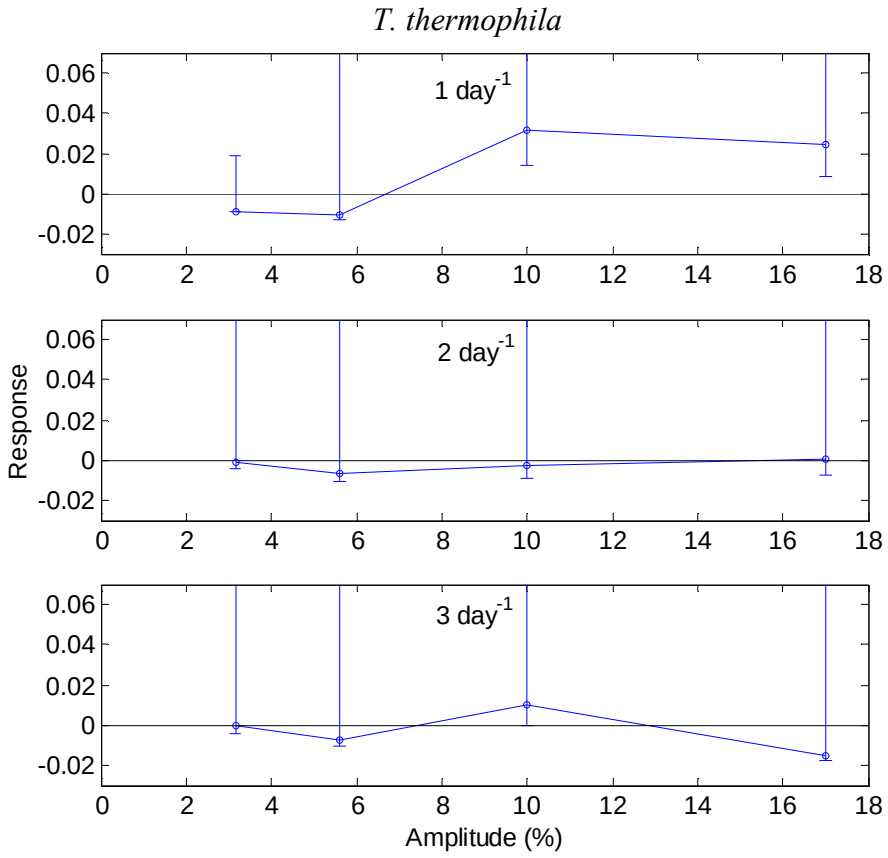


Figure 7.10 (continued)

Euglena gracilis can be entrained by stimuli with periods of 16-48 h [175]). This range may, however, depend on the contrast between light and dark period, which in most studies is large (light/dark), rather than the 17% amplitude used here.

In an initial attempt to distinguish the effects of the circadian clock from masking effects, such as photokinesis, I subjected the ecosystem of experiment 1 to a 31 h sinusoidal perturbation of the same amplitude, 17%. If the circadian clock of *C. reinhardtii* is not entrained by the illumination, the density fluctuations should show both 24 h (circadian) and 31 h (e.g., photokinesis) components. The results are shown in Figure 7.11 and 7.12. There is a clear signal in the time series of *C. reinhardtii* and *E. coli* densities at the

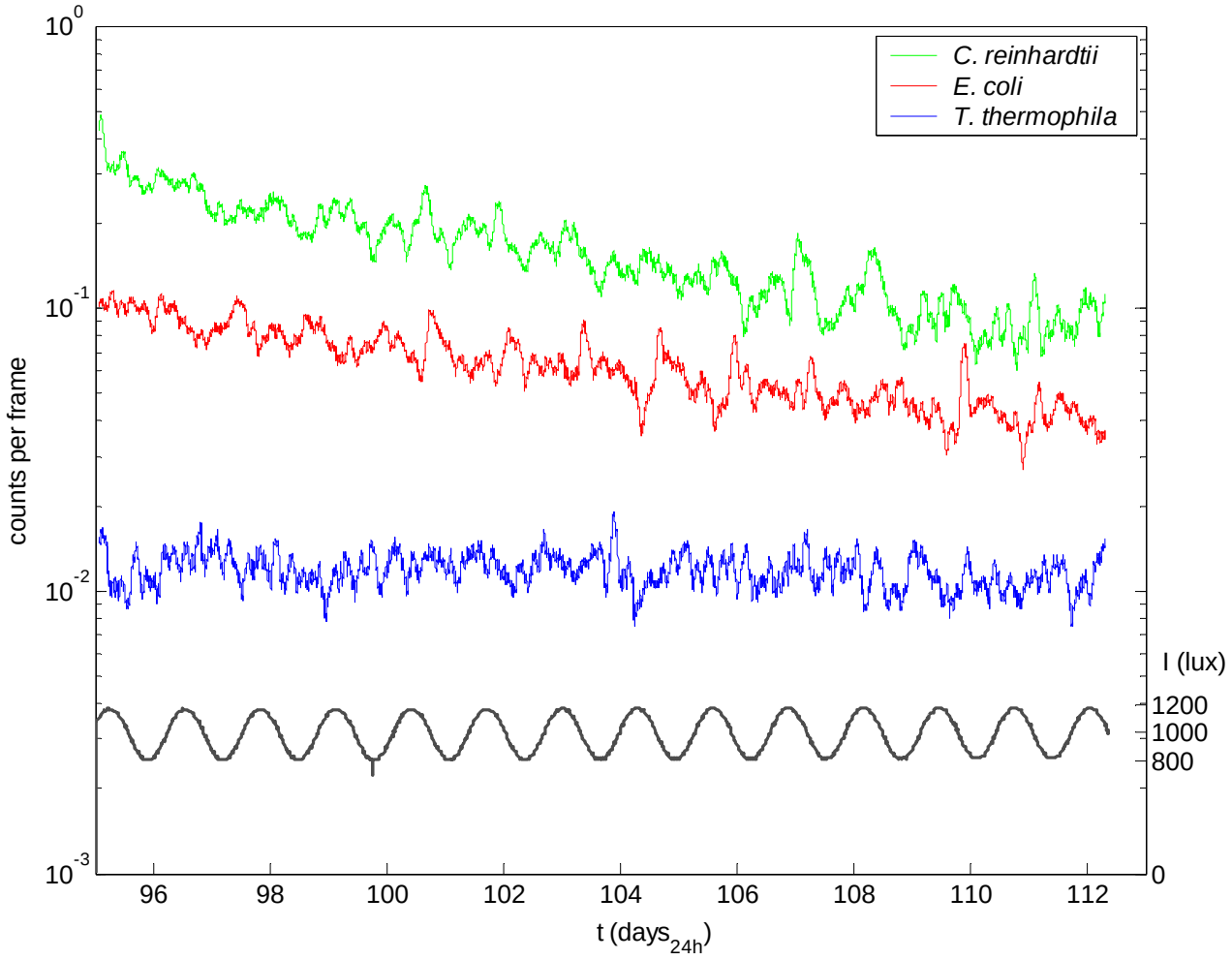


Figure 7.11: data set 1, under 31 h perturbation. Shown are 66 minute (5000 frame) moving averages for all three species (see legend).

entraining frequency and its harmonics (frequencies in Figure 7.12 are expressed in day_{31h}^{-1}, that is, per 31 h "day"). No clear signal is seen at 24 h (1.3 day_{31h}^{-1}), except remarkably, *T. thermophila* now shows a significant peak at a frequency corresponding to 26 h, although ciliates are known to exhibit circadian rhythms [84].

The power spectra in Figure 7.12 reveal that there is an increase in power of higher harmonics, 2 day_{31h}^{-1} for *C. reinhardtii* and 3 day_{31h}^{-1} for *E. coli* relative to the 24 h

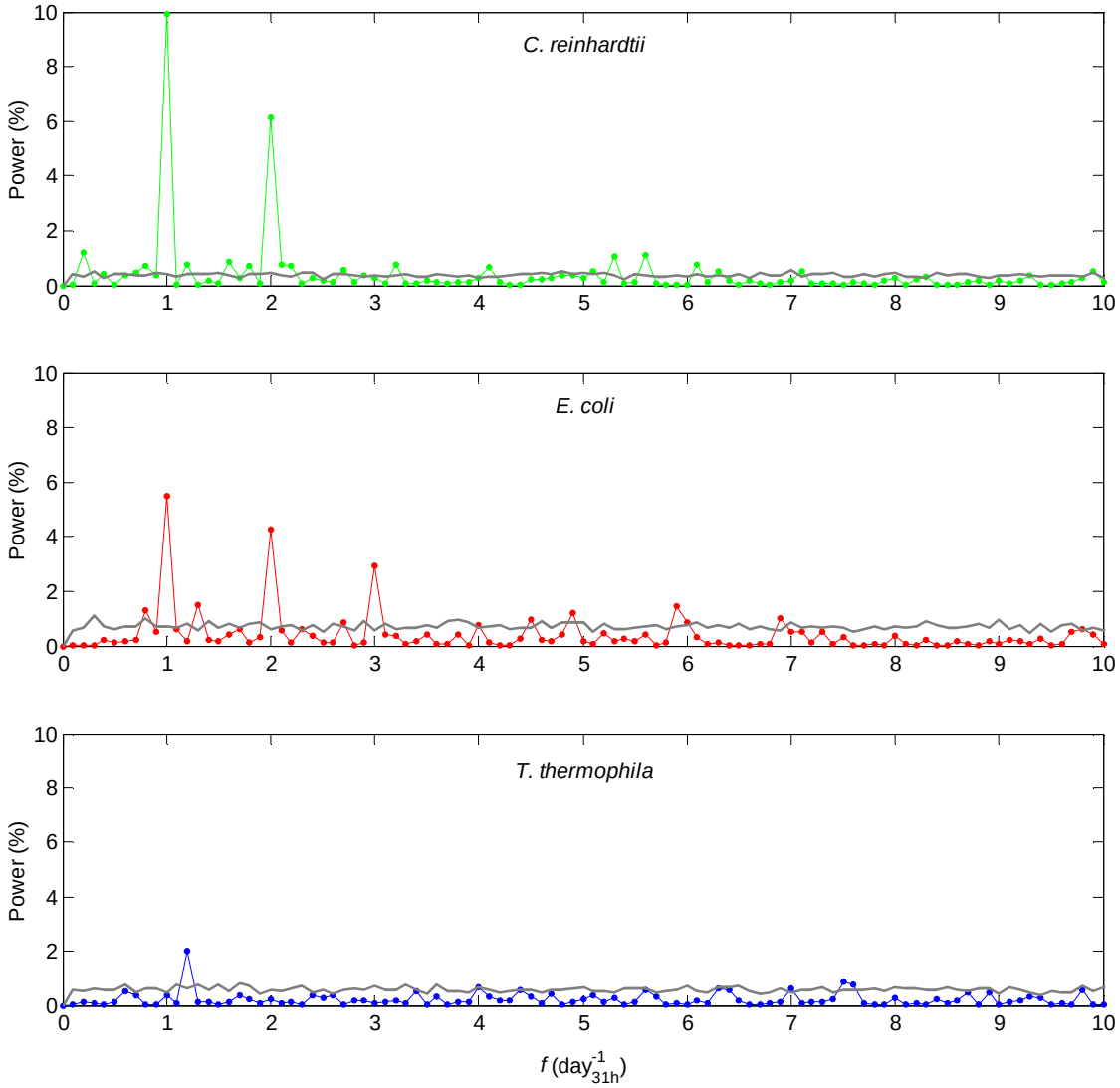


Figure 7.12: Power spectra for data set 1, under 31 h perturbation over the last 13 days (≈ 10 31 hour days). Gray lines again indicate the mean + 2 standard deviations for surrogate data. All data processing follows the description in the caption of figure 7.3. Remarkably, there is a peak in the power spectrum of *T. termophila* at about 26 hours (1.2 day_{31h}^{-1}).

perturbation, implying that the waveform does depend on the period of the illumination[70].

Both species now show phase ambiguity over some part of the perturbation cycle (Figure

[70] Since the perturbation is sinusoidal, a linear response would be sinusoidal. The sum of integer modes (see also figure 7.11 and 7.20) is not itself sinusoidal.

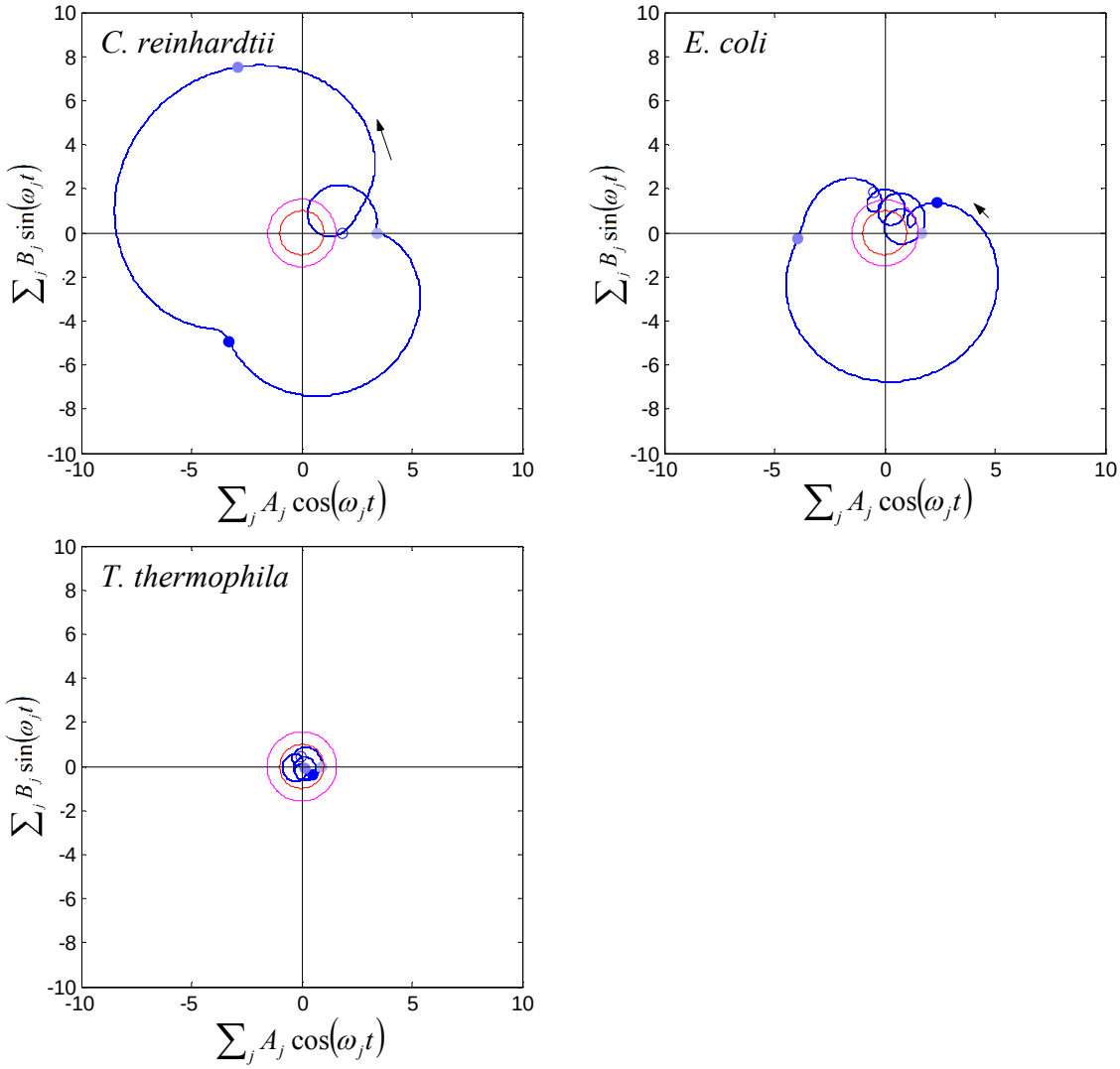


Figure 7.13: Phase plots for data set 1, under 31 h perturbation over the last 13 days (≈ 10 31 hour days). Arrows indicate the direction of progression through time, dark dots the moment of least illumination, light dots of maximum illumination, and pale blue dots the illumination midpoints. All curves were rotated to give the upward midpoint of illumination zero phase ("6 AM" if the moment of least illumination is set to 12 AM).
The red circle indicates the mean values for surrogate data, and the magenta line their mean + 1 standard deviation. The values of the sums on both axes were normalized by division by the corresponding mean absolute value for surrogate data.

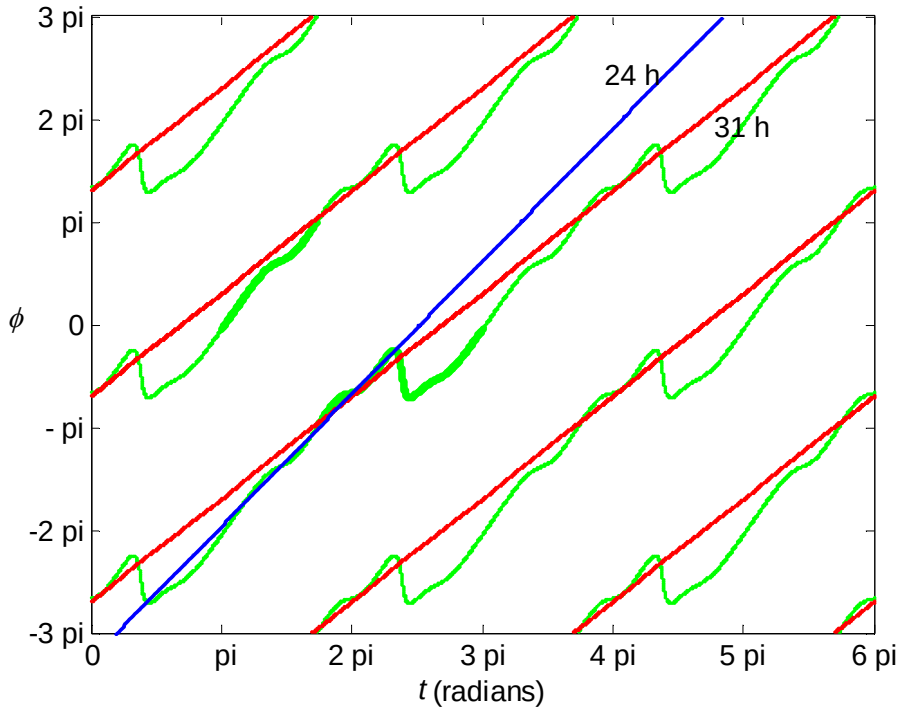


Figure 7.14: Progression of the phase over time (expressed here as the phase of illumination) for data set 1, under 31 h perturbation. Green: *C. reinhardtii*, red: illumination. Estimates of phase are based on the integer modes of day 99 to 112 (10 31 hour days) in figure 7.11. Since a phase is determined only up to $2\,k\,\pi$, with $k \in Z$, these alternative phases are shown as well. The thick green lines illustrate this by showing the inferred phase for one nominal day (not a cycle of the illumination).
In blue is shown how fast a 24 hour clock would run. This line suggests the following explanation: the *C. reinhardtii* clock runs at about a 24 h cycle, but when it gets ahead of the illumination cycle, its phase is reset.

7.13). For *C. reinhardtii*, this phase ambiguity suggests a kind of phase resetting. Figure 7.14 shows a plot of the inferred phase of the signal over time for *C. reinhardtii*. As indicated by the blue line, the intrinsic rate of increase of the phase (based on the first six integer modes, 1, 2, … day_{31h}^{-1}), appears more consistent with a 24 h day, with phase resetting just before the point of maximum illumination. The corresponding integer mode

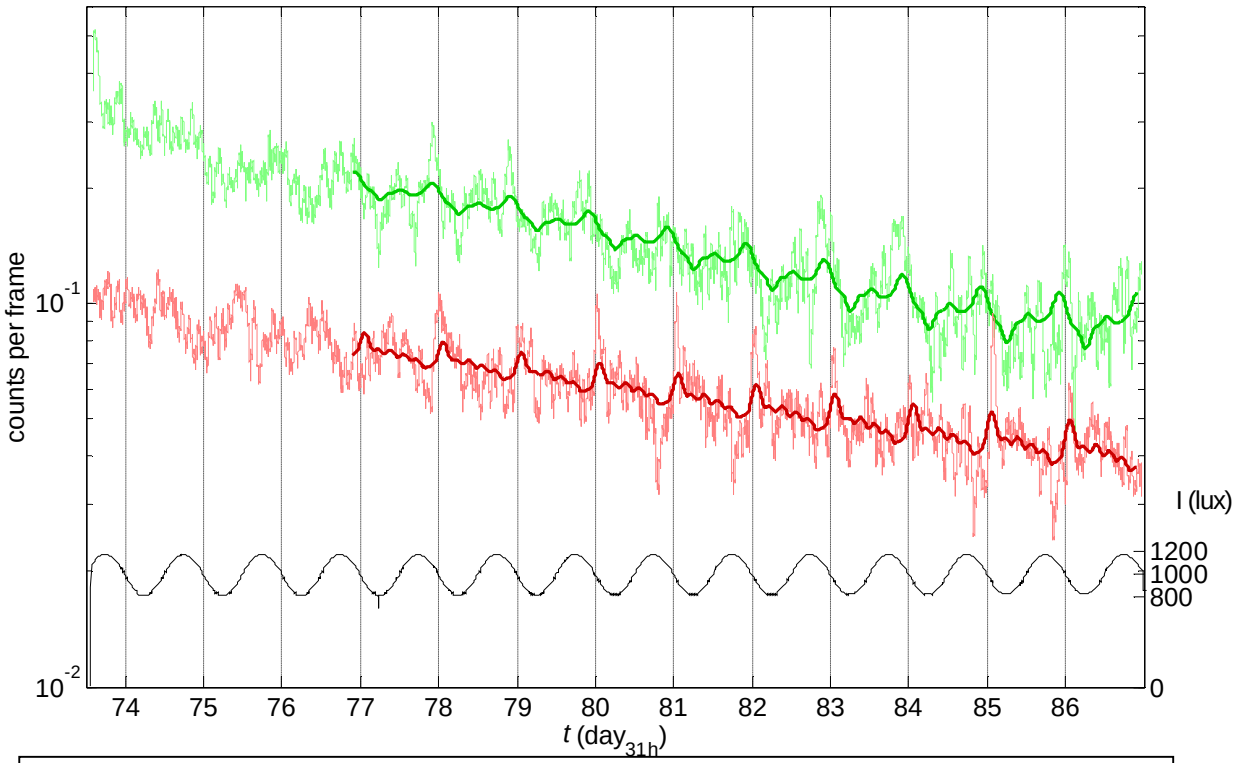


Figure 7.15: A comparison of reconstructed dynamics in data set 1 under 31 h forcing, based only on integer modes (1, 2, …, 6 day^{-1}), and observed population densities (66 minute moving averages) for *C. reinhardtii* (green) and *E. coli* (red).

approximation of the time series is shown in Figure 7.15. *In summary, both a 24 hour and 31 hour rhythm manifest themselves in the observed density fluctuations of C. reinhardtii, but not as two Fourier components. Instead, the phase of an intrinsic 24 hour rhythm appears to be reset once per cycle to match the 31 hour period of illumination.*

Temporal stability of the observed effects

I want to briefly address the temporal stability of the observed effects by two different means. First, Fourier transforms provide information on the complete signal. Information about temporal variations in the observed patterns is 'stored' in the non-integer modes of the transform. As an example, in Figure 7.16 I included each Fourier component

explaining > 1.5% of total variation in the calculation of phase plots for experiment 2, discussed under *Reproducibility and scaling*. In that experiment, I examined replication of the original results for perturbation with a 24 h period at 17% amplitude. In the *E. coli* power spectrum, peaks were broader than in the power spectrum for data set 1, and the winding number could not be reliably estimated. As shown, for each day examined, the phase *does* tend to curl twice each day, but the origin is not always included.

The other example relies on wavelet analysis introduced in chapter 6. The analyzing wavelet used is the complex Morlet wavelet, ψ, which can be considered a wave localized in time by convolution with a Gaussian (see Figure 7.17) [71]:

$$\psi(t,\tau) = C_{\alpha,\beta} \cdot e^{-\frac{1}{\alpha}\left(\frac{t-t'}{\tau}\right)^2} \cdot e^{\left(2\pi i \beta \frac{t-t'}{\tau}\right)} \qquad (7.4)$$

with C a normalization constant and α, β determining the degree of spatial and frequency localization of the wavelet. I used a common choice, $\alpha = 1$ and $\beta = 6$. The resulting transforms (see chapter 6 for more details), are shown in Figure 7.18 for the 24 h and Figure 7.19 for the 31 h perturbations in data set 1[72]. Especially for *E. coli*, the power in the integer dominant frequency bands varies over time and localization is not always precise. Most notably there is a significant shift from 2 day^{-1} to 1 day^{-1} for the 24 h perturbation after 6 days of perturbation. Also note the near disappearance of a periodic signal in *E. coli* after day110.

[71] An alternative approach, more consistent with the Fourier analysis elsewhere in this chapter, would be the use of the Wigner-Ville transform (e.g. [169]). In the presence of external forcing, however, there is no guarantee that the contributions observations decrease with the intervening lag.

[72] This graph includes some additional data not included in data set 1, which were acquired after a 0.5 day_{31h} interruption to measure densities in other ecosystems.

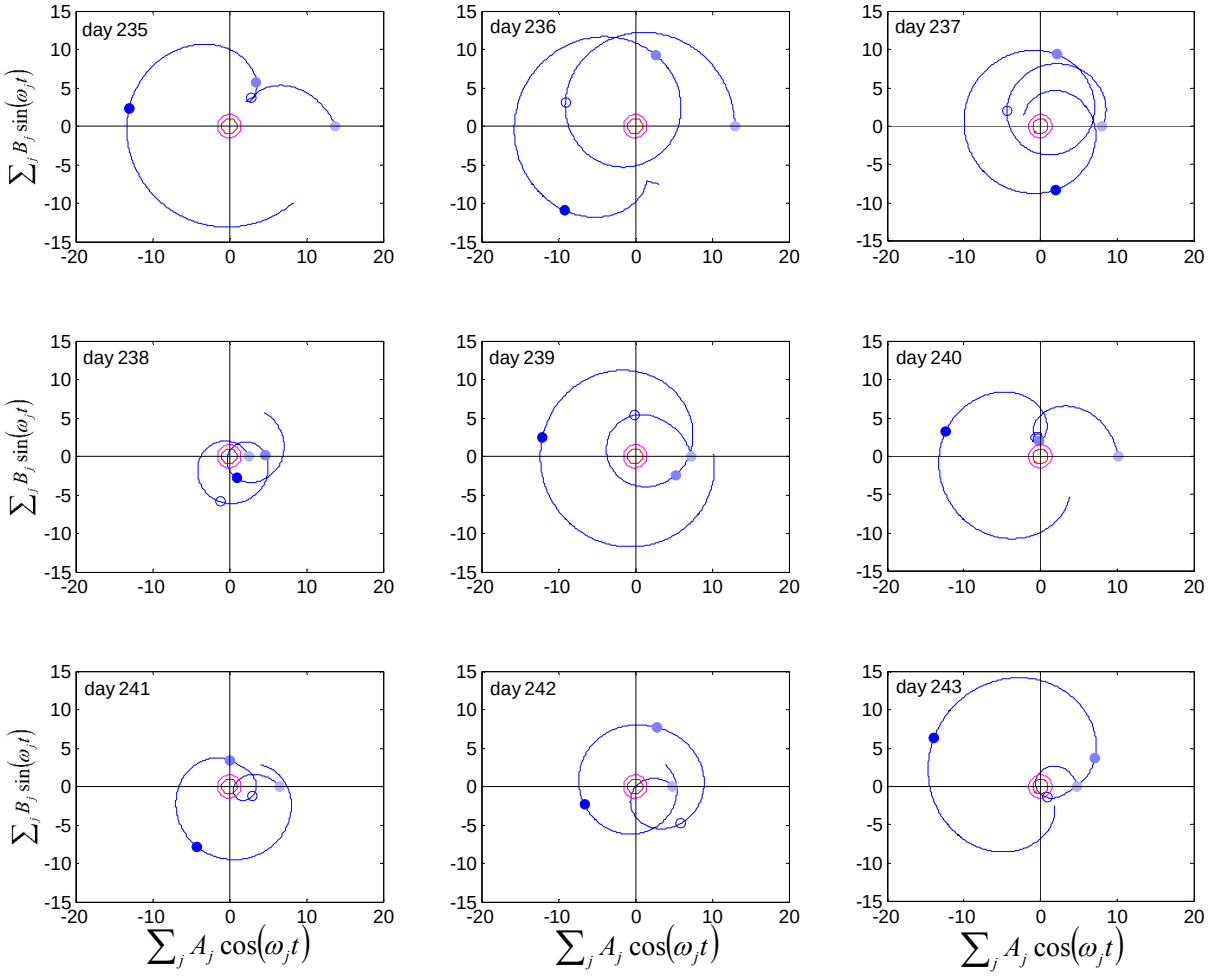


Figure 7.16: Phase plots for *E. coli* calculated for data set 2 under 24 h forcing with a 17% amplitude. The phase plots were constructed including all Fourier components with more than 1.5% of total power, not just integer modes. As a result, the phase plot is not stationary but varies from day to day. All rotation is counter clock wise. As before, the upward midpoint of illumination is set to have phase 0, and is indicated by a pale blue dot. Dark dots indicate the moment of least illumination, light dots of maximum illumination, and pale blue dots the illumination midpoints. All curves were rotated to give the upward midpoint of illumination zero phase ("6 AM" if the moment of least illumination is set to 12 AM). The red circle indicates the mean values for surrogate data, and the magenta line their mean + 1 standard deviation. The values of the sums on both axes were normalized by division by the corresponding mean absolute value for surrogate data.

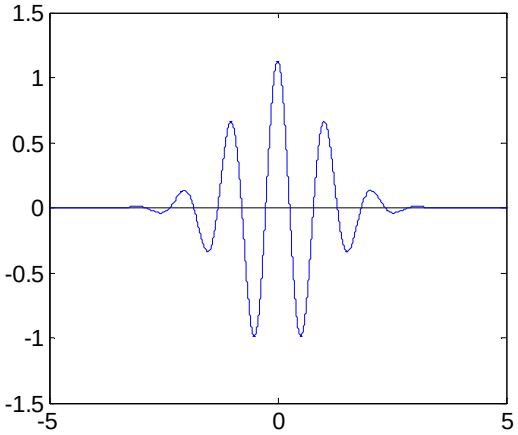


Figure 7.17: The Morlet wavelet (real component). No orthogonal decomposition is possible for this wavelet (*cf.* chapter 6).

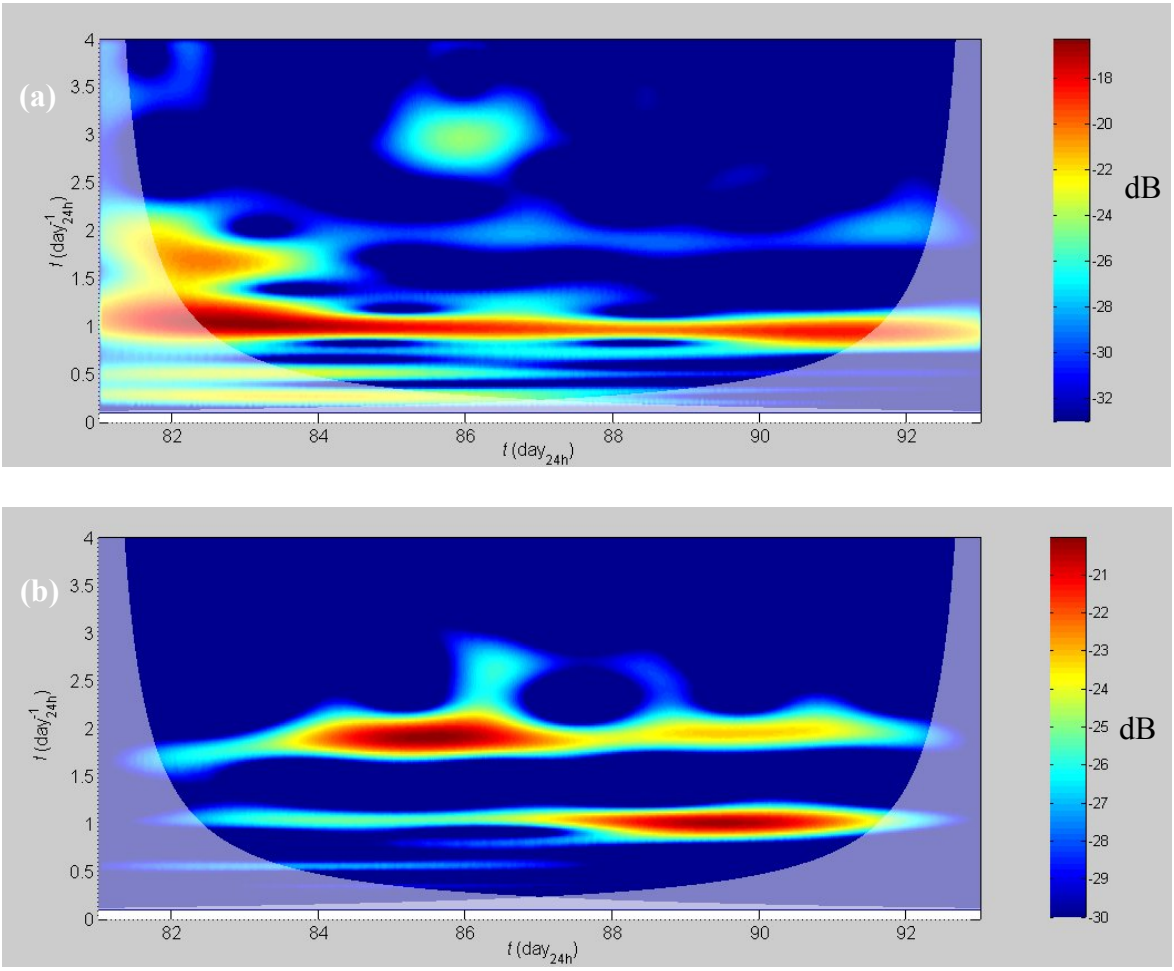


Figure 7.18: Continuous wavelet transform using the complex Morlet wavelet (see text) to detect the presence of local periodicities. (a) *C. reinhardtii* (b) *E. coli* data from experiment 1 under 24 h forcing. Shown is the local power in decibel as a function of time and frequency (that is, $10 \log_{10}\left(\left|W_n(t,\tau)\right|^2\right)$, with the frequency obtained from the time scale of the wavelet τ as $f = 1/(1.03\ \tau)$ [5]. The region of the wavelet transform affected significantly by the boundary conditions is shown shaded. The perturbation started at about 82.0 days.

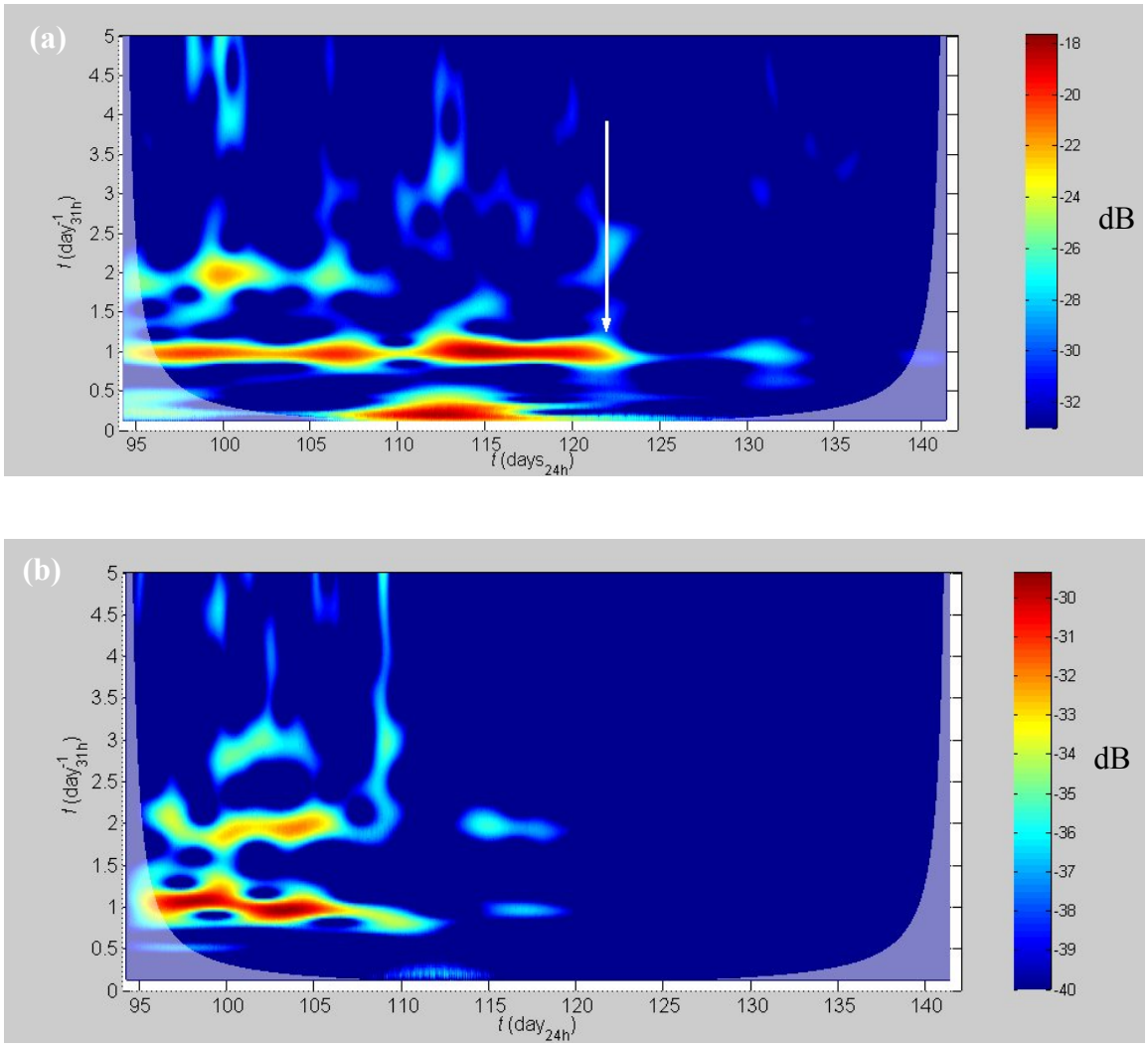


Figure 7.19: Continuous wavelet transform using the complex Morlet wavelet. (a) *C. reinhardtii* (b) *E. coli* data from experiment 1 under 31 h forcing. Forcing ceased on day 122 (white arrow in panel (a)), after which the system was illuminated at the prior mean intensity. Shown is the local power in decibel as a function of time and frequency (that is, $10 \log_{10}\left(\left|W_n(t,\tau)\right|^2\right)$, with the frequency obtained from the time scale of the wavelet τ as $f = 1/(1.03\ \tau)$ expressed per 31h day [5]. The region of the wavelet transform affected significantly by the boundary conditions is shown shaded.

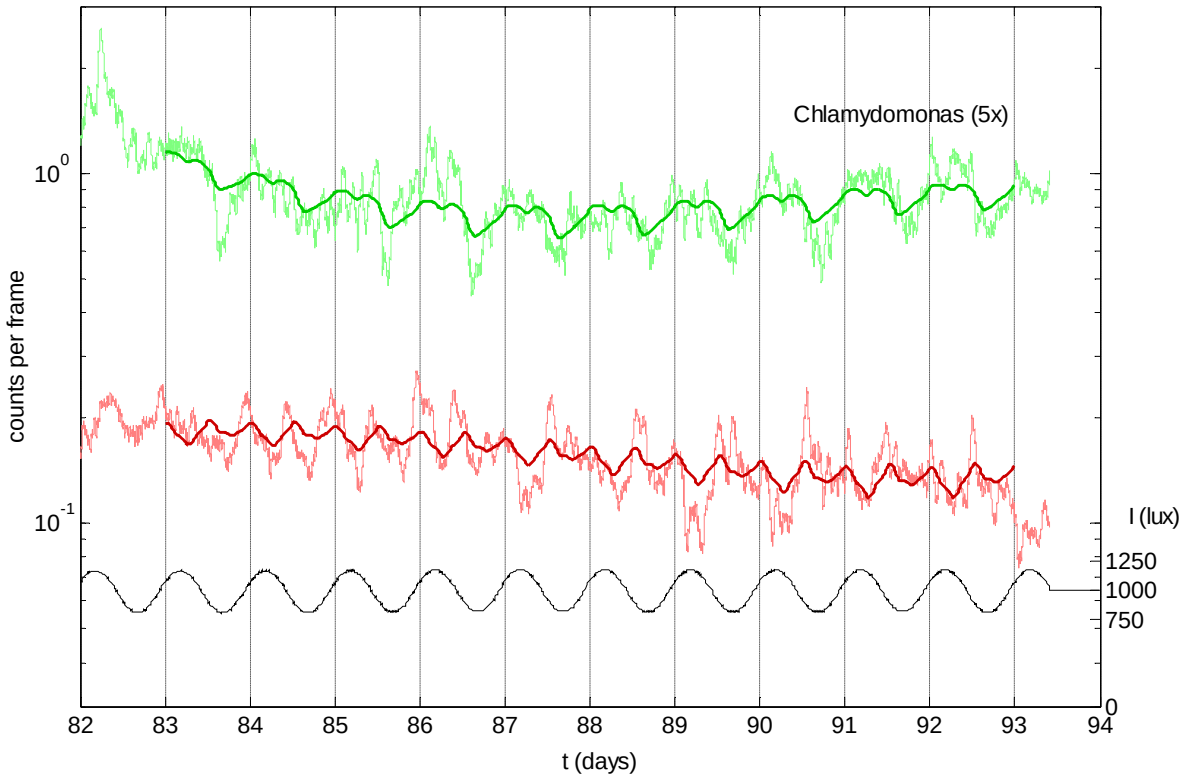
Migration versus growth

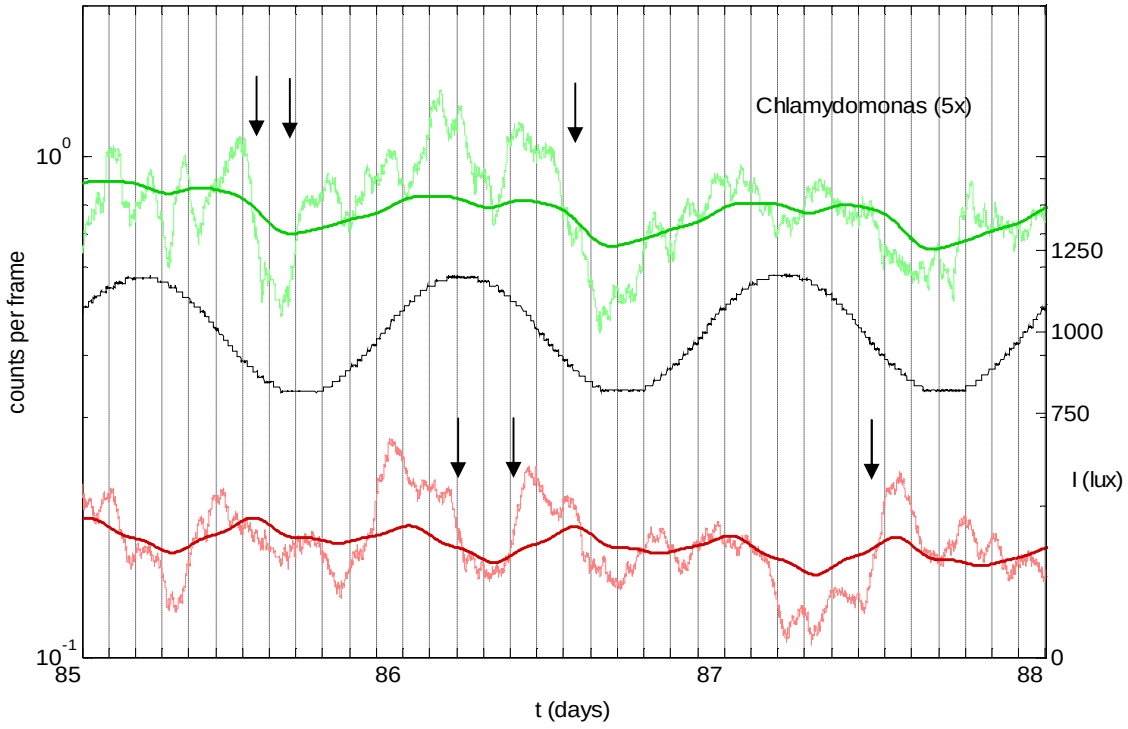
The next question is whether the density fluctuations observed under modulation of the illumination are due to spatial processes or growth and death. In particular, the motility of *C. reinhardtii* is known to depend directly on illumination (a phenomenon known as photokinesis), as well as on the phase of its circadian clock. *E. coli* displays chemotaxis and might, for example, track *C. reinhardtii* on whose photosynthesis products it presumably relies or, more indirectly, redistribute under a change in conditions. While the observed dominant frequencies for these species are 1 and 2 day^{-1}, the actual fluctuations are faster (~2 hour rise and fall times), with relative stasis in between (examples of fast fluctuations are marked with arrows in Figure 7.20). Their amplitude is typically about twofold. As a consequence, synchronized division [176] could explain fast increases in density, as long as they occur no more than once or twice a day. A mechanism for synchronized fast death is hard to imagine, but cannot be excluded.

To address the respective roles of migration and growth and death, I built a second version of the measurement apparatus ("two-point setup", chapter 4). This apparatus measures population densities of one species at two points in space and of two species each at one point in space[73]. The locations chosen were 5 mm below the center of the meniscus, and 5 mm above the center of the bottom (see the schematic in Figure 7.21). The essence of the experiment is that vertical migration would yield anticorrelated time

[73] Alternatively, it could measure densities of two species in two locations.

Figure 7.20 (next page): (a) A comparison of reconstructed dynamics based only on integer modes (1, 2, …, 6 day^{-1}) to observed population densities (66 minute moving averages) for *C. reinhardtii* and *E. coli* (data as in figure 7.1). (b) a close up view of (a) with arrows marking fast swings in local density, as described in the text.


Chlamydomonas (5x)
counts per frame
10^0
10^{-1}
I (lux)
1250
1000
750
0
82
83
84
85
86
87
88
89
90
91
92
93
94
t (days)


Chlamydomonas (5x)
counts per frame
10^0
10^{-1}
1250
1000
750
I (lux)
0
85
86
87
88
t (days)

series at the two locations, while growth and death should yield positively correlated time series.

As discussed in the appendix of chapter 4, the apparatus has improved temperature control and more isotropic illumination compared to the first apparatus. An ecosystem was constructed with the same initial conditions as the previous system, but densities were measured continuously from the start. The overall light intensity was the same as for the first two experiments and temperature was controlled at 25.0 °C.

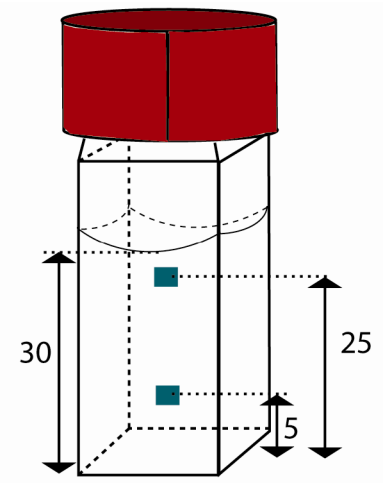


Figure 7.21: Schematic of the location of the two observation volumes, distances in mm.

The observed density time series of the experiment are shown in Figure 7.22, with the *C. reinhardtii* density measured in two locations. The size of the observation volumes has not yet been estimated (cf. chapter 4). Since different objectives were used on the top and the bottom (20x and 10x, respectively), no inferences can yet be made about the relative densities at both locations. As can be seen from the power spectra (Figure 7.23), the *C. reinhardtii* density 5 mm above the bottom has a small but significant 1 day^{-1} component. For the *C. reinhardtii* density close to the top, this component is only somewhat significant ($p \approx 0.02$), and probably affected by noise. As a consequence, no reliable

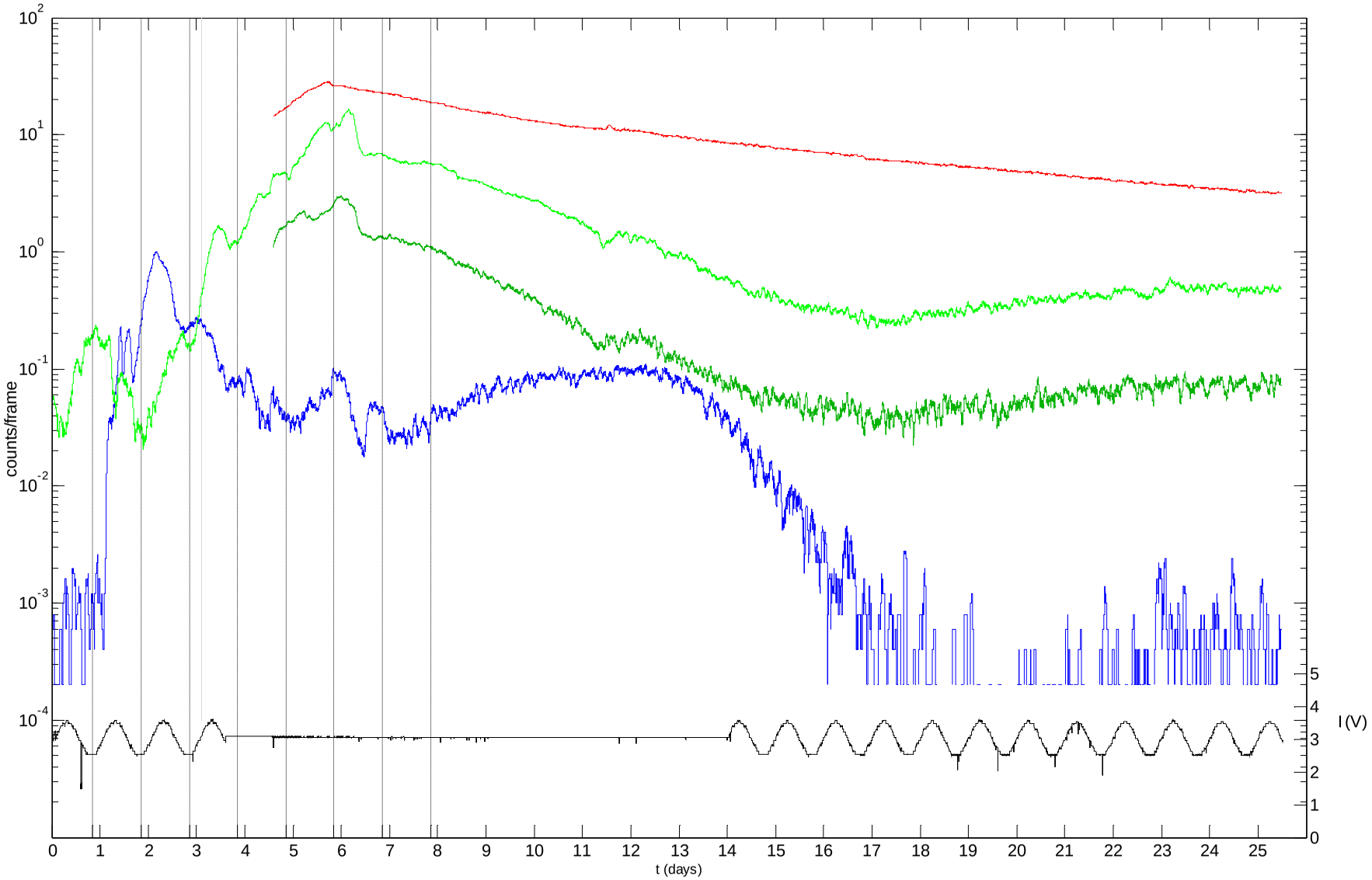


Figure 7.22 First results from the two-point measurement apparatus. Shown are 66 minute moving averages (5000 frames). *C. reinhardtii* (light green, bottom; dark green, top); *E. coli*, red, top; *T. thermophila*, blue, bottom. The illumination intensity is shown in black (in volts, yet to be converted to lux). Grid lines are shown to emphasize the effects of the initial perturbation.

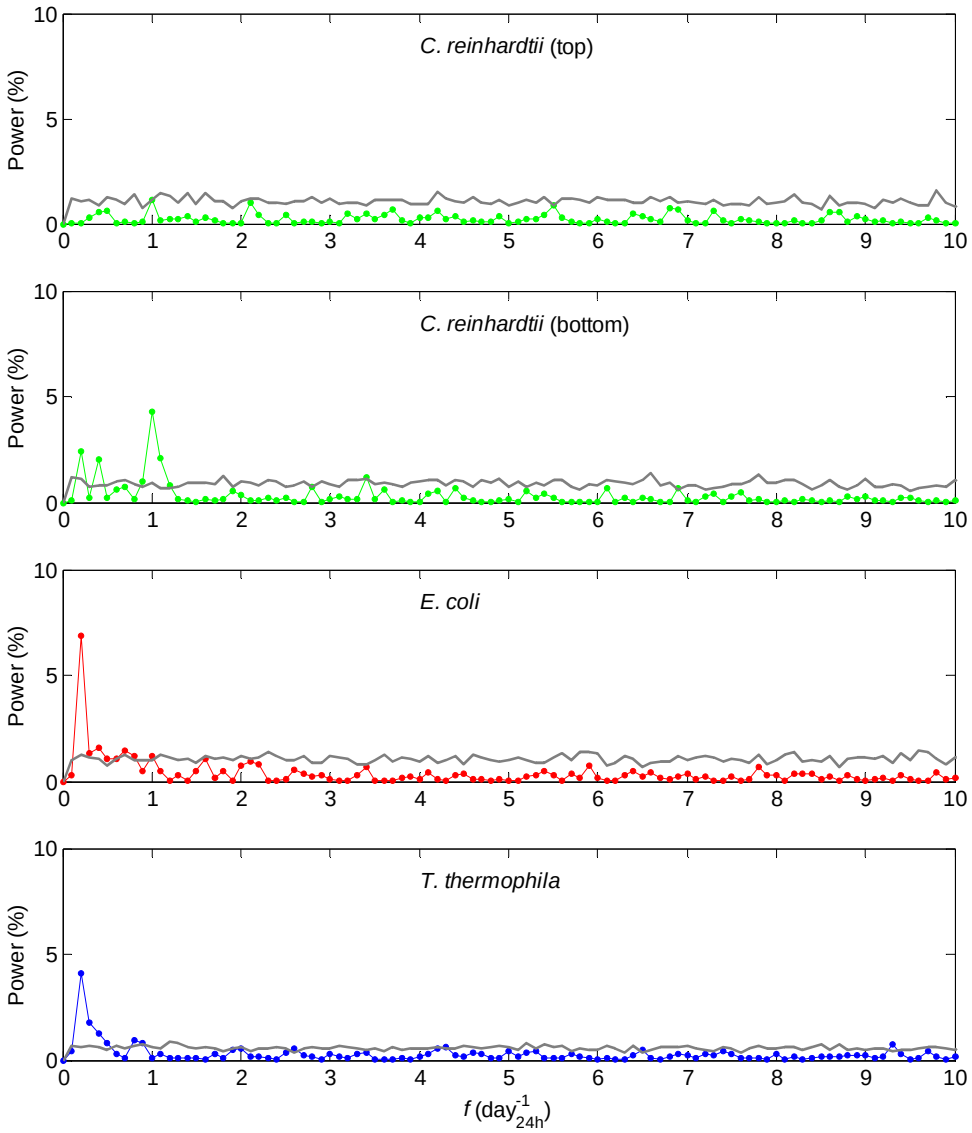


Figure 7.23: Power spectra for the two-point setup, over day 15-25. The species are indicated at the top of each panel. Gray lines again indicate the mean + 2 standard deviations for surrogate data. All data processing follows the description in the caption of figure 7.3.

inferences can be made on the correlation between the two (the observed phase difference is about 0.25 π). These observations are confirmed by the autocorrelation and crosscorrelation functions shown in Figures 7.24 and 7.25, respectively. From these

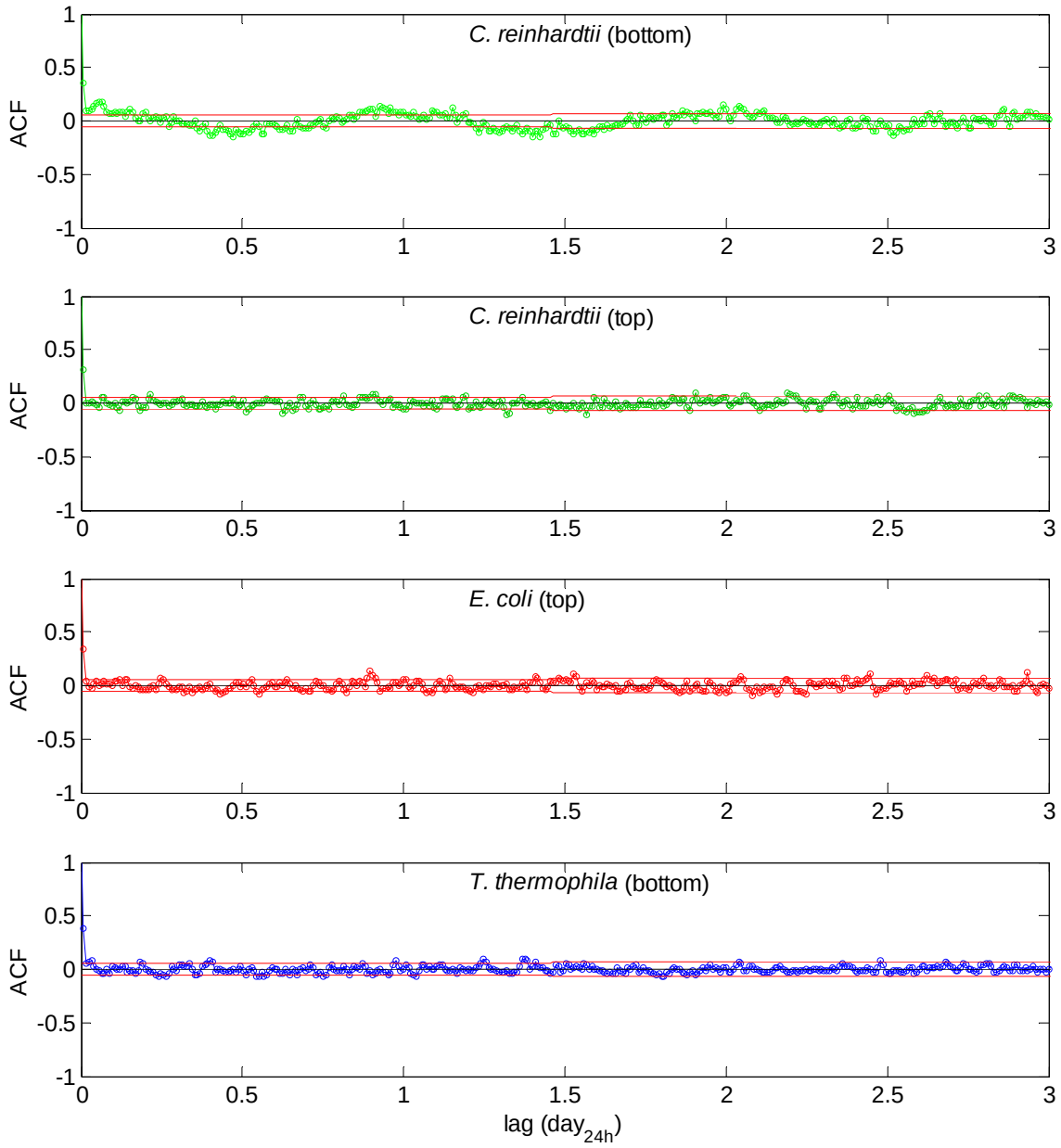


Figure 7.24: Autocorrelation functions for the two-point setup, over day 15-25. (a) *C. reinhardtii* (bottom) (b) *C. reinhardtii* (top), (c) *E. coli* (top) and (d) *T. thermophila* (bottom). The terms 'top' and 'bottom' refer to the two measurement locations within the ecosystem. The red lines indicate two standard deviations from the mean for uncorrelated data. Data treatment is described in the caption of figure 7.2.

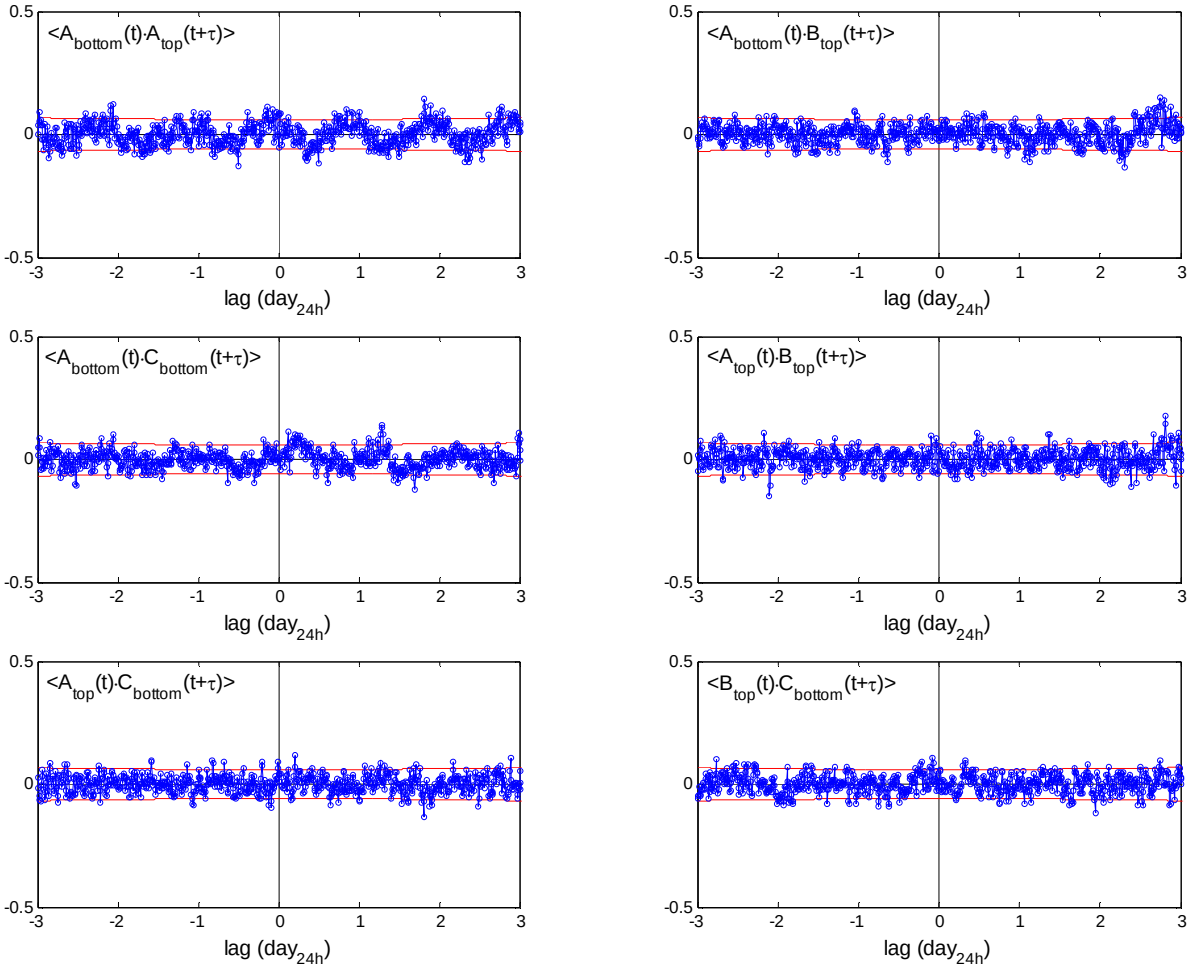


Figure 7.25: Crosscorrelation functions for the two-point setup, over day 15-25. The relevant species are indicated in each panel, with A: *C. reinhardtii*, B: *E. coli* and C: *T. thermophila*. The terms 'top' and 'bottom' refer to the two measurement locations within the ecosystem. The red lines indicate two standard deviations from the mean for uncorrelated data. Data treatment is described in the caption of figure 7.2.

results it is hard to judge why a more robust response was not observed. The main potential explanations are (1) the difference in illumination pattern, which was much more isotropic in this setup than in the 'one-point' setup, or (2) sensitivity to the perturbation may vary as the ecosystem develops. As an interesting aside, during the first few days of the experiment the ecosystem was accidentally subjected to a sinusoidal

perturbation as well, and the available time series (*C. reinhardtii* and *T. thermophila* just above the bottom) do appear to show a more robust response (see Figure 7.22).

Discussion and Conclusions

Under a modest perturbation, a 17% amplitude sinusoidal modulation of illumination (mean 1200 lux), the population densities of *C. reinhardtii* and *E. coli* showed a clear, qualitatively reproducible response. *T. thermophila* displayed a much weaker but significant response. For all three species, a modulation amplitude of at least 10% was required to see a significant effect (Figure 7.10).

The response of *E. coli* is surprising. Because *E. coli* is thought to lack a circadian clock, either direct effects of light or temperature changes, or interactions with *C. reinhardtii* are probably responsible for *E. coli*'s observed response in density. Not only do its densities respond to the perturbation, however, it also tends to double the frequency of the perturbation at 24 h, as quantified by the calculation of winding numbers (equation 7.3). As an aside, circadian patterns of locomotion activity for a large variety of metazoan taxa show two peaks, one at dawn and one at dusk [177].

Despite the clear response of *E. coli*, it appears to be impossible to infer interactions rigorously from only density fluctuations under cyclical perturbations as long as a direct response of *E. coli* to light or temperature has not been ruled out. Even if such an effect can be ruled out, however, the lag of the interaction is, strictly speaking, only determined up to an integer multiple of the period of the perturbation. In the final chapter, I will briefly consider alternative perturbation schemes.

Given the speed of response in density for *C. reinhardtii* and *E. coli*, it appears likely that most of the observed density fluctuations is due to migration within the ecosystem. Likewise, a small drop in temperature in data set 2 caused a nearly immediate twofold decrease in *C. reinhardtii* density, suggesting that spatial distribution can be very sensitive to boundary conditions. Unfortunately, the data from the two-point measurement apparatus showed a weak response, so no conclusions could be drawn about the relative roles of migration and growth and death. Apart from repeating the experiment for longer durations, the use of phototactic *C. reinhardtii* mutants [178] or chemotactic *E. coli* [179] may help discriminate the causes of the observed response.

C. reinhardtii's circadian clock likely plays a role. As for *E. coli*, there may be confounding direct effects of light and temperature, however. To rule out such effects, there are a number of options. First, one could use *C. reinhardtii* period mutants. Single and double mutants with periods from 21 to 30 h have been isolated [108, 180], and could be subjected to the same 24 h perturbation. Secondly, the circadian clock of *Euglena gracilis* can be entrained in ranges around harmonics of the period. For example, a 3h/3h light/dark cycle was shown to entrain its cell cycle with a 24 h period [176]. Direct responses, such as photokinesis, would presumably manifest themselves with a 6 h period, while effects dependent on the circadian clock would be visible on a 24 h time scale. Finally, it would be informative to study the free-running response (in constant light) after a period of entrainment. Such experiments are non-trivial, though, because the free-running response depends on the phase of the circadian clock at the moment the perturbation ceases [173, 174].

In summary, modest periodic perturbations can yield much information on the system. As was the case in the previous chapter, though, a better understanding is necessary of the main components of the observed response if we are to efficiently learn about the system. The required knowledge, however, is largely phenomenological in nature and lies within experimental reach.

Chapter 8 Conclusions and Future Directions

In the first chapter I described how both random events like mutations and demographic fluctuations and (more) deterministic factors such as species-species interactions can affect the growth and survival of species. Unfortunately, these contributions are nearly impossible to predict from the molecular composition of the organism and its environment. The necessary alternative is to address the inverse problem: the inference of species-species interactions and the effect of random events from population dynamics time series. In this inference, there are two major obstacles as discussed in chapter 1: (1) in order to move from coincidence to correlation, replicable measurements are needed, and (2) the move from correlation to causality is impeded by the multitude of potentially relevant unobserved variables.

In this thesis, I have introduced closed ecosystems as a means to address these challenges. First, I will summarize the results obtained using a model closed ecosystem. Then, I will discuss how one could proceed with the inverse problem, analytically and experimentally. Finally, I will consider the possibility of combining forward and inverse approaches to the study of, in Darwin's words, the struggle for existence.

Conclusions from the work in this thesis

In chapter 2, I argued that closed ecosystems allow the study of species-species interactions without distortions resulting from refreshment of growth medium or uncontrolled material fluxes, such as ongoing carbon fixation from the atmosphere. A simple three-species closed ecosystem (A,B,C) was introduced in chapter 3 of which the

boundary conditions, light and temperature, and initial conditions can be controlled. I showed that (1) one can start the system from an inorganic medium, (2) that each of its subsystems (AB, BC, …) can be studied, and (3) that these ecosystems can be miniaturized to about 150 μL, allowing for large scale parallel experiments. Even in this 'simple' ecosystem, with three single-cell motile microbial species considerable spatial heterogeneity and phenotypic diversity arise within weeks. Such complexity is intrinsic to biology, and it seems wiser to develop methods that allow for this complexity, or even try to make inferences about it, rather than to minimize it.

In chapter 4, I showed that population density time series from a small observation volume within a closed ecosystems can be obtained with large dynamical range and low classification error. The developed method, a modified version of selective plane illumination microscopy [22], allows for up to ten samples to be measured daily (chapter 5), or for single samples to be measured continuously for months (chapter 6 and 7).

In chapter 5, the divergence of sets of replicate ecosystems was studied. Different species showed characteristic patterns of divergence, resulting in an approximately linear increase of total divergence over ~60 days, as measured by the linear dimension of the volume spanned by the covariance matrix (of log **n**). The rate of divergence decreased markedly after about two months. The observed divergence demonstrates the importance of historical effects for the dynamics of these ecosystems. Deviations from mean dynamics three or four weeks after closure were shown to affect the state of the ecosystems up to six weeks later, confirming that these systems have idiosyncratic individual histories. Despite the variation among replicate ecosystems, the mean dynamics of sets of ecosystems could be reproduced between independent experiments.

In addition, the quantification of divergence allows for the design of experiments with a known degree of reproducibility as a function of the number of replicates and the duration of the experiment. Specifically, it takes *T. thermophila* densities about 20 days to vary twofold between replicate ecosystems, and *C. reinhardtii* and *E. coli* about 40 days.

Chapter 6 dealt with population dynamics in single ecosystems under constant light and temperature, measured continuously over several weeks. A model for fluctuations in the time series due to measurement noise was developed. The observed density time series were then shown to be strongly 'red-shifted', that is, power increased with the time scale of fluctuations. In other words, the time series were nonstationary on time scales up to the length of the experiment (~50 days). This finding was confirmed by wavelet analysis. Wavelet analysis suggests a statistically powerful, but less intuitive way of studying the correlations between density fluctuations in different species. The clearest correlations observed were between fluctuations in *C. reinhardtii* and *E. coli* on time scales of 0.5 to 4 days (Figure 6.11). I propose that the observed distribution of wavelet coefficients could be sampled to construct surrogate time series with the same wavelet power spectrum[74], and that the divergence between such surrogate data could be compared to actually observed divergence. Finally, the results from chapter 6 show that on 'intermediate' time scales of hours to days densities can be estimated with relatively little noise and only modest nonstationarity.

This last observation is exploited in chapter 7. In this chapter I explored the use of perturbations in boundary conditions to obtain information on the interaction between species. Specifically, the action of sinusoidal modulation of illumination intensity, with

[74] The same could be done, in principle, for a Fourier power spectrum.

periods of 24 and 31 hours, was studied. At amplitudes as small as 10%, population densities of *C. reinhardtii* and *E. coli* show a strong, significant response. As expected, the density of *C. reinhardtii* tracked the illumination with a strong peak in the power spectrum at 1 day^{-1}. Despite the absence of a circadian clock in *E. coli* it too showed a strong response in density, strongly suggesting an interaction with *C. reinhardtii*. Even more striking was a reproducible strong 12 hour component, implying that its densities respond nonlinearly to the perturbation. Under a 31 hour perturbation, both species showed more complex power spectra, with peaks at 1, 2 and 3 day_{31h}^{-1}, implying that (1) the response was largely the same over cycles of the perturbation and (2) the response of both species is nonlinear. The reconstruction of the phase of *C. reinhardtii* over time strongly suggests that this nonlinearity is a consequence of its circadian clock, the phase of which is reset at a discrete point every 31 h day.

The observation that most power is localized on 'integer' frequencies shows that system response can be robust for a modest perturbation. This should assuage concerns that the system varies too fast along unobserved system dimensions to reproducibly infer interactions.

However, chapter 7 also showed the challenges posed by the use of cyclical perturbations. First, intrinsic persistent rhythms, especially the circadian clock, can complicate the analysis. Secondly, both spatial dynamics and growth can contribute to observed density changes. Thirdly, causality cannot be rigorously inferred from only the response to cyclical perturbations.

In the next section I discuss how to proceed with the inverse problem based on these results.

Consequences for the inverse problem

In chapter 1, I gave the Lotka-Volterra model as an example of the phenomenological description of population dynamics. Specifically, in this model,

$$A_{ab} = -\frac{\partial r_a}{\partial n_b} = -\frac{\partial}{\partial n_b}\left(\frac{1}{n_a}\frac{\mathrm{d}n_a}{\mathrm{d}t}\right) \tag{8.1}$$

where A_{ab} is the interaction coefficient between species a and b, r_a the instantaneous growth rate of species a, and $n_a(t)$ and $n_b(t)$ are time series of their population densities. Unfortunately, partial derivatives are not directly observed, but, at best, total derivatives including the effects of other variables:

$$\frac{\partial}{\partial n_b} = \frac{\mathrm{d}}{\mathrm{d}n_b} - \sum_x \frac{\partial x}{\partial n_b}\cdot\frac{\partial}{\partial x} \tag{8.2}$$

as in equation 1.5, with x other observables, most of which will not be observed.

Closure of an ecosystem limits the complications caused by these unobserved variables. The unobserved variables, x, can be partitioned into a few classes. First, there are factors external to the ecosystem, the boundary conditions. In closed ecosystems, these can be known and controlled. Secondly, there are the densities of other species (e.g., species c when estimating A_{ab}), which can be measured using the method described in chapter 4. Then there are chemical constituents of the medium. In a closed ecosystem there are no material fluxes across the boundaries of the system. As a consequence, any change in the composition of the medium can be considered a passive consequence of the densities and phenotypes of the organisms present, and the laws of chemistry. As such, they are part of the very interactions that need to be quantified. Finally, there are two classes of variables

that are harder to deal with. The first is phenotypic composition. The second class of variables is spatial.

As the phenotypic composition of a population changes, its internal and interspecies interactions are affected. The difficulty phenotypic change poses, depends on its speed. Ideally, interactions change sufficiently little or sufficiently slowly that their main features can be identified from time series. Change in interactions on the time scale of population dynamics has been observed in chemostat experiments however [181]. While a separation of time scales is conceptually convenient, it remains to be seen whether it is necessary.

The last class of unobserved variables described the spatial distribution of organisms and, in particular, of distinct populations on surfaces. These populations can rival the population in the medium in size and have different phenotypes. There may, again, be reasonable approximations, such as similar interactions of wall and medium populations with other species and/or antagonistic interactions between the wall and medium populations. Or, alternatively, changes in one population may be much slower than in the other. In general, however, it is not yet clear whether any such approximation can be made.

I now return to equation 8.1. As seen in chapter 6, Figure 6.9, the correlation coefficients between species densities are weak and not stable over entire time series. In addition, the time derivative in equation 8.1 can not be estimated reliably[75]. Moreover, even strong

[75] These slopes are hard to estimate. For example, the slope of a moving average is locally dominated by noise. Likewise I considered a low pass filter motivated by figure 6.5. However, the power spectrum falls of as about f^{-2}, so the average absolute value of the Fourier transform falls of as about f^{-1}, implying that

relationships between the different populations can, in principle, be obscured by the additional variables *x*. As a consequence of this, equation 8.1 can not be applied directly to time series.

To estimate partial derivatives, the dynamics need instead to be mapped onto a phenomenological model [151]. The question is how observed population dynamics can be mapped onto models in the presence of unobserved variables. Can anything be said about these variables? I will consider two approaches: a nonlinear dynamics approach and a rigorous statistical approach.

First, in the theory of deterministic nonlinear dynamics there exists a basic conceptual answer to the problem of unobserved variables. If only a few of the dimensions of a deterministic nonlinear system are observed, $\{s_{obs}(t)\}$, inferences about the true number of relevant dynamical variables and their behavior can be made by finding a proper embedding dimension. This entails, in practice, finding a space of dimension d_E, and a time lag τ, such that the trajectory of $\{(s_{obs}(t), s_{obs}(t-\tau), ..., s_{obs}(t-(d_E-1)\cdot\tau))\}$ does not intersect itself (see ref. [158] for an account of the choice of d_E and τ)[76]. Unfortunately, the theorem is for infinite time series without measurement noise and dynamics without dynamical noise, and it is doubtful that any of this methodology can be applied to population density time series. It does provide a positive example of the inference of the existence and role of unobserved dynamical variables.

Another perspective, more suitable for stochastic processes, is provided by Bayesian statistics (see appendix for details). Bayes' theorem, applied to data, captures the essence

---

the average contribution of a Fourier component to the derivative is approximately frequency-independent, i.e. every term is approximately equally important.

[76] The original embedding theorem is rather vague about the choice of an embedding, but the presented method is the most common one.

of an inverse approach: a basic model, possibly including unobserved system variables, is formulated for the data, such that the model describes a probability distribution over the space of possible data (here the space of possible population dynamics trajectories). This model can include the description of both observed and unobserved system variables.

By Bayes' theorem, the probability of the data given a set of parameters is proportional to the probability of those parameters given the data[77]. In other words, the observed density time series can provide information on both the interaction coefficients in a model and variation in unobserved system variables included in the model. This information is contained in the posterior distribution of model parameters and system states given the data. The important open questions are (1) to what extent the data constrain the estimates of parameters and unobserved variables, and (2) what the best choice of system perturbations is to maximize the desired information in the posterior distribution.

The posterior distribution on the parameters and system states given the data can also be used to evaluate the likelihood of any number of scenarios, such as whether an interaction coefficient is significantly different from 0, or whether a species will be driven to extinction.

Ultimately, the question is what the limits are of an inverse approach and what data are needed to achieve these limits. Imagine, for example, that for a closed ecosystem we had knowledge of $N_a(\vec{x},t)$ for all species *a*, the positions $\vec{x}$ of the entire ecosystem volume and all $t \in [0,\infty)$, with an individual at a position *x* represented by a δ function. What could be inferred? The only unknowns left are the phenotypes of each organism[78] and the chemical composition of the ecosystem, which should be a *consequence* of the initial

[77] That is, assuming the model has the right structure. Alternatively, different models can be compared.
[78] That is, the phenotype is –here- broadly understood as the entire make-up an organism.

conditions and the phenotypic history of individuals in the ecosystem. Description of organisms in terms of phenotypes requires a certain degree of coarse-graining of cellular or organismal states. Does there need to be a minimum degree of heritability of such phenotypes in order to make reliable inferences, that is a certain stability of possible interactions over multiple generations? Can the space of phenotypes have arbitrary topology, or can only a finite number of states, say much smaller than the total population, or a continuous distribution of phenotypes with certain properties (e.g. smoothness) be described? Does there need to be a longest time scale of interaction?

Improvements in experimental design

I have shown that the study of interactions is simplified by the use of closed ecosystems. Progress can also be made in the design of experiments. There are three general strategies for obtaining information on interactions in a closed ecosystem: (1) from observation of dynamics under constant boundary conditions for an individual ecosystem, (2) from observation of the response of individual ecosystems to perturbation of boundary conditions, and (3) from variation of initial conditions across a set of ecosystems. This last approach was advocated by Pascual and Kareiva [182]. It probably is a powerful approach to identifying the main features of interactions, because system states with very different population densities can be explored, in which different interaction coefficients can have a dominant effect on dynamics. However, radically different species densities probably lead to different chemical compositions such that the systems rapidly become incomparable, precluding long-term studies. The use of small perturbations in boundary conditions is, hence, preferable.

The measurement of dynamics under constant boundary conditions, as in chapter 6, does not allow for much methodological improvement, and it remains to be seen what information such time series provide about species-species interactions.

The perturbation of boundary conditions can be optimized in several respects. First, data gathered earlier in the same experiment can suggest how to perturb the system to gain the most new information on interactions of interest ([183], see appendix).

Secondly, the perturbations in chapter 7 were cyclical. The primary advantage of such perturbations is that the results of many cycles can be averaged, increasing statistical power. The disadvantage of cyclical perturbations, however, is that they don't allow for unambiguous determination of causality. A sequence of perturbations of randomized duration might circumvent this issue, while still retaining some of the averaging power of cyclical perturbations.

Additional experiments

Several kinds of experimental information could strengthen our approach. First, information on the spatial distribution of each species over time would be valuable. Interactions can depend both on location within the ecosystem and distance. Spatial information comes in two flavors: information about spatial structure within the liquid medium and about populations growing on surfaces. So far, I have assumed that the observation volume, 5 mm below the meniscus is representative of the entire closed ecosystem. In other words I assumed that each species is uniformly distributed, or at least that densities fluctuate proportionally, throughout the liquid phase of the ecosystem. The results in chapter 7 show that, while this may be true in the long run (Figure 7.12), it

probably is not true for the short term response to perturbations in light and temperature. Interactions can vary in the medium (e.g., shading effects between algae, the formation of oxygen gradients, decomposition at the bottom). Surface populations differ phenotypically from populations in the medium, as shown in chapter 3. Efforts are currently under way in the laboratory to image the three-dimensional distribution of organisms.

The importance of *gross* growth rates (i.e., division rates) has surfaced repeatedly in this thesis. For instance, to really demonstrate the persistence of species in a closed ecosystem, one needs to show survival over many generations, not just for long times. Since a predator is present in the system, *T. thermophila*, which does not survive for more than two or three weeks under starvation (Figure 3.4), it is likely that there is substantial growth and death of both *T. thermophila* and *E. coli*. This cannot be inferred, however, from the observed population density time series, since these only show *net* population changes. Likewise, demographic fluctuations and the rate of phenotypic change are dependent on gross growth rates, as is (in most models) the rate at which population densities respond to perturbations and demographic fluctuations. Possibly, gross growth rates can be estimated non-invasively by placing the gene for a fluorescent protein with time-dependent emission properties [184] or a degradation tag under the control of a cell-cycle dependent promoter.

Finally, a better grasp of the effects of light and temperature, as well as an extension of the number of 'knobs' with which to perturb the system, would be welcome. As shown, our model closed ecosystem responds sensitively to such perturbations. A good understanding of the effects of light and temperature would allow addressing specific species-species interactions by causing specific changes in the densities of each species.

There appear to be many genetic tools available to extend the control afforded by light and temperature. For example, temperature-sensitive mutants are available for many genes and have long been used in genetics. Light-based genetic tools are much more rare. For *E. coli*, mutants have been isolated which can be killed by blue light, but have only limited growth defects in darkness [185-187]. A flexible light-controllable gene expression system has been developed by Shimizu-Sato and colleagues [188] based on a plant photoreceptor, phyB. The binding of this receptor to a transcription factor, PIF3, is dependent on the state of its chromophore. The state of this chromophore can be controlled using red and near-infrared illumination (peak absorption at 664 and 748 nm, respectively) to turn transcription on and off, respectively. Genes necessary for the production of the chromophore, a tetrapyrrole, can be introduced into bacteria.

Mixing forward and inverse approaches

The approach in this thesis is different from classical genetics approaches, in which either a candidate gene is knocked out and the resulting phenotype studied, or a screen or selection experiment is conducted to find mutations resulting in a phenotype of interest. There are several ways, however, in which the methodology in this thesis can be combined with classical genetics approaches. First, one can simply see which mutations thrive in a closed ecosystem. By sequencing of small samples, genotype dynamics could be followed over time and linked to inferred changes in interaction patterns. Secondly, one could perform competition experiments between strains with different genotypes within an ecosystem, or between −otherwise replicate− sets of ecosystems. The latter experiment would, ideally, connect differences in inferred interactions to genetic

differences. The observed phenotypes corresponding to different genotypes would be much richer in features than they would be in a screen or selection experiment.

Mutations occur in a dynamic context, not in an idealized background. In effect, when studying the survival of mutants over long times, mutations in genotype neighborhoods rather than individual mutations are compared. In other words, the fitness of a mutant is not just dependent on the mutation of interest but also on any other mutations it interacts with, the effects of mutations which occurred in other species in the ecosystem and the response of other species and strains. In effect, one obtains a much more subtle description of phenotypes than provided by conventional techniques and an actual connection of phenotype to fitness.

The feasibility of reproducible long-term studies, allowed by closed ecosystems, raises the question though: why limit ourselves to the study of phenotypes we can imagine? In an inverse approach the system tells us which traits are important. Closure of the system allows us to actually listen.

Appendix 8A.1: Bayesian approach to inverse problems

Basis

The basis of bayesian statistics is Bayes' theorem:

$$P(A|B)\cdot P(B) = P(B|A)\cdot P(A) = P(A\ \&\ B) \tag{8A.1}$$

For two possible events, this states that the probability that A will happen given that B will happen multiplied by the probability that B will happen equals the probability that B will happen given that A will happen multiplied by the probability that A will happen (and both products equal the probability both A and B will happen). Now, if A is a set of parameter values, and B is a particular set of observations, the bayesian formalism can be used to update our expectations (belief) about the true parameter values based on observations [189].

Implementation

Let x_t be the state of our system at time t, s_t the measurement(s) done on the system at time t (e.g., the measured populations time series), and θ a set of parameters, including both dynamical parameters and properties of the measurements (e.g. the variance). Bayes' theorem implies for the joint posterior probability distribution on the initial conditions and parameters:

$$P(\theta, x_0|\{s_t\}) \propto P(\{s_t\}|\theta, x_0)\cdot P^o(x_0)\cdot P^o(\theta) \tag{8A.2}$$

with x_0 the initial state of the (eco)system.

$$P(\{s_t\}|x_0,\theta) = \int_X d\{x_t\} P(\{s_t\}|\{x_t\},\theta) \cdot P(\{x_t\}|x_0,\theta) \tag{8A.3}$$

where the first factor in equation 8A.3 describes the measurements, and the second the real system dynamics. The proportionality constant in equation 8A.2,

$$\left[\int_\Theta d\theta \int_X d\{x_t\}\, P(\{s_t\}|\{x_t\},\theta) \cdot P(\{x_t\}|x_0,\theta) \cdot P^o(x_0) \cdot P^o(\theta) \right]^{-1} \tag{8A.4}$$

is prohibitively difficult to evaluate. There are two ways to address this dilemma.

In the most rigorous approach, called the Markov Chain Monte Carlo (MCMC) method [190], the same trick is used as in statistical mechanics: a Monte Carlo walk is performed over the space $X \times \Theta$ (with X the space of underlying system states from which $\{x_t\}$ is drawn, and Θ the parameter space from which to estimate θ). Instead of the ratio of Boltzmann factors used in Monte Carlo methods in physics, acceptation of transitions is based on the ratio of posterior probabilities assigned to the data based on equation 8A.2. The Monte Carlo walk has as its stationary distribution the true posterior distribution of θ. So from a sufficiently long Monte Carlo walk, equation 8A.2 can be evaluated without normalizing the distribution. Brooks' group has developed a software package (WinBUGS, http://www.mrc-bsu.cam.ac.uk/bugs) that integrates MCMC and analytical methods.

The MCMC method is, however, computationally intensive and equilibration of the algorithm is not guaranteed for finite time scales. The alternative approach is similar to a maximum likelihood approach: the posterior distribution (equation 8A.2) can be stated in exponential form, such that the distribution will be sharply peaked on the argument of the

exponential. Our lab is collaborating with R. Monasson and S. Cocco (ENS), who have experience with the analytical treatment of such maximization [191]. The Hessian of the argument of the exponential can be estimated in order to obtain error estimates.

Optimal experimental design

In principle, the bayesian framework suggests how to optimally design experiments. Design of an experiment implies selecting its initial conditions, x_0, duration, and boundary conditions (light and temperature over time).

Given a prior distribution for system and measurement parameters, possibly based on previous experiments, there is an expectation of what will be observed, and hence, of what can be learnt from the observations given the experimental design.

To be precise, given a prior distribution of system and measurement parameters θ, $P^o(\theta)$, the distribution of observations, prior to the actual experiment, is

$$P_{x_0}(\{s_t\}) = \int_\Theta d\theta \, P_{x_0}(\{s_t\}|\theta) P^o(\theta) \tag{8A.5}$$

Next, each actual observation $\{s_t\}$ would lead to a *posterior*, or updated, distribution on system and measurement parameters, $P_{x_0}(\theta|\{s_t\})$. So, effectively, $P^o(\theta)$, and a choice of initial and boundary conditions also induce a probability distribution on the space of posterior probability distributions for θ. That is, there is a

$$P_{x_0}\left[P(\theta|\{s_t\})|P^o(\theta)\right] \tag{8A.6}$$

In other words, the experimental design, here simply indicated by the subscript x_0, determines what can be learnt from the experiment about θ. How much we can expect to learn from an experiment, can be quantified using information theory [153], which fits naturally with bayesian statistics. With each posterior distribution there is an associated posterior uncertainty,

$$H\left[P\left(\theta\middle|\{s_t\}\right)\right] = -\int_{\Theta} d\theta\, P(\theta)\log P(\theta) \tag{8A.7}$$

and the information gained, $I = H^o - H$, equals the prior minus the posterior uncertainty in θ. Hence, for each set of initial conditions, the expected information gain can be calculated, and an optimal set of initial conditions chosen.

The above scheme was implemented by Póczos and Lőrincz for the study of neural networks for which only the input can be given (the experimental design) and the output observed [183]. For deterministic systems, Emery and colleagues used a similar approach [192-194], based on Fisher information.

Bibliography

1. Maguire, B.J., *Some Patterns in Post-Closure Ecosystem Dynamics (Failure)*, in *Microcosms in Ecological Research*, J.P.J. Giesy, Editor. 1980, US DOE: Washington, DC. p. 319-332.

2. Brockwell, P.J. and R.A. Davis, *Introduction to Time Series and Forecasting*. Springer Texts in Statistics, ed. G. Casella. 2002, New York: Springer-Verlag.

3. Hamilton, J.D., *Time Series Analysis*. 1994, Princeton, NJ: Princeton University Press.

4. Kraskov, A., H. Stogbauer, and P. Grassberger, *Estimating Mutual Information.* Phys Rev E Stat Nonlin Soft Matter Phys, 2004. **69**(6 Pt 2): p. 066138.

5. Torrence, C. and G.P. Compo, *A Practical Guide to Wavelet Analysis.* Bulletin of the American Meteorological Society, 1998. **79**(1): p. 61-78.

6. Anon., *The Ecosphere(R) Handbook*, Ecosphere Associates, Inc. p. 12.

7. Hanson, J.A. and H.P. Stevens, *Experimental Ecosystems Sealed in Glass*. 1984, Jet Propulsion laboratory (NASA/CalTech): Pasadena, CA.

8. Sugiura, K., Y. Kurihara, and E. Wada. *Characteristics of a Materially Closed Microbial Ecosystem*. in *11th ISAS Space Utilization Symposium*. 1994. Tokyo.

9. Gorden, R.W., R.J. Beyers, E.P. Odum, and R.G. Eagon, *Studies of a Simple Laboratory Microecosystem - Bacterial Activities in a Heterotrophic Succession.* Ecology, 1969. **50**(1): p. 86-&.

10. Weatherby, A.J., P.H. Warren, and R. Law, *Coexistence and Collapse: An Experimental Investigation of the Persistent Communities of a Protist Species Pool.* Journal of Animal Ecology, 1998. **67**(4): p. 554-566.

11. Jain, R. and I. Chlamtac, *The P2 Algorithm for Dynamic Calculation of Quantiles and Histograms without Storing Observations.* Communications of the Acm, 1985. **28**(10): p. 1076-1085.

12. Darwin, C., *The Origin of Species*. Reprint of the 1976 issue of the 1968 edition by Penguin Books ed. 1979, New York: Gramercy Books.

13. Ebeling, W. and R. Feistel, *Theory of Self-Organization and Evolution - the Role of Entropy, Value and Information.* Journal of Non-Equilibrium Thermodynamics, 1992. **17**(4): p. 303-332.

14. Lunzer, M., S.P. Milter, R. Felsheim, and A.M. Dean, *The Biochemical Architecture of an Ancient Adaptive Landscape.* Science, 2005. **310**(5747): p. 499-501.

15. Price, N.D., J.L. Reed, and B.O. Palsson, *Genome-Scale Models of Microbial Cells: Evaluating the Consequences of Constraints.* Nature Reviews Microbiology, 2004. **2**(11): p. 886-897.

16. Edwards, J.S., R.U. Ibarra, and B.O. Palsson, *In Silico Predictions of Escherichia Coli Metabolic Capabilities Are Consistent with Experimental Data.* Nature Biotechnology, 2001. **19**(2): p. 125-130.

17. Dekel, E. and U. Alon, *Optimality and Evolutionary Tuning of the Expression Level of a Protein.* Nature, 2005. **436**(7050): p. 588-592.

18. Schrödinger, E., *What Is Life?* Canto ed. 1967, Cambride, UK: Cambridge University Press.

19. Kussell, E. and S. Leibler, *Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments.* Science, 2005. **309**(5743): p. 2075-2078.

20. Turchin, P., *Complex Population Dynamics: A Theoretical/Empirical Synthesis* Monographs in Population Biology. 2003: Princeton University Press.

21. Kawabata, Z., K. Matsui, K. Okazaki, M. Nasu, N. Nakano, and T. Sugai, *Synthesis of a Species-Defined Microcosm with Protozoa.* Journal of Protozoology Research, 1995. **5**(1): p. 23-26.

22. Huisken, J., J. Swoger, F. Del Bene, J. Wittbrodt, and E.H.K. Stelzer, *Optical Sectioning Deep inside Live Embryos by Selective Plane Illumination Microscopy.* Science, 2004. **305**(5686): p. 1007-1009.

23. Beyers, R.J. and H.T. Odum, *Ecological Microcosms*. Springer Advanced Texts in Life Sciences, ed. D.E. Reichle. 1993, New York: Springer-Verlag.

24. Whittaker, R.H. and P.P. Feeny, *Allelochemics - Chemical Interactions between Species.* Science, 1971. **171**(3973): p. 757-&.

25. Shaffer, J., *Stability in Closed Ecological Systems: An Examination of Material and Energetic Parameters*, in *Microbiology*. 1991, University of Hawaii.

26. Jessup, C.M., R. Kassen, S.E. Forde, B. Kerr, A. Buckling, P.B. Rainey, and B.J.M. Bohannan, *Big Questions, Small Worlds: Microbial Model Systems in Ecology.* Trends In Ecology & Evolution, 2004. **19**(4): p. 189-197.

27. Konig, B., M. Dunne, and K. Slenzka, *Earth Life Support for Aquatic Organisms, System and Technical Aspects.* Space Life Sciences: Closed Ecological Systems: Earth and Space Applications, 2001. **27**(9): p. 1523-1528.

28. Nitta, K., K. Otsubo, and A. Ashida, *Integration Test Project of Ceef - a Test Bed for Closed Ecological Life Support Systems.* Life Sciences: Space Life Support Systems and the Lunar Farside Crater Saha Proposal, 2000. **26**(2): p. 335-338.

29. Taub, F.B., *Closed Ecological Systems.* Annual reviews of Ecology and Systematics, 1974: p. 139-160.

30. Gitelson, I.I., G.M. Lisovsky, and R.D. MacElroy, *Manmade Closed Ecological Systems*. Earth Space Institute Book Series, ed. P. Kleber. Vol. 9. 2003, London and New York: Taylor & Francis. 402.

31. Allen, J. and M. Nelson, *Biospherics and Biosphere 2, Mission One (1991-1993).* Ecological Engineering, 1999. **13**(1-4): p. 15-29.

32. Morowitz, H., J.P. Allen, M. Nelson, and A. Alling, *Closure as a Scientific Concept and Its Application to Ecosystem Ecology and the Science of the Biosphere*, in *Space Life Sciences: Gravity-Related Effects on Plants and Spaceflight and Man-Made Environments on Biological Systems*. 2005. p. 1305-1311.

33. Walter, A. and S.C. Lambrecht, *Biosphere 2 Center as a Unique Tool for Environmental Studies.* Journal of Environmental Monitoring, 2004. **6**(4): p. 267-277.

34. Cooke, G.D., R.J. Beyers, and E.P. Odum, *The Case for the Multispecies Ecological System with Special Reference to Succession and Stability*. 1968, N.A.S.A.: Washington, D.C. p. 129-139.

35. Hanson, J.A., *Workshop on Closed-System Ecology, Held at The "California-Institute-of-Technology, Pasadena, California, USA, During 18-22 January 1982.* Environmental Conservation, 1982. **9**(4): p. 357-359.

36. Beyers, R.J., *Metabolism of 12 Aquatic Laboratory Microecosystems.* Ecological Monographs, 1963. **33**(4): p. 281-&.

37. Cooke, G.D., *Pattern Autotrophic Succession in Laboratory Microcosms.* Bioscience, 1967. **17**(10): p. 717-&.

38. Kurihara, Y., *Studies of the Interaction in a Microcosm.* Scientific reports of Tôhoku University Series IV, 1978. **37**: p. 161-177.

39. Sugiura, K., *The Use of an Aquatic Microcosm for Pollution Effects Assessment.* Water Research, 1996. **30**(8): p. 1801-1812.

40. Sugiura, K., *A Materially-Closed Aquatic-Ecosystem: A Useful Tool for Determining Changes of Ecological Processes in Space.* Biological Sciences in Space, 1998. **12**(2): p. 115-118.

41. Obenhuber, D.C., *The Persistence of Life Measured by Carbon Cycling in Closed Ecological Systems*, in *Microbiology*. 1986, University of Hawaii. p. 129.

42. Obenhuber, D.C. and C.E. Folsome, *Carbon Recycling in Materially Closed Ecological Life Support Systems.* Biosystems, 1988. **21**(2): p. 165-173.

43. Wright, D.H., C.E. Folsome, and D. Obenhuber, *Competition and Efficiency in Closed Fresh-Water Algal Systems - Tests of Ecosystem Design Principles.* Biosystems, 1985. **17**(3): p. 233-239.

44. Warren, P.H., R. Law, and A.J. Weatherby, *Mapping the Assembly of Protist Communities in Microcosms.* Ecology, 2003. **84**(4): p. 1001-1011.

45. Huisman, J., *Population Dynamics of Light-Limited Phytoplankton: Microcosm Experiments.* Ecology, 1999. **80**(1): p. 202-210.

46. Passarge, J., S. Hol, M. Escher, and J. Huisman, *Competition for Nutrients and Light: Stable Coexistence, Alternative Stable States, or Competitive Exclusion?* Ecological Monographs, 2006. **76**(1): p. 57-72.

47. Kearns, E.A., *Efficiency of Energy Utilization in Thermodynamically Closed Ecosystems*, in *Microbiology*. 1983, University of Hawaii. p. 139.

48. Brittain, A.M., *The Metabolic Diversity, Biological Activity, and Stability of the Steady State Condition in Closed Ecosystems*, in *Microbiology*. 1993, University of Hawaii. p. 145.

49. Folsome, C.E. and J.A. Hanson, *Xii: The Emergence of Materially-Closed-System Ecology*, in *Ecosystem Theory and Applications*, N. Polunin, Editor. 1986, John Wiley & Sons.

50. Kearns, E.A. and C.E. Folsome, *Measurement of Biological-Activity in Materially Closed Microbial Ecosystems.* Biosystems, 1981. **14**(2): p. 205-209.

51. Anon., *Services & Facilities Annual Report, the Ecotron*. 2007, NERC Centre for Population Biology: London. p. 4.

52. Nitta, K., *The Ceef, Closed Ecosystem as a Laboratory for Determining the Dynamics of Radioactive Isotopes*, in *Space Life Sciences: Closed Ecological Systems: Earth and Space Applications*. 2001. p. 1505-1512.

53. Sterner, R.W. and J.J. Elser, *Ecological Stoichiometry*. 1st ed. 2002, Princeton, NJ: Princeton University Press.

54. Loladze, I., Y. Kuang, J.J. Elser, and W.F. Fagan, *Competition and Stoichiometry: Coexistence of Two Predators on One Prey.* Theoretical Population Biology, 2004. **65**(1): p. 1-15.

55. Pisman, T.I., N.S. Pechurkin, and L.A. Somova, *Competition between Links In "Producer-Consumer" Trophic Chains in an Aquatic Closed System with Spatially Separated Components*, in *Space Life Sciences: Closed Ecological Systems: Earth and Space Applications*. 2001. p. 1599-1603.

56. Cadotte, M.W., J.A. Drake, and T. Fukami, *Constructing Nature: Laboratory Models as Necessary Tools for Investigating Complex Ecological Communities.* Advances In Ecological Research, Vol. 37: Population Dynamics And Laboratory Ecology, 2005. **37**: p. 333-353.

57. Brittain, A.M. and D.M. Karl, *Catabolism of Tritiated-Thymidine by Aquatic Microbial Communities and Incorporation of Tritium into Rna and Protein.* Applied And Environmental Microbiology, 1990. **56**(5): p. 1245-1254.

58. Matsui, K., S. Kono, A. Saeki, N. Ishii, M.G. Min, and Z. Kawabata, *Direct and Indirect Interactions for Coexistence in a Species-Defined Microcosm.* Hydrobiologia, 2000. **435**(1-3): p. 109-116.

59. Elena, S.F. and R.E. Lenski, *Evolution Experiments with Microorganisms: The Dynamics and Genetic Bases of Adaptation.* Nature Reviews Genetics, 2003. **4**(6): p. 457-469.

60. Lenski, R.E. and M. Travisano, *Dynamics of Adaptation and Diversification - a 10,000-Generation Experiment with Bacterial-Populations.* Proceedings Of The National Academy Of Sciences Of The United States Of America, 1994. **91**(15): p. 6808-6814.

61. Fredrickson, A.G., *Behavior of Mixed Cultures of Microorganisms.* Annual Review of Microbiology, 1977. **31**: p. 63-87.

62. Fredrickson, A.G. and G. Stephanopoulos, *Microbial Competition.* Science, 1981. **213**(4511): p. 972-979.

63. Grivet, J.P., A.M. Delort, and J.C. Portais, *Nmr and Microbiology: From Physiology to Metabolomics.* Biochimie, 2003. **85**(9): p. 823-840.

64. Boschker, H.T.S. and J.J. Middelburg, *Stable Isotopes and Biomarkers in Microbial Ecology.* Fems Microbiology Ecology, 2002. **40**(2): p. 85-95.

65. Chiarappa-Zucca, M.L., K.H. Dingley, M.L. Roberts, C.A. Velsko, and A.H. Love, *Sample Preparation for Quantitation of Tritium by Accelerator Mass Spectrometry.* Analytical Chemistry, 2002. **74**(24): p. 6285-6290.

66. Fox, S.D., et al., *A Comparison of Mu Lc/Electrospray Ionization-Ms and Gc/Ms for the Measurement of Stable Isotope Enrichment from a [H-2(2)]-Glucose Metabolic Probe in T-Cell Genomic DNA.* Analytical Chemistry, 2003. **75**(23): p. 6517-6522.

67. Daly, K.L., R.H. Byrne, A.G. Dickson, S.M. Gallager, M.J. Perry, and M.K. Tivey, *Chemical and Biological Sensors for Time-Series Research: Current Status and New Directions.* Marine Technology Society Journal, 2004. **38**(2): p. 121-143.

68. Hegreness, M., N. Shoresh, D. Hartl, and R. Kishony, *An Equivalence Principle for the Incorporation of Favorable Mutations in Asexual Populations.* Science, 2006. **311**(5767): p. 1615-1617.

69. Millie, D.F., O.M.E. Schofield, G.J. Kirkpatrick, G. Johnsen, and T.J. Evens, *Using Absorbance and Fluorescence Spectra to Discriminate Microalgae.* European Journal of Phycology, 2002. **37**(3): p. 313-322.

70. Wilhelm, C. and L. Manns, *Changes in Pigmentation of Phytoplankton Species During Growth and Stationary Phase - Consequences for Reliability of Pigment-Based Methods of Biomass Determination.* Journal Of Applied Phycology, 1991. **3**(4): p. 305-310.

71. Davis, C.S., Q. Hu, S.M. Gallager, X. Tang, and C.J. Ashjian, *Real-Time Observation of Taxa-Specific Plankton Distributions: An Optical Sampling Method.* Marine Ecology-Progress Series, 2004. **284**: p. 77-96.

72. Joeris, K., J.G. Frerichs, K. Konstantinov, and T. Scheper, *In-Situ Microscopy: Online Process Monitoring of Mammalian Cell Cultures.* Cytotechnology, 2002. **38**(1-2): p. 129-134.

73. Genin, A., J.S. Jaffe, R. Reef, C. Richter, and P.J.S. Franks, *Swimming against the Flow: A Mechanism of Zooplankton Aggregation.* Science, 2005. **308**(5723): p. 860-862.

74. Hebert, P.D.N., S. Ratnasingham, and J.R. deWaard, *Barcoding Animal Life: Cytochrome C Oxidase Subunit 1 Divergences among Closely Related Species.* Proceedings of the Royal Society of London Series B-Biological Sciences, 2003. **270**: p. S96-S99.

75. Becker, S., M. Fahrbach, P. Boger, and A. Ernst, *Quantitative Tracing, by Taq Nuclease Assays, of a Synechococcus Ecotype in a Highly Diversified Natural Population.* Applied and Environmental Microbiology, 2002. **68**(9): p. 4486-4494.

76. Stubner, S., *Enumeration of 16s Rdna of Desulfotomaculum Lineage 1 in Rice Field Soil by Real-Time Pcr with Sybrgreen (Tm) Detection.* Journal of Microbiological Methods, 2002. **50**(2): p. 155-164.

77. Shikano, S. and Z. Kawabata, *Effect at the Ecosystem Level of Elevated Atmospheric Co2 in an Aquatic Microcosm.* Hydrobiologia, 2000. **436**(1-3): p. 209-216.

78. Fuma, S., et al., *Ecological Effects of Various Toxic Agents on the Aquatic Microcosm in Comparison with Acute Ionizing Radiation.* Journal of Environmental Radioactivity, 2003. **67**(1): p. 1-14.

79. Fuma, S., et al., *Simple Aquatic Microcosm for Ecotoxicity Screening at the Community Level.* Bulletin of Environmental Contamination and Toxicology, 2000. **65**(6): p. 699-706.

80. Matsui, K., N. Ishii, and Z. Kawabata, *Microbial Interactions Affecting the Natural Transformation of Bacillus Subtilis in a Model Aquatic Ecosystem.* Fems Microbiology Ecology, 2003. **45**(3): p. 211-218.

81. Harris, E.H., *The Chlamydomonas Sourcebook*. 1989, San Diego: Academic Press, Inc.

82. *The Biology of Euglena*, D.E. Buetow, Editor. 1968, Academic Press, Inc.: New York and London.

83. Neidhardt, F.C., J.L. Ingraham, and M. Schaechter, *Physiology of the Bacterial Cell*. 1990, Sunderland, MA: Sinauer Associates.

84. Asai, D.J. and J.D. Forney, eds. *Tetrahymena Thermophila*. Methods in Cell Biology, ed. L. Wilson and P. Matsudaira. Vol. 62. 2000, Academic Press: San Diego.

85. Laakso, J., K. Loytynoja, and V. Kaitala, *Environmental Noise and Population Dynamics of the Ciliated Protozoa Tetrahymena Thermophila in Aquatic Microcosms.* Oikos, 2003. **102**(3): p. 663-671.

86. Muela, A., J.M. Garcia-Bringas, I. Arana, and I. Barcina, *The Effect of Simulated Solar Radiation on Escherichia Coli: The Relative Roles of Uv-B, Uv-a, and Photosynthetically Active Radiation.* Microbial Ecology, 2000. **39**(1): p. 65-71.

87. *Certificate of Analysis*. [cited; Available from: www.bd.com/ds/technicalCenter/typicalAnalysis/typ-proteose_peptone3.pdf.

88. Taub, F.B. and A.M. Dollar, *A Chlorella-Daphnia Food-Chain Study - the Design of a Compatible Chemically Defined Culture Medium.* Limnology and Oceanography, 1964. **9**(1): p. 61-74.

89. Miller, J.H., *A Short Course in Bacterial Genetics*. 1992, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

90. Crutchfield, A.L.M., K.R. Diller, and J.J. Brand, *Cryopreservation of Chlamydomonas Reinhardtii (Chlorophyta).* European Journal of Phycology, 1999. **34**(1): p. 43-52.

91. Finkel, S.E. and R. Kolter, *Evolution of Microbial Diversity During Prolonged Starvation.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(7): p. 4023-4027.

92. Macarthur, R.H. and E.O. Wilson, *Equilibrium-Theory of Insular Zoogeography.* Evolution, 1963. **17**(4): p. 373-&.

93. Preston, F.W., *Canonical Distribution of Commonness and Rarity .1.* Ecology, 1962. **43**(2): p. 185-&.

94. Preston, F.W., *Canonical Distribution of Commonness and Rarity .2.* Ecology, 1962. **43**(3): p. 410-&.

95. Brose, U., A. Ostling, K. Harrison, and N.D. Martinez, *Unified Spatial Scaling of Species and Their Trophic Interactions.* Nature, 2004. **428**(6979): p. 167-171.

96. Hubbell, S.P., *The Unified Neutral Theory of Biodiversity and Biogeography*. Monographs in Population Biology, ed. S.A. Levin and H.S. Horn. 2001, Princeton: Princeton University Press.

97. Maharjan, R., S. Seeto, L. Notley-McRobb, and T. Ferenci, *Clonal Adaptive Radiation in a Constant Environment.* Science, 2006. **313**(5786): p. 514-517.

98. Treves, D.S., S. Manning, and J. Adams, *Repeated Evolution of an Acetate-Crossfeeding Polymorphism in Long-Term Populations of Escherichia Coli.* Molecular Biology and Evolution, 1998. **15**(7): p. 789-797.

99. Zambrano, M.M., D.A. Siegele, M. Almiron, A. Tormo, and R. Kolter, *Microbial Competition- Escherichia Coli Mutants That Take over Stationary Phase Cultures.* Science, 1993. **259**(5102): p. 1757-1760.

100. Zinser, E.R. and R. Kolter, *Prolonged Stationary-Phase Incubation Selects for Lrp Mutations in Escherichia Coli K-12.* Journal of Bacteriology, 2000. **182**(15): p. 4361-4365.

101. Rainey, P.B. and M. Travisano, *Adaptive Radiation in a Heterogeneous Environment.* Nature, 1998. **394**(6688): p. 69-72.

102. Rainey, P.B. and K. Rainey, *Evolution of Cooperation and Conflict in Experimental Bacterial Populations.* Nature, 2003. **425**(6953): p. 72-74.

103. Gause, G.F., *The Struggle for Existence*. 1934, Dover Books

104. Hardin, G., *Competitive Exclusion Principle.* Science, 1960. **131**(3409): p. 1292-1297.

105. Rainey, P.B., A. Buckling, R. Kassen, and M. Travisano, *The Emergence and Maintenance of Diversity: Insights from Experimental Bacterial Populations.* Trends in Ecology & Evolution, 2000. **15**(6): p. 243-247.

106. Sineshchekov, O.A., K.H. Jung, and J.L. Spudich, *Two Rhodopsins Mediate Phototaxis to Low- and High-Intensity Light in Chlamydomonas Reinhardtii.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(13): p. 8689-8694.

107. Wakabayashi, K. and S.M. King, *Modulation of Chlamydomonas Reinhardtii Flagellar Motility by Redox Poise.* Journal of Cell Biology, 2006. **173**(5): p. 743-754.

108. Bruce, V.G., *Mutants of Biological Clock in Chlamydomonas-Reinhardi.* Genetics, 1972. **70**(4): p. 537-&.

109. Goto, K. and C.H. Johnson, *Is the Cell-Division Cycle Gated by a Circadian Clock - the Case of Chlamydomonas-Reinhardtii.* Journal of Cell Biology, 1995. **129**(4): p. 1061-1069.

110. Hasman, H., M.A. Schembri, and P. Klemm, *Antigen 43 and Type 1 Fimbriae Determine Colony Morphology of Escherichia Coli K-12.* Journal of Bacteriology, 2000. **182**(4): p. 1089-1095.

111. Kjaergaard, K., M.A. Schembri, H. Hasman, and P. Klemm, *Antigen 43 from Escherichia Coli Induces Inter- and Intraspecies Cell Aggregation and Changes in Colony Morphology of Pseudomonas Fluorescens.* Journal of Bacteriology, 2000. **182**(17): p. 4789-4796.

112. Shikano, S., L.S. Luckinbill, and Y. Kurihara, *Changes of Traits in a Bacterial Population Associated with Protozoal Predation.* Microbial Ecology, 1990. **20**(1): p. 75-84.

113. Corno, G. and K. Jurgens, *Direct and Indirect Effects of Protist Predation on Population Size Structure of a Bacterial Strain with High Phenotypic Plasticity.* Applied and Environmental Microbiology, 2006. **72**(1): p. 78-86.

114. Goehring, N.W. and J. Beckwith, *Diverse Paths to Midcell: Assembly of the Bacterial Cell Division Machinery.* Current Biology, 2005. **15**(13): p. R514-R526.

115. Branda, S.S., A. Vik, L. Friedman, and R. Kolter, *Biofilms: The Matrix Revisited.* Trends in Microbiology, 2005. **13**(1): p. 20-26.

116. Reisner, A., J.A.J. Haagensen, M.A. Schembri, E.L. Zechner, and S. Molin, *Development and Maturation of Escherichia Coli K-12 Biofilms.* Molecular Microbiology, 2003. **48**(4): p. 933-946.

117. Preer, J.R., *Epigenetic Mechanisms Affecting Macronuclear Development in Paramecium and Tetrahymena.* Journal of Eukaryotic Microbiology, 2000. **47**(6): p. 515-524.

118. Bleyman, L.K., M.P. Baum, P.J. Bruns, and E. Orias, *Mapping the Mating Type Locus of Tetrahymena-Thermophila - Meiotic Linkage of Mat to the Ribosomal-Rna Gene.* Developmental Genetics, 1992. **13**(1): p. 34-40.

119. Nelsen, E.M., *Transformation in Tetrahymena-Thermophila - Development of an Inducible Phenotype.* Developmental Biology, 1978. **66**(1): p. 17-31.

120. Hill, R.J., *Thermotropic Changes Occur at Fixed Temperatures in Tetrahymena in Spite of Variations in Culture Temperature.* Biochimica Et Biophysica Acta, 1983. **732**(1): p. 11-23.

121. Goto, M., K. Ohki, and Y. Nozawa, *Evidence for a Correlation between Swimming Velocity and Membrane Fluidity of Tetrahymena Cells.* Biochimica Et Biophysica Acta, 1982. **693**(2): p. 335-340.

122. Platt, J.R., *Bioconvection Patterns in Cultures of Free-Swimming Organisms.* Science, 1961. **133**(346): p. 1766-&.

123. Giese, A., *Blepharisma: The Biology of a Light-Sensitive Protozoan*. 1973, Stanford, CA: Stanford University Press.

124. Teplitski, M., et al., *Chlamydomonas Reinhardtii Secretes Compounds That Mimic Bacterial Signals and Interfere with Quorum Sensing Regulation in Bacteria.* Plant Physiology, 2004. **134**(1): p. 137-146.

125. Bassler, B.L., *Small Talk: Cell-to-Cell Communication in Bacteria.* Cell, 2002. **109**(4): p. 421-424.

126. Yao, Y., M.A. Martinez-Yamout, T.J. Dickerson, A.P. Brogan, P.E. Wright, and H.J. Dyson, *Structure of the Escherichia Coli Quorum Sensing Protein Sdia: Activation of the Folding Switch by Acyl Homoserine Lactones.* Journal of Molecular Biology, 2006. **355**(2): p. 262-273.

127. Lutz, R. and H. Bujard, *Independent and Tight Regulation of Transcriptional Units in Escherichia Coli Via the Lacr/O, the Tetr/O and Arac/I-1-I-2 Regulatory Elements.* Nucleic Acids Research, 1997. **25**(6): p. 1203-1210.

128. Manen, D., G.X. Xia, and L. Caro, *A Locus Involved in the Regulation of Replication in Plasmid Psc101.* Molecular Microbiology, 1994. **11**(5): p. 875-884.

129. Murdoch, W.W., *Population Regulation in Theory and Practice - the Robert-H-Macarthur-Award-Lecture Presented August 1991 in San-Antonio, Texas, USA.* Ecology, 1994. **75**(2): p. 271-287.

130. Stenseth, N.C., *Population Cycles in Voles and Lemmings: Density Dependence and Phase Dependence in a Stochastic World.* Oikos, 1999. **87**(3): p. 427-461.

131. Abu-Absi, N.R., A. Zamamiri, J. Kacmar, S.J. Balogh, and F. Srienc, *Automated Flow Cytometry for Acquisition of Time-Dependent Population Data.* Cytometry Part A, 2003. **51A**(2): p. 87-96.

132. Balagadde, F.K., L.C. You, C.L. Hansen, F.H. Arnold, and S.R. Quake, *Long-Term Monitoring of Bacteria Undergoing Programmed Population Control in a Microchemostat.* Science, 2005. **309**(5731): p. 137-140.

133. Stanislas, M. and J.C. Monnier, *Practical Aspects of Image Recording in Particle Image Velocimetry.* Measurement Science & Technology, 1997. **8**(12): p. 1417-1426.

134. Fuchs, E., J.S. Jaffe, R.A. Long, and F. Azam, *Thin Laser Light Sheet Microscope for Microbial Oceanography.* Optics Express, 2002. **10**(2): p. 145-154.

135. Engelbrecht, C.J. and E.H.K. Stelzer, *Resolution Enhancement in a Light-Sheet-Based Microscope (Spim).* Optics Letters, 2006. **31**(10): p. 1477-1479.

136. Grossman, A.R., M. Lohr, and C.S. Im, *Chlamydomonas Reinhardtii in the Landscape of Pigments.* Annual Review of Genetics, 2004. **38**: p. 119-173.

137. Henderson, I.R., M. Meehan, and P. Owen, *Antigen 43, a Phase-Variable Bipartite Outer Membrane Protein, Determines Colony Morphology and Autoaggregation in Escherichia Coli K-12.* Fems Microbiology Letters, 1997. **149**(1): p. 115-120.

138. Shaner, N.C., R.E. Campbell, P.A. Steinbach, B.N.G. Giepmans, A.E. Palmer, and R.Y. Tsien, *Improved Monomeric Red, Orange and Yellow Fluorescent Proteins Derived from Discosoma Sp Red Fluorescent Protein.* Nature Biotechnology, 2004. **22**(12): p. 1567-1572.

139. Chalfie, M., Y. Tu, G. Euskirchen, W.W. Ward, and D.C. Prasher, *Green Fluorescent Protein as a Marker for Gene-Expression.* Science, 1994. **263**(5148): p. 802-805.

140. Cui, B.W., Y.F. Liu, and M.A. Gorovsky, *Deposition and Function of Histone H3 Variants in Tetrahymena Thermophila.* Molecular and Cellular Biology, 2006. **26**(20): p. 7719-7730.

141. Cormack, B.P., R.H. Valdivia, and S. Falkow, *Facs-Optimized Mutants of the Green Fluorescent Protein (Gfp).* Gene, 1996. **173**(1): p. 33-38.

142. Haddad, A. and A.P. Turkewitz, *Analysis of Exocytosis Mutants Indicates Close Coupling between Regulated Secretion and Transcription Activation in Tetrahymena.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(20): p. 10675-10680.

143. Gerami-Nejad, M., J. Berman, and C.A. Gale, *Cassettes for Pcr-Mediated Construction of Green, Yellow, and Cyan Fluorescent Protein Fusions in Candida Albicans.* Yeast, 2001. **18**(9): p. 859-864.

144. Horowitz, P. and W. Hill, *The Art of Electronics*. 2 ed. 1989, Cambridge, UK: Cambridge University Press.

145. Haldimann, A. and B.L. Wanner, *Conditional-Replication, Integration, Excision, and Retrieval Plasmid-Host Systems for Gene Structure-Function Studies of Bacteria.* Journal of Bacteriology, 2001. **183**(21): p. 6384-6393.

146. Baba, T., et al., *Construction of Escherichia Coli K-12 in-Frame, Single-Gene Knockout Mutants: The Keio Collection.* Molecular Systems Biology, 2006. **2**.

147. Liu, Y.F., K. Mochizuki, and M.A. Gorovsky, *Histone H3 Lysine 9 Methylation Is Required for DNA Elimination in Developing Macronuclei in Tetrahymena.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(6): p. 1679-1684.

148. Perry, J.N., ed. *Chaos in Real Data: Analysis of Non-Linear Dynamics from Short Ecological Time Series* 2000, Kluwer Academic Publishers: Dordrecht, The Netherlands.

149. Strogatz, S.H., *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering* 1994: Perseus Books Publishing, LLC.

150. Ellner, S. and P. Turchin, *Chaos in a Noisy World - New Methods and Evidence from Time-Series Analysis.* American Naturalist, 1995. **145**(3): p. 343-375.

151. Beninca, E., et al., *Chaos in a Long-Term Experiment with a Plankton Community.* Nature, 2008. **451**(7180): p. 822-U7.

152. Efron, B. and G. Gong, *A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation.* American Statistician, 1983. **37**(1): p. 36-48.

153. Clover, T.M. and J.A. Thomas, *Elements of Information Theory*. 1991, New York: John Wiley & Sons, Inc.

154. Imhof, M. and C. Schlotterer, *E-Coli Microcosms Indicate a Tight Link between Predictability of Ecosystem Dynamics and Diversity.* Plos Genetics, 2006. **2**: p. 966-971.

155. Massin, N. and A. Gonzalez, *Adaptive Radiation in a Fluctuating Environment: Disturbance Affects the Evolution of Diversity in a Bacterial Microcosm.* Evolutionary Ecology Research, 2006. **8**(3): p. 471-481.

156. Van Kampen, N.G., *Stochastic Processes in Physics and Chemistry*. 2nd ed. 1992, Amsterdam: Elsevier.

157. Hegger, R., H. Kantz, and T. Schreiber, *Practical Implementation of Nonlinear Time Series Methods: The Tisean Package.* Chaos, 1999. **9**(2): p. 413-435.

158. Abarbanel, H.D.I., T.W. Frison, and L.S. Tsimring, *Obtaining Order in a World of Chaos.* Ieee Signal Processing Magazine, 1998. **15**(3): p. 49-65.

159. Rosso, O.A., S. Blanco, J. Yordanova, V. Kolev, A. Figliola, M. Schurmann, and E. Basar, *Wavelet Entropy: A New Tool for Analysis of Short Duration Brain Electrical Signals.* Journal of Neuroscience Methods, 2001. **105**(1): p. 65-75.

160. Maraun, D., J. Kurths, and M. Holschneider, *Nonstationary Gaussian Processes in Wavelet Domain: Synthesis, Estimation, and Significance Testing.* Physical Review E, 2007. **75**(1).

161. Mallat, S., *A Wavelet Tour of Signal Processing*. 1999, San Diego, CA: Academic Press.

162. Jawerth, B. and W. Sweldens, *An Overview of Wavelet-Based Multiresolution Analyses.* Siam Review, 1994. **36**(3): p. 377-412.

163. Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C*. Second ed. 1992, Cambridge, UK: Cambridge University Press.

164. Eddison, J.C. and J.G. Ollason, *Diversity in Constant and Fluctuating Environments.* Nature, 1978. **275**(5678): p. 309-310.

165. Ollason, J.G., *Freshwater Microcosms in Fluctuating Environments.* Oikos, 1977. **28**(2-3): p. 262-269.

166. Fontaine, C. and A. Gonzalez, *Population Synchrony Induced by Resource Fluctuations and Dispersal in an Aquatic Microcosm.* Ecology, 2005. **86**(6): p. 1463-1471.

167. Grover, J.P., D. McKee, S. Young, H.C.J. Godfray, and P. Turchin, *Periodic Dynamics in Daphnia Populations: Biological Interactions and External Forcing.* Ecology, 2000. **81**(10): p. 2781-2798.

168. McQuarrie, D.A., *Statistical Mechanics*. 2000, Sausalito, CA: University Science Books.

169. Honerkamp, J., *Statistical Physics: An Advanced Approach with Applications*. 1998, Berlin: Springer Verlag.

170. Kusse, B. and E. Westwig, *Mathematical Physics*. 1998, New York: John Wiley & Sons, Inc.

171. Johnson, C.H., S.S. Golden, M. Ishiura, and T. Kondo, *Circadian Clocks in Prokaryotes.* Molecular Microbiology, 1996. **21**(1): p. 5-11.

172. Lipan, O. and W.H. Wong, *The Use of Oscillatory Signals in the Study of Genetic Networks.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(20): p. 7063-7068.

173. Kreuels, T., R. Joerres, W. Martin, and K. Brinkmann, *System-Analysis of the Circadian-Rhythm of Euglena-Gracilis .2. Masking Effects and Mutual Interactions of Light and Temperature Responses.* Zeitschrift Fur Naturforschung C-a Journal of Biosciences, 1984. **39**(7-8): p. 801-811.

174. Lork, W., T. Kreuels, W. Martin, and K. Brinkmann, *System-Analysis of the Circadian-Rhythm of Euglena-Gracilis .1. Linearities and Non-Linearities in the Response to Temperature Signals.* Zeitschrift Fur Naturforschung C-a Journal of Biosciences, 1982. **37**(11-1): p. 1240-1252.

175. Aschoff, J. and H. Pohl, *Phase Relations between a Circadian-Rhythm and Its Zeitgeber within Range of Entrainment.* Naturwissenschaften, 1978. **65**(2): p. 80-84.

176. Malinowski, J.R., D.L. Lavalmartin, and L.N. Edmunds, *Circadian Oscillators, Cell-Cycles, and Singularities - Light Perturbations of the Free-Running Rhythm of Cell-Division in Euglena.* Journal of Comparative Physiology B-Biochemical Systemic and Environmental Physiology, 1985. **155**(2): p. 257-267.

177. Aschoff, J., *Circadian Activity Pattern with 2 Peaks.* Ecology, 1966. **47**(4): p. 657-&.

178. Horst, C.J. and G.B. Witman, *Ptx1, a Nonphototactic Mutant of Chlamydomonas, Lacks Control of Flagellar Dominance.* Journal of Cell Biology, 1993. **120**(3): p. 733-741.

179. Weis, R.M. and D.E. Koshland, *Reversible Receptor Methylation Is Essential for Normal Chemotaxis of Escherichia-Coli in Gradients of Aspartic-Acid.* Proceedings of the National Academy of Sciences of the United States of America, 1988. **85**(1): p. 83-87.

180. Bruce, V.G. and N.C. Bruce, *Diploids of Clock Mutants of Chlamydomonas-Reinhardi.* Genetics, 1978. **89**(2): p. 225-233.

181. Yoshida, T., L.E. Jones, S.P. Ellner, G.F. Fussmann, and N.G. Hairston, *Rapid Evolution Drives Ecological Dynamics in a Predator-Prey System.* Nature, 2003. **424**(6946): p. 303-306.

182. Pascual, M.A. and P. Kareiva, *Predicting the Outcome of Competition Using Experimental Data: Maximum Likelihood and Bayesian Approaches.* Ecology, 1996. **77**(2): p. 337-349.

183. Póczos, B. and A. Lőrincz (2008) *D-Optimal Bayesian Interrogation for Parameter and Noise Identification of Recurrent Neural Networks*. ArXiv **Volume**, 0801.1883v1

184. Yanushevich, Y.G., N.G. Gurskaya, D.B. Staroverov, S.A. Lukyanov, and K.A. Lukyanov, *A Natural Fluorescent Protein That Changes Its Fluorescence Color During Maturation.* Russian Journal of Bioorganic Chemistry, 2003. **29**(4): p. 325-329.

185. Nakahigashi, K., K. Miyamoto, K. Nishimura, and H. Inokuchi, *Isolation and Characterization of a Light-Sensitive Mutant of Escherichia-Coli K-12 with a Mutation in a Gene That Is Required for the Biosynthesis of Ubiquinone.* Journal of Bacteriology, 1992. **174**(22): p. 7352-7359.

186. Miyamoto, K., K. Nakahigashi, K. Nishimura, and H. Inokuchi, *Isolation and Characterization of Visible Light-Sensitive Mutants of Escherichia-Coli K12.* Journal of Molecular Biology, 1991. **219**(3): p. 393-398.

187. Nakahigashi, K., K. Nishimura, K. Miyamoto, and H. Inokuchi, *Photosensitivity of a Protoporphyrin-Accumulating, Light-Sensitive Mutant (Visa) of Escherichia-Coli K-12.* Proceedings of the National Academy of Sciences of the United States of America, 1991. **88**(23): p. 10520-10524.

188. Shimizu-Sato, S., E. Huq, J.M. Tepperman, and P.H. Quail, *A Light-Switchable Gene Promoter System.* Nature Biotechnology, 2002. **20**(10): p. 1041-1044.

189. Bremer, C.L. and D.T. Kaplan, *Markov Chain Monte Carlo Estimation of Nonlinear Dynamics from Time Series.* Physica D-Nonlinear Phenomena, 2001. **160**(1-2): p. 116-126.

190. Brooks, S.P., *Markov Chain Monte Carlo Method and Its Applications.* The Statistician, 1998. **47**(1): p. 69-100.

191. Cocco, S. and R. Monasson, *Reconstructing a Random Potential from Its Random Walks.* Epl, 2008. **81**.

192. Emery, A.F., *Using the Concept of Information to Optimally Design Experiments with Uncertain Parameters.* Journal of Heat Transfer-Transactions of the Asme, 2001. **123**(3): p. 593-600.

193. Emery, A.F., A.V. Nenarokomov, and T.D. Fadale, *Uncertainties in Parameter Estimation: The Optimal Experiment Design.* International Journal of Heat and Mass Transfer, 2000. **43**(18): p. 3331-3339.

194. Emery, A.F., B.F. Blackwell, and K.J. Dowding, *The Relationship between Information, Sampling Rates, and Parameter Estimation Models.* Journal of Heat Transfer-Transactions of the Asme, 2002. **124**(6): p. 1192-1199.