2015

Mathematical modeling of Emiliania huxleyi and a host-specific virus

Julia Middleton
Colby College

Mathematical Modeling of *Emiliania huxleyi* and a Host-Specific Virus

Julia Middleton

Colby College

Mathematical Modeling of *Emiliania huxleyi* and a Host-Specific Virus

An Honors Thesis

Presented to

The Faculty of the Department of Biology

Colby College

in partial fulfillment of the requirements for the

Degree of Bachelor of Arts with Honors

by

Julia E. Middleton

Waterville, ME

18 May, 2015

Advisor: Nick Record _____

Reader: Cathy Bevier _____

Reader: Jim Scott _____

It is important that students bring a certain ragamuffin, barefoot irreverence to their studies; they are not here to worship what is known, but to question it.

-Jacob Bronowski

Acknowledgements

I had my first taste of real research during my semester at Bigelow Laboratory in the fall of my junior year. The semester began with some confusion over who was a chemistry major in our little group and led to me working on a computer modeling project on which I felt I'd be little help. Luckily for me, this turned out to be the best clerical error ever. Working with my advisor, Nick Record, has helped me gain a real appreciation for both the hard work of research and the times when really all you need to is take a walk to stretch your legs and talk about anything but research. Nick's patience and advice helped me through innumerable coding struggles and his feedback have been invaluable.

I also want to thank my two readers, Cathy Bevier and Jim Scott who have given me excellent advice throughout the writing process. I can't count number of times I've ended up in Cathy's office talking through last minute changes to a draft. Her calming influence when I'd been up writing through the night was always appreciated. I would be remiss to not acknowledge Cat Collins here, as her support and enthusiasm for ecological modeling was the driving force that first ignited my interest in the union of math and ecology.

I want to extend a thank you to all the faculty who have been inspirational throughout my time at Colby and have fostered a true eagerness for information and understanding. Their influence has led me to actively pursue my interests and dive head first into subjects. Possibly more importantly, they've taught me that being wrong is often more interesting than being right, as it opens new pathways for exploration.

Finally I thanks my wonderful friends. Long nights in Keyes were made so much more enjoyable because of their presence!

Contents

1 Abstract

The world's oceans provide the basis for life on the planet. One microscopic algae, the coccolithophores, and *Emiliania huxleyi* in particular, is a major source of carbon drawdown in the context of the global carbon cycle and account for a significant amount of the primary production in oceanic ecosystems. We know that the oceans are packed with marine viruses and they have an important role in the rise and fall of plankton populations but current mathematical models do not accurately account for virus-host interactions when predicting plankton blooms. Therefore I am using model optimization and comparison techniques to evaluate current models and reassess how the virus-host system should be described. Analysis of *in situ* blooms and mesocosm data has revealed previously unreported trends occurring over the course of an *E. huxleyi* bloom. Inclusion of virus-host interactions in models describing the calcifying coccolithophore *E. huxleyi* may give insight into global carbon cycling, as large-scale blooms are known to draw down significant amounts of atmospheric carbon.

2 Functional form analysis of marine virus-host interactions

2.1 Introduction

The world's oceans provide the basis for life on the planet. Tiny organisms called phytoplankton provide the foundation for the majority of life in the oceans. Although these tiny organisms are microscopic individually, they periodically undergo rapid growth during which they can be seen from space. These rapid periods of growth and drastic increases in population size are called phytoplankton blooms. Blooms are major sources of carbon drawdown in the context of the global carbon cycle and account for a significant amount of the primary production in oceanic ecosystems (Antoine et al. 1996, Pan et al. 2015). The influence and behavior of phytoplankton has been studied extensively for decades. More recently, the importance of marine viruses and their role in phytoplankton bloom dynamics has come to the forefront of oceanographic ecosystem research. One variety of phytoplankton, called coccolithophores, provide an interesting study organism as they are distinctive in being both phytoplankton that draw down carbon through photosynthetic reactions and calcifying organisms capable of precipitating calcium carbonate out of the water column. This calcium carbonate is used to form protective coccoliths that protect the organism from predation. The coccolithophore *Emiliania huxleyi* is of particular interest, as interactions between this species and marine viruses have been examined on a variety of spatial scales (Martínez et al. 2006, Wilson et al. 2002). Although early studies found viral lysis may account for between 25 to 100% of the net mortality of the phytoplankton such as *Emiliania huxleyi* (Bratbak et al. 1993), the traditional view of grazing and nutrient limitation as the primary mode of bloom termination has persisted (Evans et al. 2003). More recently, Jacquet et al. (2002) found phytoplankton mortality due to viral infection may be of a similar magnitude to grazing. Despite its apparent importance, only a handful of comprehensive models describe virus-host interactions in a

marine setting (Michaloudi et al. 2009, Turner 2014, Watras et al. 1985), with only a handful of functional forms describing infection and lysis rates (Bratbak et al. 1998, Chattopadhyay and Pal 2002, Rhodes et al. 2008, Thyrhaug et al. 2003, Weitz and Dushoff 2008). Conventional nutrient-phytoplankton-zooplankton (NPZ) models have a variety of functional forms whose responses and effect on population dynamics are well documented (Franks 2002, Gentleman et al. 2003, Tian 2006). A functional form accounts for the interaction between two populations in an ecosystem model. In order to facilitate further work on marine-virus host systems, we endeavor here to outline the functional forms currently used for virus-host dynamics. The most common model used to explore phytoplankton dynamics is the Nutrient-Phytoplankton-Zooplankton (NPZ) model. Franks (2002) created a comprehensive catalogue and critique of the functional forms used within a general set of NPZ equations:

Susceptible phyto $\frac{dP}{dt} = f(I)g(N)P - h(P)Z - i(P)P$

Zooplankton $\frac{dZ}{dt} = \gamma_e Z h(P) - j(Z)Z$

Nutrients $\frac{dN}{dt} = -f(I)g(N)P + (1-\gamma_e)Zh(P) + i(P)P + j(Z)Z$

An outline of these transfer functions can be found in Table 4. The models shown here cover a range of complexity and highlight the dearth of complexity specifically in the functional forms used to model marine virus-host interactions. Where NPZ models have many terms well studied and fit to data, virus-host interactions have a large number of theoretical forms that have not been validated with data (Franks 2002, Bratbak et al. 1998).

In the current review I explored how marine virus-host interactions have been modeled. Many NPZ models account for interactions, such as the response of phytoplankton to light or grazing pressure, that are well described by the literature. As these interactions are extensively characterized, I will not discuss them here. However, I do

Susceptible phyto	$\frac{dP}{dt} = f(I)g(N)P - h(P)Z - i(P)P - f(V)PV$
Infected phyto	$\frac{dI}{dt} = f(V)PV - m(I)I - h(I)Z$
Zooplankton	$\frac{dZ}{dt} = \gamma_e Z(h(P) + h(I)) - j(Z)Z$
Viruses	$\frac{dV}{dt} = m(I)f(b)I - m(V)V$
Nutrients	$\frac{dN}{dt} = -f(I)g(N)P + (1 - \gamma_e)Z(h(P) + h(I)) + i(P)P + j(Z)Z + m(I)I + m(V)V$

Table 1: The base model incorporating an infected phytoplankton population and a virus into them common NPZ framework. Table 4 outlines the terms on the right hand side of the equations, including those originating from Franks (2002) and the virus-host forms included here.

make comments with regards to models that include populations involved in viral interactions, such as the inclusion of an infected population. The populations relevant to viral infections, that is the infected phytoplankton and viruses, have been built into the original Franks (2002) base model (Table 1). This model is considered a general framework from which any model under consideration can be derived simply by setting certain interaction terms to zero. I will also discuss specific reasoning behind the functional forms and populations included, and the validity of current functional forms.

2.2 What state variables should be included?

The seven models reviewed here contained four state variables specifically relevant to virus-host modeling: a phytoplankton population susceptible to infection (P), an infected phytoplankton population (I), a virus population (V), and, in one case, a viral inhibitor population (S). Of these seven models, five included both a susceptible and an infected phytoplankton population, giving a direct view into the population dynamics of both healthy and incapacitated phytoplankton. Of these models only three of seven also included a specific virus population. In two cases no infected population was included and viral infection was modeled entirely as a mortality term within the susceptible population. A summary of the state variables included for each model can be found in Table 2.

Currently, there has not been agreement on what state variables are necessary to include in models focused on marine virus-host interactions.

Model	Population							
	Susceptible Phyto	Infected Phyto	Viruses	Grazers	Nutrients	Inhibitor	Detritus	Fit to data?
Bratbak et al (1998)	✓	✓	✓					✓
Chattopadhyay et al (2002)	✓	✓		✓				
Thyrhaug et al (2003)	✓		✓			✓		
Singh et al (2004)	✓	✓		✓				
Weitz and Dushoff (2008)	✓		✓					
Rhodes et al (2008)	✓	✓	✓	✓	✓		✓	
Rhodes et al (2010)	✓	✓		✓				

Table 2: Summary of the components included in each model. There are several ways in which the models express each population, which are outlined in the model equations in the following subsections.

2.3 What functional forms are used to describe virus-host interactions?

Within the seven models reviewed here there were two ways in which the function describing viral infection was described. This functional form joins three of the state variables together: susceptible phytoplankton, infected phytoplankton, and viruses. The type of functional form used can greatly influence the dynamics, and therefore great care must be taken when choosing how to describe relationships between populations. The choices are currently limited in the case of marine virus-host systems. Table 5 outlines the functional forms used to describe viral infection by the seven models under review. In all cases either a linear or a logistic interaction was used to describe the infection of hosts by virulent viruses. The linear form is used throughout virus-host models and is the simplest description of the infection term. The logistic response was introduced by Thyrhaug *et al.* (2003) as a way to examine the effect of an inhibitor. Those authors argued that the release of an inhibitor during lysis by a phytoplankton population under attack explained

Variable	Lab	Lab 2	Graze	Meso 2000	Meso 2003	Meso 2008	DISCO 1999
Total *E.hux*	✓	✓	✓	✓	✓	✓	✓
Infected *E.hux*	✓	✓	✓				
EhV	✓	✓		✓	✓	✓	✓
Grazers			✓				
Nutrients				✓	✓	✓	✓
Light			✓	✓			
Depth Salinity Temp							✓
Resolution (time)	2-24 hrs	1-24hrs	2 hrs	Daily	Daily	6 hrs	Daily
Time Span	74 hrs	48hrs	8 hrs	18 days	12 days	17 days	17 days

Table 3: Summary of what populations and other data were collected in each experiment.

Term	Meaning	Refer to Franks 2002	Reference
f(I)	Phytoplankton response to light	✓	
g(N)	Phytoplankton nutrient uptake rate	✓	
h(P)	Zooplankton grazing of susceptible	✓	
h(i)	Zooplankton grazing of infected	see Table 6	
i(P)	General phytoplankton mortality term accounting for organisms not included in the model	✓	
j(Z)	General zooplankton mortality term accounting for organisms not included in the model	✓	
m(I)	Mortality of infected due to viral pressure		Mortality of infected phytoplankton due to viral activity is described as linear (m_i) in current models
f(V)	Viral infectivity of susceptibles		see Table 5
f(b)	Burst size (number of viruses released per cell lysis)		Burst size is described linear (β) in current models
m(V)	General viral mortality term		Viral mortality is described linear (m_v) in current models

Table 4: Overview of the functional forms previously described by Franks (2002) and forms described in this review. The three virus-related terms that are described as linear in the majority of current models may be better described using other functional forms. The basis for these changes will be discussed.

Functional form	Description	References
λ	Linear response with one general infectivity term	Chattopadhyay and Pal 2002, Singh et al. 2004, Rhodes et al. 2008, Rhodes and Martin 2010
$c\sigma_1\sigma_2$*	Linear response with component variables defined	Bratbak et al. 1998
$\frac{\sigma_1}{1+S}c\sigma_2$	Logistic response to an inhibitor S	Thyrhaug et al. 2003
$\lambda(1-\frac{aP}{k})$	Fractional decrease in infection near stationary phase	Weitz and Dushoff 2008

Table 5: Functional forms of *f(V)* from the models under review, the infection rate of phytoplankton by a virus. *Some of these forms include the maximum virus-host contact rate c, the fraction of viruses adsorbed by the host σ_1, the fraction of infective viruses σ_2, and the phytoplankton carrying capacity k.

the trends seen in a long term infection experiment. The presence of an inhibitor decreases the infection rate as the lysis rate increases, leading to a negative feedback loop. Weitz and Dushoff (2008) argued that the decrease in infection was due to a fractional reduction in lysis when phytoplankton reached the stationary phase. The reduction in infection was attributed to a reduction in overall growth and a decrease in susceptibility of the host, possibly combined with the effects of some of the host entering a stationary or refractory period.

The other transfer functions are max phytoplankton growth μ, infected phytoplankton mortality m_{inf}, number of viruses released per lysis b, and viral mortality m_v. Notice that the viral infection term is equivalent to the viral growth term multiplied by a scalar, the number of viruses released per lysis b. Notice that in some cases more than one parameter is used to describe the linear term. While the terms end up being mathematically identical, this format often aids in parameterization by laboratory data. Although these functional forms offer many options for modeling the virus-host interaction,

Functional form	Description	References
g_i	Linear grazing term, typically higher than the susceptible grazing term	Chattopadhyay and Pal 2002, Singh et al. 2004
$\frac{g_i \epsilon I(P+I)}{g_i+\epsilon(P+I)^2}$	Increased grazing at high phytoplankton density, saturating	Rhodes et al. 2008

Table 6: Some functional forms of *m(I)*, the grazing rate of infected phytoplankton. Some of these forms include the slope of the grazing function ϵ,

none have been sufficiently validated against data to enable differentiation among them (Table 2). Aside from the linear term, which has been fit to data in one model, these functional forms have not been used in models that were validated through data fitting. Other functional forms, while justified on theoretical grounds, have not been validated against measured population dynamics.

2.4 How have these models been used?

Marine virus-host models have been used to probe a range of questions that span across disciplines. On the biological side, virus-host models have been used to study the role of viruses in the collapse of phytoplankton blooms (Bratbak et al. 1998), influence of viruses on phytoplankton and grazer abundance (Singh et al. 2004), the influence of a delay term (mimicking an incubation period) (Levin et al. 1977), and the role of viruses in maintaining marine microbial diversity (Thingstad 2000, Weinbauer and Rassoulzadegan 2004). Virus-host models have also been used to explore the impact of viral interactions on nutrient cycling (Tyrrell and Taylor 1996), the implications of inhibitory chemicals released during phytoplankton lysis (Thyrhaug et al. 2003), and the effect of eutrophication on viral infection (Rhodes and Martin 2010). Studies have also focused on model parameter space and dynamical properties, determining parameter ranges that allow populations to stabilize to an equilibrium level (Beretta and Kuang 1998) and what changes in parameters caused a Hopf bifurcation and lead to limit cycle oscillations in the populations (Siekmann and

Malchow 2008). Understanding the parameter space of a model is essential for accurate interpretation of results. Additionally, one of the major goals of many studies is to find interesting behavior in a theoretical model that can be used to fuel hypotheses. Fully evaluating the range of responses that can be described by a model allows model builders to identify and explore this behavior.

2.5 Discussion

Current models offer an excellent starting place for future research and highlight the need for continued in-depth analysis of what functional forms best describe an infection interaction. Additionally, some interactions that are currently modeled as linear terms may offer points of interest that can be further expanded into functional forms better suited to describing the trends seen in the real world. The current collection of models does not consider a number of ecological processes and strategies known to be important in virus ecology. For example, all the models are working under the assumption that the virus-host interaction functions with a lytic virus cycle. Even the models that incorporate a time delay in the infected population are using a lytic type virus, since there is no ongoing release of viruses (as would happen with a chronic or lysogenic virus). Furthermore, although viral mortality $m(I)$ and the number of cells released during cell lysis (burst size) $f(b)$ are represented here as constant parameters, there is evidence that they may be better described as non-linear terms (Bratbak et al. 1998, Danovaro et al. 2011, Suttle 1992, Wilhelm et al. 2003). Currently, these functional forms are not well described in marine virus-host models. In order to fully understand virus-host dynamics and the ideal way to model this interaction, these models and their variations must be assessed.
In order to facilitate future research, I have compiled a list of current model assumptions that may be explored:

- The majority of marine virus-host models in current literature assume that viruses are lytic (also called virulent) phages. As such, phages are not modeled with any

extended intracellular phase. Virus-host models holding this assumption do not account for viruses capable of inserting their genome into the host genome (Silander et al. 2005, Weinbauer 2004b)

- The role of UV degradation is not considered in the virus mortality term, although UV is known to damage viral DNA (Wilhelm et al. 2003). Although there is a linear term for mortality, irradiation should be considered because viruses function across a depth range that is known to have a steep light attenuation gradient. This will be particularly important when virus equations are added to global biogeochemical models.

- Most models also assume that once a host is infected there is no chance of recovery, although there is evidence that recovered populations occur and are resistant to infection (Thyrhaug et al. 2003).

- Most current marine virus-host models assume all phytoplankton mortality is due to viral infection, despite known effects of nutrient limitation and grazing.

While studying virus-host interactions may be the topic of interest, using models that do not consider the role of grazing and nutrient limitation may lead to biased conclusions.

One of the major gaps identified by this review is the general absence of model validation against data (Table 2). If we hope to ultimately incorporate virus-host dynamics into ocean ecosystem models, some model selection is needed, which requires comparison to observed dynamics. To this end, I have compiled a database of *E. huxleyi* - EhV dynamics measured at scales ranging from lab to ocean basin, with the goal of evaluating each model against observed patterns.

3 *Emiliania huxleyi* bloom analysis using basic models

3.1 Introduction

I studied the bloom structure of the coccolithophore *Emiliania huxleyi*, as this species is of particular interest due to its role in the global carbon cycle (Pan et al. 2015). The general structure of an *E. huxleyi* bloom gives insight into the forces controlling the population. Phytoplankton blooms often take on a characteristic shape distinguished by an initial growth phase leading into a population boom followed by a rapid die-off (Castberg et al. 2001, Harris 1996). Both phases are well described in studies that do not take into account viruses. An influx of nutrients combined with stratification due to heat initiating growth and a combination of nutrient depletion and grazer population increase subsume the phytoplankton population. As viruses become more prevalent in literature, thanks to improved sampling techniques, studies have begun to attribute 25 to 100% of the mortality term of phytoplankton to viruses (Brussard et al. 1996). Suttle (2007) points out that 90% of the ocean's biomass is composed of microorganisms, such as *E. huxleyi*, of which approximately 20% is estimated to be killed daily by viruses. This large scale view of viral interactions in the world's oceans highlights the need for models that can accurately reproduce the effects of viruses on important species like *E. huxleyi*.

Recent findings have identified that mortality due to viral infection may play a significant role in phytoplankton population decline in a bloom scenario (Evans et al. 2003, Jacquet et al. 2002). Furthermore, Martínez et al. (2007) found the presence or absence of grazer species failed to cause a significant difference in bloom development and termination. They suggest that this indicates the contribution of grazing to the decline of *Emiliania huxleyi* blooms may be small in some scenarios. Field and lab studies examining other prototypical systems have found a range of viral interactions in bloom dynamics. Brussaard et al. (2008) report that viruses only controlled bloom dynamics in one out of

six study populations, while Schwierzke et al. (2010) postulated that viruses had a controlling effect in the population density of *Prymnesium parvum*. Llewellyn et al. (2008) suggested that at the very least viral infection was occurring, although they could not comment on the overall influence of viruses on the bloom dynamics.

Here I analyzed population abundance data from four data sets to determine if host-specific *E. huxleyi* viruses (EhVs) were playing a role in the bloom dynamics of *E. huxleyi*. These populations ranged in size from those in contained *in situ* mesocosm experiments to measurements taken during a large scale cruise of the North Sea (Martínez et al. 2006, Martínez et al. 2007, Martínez et al. 2012, Wilson et al. 2002). My main goal was to determine the relationships between EhV and *E. huxleyi* dynamics in order to better understand the functional forms that should be included in a descriptive model.

In addition to determining if viruses play a significant role in the bloom structure of *E. huxleyi*, this analysis gives insight into which processes are essential to include when modeling phytoplankton populations based on the overall structure of *E. huxleyi* blooms and the variation seen in parameter estimations. Trends in bloom dynamics may also illuminate the appropriate functional forms needed to accurately describe EhV - *E. huxleyi* interactions. I used the most basic forms of phytoplankton growth to determine how closely the bloom dynamics adhered to conventional model equations. This set up also allowed for an estimation of certain fundamental bloom parameters, such as phytoplankton growth, general phytoplankton mortality, viral infection rate, viral adsorption rate, and general viral mortality. Although these were estimates, they served as a useful frame of reference for comparing the orders of magnitude of terms. Separate analysis of the functional response of bloom stages gave insight into deviations from the traditionally accepted bloom dynamics. In this case, variation may indicate viral pressure on the *E. huxleyi* population.

3.2 Methods

Analysis of bloom dynamics was implemented using data from a variety of sources (Table 2). Because these data were collected during different studies, each had different data available (Table 3). I performed a parameter estimation and tested for how well the most basic functional form fit data. Parameter estimation was carried out using the first-order (linear) polynomial fit function in MatLab. Raw data were used for parameter estimation. The code files used to carry out this analysis are *percap*, *percapgen*, and *percapitamanual* (Appendix). Two analyses of linearity were conducted. The first assessed the traditional bloom dynamics using a signal break. For the functional response analysis, in order to determine when a linear section of the growth phase ended, the dividing point between the two halves of the bloom (break) was moved through the time series until the linear segment with the greatest significance was found. For this analysis a break represents the day in the time series at which the data set was split. Following this, I fit each half of the data set with a linear regression. This analysis was performed after a characteristic segmented pattern was observed in the preliminary scatter plots of per capita growth. Data centered in time used an average of the point on either side of a time step, rather than the point at that time step in order to reduce noise. Both time centered and uncentered data were analyzed for the break point, but smoothed data were considered more strongly since they account for noise in the data sets. For analyzing the three visually identified segments, the sections were defined as follows: from the first point in the data set to the greatest per capita growth rate value, from the greatest per capita growth to the greatest population abundance, and finally from the greatest population abundance to the terminal data point. These segments were representative of the three linear segments in the majority of data sets. The m-files associated with the linearity analysis were *breakpoint* for the two linear segments and *threebreak* and *threebreakmanual* for the secondary analysis (Appendix).

3.3 Does the basic model accurately predict *E. huxleyi* and EhV population trends?

The most basic virus-host system can be represented by:

$$\frac{dP}{dt} = \mu P - mP^2$$

$$\frac{dV}{dt} = \beta\phi PV - nV$$

where μ and m are the *E. huxleyi* growth and mortality parameters, respectively. The β and ϕ parameter combination accounts for the viral infection rate and the adsorptivity, although they cannot be resolved in this model. The n term accounts for viral mortality, which is typically attributed to degradation by UV-radiation (Fuhrman 1999, Suttle 1992, Wilhelm et al. 2003). These equations describe a system in which all phytoplankton are producing viruses, but the phytoplankton mortality term is not linked to virus population size. This can be considered a situation in which a non-lethal virus is infecting the phytoplankton population or one in which viral mortality is very small compared to other sources or mortality. While this assumption is obviously untrue, examining the dynamics of this simplified model can still give insight into how well these functional forms describe the system. Linear regression analysis was used to determine how well the model was able to fit to experimental data from three mesocosm experiments in which both *E. huxleyi* and EhV populations were tracked over the course of two to three weeks (Martínez et al. 2006, Martínez et al. 2007). Plotting per capita changes in the *E. huxleyi* and EhV populations against the abundance of *E. huxleyi* provides a simple yet powerful method for estimating parameter values and understanding population behavior. When populations are rearranged to represent per capita changes, our equations become:

$$\frac{1}{P}\frac{dP}{dt} = \mu - mP$$

$$\frac{1}{V}\frac{dV}{dt} = \beta\phi P - n$$

Notice that this rearrangement gives equations with representation of parameters as slopes and intercepts. I plotted the per capita change in *E. huxleyi* population against *E. huxleyi* abundance, which isolates the mortality term m as the slope coefficient and the intercept μ as the *E. huxleyi* growth term (3). Similarly, I plotted the per capita change in EhV against the abundance of *E. huxleyi* to isolate the infectivity ($\beta\phi$) as the slope and the intercept as viral mortality n. In addition to these easily interpreted regressions, two other per capita plots can be examined: per capita change in EhV against EhV abundance and per capita change in *E. huxleyi* against EhV abundance. Because neither per capita equation accounts for population change based on viral abundance, neither set of plots gives clear parameter estimates. These graphs reveal how the two populations are interacting. Significant linear regressions for these later plots are indicative of a relationship between viral interactions and population dynamics, rather than any specific estimation of parameter values. Examining these other relationships can give insight into the functional forms underlying the virus-host interactions beyond the simplified model shown here.

In this case, the significance of the linear regression gives some clue to the validity of the parameter estimation. Mesocosms with significant fits may have parameter values near the real value, while poor fits with no significance may indicate that the model equations are not correct or should be non-linear. For these basic model equations, few of the linear regressions returned significant correlations (Fig 1). Upon initial examination, there may appear to be many significant correlations. However, notice that half of the significant values fall within the two per capita plots that do not give variable estimates. Rather, these plots show that there is a strong correlation between the abundance of *E. huxleyi* and the per capita change of one of the populations. The strong negative correlation between per capita *E. huxelyi* growth and viral abundance indicates viral increase is directly related to the phytoplankton population and supports the possibility of strong viral interactions in all three mesocosm populations (Fig 1). This relationship also appears in the estimates of the

β, ϕ, and n parameters, which have the highest frequency of significant regressions. The positive values of the significant correlations indicate that high levels of viral infection were taking place in the 2000 and 2003 mesocosm experiments (Fig 1). In the 2003 mesocosms there was a strong negative correlation between per capita growth of *E. huxleyi* and the abundance of *E. huxleyi*. This trend may be due to strong density dependence in the phytoplankton population, with growth rate decreasing at high densities due to nutrient limitation or shading. The significant values found in a handful of the 2008 mesocosm may be indicative of both μ and m being significant. Parameters arising from significant correlations may represent accurate estimates of the true parameter values. Ultimately, the per capita scatter plots indicate that the relationship is definitely non-linear.

Examining the distributions of parameter values found using the basic virus-host model can also give insight into their certainty. Accurately estimated variables are expected to follow a normal distribution, centered on the true mean. The significant values have been plotted in histograms (Fig 2). Of the four parameters estimated, only the maximum *E. huxleyi* growth, μ, shows a normal distribution, though the EhV mortality, n, is arguably normal. Both burst size, β, and infection rate, ϕ, show very non-normal distributions, indicating that the basic model may not have the capacity to accurately estimate those values. Specific values of the parameter estimates can be found in Table 7.

Similar analysis was performed on data taken during the dimethyl sulphide biogeochemistry of a coccolithophore bloom (DISCO) in the North Sea during June 1999 (Wilson et al. 2002). The samples of interest, *E. huxleyi* and EhV, were collected using Niskin bottles mounted on a rosette, which collect water at both set and manually triggered depths through the water column, and samples were counted using flow cytometry. Samples were taken between zero and 180 meters. Rather than look at this entire depth scale at once, each rosette cast was broken into fourteen depth slices, with averages calculated for each layer. The depth layers ranged between two and 69 meters, with thiner slices (higher resolution) accounting for the greater variability near the surface,

while thicker slices are used to account for depths below the photosynthetic zone. Only one form of the basic model fit well to the data, and even then the correlation appeared at only one depth. Per capita change in *E. huxleyi* was negatively influenced by the viral abundance in between zero and two meters (Fig 4). This finding is not surprising, as subsurface layers further from the sun do not contain appreciable amounts of photosynthetic algae. Although they were not found to be significant, the parameter estimates based off of the DISCO 1999 data can be found in Table 7.

Another interesting finding from the basic model regressions was identification of strong linear trends in the bloom structure across all three mesocosms. These trends appear in the per capita growth rate plots of *E. huxleyi*. Notice in the plots there appears to be a linear phase during the growth of the *E. huxelyi* population but breaks down at high abundances (Fig 6). This may be indicative of viruses strongly influencing the bloom dynamics at high abundances, but contributing an effect that is too weak to overcome the initial growth phase. Furthermore, the distinct nature of the trend may indicate that the parameters describing the populations may vary throughout the lifetime of a *E. huxleyi* bloom. Specifically, viral interactions may be controlling these changes in bloom dynamics, in a similar manner to varying growth rate as seen in logistic models. Piecewise analysis of the bloom will be discussed later in the chapter.

Unlike the three mesocosm experiments, the DISCO 1999 data does not exhibit a clear break down of linearity for the *E. huxleyi* population (Fig 3(d)). The large spatial scale over which the data were collected may have increased the number of other factors influencing the population beyond those present in the mesocosms. Furthermore, the overall *E. hux* mortality for the data was very low to non-existent (Table 7). The DISCO 1999 dataset was taken without experimental controls and represents a true ecosystem. As such, the poor representation of the data by our simplistic model is not surprising.

3.4 Does the bloom follow traditionally accepted stages?

Piecewise analysis was used to determine if the two classically accepted linear stages of bloom growth appeared in the mesocosm (Martínez et al. 2006, Martínez et al. 2007) and DISCO cruise (Martínez et al. 2012, Wilson et al. 2002) *E. huxleyi* populations. All mesocosm experiments and the DISCO cruise data displayed stages of linearity that split the bloom nearly equally in half between a high growth phase and a high mortality phase. However, under this analysis only one phase of the bloom was significantly linear. In order to determine if the first day of bloom decline was the same between years, the average break point (day) was calculated for data centered in time. This day indicates the time at which the bloom dynamics changed from a growing to a diminishing population. The 2000 mesocosms had populations that began to decline on day ten, 2003 saw the average first day on decline on day seven, 2008 had bloom decline beginning on day twelve, on average, and for DISCO 1999 it was day 9.5 (one time series, no average). The average length of these time series was 16 days. Data collection for each time series began at different stages before the initiation of the bloom under observation. Despite this, all series captured the full span of the bloom. The variation in observation period may account for the differences seen in the break point days. The five day range in the initiation of bloom decline indicates variability in the factors controlling *E. huxleyi* mortality of blooms which may be attributed to nutrient depletion, grazers, or virus interactions. In the mesocosm experiments all treatments were run with nutrient supplements, indicating that nutrient limitation was unlikely (Martínez et al. 2006). Similarly, although the DISCO cruise measured *E. huxleyi in situ*, the population moved throughout the North Sea and therefore may have been exposed to new influx of nutrients as the bloom moved (Wilson et al. 2002).

Significant correlations were found for at least one linear phase of the bloom for almost every trial, regardless of the year in which the mesocosms was run. To build a clearer picture of what this means, *E. huxleyi* mortality was calculated as a running mean throughout the length of each bloom. Obvious changes in mortality may be due to some

other influence outside the basic *E. huxelyi* growth trend on the bloom dynamics. Because the slope (m) was calculated over a five time step period in order to reduce noise, it greatly reduced the number of points in each plot. Despite this, some interesting trends appeared. The mortality term was plotted with *E. huxelyi* abundance to see if any trends in mortality aligned with a specific period across blooms. If mortality is attributed solely to grazers, it appears that a basic linear mortality term does not accurately capture the mortality of *E. hux*. This can be seen in the positive (no mortality) value of the running mean for mortality during the termination period of the bloom (when the most negative term would be expected, to account for the die off) (Fig 5(a)). However, viral interactions may be able to explain this trend, as viral mortality is expected to be high during high abundance of *E. huxleyi* (corresponding with proliferation through lysis), but decrease with decreasing host abundance (Danovaro et al. 2011, Suttle 2007). As such, the trends seen in the running mean of overall mortality may actually encourage the inclusion of viruses in any modeling attempt, as the same trend would not be seen if only grazers were present. Although it is possible to find any number of piecewise correlations, a combination of visual analysis and output from a single-break correlation analysis indicated that three distinct sections of linearity may exist across the blooms analyzed.

The single-break piecewise analysis revealed a significant break point at which bloom dynamics changed. However, the lack of significant linearity in both phases of the bloom, combined with visual examination of the data, exposed the possibility of three main linear phases within the bloom structure. The same structure appeared across the majority of blooms when analyzing the per capita growth of *E. huxleyi* against the abundance of *E. huxleyi*. The three stages identified included an initial steep increase in per capita growth when the population size was small, followed by a decrease in per capita growth as the population reached larger abundances. Finally, as the population size decreased, per capita growth became negative, as expected during the termination stage of a bloom. Nutrient depletion has been ruled out as a contributing factor to these trends, as discussed. As such,

grazer and viral interactions are likely drivers behind the observed trend. While analyzing the data using a piecewise analysis with two sections, it became clear that the majority of the blooms occurring in the mesocosm were better characterized by three distinct sections (Fig. 6). Significance of the apparent trend was assessed using a second piecewise analysis from which correlations were calculated for these sections (Fig 7). The novel finding of three linear phases in the per capita growth rate of *E. huxleyi* leads me to conclude that variable growth and mortality terms must be included in a comprehensive virus-host model. Variation in *E. huxleyi* growth rate likely stems from density dependence, which can be succinctly modeled by a logistic growth term with the imposed carrying capacity based on nutrient availability. Following my current assumption that variations in mortality are due to viral interactions, I aim to develop a functional form describing virus-host interactions that can accurately describe the trends seen in the mesocosms and the DISCO cruise populations.

4 Model validation using basic models

4.1 Introduction

During the early 1900s two revolutionary minds, Lotka and Volterra, brought together the realms of mathematics and biology in an attempt to describe the rate of change of numbers of individuals (Lotka 1927, Volterra 1937). Their ideas are still used today as foundational models in computational biology. There is no better place to begin a foray into the realm of model validation than with the Lotka-Volterra model, as it represents the most simplistic model that accounts for population interactions. As such, I begin here with the a Lotka-Volterra model describing the interaction between *E. huxleyi* (P) and EhV (V):

$$\frac{dP}{dt} = \mu P - \phi PV$$

$$\frac{dV}{dt} = \beta \phi VP - nV$$

Unlike the basic equations examined in Chapter 2, the purpose of this model is not to

determine the relationship between *E. huxleyi* and EhV. Rather, I use this model to see how well the most basic predictive model fits to experimental data collected during the mesocosm experiments run by Martínez et al. (2007) and during the DISCO 1999 cruise (Wilson et al. 2002). The major revision to the original model lies in the linear transfer term (ϕPV), which represents viral pressure (here modeled in a manner similar to predation) on the *E. huxleyi* population. Taken alone, this model accounts for the biological virus-host interactions under examination but does not describe any physical forces on the system. This unforced version of the model was fit to the mesocosm data.

In order to capture some aspect of the dynamic nature of the physical environment the bloom measured by the DISCO cruise I included one-dimensional physical forcing that accounts for the changing depth of the mixed layer. The mixed layer is maintained during warmer months due to stratification through solar heating. This layer represents the photosynthetically active layer of the ocean where all of the population dynamics being examined in this study occur. Although the one-dimensional model accounts for vertical changes in the water column it does not account for three-dimensional physical dynamics such as advection. However, including the mixed layer allows for greater description of the population dynamics, as shoaling up of both *E. huxleyi* and EhV populations can significantly change population densities. This layer plays a critical role in the development of phytoplankton blooms, leading to my decision to include it in the Lotka-Volterra model.

As with any model, it is important to consider major assumptions. In this case, I assumed that all *E. huxleyi* mortality was due to lysis by EhV, that the population was not nutrient limited, and that *E. huxleyi* were not able to recover from viral infection. In the physical forcing model I assumed the population would not be condensed during shoaling, with individuals sinking out of the bloom as the mixed layer depth changed. Additionally, the possibility of viral decay due to UV radiation and the fact that *E. huxleyi* may actually suffer under strong light conditions, such as those near the surface, were not considered.

4.2 Methods

I began validating the Lotka-Volterra model by parameterizing the model to the mesocosm and DISCO data sets. I designed a model-fitting algorithm that utilizes the *fminsearch* functionality in MatLab. This function finds the minimum of an unconstrained multivariable system using a derivative free method. In order to find the global minimum of each system I iterated the search, thereby allowing the fit to improve with each iteration. The fit of each iteration was quantified as an output value which described both the root-mean-square (RMS) between the model output and raw data and the correlation between both the model populations and the experimental populations. I included both RMS and R^2 in the fit in order to capture both population magnitude (RMS) and bloom dynamics (R^2) in the model fit. Once model fits were performed the model populations were plotted in order to examine if they exhibited proper bloom dynamics. This type of fitting was performed for both the three years of mesocosm data and the DISCO cruise data.

For the DISCO data analysis, I began with an unforced Lotka-Volterra model to examine the model fit without physical dynamics. I then added the Evans and Parslow (1985) 1-D physical model, combined with mixed layer depth data collected during the cruise in order to incorporate shallowing and deepening of the mixed layer. This allowed for a more accurate representation of population dynamics. Because the Evans-Parslow model assumes a perfectly mixed surface layer, the depth slices from the DISCO layer that were above the initial mixed layer depth were averaged in order to find the initial population values. Furthermore, the start date from which to run the model was determined by finding the first time both *E. hux* and EhVs were measured in the DISCO data for the averaged mixed layer values. The same model fitting algorithm was used to parameterize the DISCO Lotka-Volterra model. I plotted the model output populations alone and within the water column in order to highlight the influence of the mixed layer depth on the population density.

4.3 How well does the Lotka-Volterra model fit data?

The Lotka-Volterra output for the basic model without physical forcing showed bloom dynamics that cycled on a time-scale much faster than the experimental DISCO data at nearly every depth (Fig 8). The deepest measured populations show a lack of cycling mainly due to the scarcity of data with which to parameterize the model, as *E. huxleyi* are not typically found below the phototrophic zone. Initial depth slices broke the water column into segments in a manner that did not isolate the mixed layer specifically, but did have a higher resolution (thinner slices of the water column) in the upper regions. This allowed me to look closely at the dynamics of the *E. huxleyi* bloom over a range of discrete depths. While the coupled cycling of the two populations is ideal, it does not follow the trends seen in the DISCO data, and therefore indicates that there may be other factors at play that strongly influence the two populations.

It is probable that the depth of the mixed layer has a strong influence on the dynamics of *E. huxleyi* and EhV. When changes in the mixed layer depth were included, the Lotka-Volterra model output showed that the physical forcing from the Evans & Parslow model had population increases (seen around June 25th in the model data). However, a much larger bloom centered on June 30 also appeared in the model output (Fig 9). The small population increase seen in the model on June 25th matches in both timing and size with the DISCO data, but the following large bloom far exceeds any population sizes seen within DISCO. The large bloom seen in the model may be due to the lack of a population cap and no nutrient limitation on the model populations. The bar seen below the surface layer represents *E. huxleyi* that have sunk out of the surface layer. This was added in order to better visualize the movement of the entire population throughout the course of the bloom. In order to determine how well the model fit with the experimental data, I ran linear regressions for both the *E. huxelyi* and EhV populations. I found that while neither population had a statistically significant correlation with the DISCO data, the *E. huxelyi* population very nearly matched the the DISCO data (despite the large bloom

seen on the 30th in the model) (p_{ehux} = 0.0539,r_{ehux} = 0.3823, p_{ehv}= 0.8080, r_{ehv}= 0.0512).

Only the unforced Lotka-Volterra model was considered for the mesocosms, as the populations were confined to the top of the water column by the experimental set up (Martínez et al. 2007). In the three years of mesocosm data, I found the model output to fit to the mescosm data very well, with 23 out of 24 of the correlations found to be significant. However, it is important to notice that even though these significance values seem very good, the actual R^2 values range greatly (Fig 13). The model can be considered to fit well to the data only in the case that the fit is both statistically significant and has an R^2 value near one (I set the cut off at 0.9). Of the 24 mescosms modeled, only nine met this criteria (starred in Figs 14, 15, 16). This represents successful use of this model, yet similar to the parameter estimations found in Chapter 2, this model did not give parameter estimations that were normally distributed (Fig 12). I did find somewhat normal distribution for the value of μ, but it was unique in that regard. The other parameters were similar to those found in Chapter 2, where both ϕ and β were practically bimodal, with no obvious mean.

Despite the shortcomings found in the model, the output predicts the blooms dynamics surprisingly well. The timing of the *E. huxelyi* bloom in the DISCO data very well predicted by the shoaling up of the mixed layer (Fig 10). However, despite the somewhat accurate prediction of the *E. huxleyi* population, the LV model predicts the EhV population quite poorly (Fig 11). Within the timeline of the the DISCO data, the model EhV population decreases and does not increase in response to the *E. huxleyi* increase around June 25th. Unfortunately, this indicates that the current modeling of virus-host interactions are not well described by the Lotka-Volterra model. I believe that the similarity between the model *E. huxleyi* and the *E. huxleyi* population recorded during the DISCO cruise may be due solely to the incorporation of the mixed depth physical forcing model. While it is gratifying to find that the physical model aids in the fitting capability of the model, it is unfortunate that the viral population does not also improve due to this addition. It is possible that the addition of the mixed layer dynamics does not influence

the EhV model output significantly because the factor affecting the virus, UV degradation of the capsid, was not included in this version of the Lotka-Volterra model. Furthermore, it is possible that *E. huxleyi* experiences variable effects due to changing light conditions as shoaling occurs.

4.4 What influence does light have on the model populations?

As the influence of light on *E. huxleyi* and EhV has been documented, I created a model that accounted for light effects on the populations. First, I examined the effects of light on the *E. huxleyi* population (Fig 19). Rather than having a fixed growth rate of *E. huxleyi*, I based the growth rate on the light level. I found that neither population had a statistically significant correlation with the DISCO data (p_{ehux} = 0.4844,r_{ehux} = -0.1435, p_{ehv}= 0.3077, r_{ehv}= -0.2547). Although, the relationships were not significant, there seems to be some indication that the model predicted the data to some degree: There is an increase in *E. huxleyi* on the 23-24th in the model occurred near the time of the bloom in the surface layer from the 23-28th in the data (Fig 17). The large increase on the 27th in the model may match with the subsurface maximum seen on the 28-29th in the data. However, although the model has some similarities to the data, the relationship is not strong enough for the model to be considered successful. The results for the EhV population were similar: although there were shadows of the dynamics seen in the data, the relationships were not very strong (Fig 18).

In addition to light affecting phytoplankton growth, UV radiation induces damage in native marine viruses (Danovaro et al. 2011, Wilhelm et al. 2003). Suttle et al (1992) found that this relationship could be represented by using depth as a proxy for light. Inputting their equation ($d_{r0} * e^{-kz}$), which uses the mortality by light at the surface (d_{r0}), the attenuation of light based on turbidity (k), and depth (z), I could more accurately model virus mortality. This term was added to a baseline term that accounted for other causes of virus removal, such as predation (Suttle 1992). Even though I tried a range of

light attenuation constants, the addition of the viral mortality term did not greatly affect the virus population's dynamics (Fig 21). Furthermore, the correlation between the model virus population and the real virus population remained strongly negative, while the correlation for the *E. hux* population barely changed
($r_{ehux} = -0.1435, p_{ehux} = 0.4844, r_{ehv} = -0.5662, p_{ehv} = 0.0026$). This indicates that the decrease in the virus population seen around day 25 in the field data is likely not due to viral decay by UV radiation.

Because this model had a somewhat close resemblance to real data, an optimization function was run varying the μ (max phytogrowth), ϕ (infection rate), β (burst size), and m_v (viral mortality) parameters using the *E. hux* population for fitting. The optimized parameters found were $\mu = 49.56$ day^{-1}, $\phi = 2.82 * 10^{-5}$ day^{-1}, $\beta = 4208.63$ viruses cell^{-1}, and $m_v = 12.15$viruses day^{-1}. Although the addition of light aided in accurately modeling the EhV - *E. huxleyi*, the populations still were not predicted with high accuracy. This indicates that additional modifications may need to be added to the interaction. I believe that modifying the infection term will allow the model to capture the exchange between the virus and *E. huxleyi* populations more accurately, thereby allowing us to better capture the overall bloom dynamics.

5 Implications and Future Research

This research has highlighted the need for further exploration of the interactions between EhV and *Emiliania huxleyi*. With the data available, I have highlighted the occurrence of a three-phase bloom structure that does not follow conventional bloom dynamics. The first of the three phases appears to exhibit a strong Allee effect, which is unexpected considering the asexual nature of *E. huxleyi*. Although the mechanisms behind this apparent Allee effect is currently unknown, the functional form of *E. huxleyi* growth could be changed to reflect the trend. Changing the growth term of *E. huxleyi* would account for the trends seen in the first phase of the population dynamics. I believe that accurately capturing these novel bloom dynamics could give insight into the interactions between marine viruses and their hosts. Including a separate infected population may aid in capturing these dynamics. Ideally, future models will include infectivity terms that capture aspects of virus life-cycles, such as infection type (lytic vs chronic) and differential infectivity throughout the lifetime of the bloom. Future field, mesocosm, and lab studies would be well advised to collect infected population data, in addition to the usual virus and susceptible host data, in order to aid in the expansion of these models. Bench studies specifically aimed at analyzing the virus-host infection dynamics may also give insight into the infectivity transfer term.

A functional model that can accurately predict the influences of viral infection on *E. huxleyi* could play a critical role in our understanding of global carbon cycling and may have applications in climatological research. The inclusion of a infected population in the virus-host model could allow for the prediction of differential carbon draw down during the course of an *E. huxleyi* bloom. These effects would necessarily need to be propagated over a long time scale, as the immediate effect of a decrease in calcification would appear as an decrease in atmospheric CO_2. However, when looked at on a long time-scale, the decrease in calcification should lead to a net decrease in CO_2 in the system. Although coccolithophores are known to have a significant effect on the carbonate buffer system, the effects of viral infection are not accounted for in current cycling models. Inclusion of viral

pressure may give a better view of how coccolithophores interact with the carbonate system. Ultimately, these types of models may have the potential to provide a clearer view of how viral interactions with microscopic algae influence global carbon cycling and modern climate change.

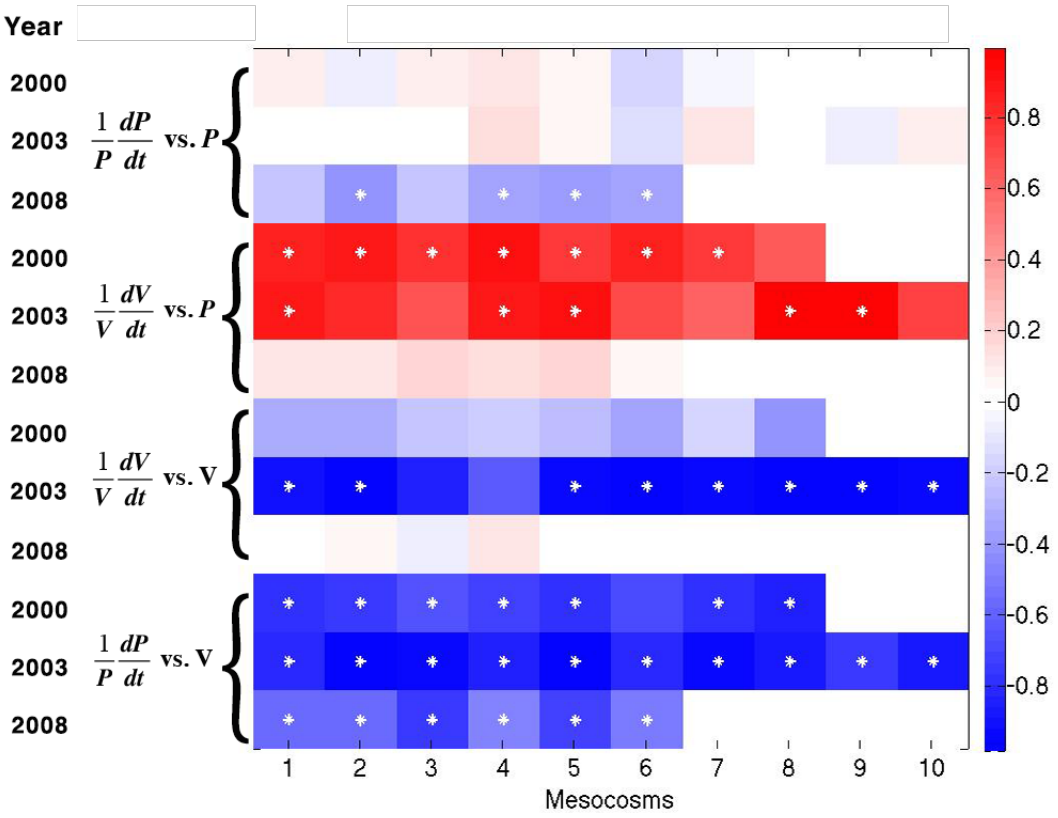
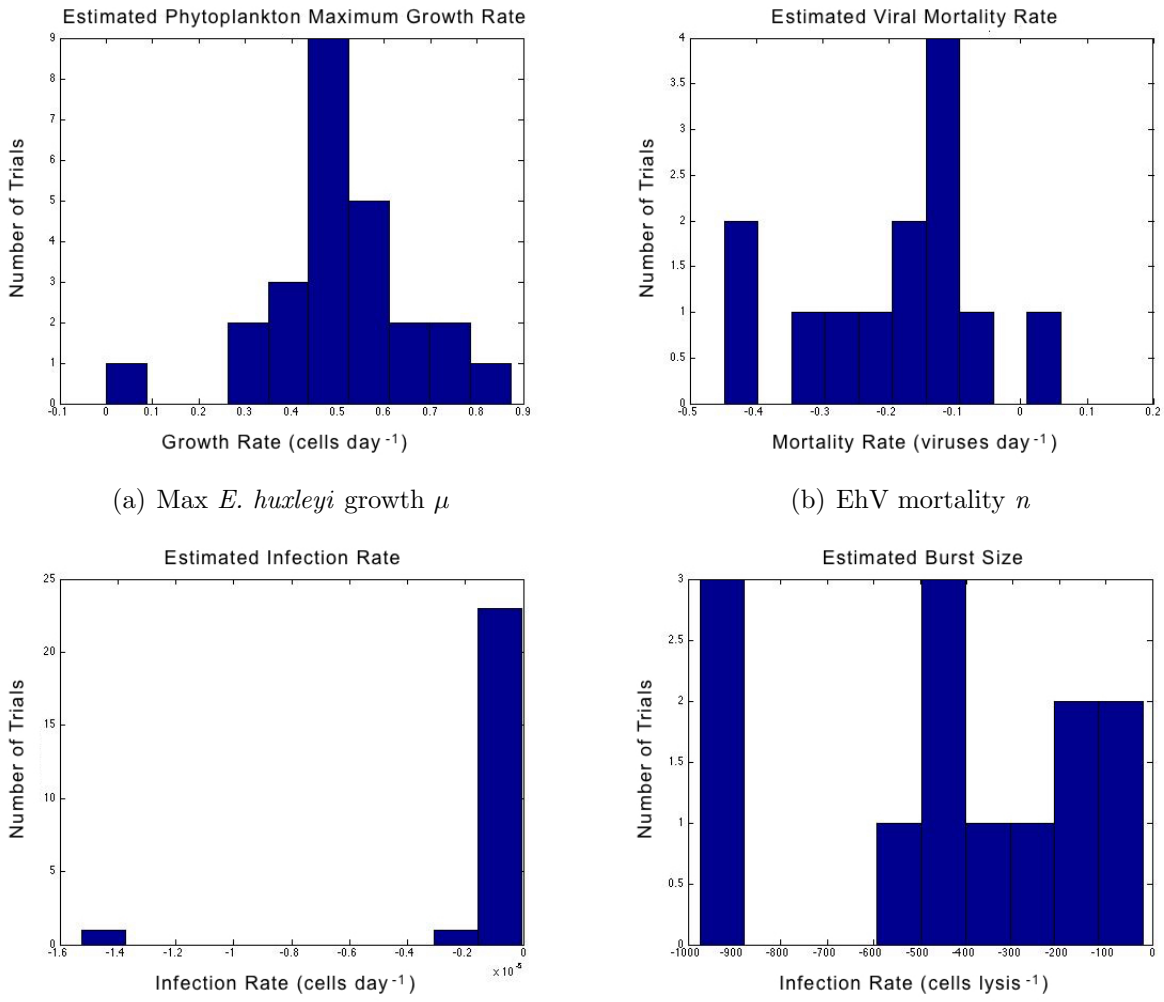


Figure 1: R^2 values for linear regression of the basic virus-host model fit to three separate years of mesocosm data. White stars represent regressions with $p < 0.05$, while blank white squares indicate absence of those mesocosms in a particular year.



(a) Max *E. huxleyi* growth μ

(b) EhV mortality n

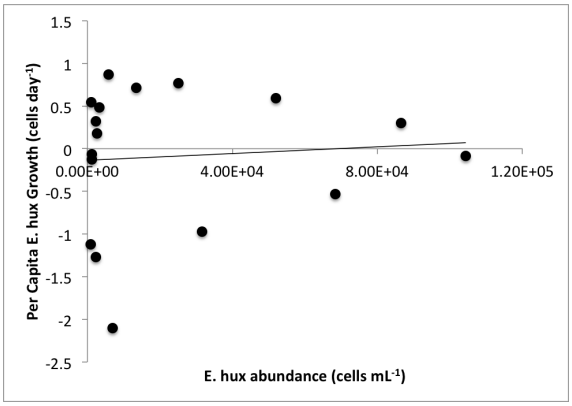
(c) Infection rate ϕ

(d) Burst size β

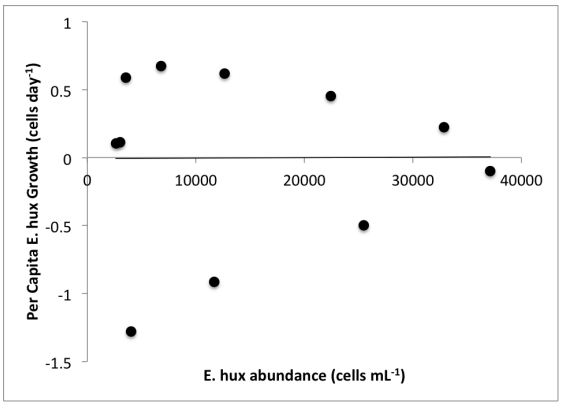
Figure 2: Histograms depicting the distribution of estimates for each variable.

Table 7: Parameter estimates for max phytoplankton growth, infection rate, burst size, and viral mortality. Bold values indicate parameters that were estimated from relationships that were statistically significant. The averages shown were calculated using only data that was statistically significant. Because only the surface depth slice (0-2m) had significant relationships for the DISCO data, only those value is shown.

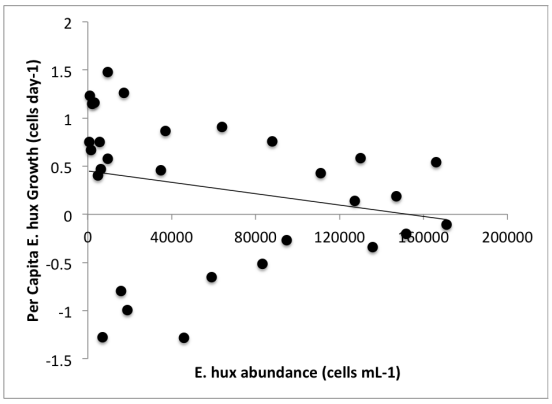
Mesocosm		1	2	3	4	5	6	7	8	9	10	Average
2000												
	μ	**0.4492**	**0.3316**	**0.4596**	**0.7739**	**0.3280**	0.8448	**0.4048**	**0.8736**			*0.5172*
	ϕ	**-4.9673e-08**	**-6.6292e-08**	**-1.0138e-07**	**-9.0533e-08**	**-6.0197e-08**	-1.0455e-07	**-6.4203e-08**	**-9.5578e-08**			*-7.5408e-08*
	β	**428.3012**	**479.4723**	**164.5834**	**186.2012**	**906.4192**	**107.0545**	**539.0774**	122.8366			*111.59*
	m	**-0.1027**	**-0.1004**	**-.1674**	**-0.0979**	**-0.3457**	**0.0604**	**-0.1835**	0.1345			*-0.1339*
2003												
	μ	**0.5485**	**0.5151**	**0.3946**	**0.4703**	**0.5017**	**0.4162**	**0.7221**	**0.6043**	**0.5443**	**0.6234**	*0.5341*
	ϕ	**-4.8779e-08**	**-1.0199e-07**	**-1.4513e-07**	**-6.51438e-08**	**-8.0693e-08**	**-1.1990e-07**	**-9.0849e-08**	**-9.6482e-08**	**-1.2551e-07**	**-1.0581e-07**	*-9.8029e-08*
	β	**973.6430**	197.9625	149.8994	**938.9601**	**456.8545**	156.3896	248.8043	**326.0587**	**230.2417**	339.7791	*585.1516*
	m	**-0.2369**	-0.1184	0.0771	**-0.4482**	**-0.2635**	0.1943	0.1203	**-0.0981**	**-0.0456**	0.0895	*-0.2185*
2008												
	μ	**0.5031**	**0.4504**	**0.5058**	**0.5294**	**0.5296**	**0.6398**					*0.4583*
	ϕ	**-1.1770e-07**	**-1.33461e-07**	**-5.9963e-08**	**-3.0179e-07**	**-9.1221e-08**	**-5.7509e-07**					*-4.3076e-07*
	β	-35.4946	-72.4978	-86.4532	-52.8429	-59.1459	-8.0243				-2379.6989	
	m	0.0746	-0.0821	0.0969	-0.1957	0.0759	0.1327				-0.6880	
DISCO												
1999												
	μ	**1.7152**										
	ϕ	**-1.5246e-05**										
	β	**20.2591**										
	m	**-0.4434**										



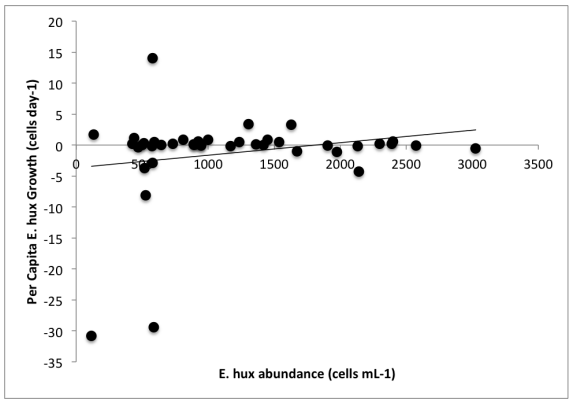
(a) 2000 (Mesocosm 1)



(b) 2003 (Mesocosm 3)



(c) 2008 (Mesocosm 3)



(d) DISCO 1999

Figure 3: Plot of per capita growth of *E. huxleyi* against *E. huxleyi* abundance. Note that these plots are for the centered in time data, which means the dPdt value was found using points from either side of the target data point. This was done in order to reduce noise. The DISCO 1999 data was centered in time to reduce noise.

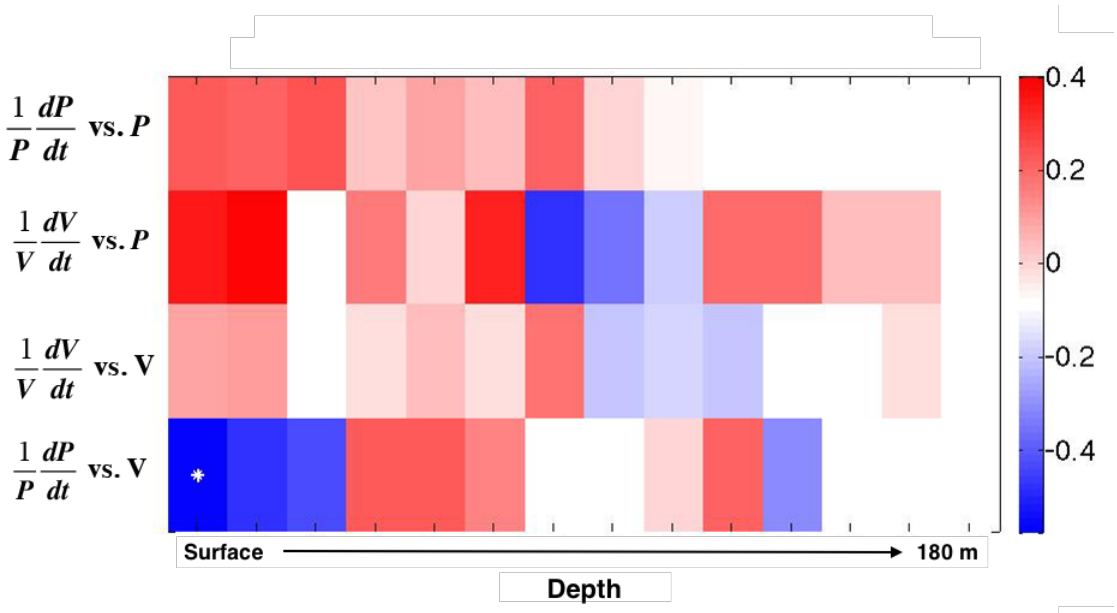
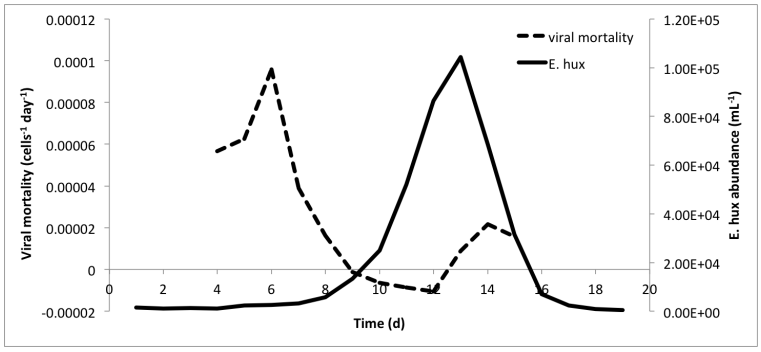
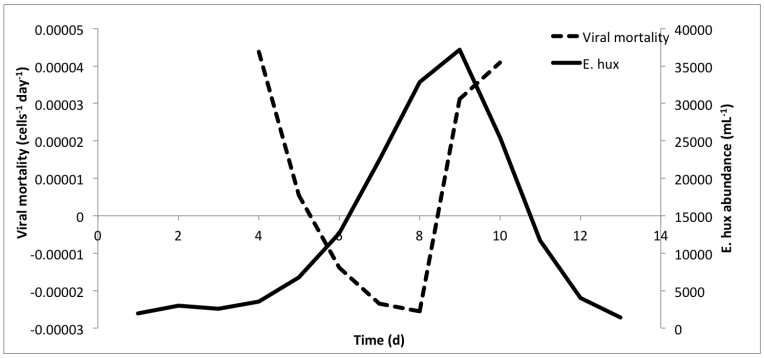


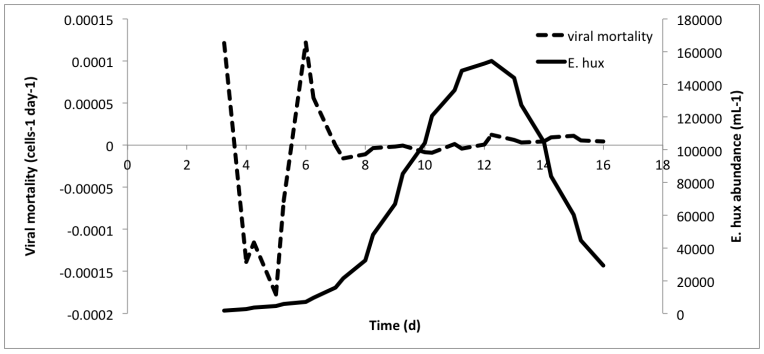
Figure 4: Linear regression significance for the basic virus-host model fit to the DISCO 1999 cruise in the North Sea. The second set represents R^2 values. White stars represent regressions with $p < 0.05$, while blank white squares indicate absence of data at those depths in a particular year.



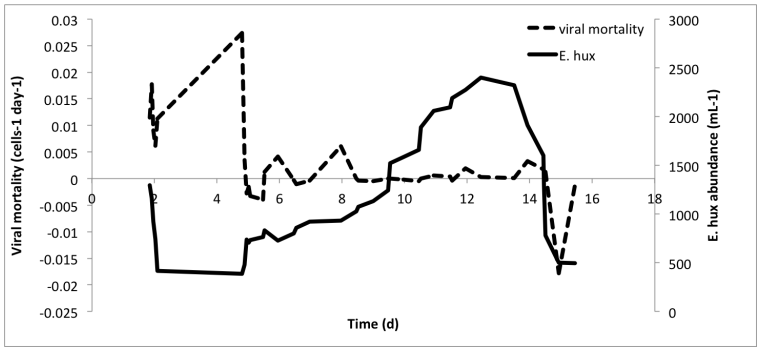
(a) 2000 (Mesocosm 1)



(b) 2003 (Mesocosm 3)

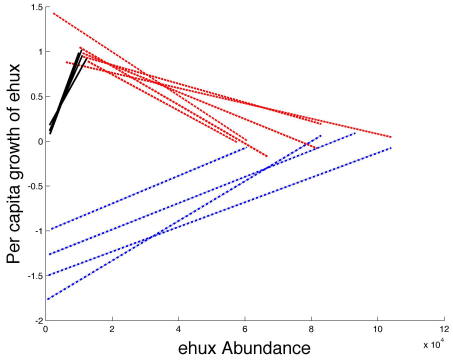


(c) 2008 (Mesocosm 3)

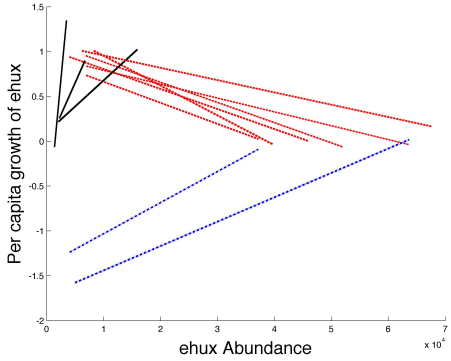


(d) DISCO 1999

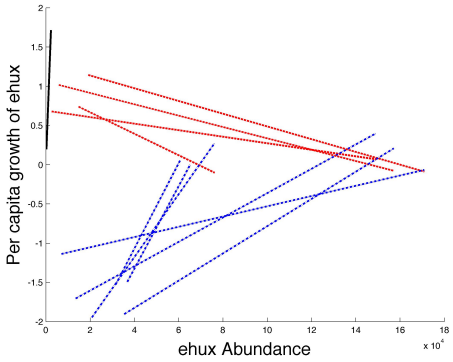
Figure 5: Plots of the running mean for the mortality term (dotted line) (calculated from the slope of the per capita growth of *E. huxleyi* v. abundance of *E. huxleyi* graphs) and the *E. huxelyi* abundance (solid line).



(a) 2000



(b) 2003



(c) 2008

Figure 6: Three-section piecewise analysis of the mesocosm data. All mesocosms from a year are plotted together to highlight the universality of the bloom segmentation. Only regressions found to be significant are plotted.

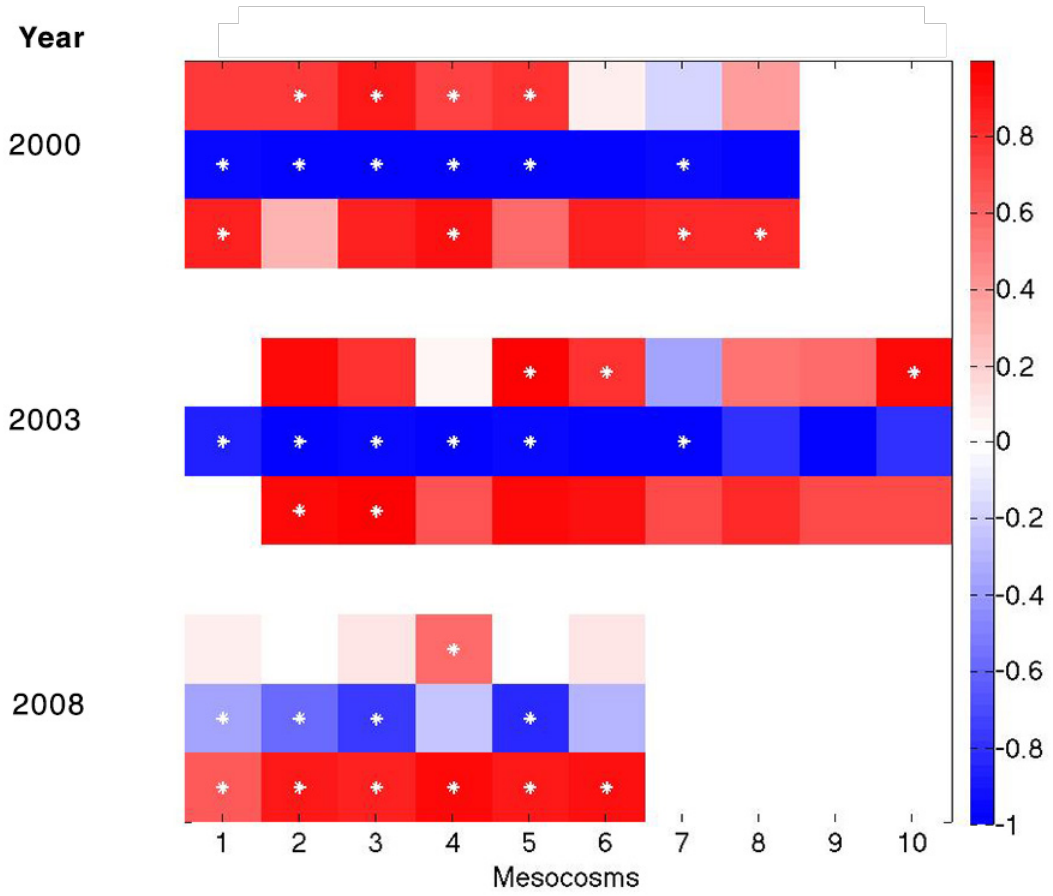
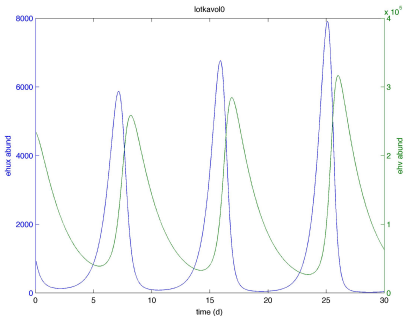
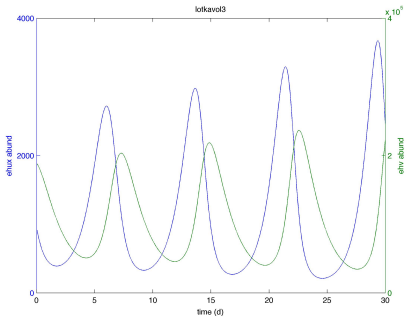


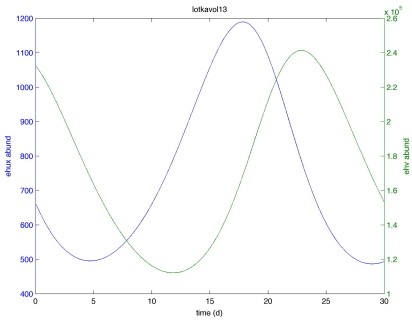
Figure 7: Linear regression significance and correlation for the three-section piecewise virus-host model. R^2 values are represented by color, while white stars represent regressions with $p < 0.05$. Blank white squares indicate absence of data in a particular year.



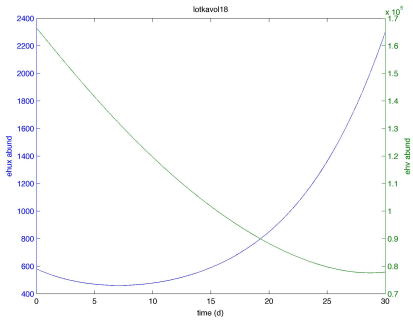
(a) 0-2 m



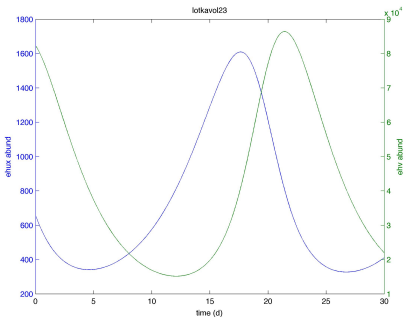
(b) 3-12m



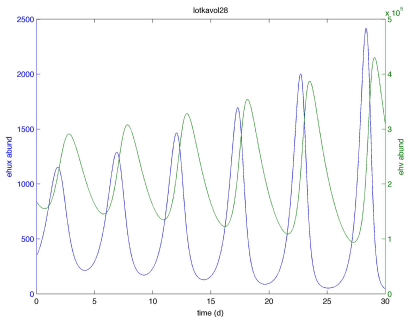
(c) 13-17m



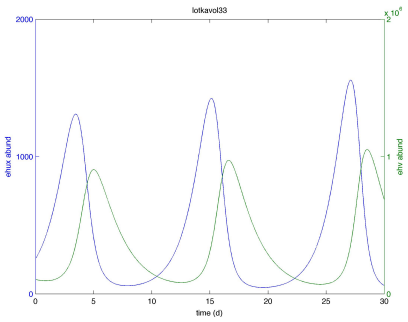
(d) 18-22m



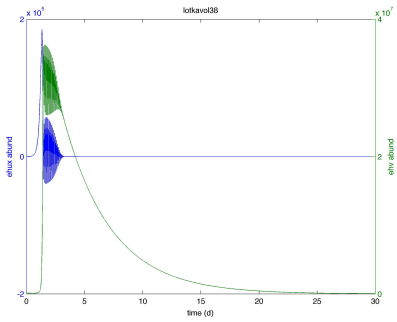
(e) 23-27m



(f) 28-32m

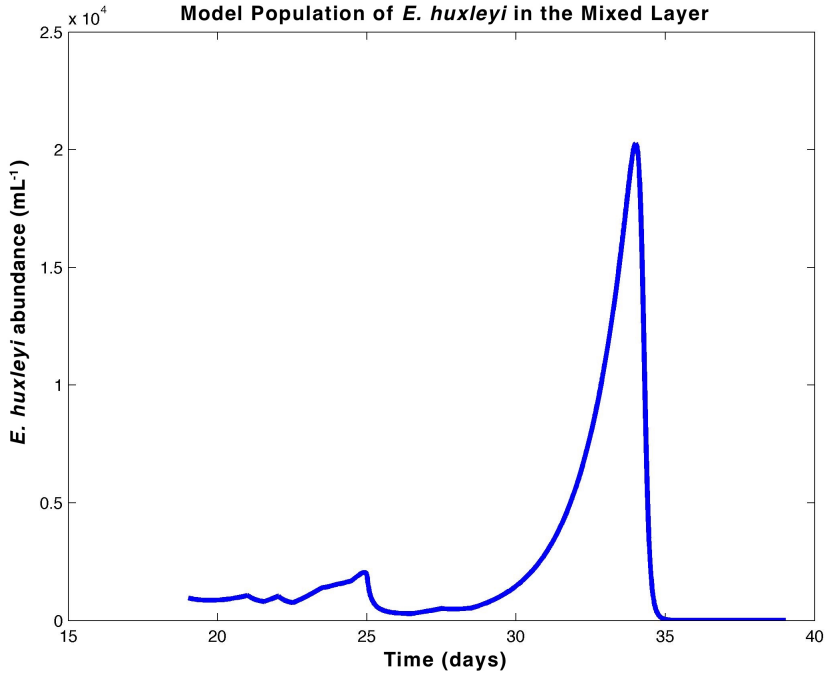


(g) 33-37m

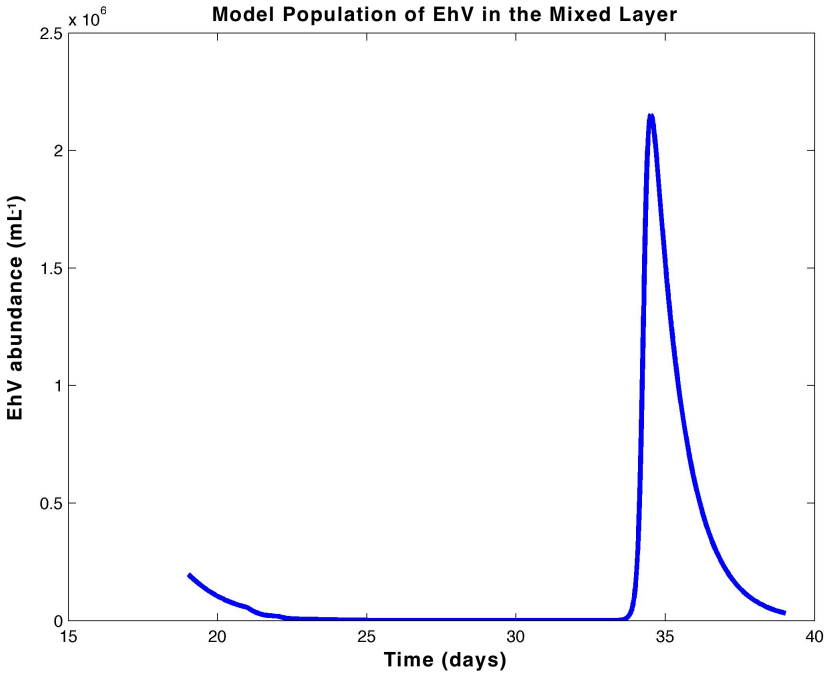


(h) 38-180m

Figure 8: Unforced Lotka-Volterra model output describing the interactions between *E. huxleyi* and EhV. This model does not account for viral degradation by UV.



(a) Model *E. huxleyi* Population



(b) Model EhV Population

Figure 9: Model output from the Lotka-Volterra virus-host model within the Evans & Parlsow physical forcing model for the mixed depth layer.

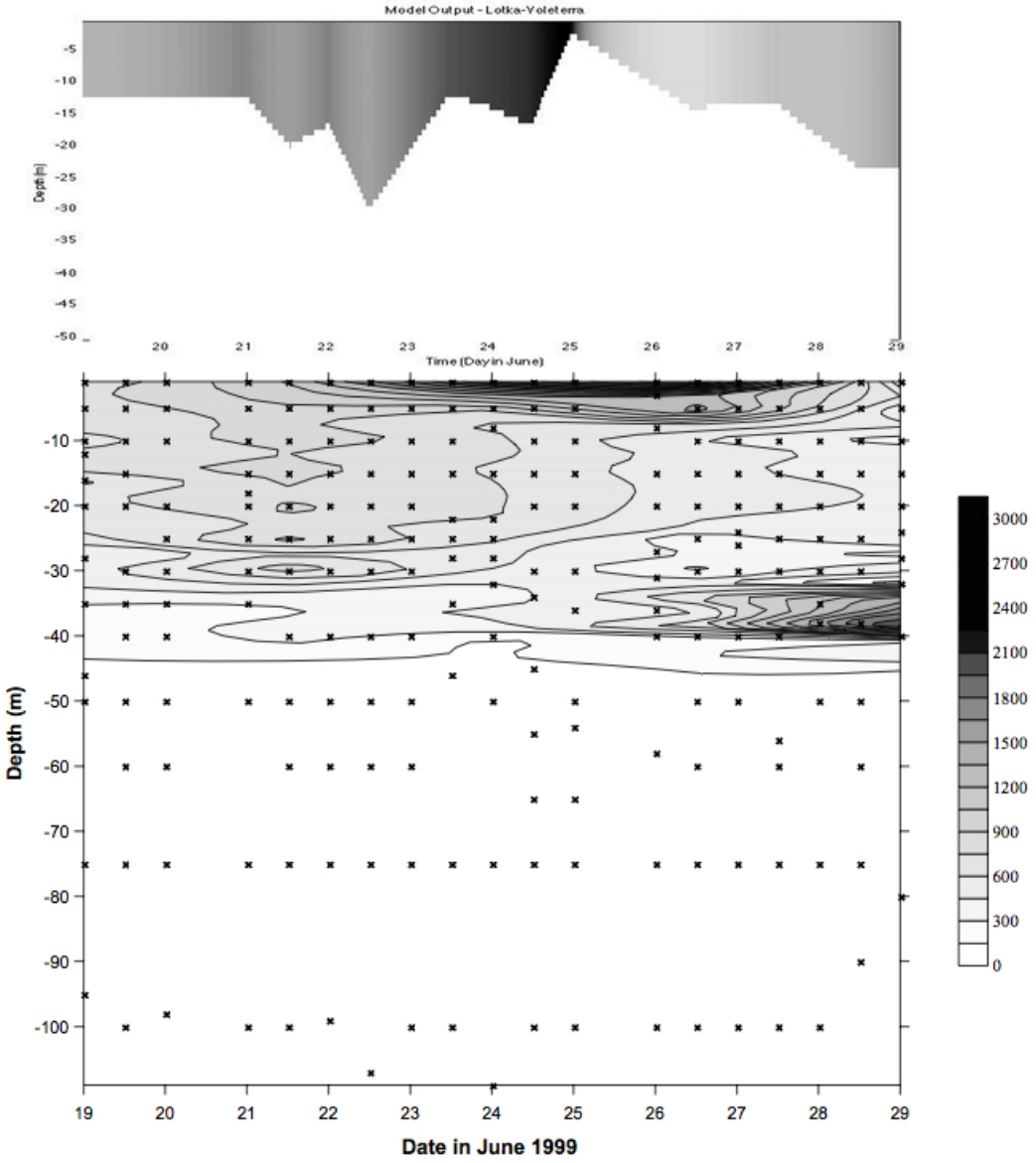


Fig. 5. Contour plot of coccolithophore concentrations (cells ml^{-1}), determined by AFC, throughout the course of the study. Crosses indicate the depth samples were collected.

Figure 10: Model output of the Lotka-Volterra virus-host model within the Evans & Parlsow physical forcing model compared to the DISCO 1999 data (Wilson et al. 2002). This is for the *E. hux* population.

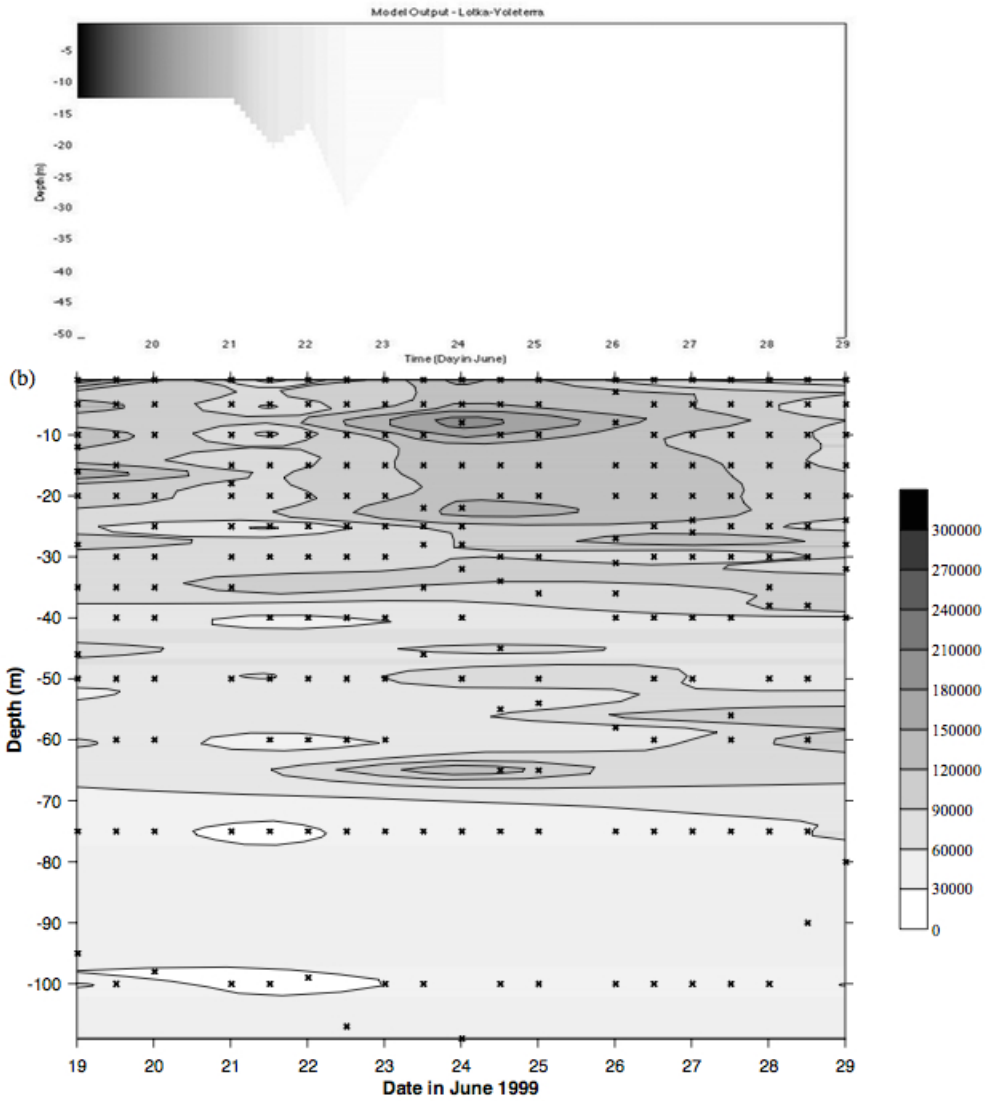
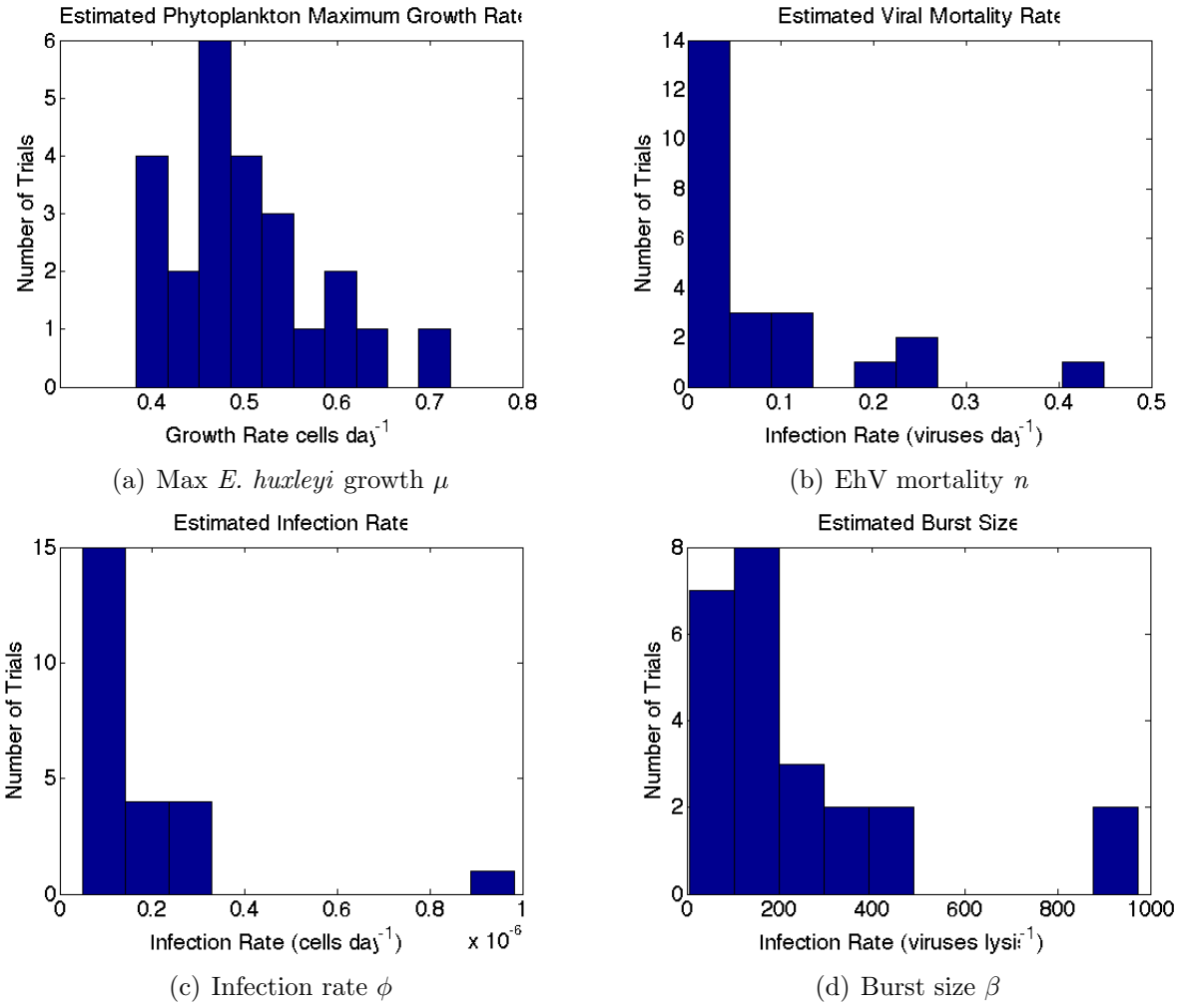


Fig. 4. Contour plots of total virus (a) and large-virus (b) concentrations (virus particles ml^{-1}), determined by AFC, throughout the course of the study. Crosses indicate the depth samples were collected.

Figure 11: Model output of the Lotka-Volterra virus-host model within the Evans & Parlsow physical forcing model compared to the DISCO 1999 data (Wilson et al. 2002). This is for the EhV population.



(a) Max *E. huxleyi* growth μ

(b) EhV mortality n

(c) Infection rate ϕ

(d) Burst size β

Figure 12: Histograms depicting the occurrences of estimates for each variable using the Lotka-Volterra model with Evans & Parslow physical forcing.

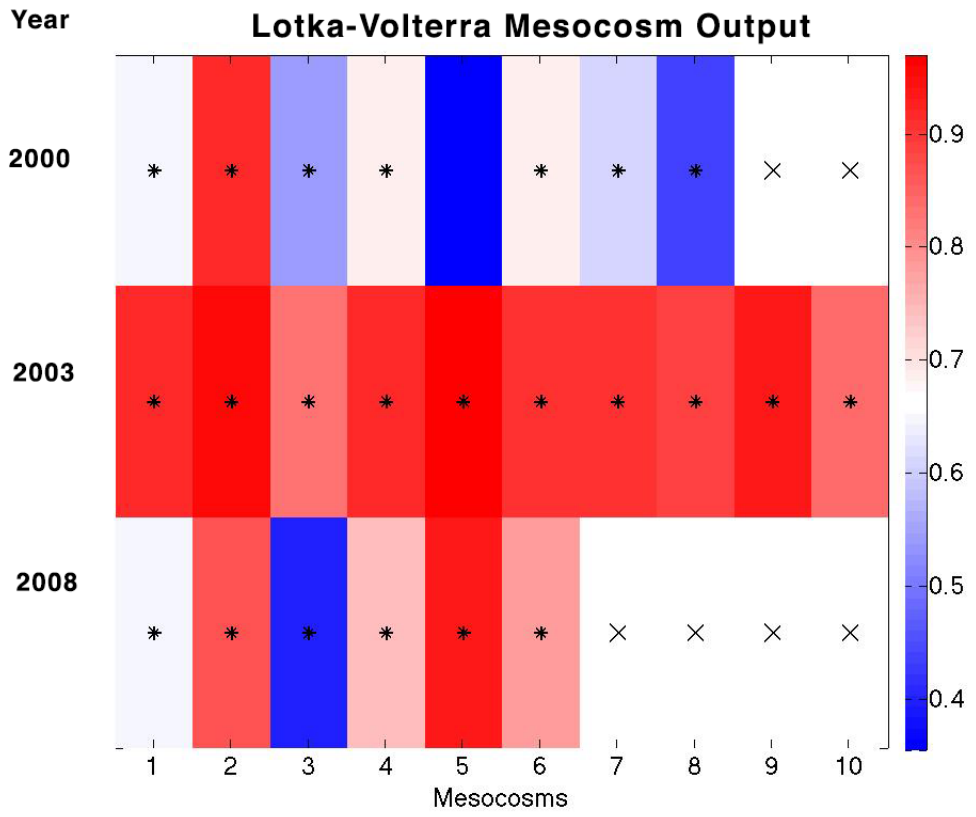
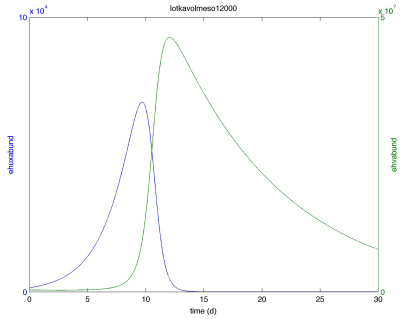
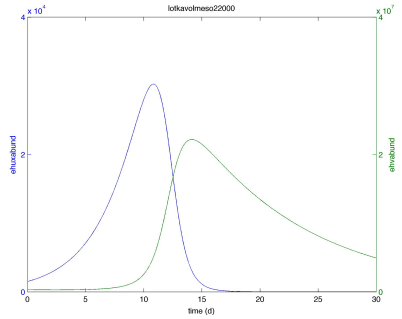


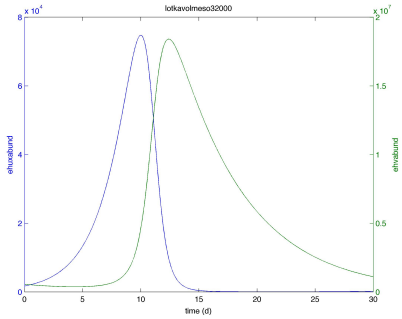
Figure 13: Correlation and significance data for the Lotka-Volterra virus-host model output based on fits to mesocosm data. Black stars represent regressions with $p < 0.05$, while blank white squares with a X indicate absence of those mesocosms in a particular year.



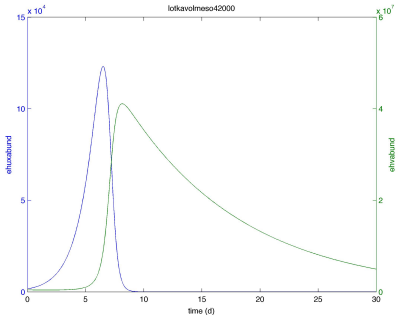
(a) Mesocosm 1



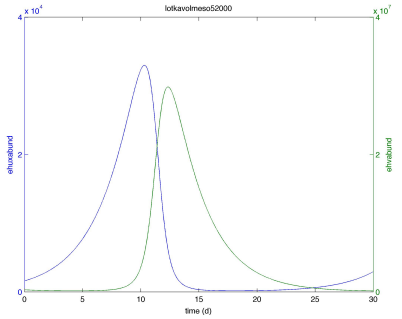
(b) Mesocosm 2*



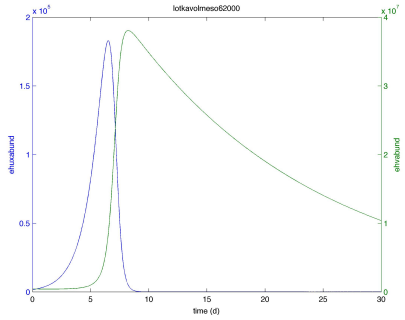
(c) Mesocosm 3



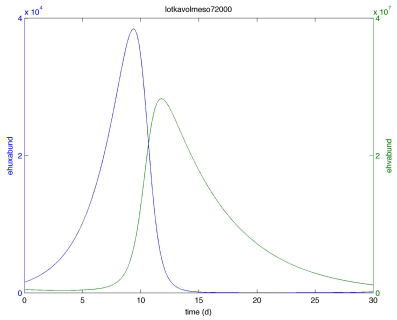
(d) Mesocosm 4



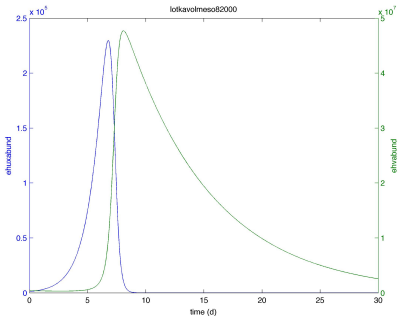
(e) Mesocosm 5



(f) Mesocosm 6

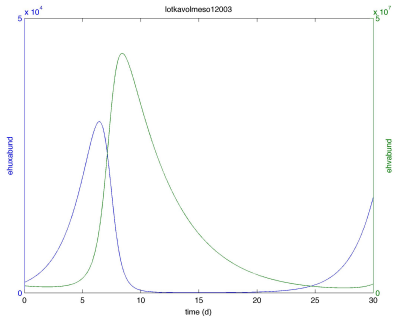


(g) Mesocosm 7

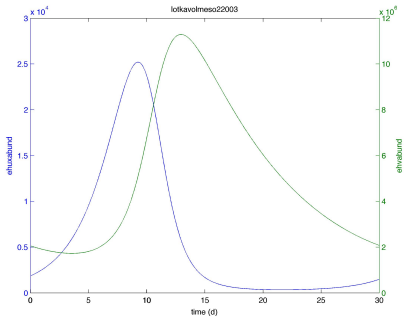


(h) Mesocosm 8

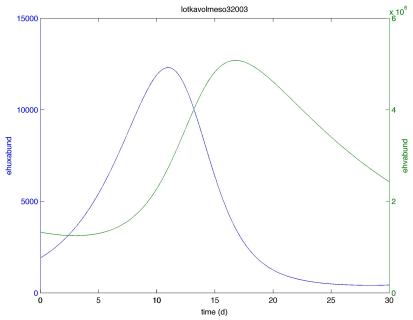
Figure 14: Unforced Lotka-Volterra model output describing the interactions between *E. huxleyi* and EhV fit to the 2000 mesocosm data. This model does not account for viral degradation by UV. A star in the subtitle indicates that the model fit was both statistically significant and had a R^2 value greater than 0.9.



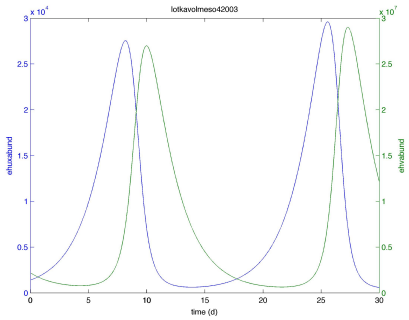
(a) Mesocosm 1*



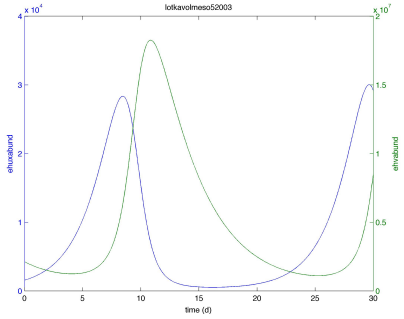
(b) Mesocosm 2*



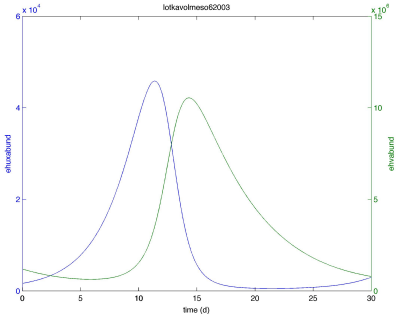
(c) Mesocosm 3



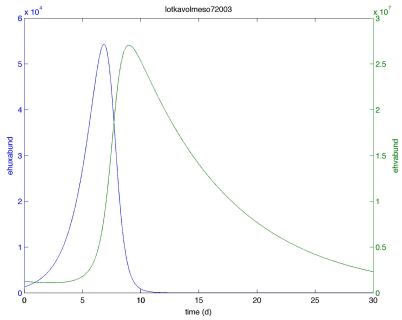
(d) Mesocosm 4*



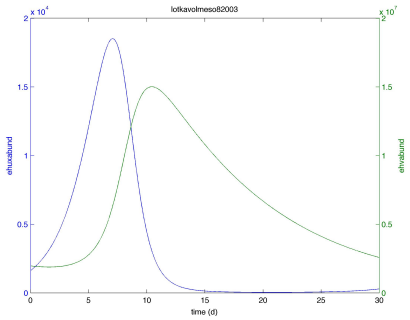
(e) Mesocosm 5*



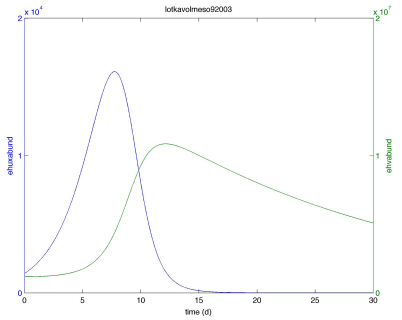
(f) Mesocosm 6*



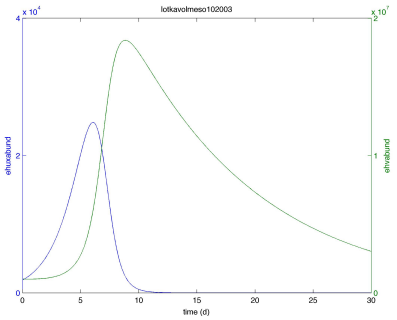
(g) Mesocosm 7*



(h) Mesocosm 8

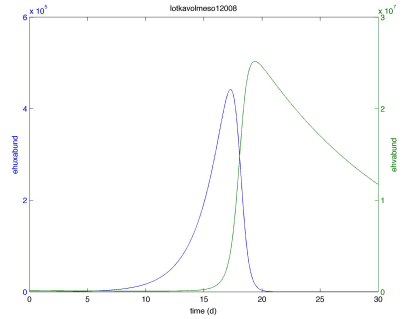


(i) Mesocosm 9*

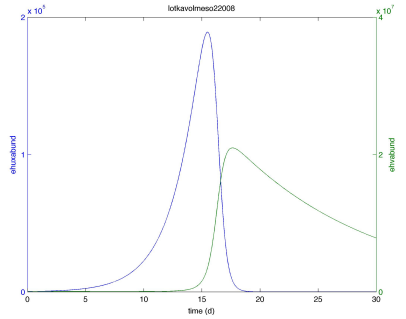


(j) Mesocosm 10

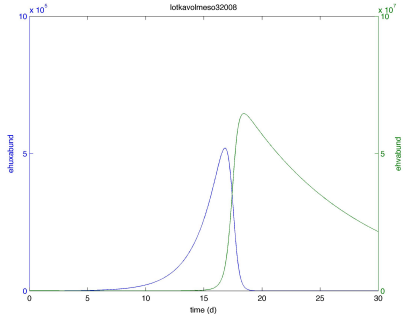
Figure 15: Unforced Lotka-Volterra model output describing the interactions between *E. huxleyi* and EhV fit to the 2003 mesocosm data. This model does not account for viral degradation by UV. A star in the subtitle indicates that the model fit was both statistically significant and had a R^2 value greater than 0.9.



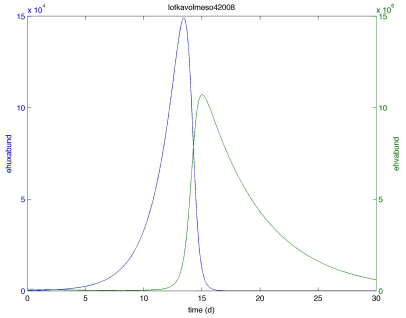
(a) Mesocosm 1



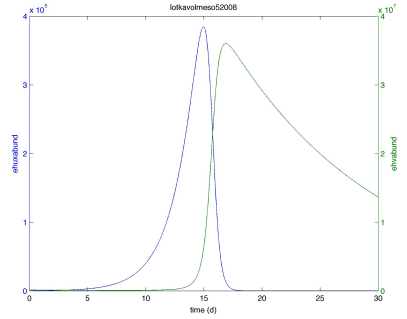
(b) Mesocosm 2



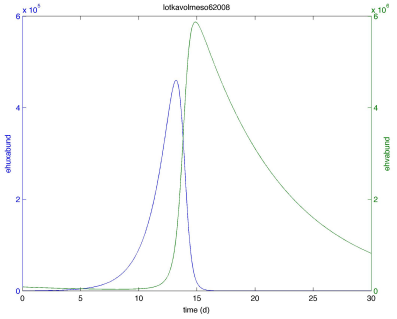
(c) Mesocosm 3



(d) Mesocosm 4



(e) Mesocosm 5*



(f) Mesocosm 6

Figure 16: Unforced Lotka-Volterra model output describing the interactions between *E. huxleyi* and EhV fit to the 2008 mesocosm data. This model does not account for viral degradation by UV. A star in the subtitle indicates that the model fit was both statistically significant and had a R^2 value greater than 0.9.

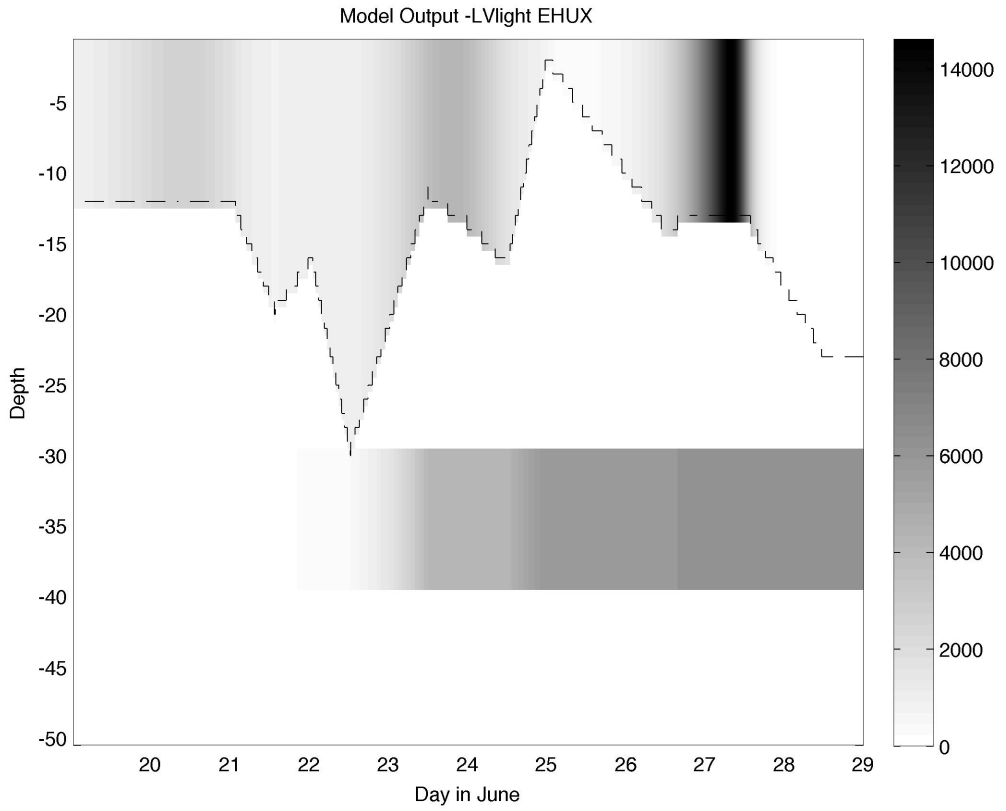


Figure 17: Model output of the Lotka-Volterra virus-host model with light affecting *E hux* growth within the Evans & Parlsow physical forcing model compared to the DISCO 1999 data (Wilson et al. 2002). This is for the *E. hux* population.

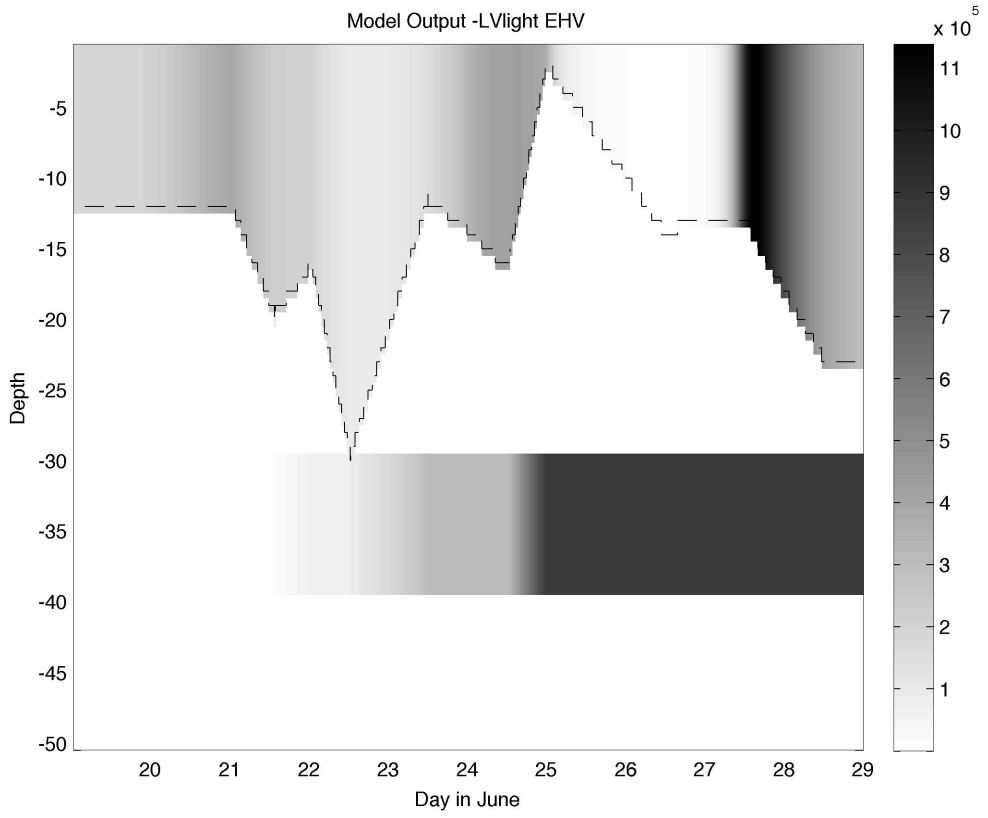
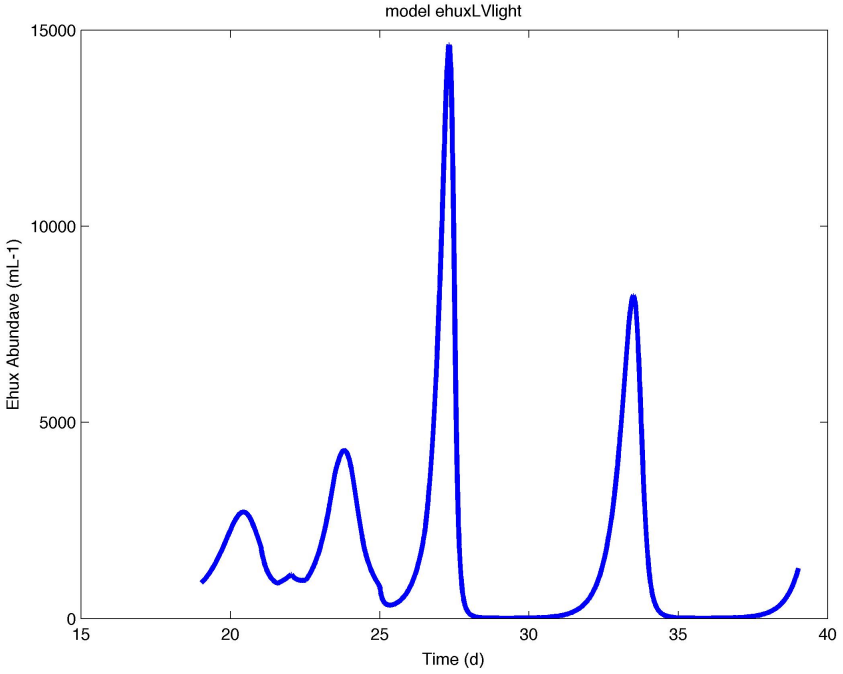
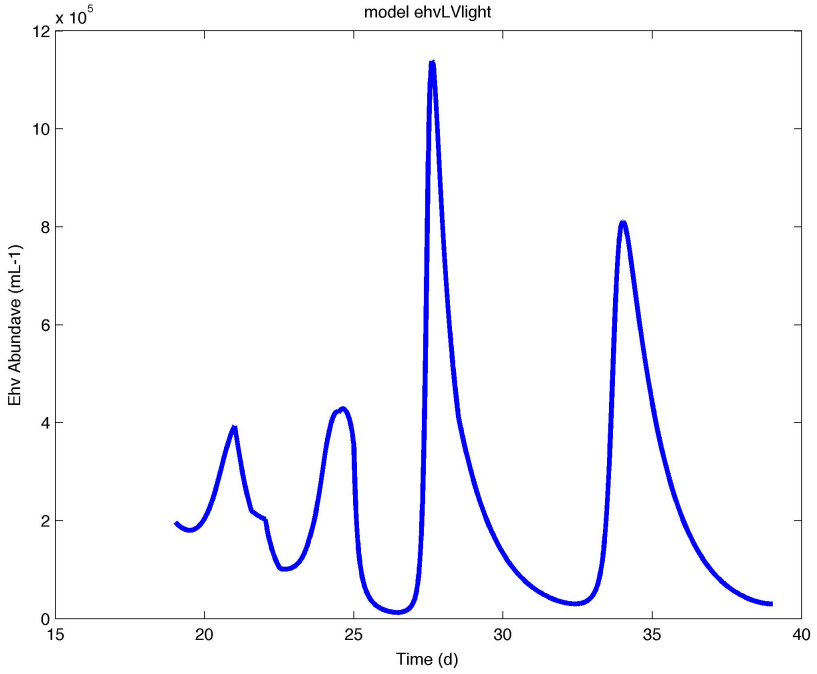


Figure 18: Model output of the Lotka-Volterra with light virus-host model within the Evans & Parlsow physical forcing model compared to the DISCO 1999 data (Wilson et al. 2002). This is for the EhV population.



(a) Model E. hux



(b) Model EhV

Figure 19: Model output of the Lotka-Volterra virus-host model with light affecting *E hux* growth within the Evans & Parlsow physical forcing model.

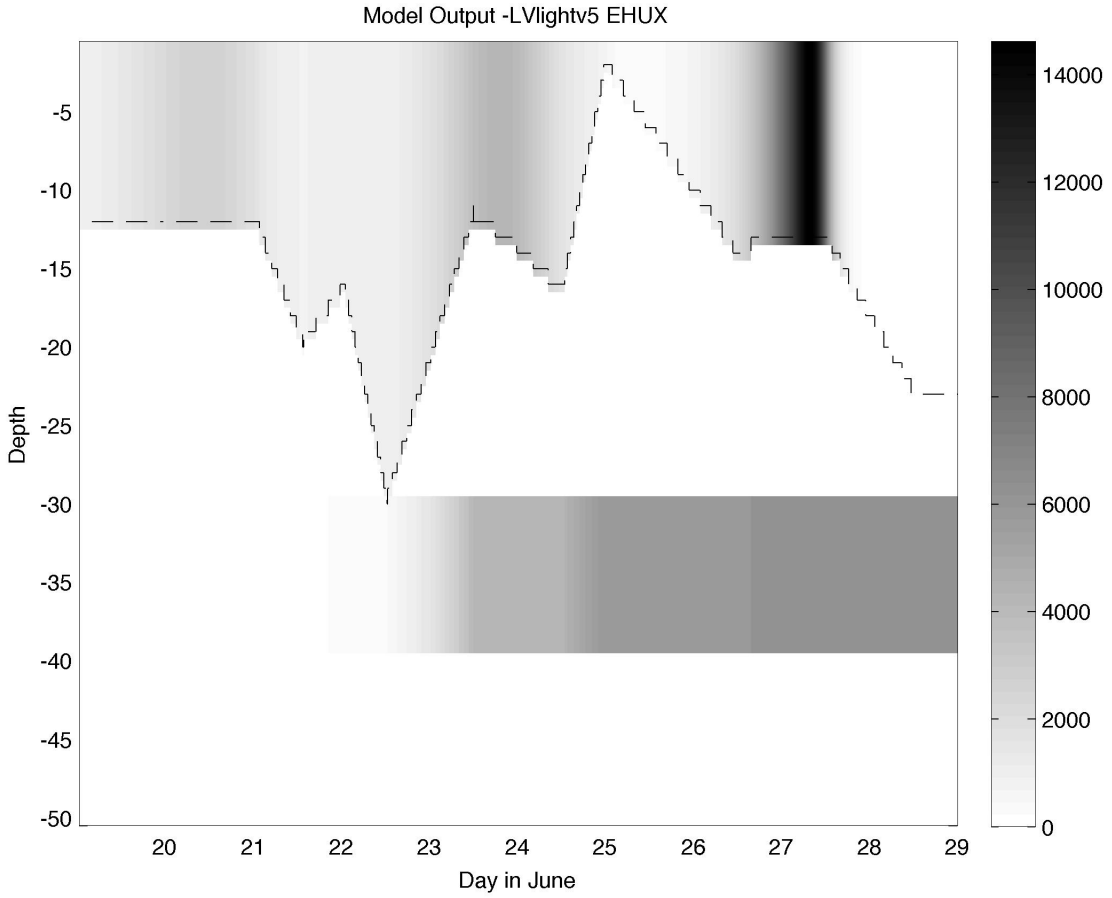


Figure 20: Model output of the Lotka-Volterra virus-host model with light affecting *E hux* growth within the Evans & Parlsow physical forcing model. This is for the *E. hux* population.

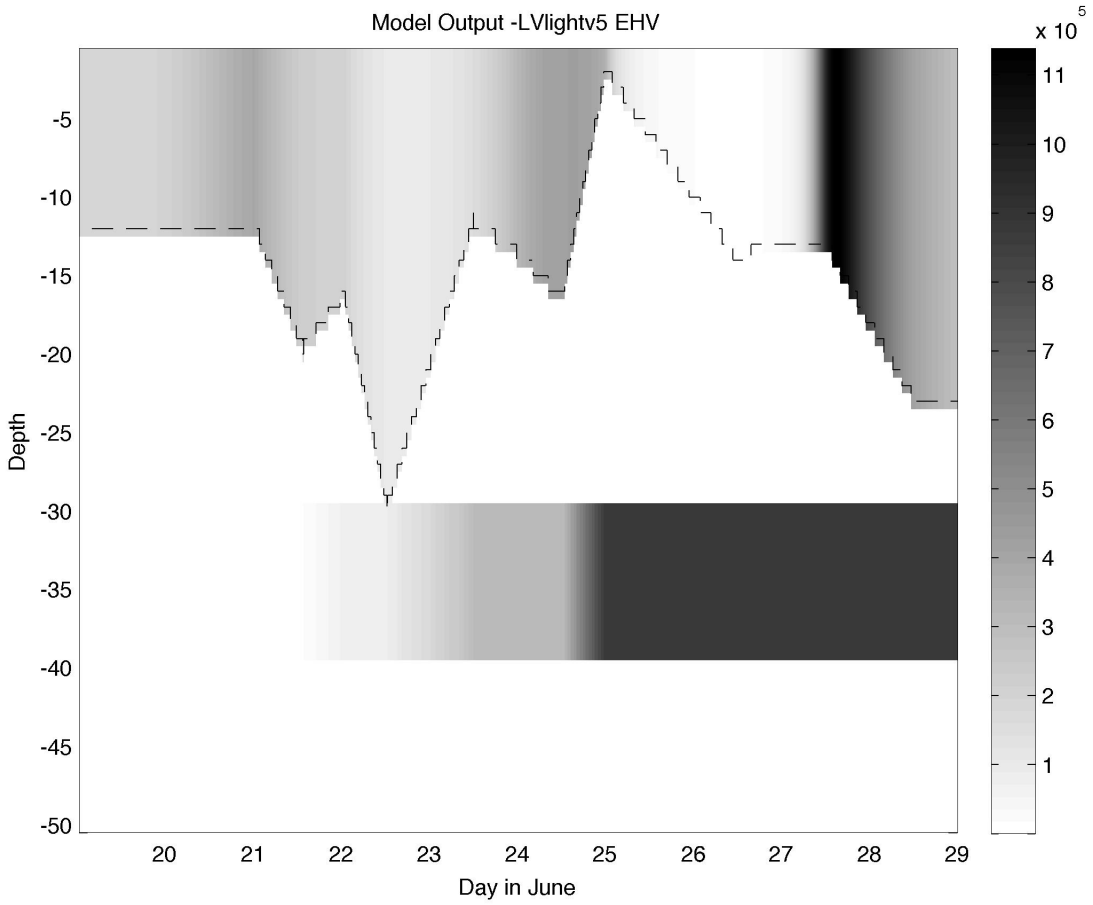


Figure 21: Model output of the Lotka-Volterra with light virus-host model within the Evans & Parlsow physical forcing model. This is for the EhV population.

6 Appendix

Most of the figures in this document were created using MatLab and have associated m-files. Additionally, model runs were performed using m-files. I will use this section in order to keep track of these m-files and explain what files do what, which m-files are associated with each other, and any other information that is related to the virus dynamics project.

6.1 Per capita changes in populations

These m-files go with the Phytoplankton growth and mortality (Section 3) and viral infection, burst size, and mortality (Section 5) sections. The functions were also later used to examine the infected populations (Section 6). The associated m-files are: *percap* and *percapitamanual*. These functions find the change per capita of one population compared to the abundance of another population (though both populations can be the same). Note that for both functions the variables used must be loaded into the funfction, which is currently not written in. For specifically the mesocosms used in this study, with the related file-naming system, the *percapitagen* m-file can be used.

6.2 Break-point analysis

These m-files go the the Breakpoint Analysis section (4). There is the first breakpoint analysis in which only one breakpoint was found, which uses *breakpoint*. The *three break* and *threebreakmanual* m-files are used for finding three break points (between the start and the greatest y-value point, between the greatest y-value point and the greatest x-value point, and between the greatest x-value point and the end of the data set). The difference between the two three break files is that the *threebreakmanual* file is not set up to loop through the mesocosm files specifically, while the *threebreak* file does loop through specifically the mesocosms and their associated files in the way they are named in the

VirusData folder.

6.3 Running means

The Running mean of *E. hux* mortality (Section 4.2) section used the *slopechange* and *slopechangeDISCO* files. The *slopechange* m-file is also set up to loop through specifically the mesocosm files as they are currently labelled.

6.4 Estimating variables

The *varest* (set up for looping through mesocosms) and *varestDISCO* files are used to estimate variables.

6.5 Lotka Volterra model

The Lotka Volterra model can be found in the *lotkavol* file. The *lotkavol_eval* file uses the Lotka Volterra model with input start parameters and uses these parameters as a start point for finding the parameters that best fit the model output to data. There is a script file (*lotkavol_fit*) which runs the eval function and also finds the correlation between the model output (using the best fit parameters) and real data. NOTE that the script file is currently written so that the file being evaluated must be manually written in the load lines.

6.6 Mixed layer depth

In order to run the Evans & Parslow model, the rate of change of the mixed depth must be know. We used the DISCO 1999 cruise data to calculate these values. The *mixinterp* script calculates these values.

6.7 Evans & Parslow 1D model and evaluating existing models

In order to evaluate existing models they were run within the Evans & Parslow 1D model, which simulates the effects of a changing mixed layer depth and light level. The complete E&P model can be viewed in the *Epode* m-file. This model has been revised to use a Lotka-Volterra model in the *EPodeLV* file. In order to run this model, the *EPode_solver* is used. The model being tested can be changed within the *EPode_solver* function. The *EPplot* function plots the density of a population (found using the model) above the mixed layer depth. The *EPplot2* function is the same function but includes densities below the mixed layer.

Within these files, the *Epode_setparams* function can be used to change some of the parameters. Currently it is set to change the values of μ (phytoplankton growth), ϕ (infection rate), β (burst size), and m_v (virus death)). The input "change" is a 1x4 vector where the entries are $[\mu\phi\beta\ m_v]$.

The *EPodeLVlightv* file includes light effects in both *E. hux* growth and EhV mortality.

6.8 Fitting to data

Currently there is one version of the fit function that works with the Epode set of functions. It was written for the EPodeLVlightv function, but can easily be modified to work with other models.

The set uses the *EPodeLVlightv_fit* function, which takes the *EPodeLVlightv_eval* function into an fminsearch. Although these functions are optimising parameters for the EPodeLVlightv function, they use the *EPLVlighvMOD* function, which is essentially the same function, but allows for the cycling of selected parameters within the fminsearch function.

6.9 Miscellaneous

These are some files that are not necessary, but sometimes just make things easier.

modeloutput_processing plots relevant plots for the data, including the abundance of populations against depth and time, calculating and plotting correlations, and other useful pieces of information.

deg2rad converts degrees to radians.

makepos makes all values in a matrix positive. This is specifically useful for arrays of estimated parameters (ex. although phi is always estimated as being negative, the negative is built into the model and therefore should not be part of the estimated parameter.

matchtimes finds all the matching time points between two time series that have different times and pulls the matching values from two related files.

plotunequal similar to matchtimes, but can use a tolerance (how close the two times must be) to locate time values. It also plots the matched values. This is useful for plotting the model against data (since the model usually has MANY more points).

removenans removes NaN values from data sets and also removes the corresponding values in a data set of the same length.

References

David Antoine, Jean-Michel André, and André Morel. Oceanic primary production: 2. estimation at global scale from satellite (coastal zone color scanner) chlorophyll. *Global Biogeochemical Cycles*, 10(1):57–69, 1996.

Edoardo Beretta and Yang Kuang. Modeling and analysis of a marine bacteriophage infection. *Mathematical Biosciences*, 149:57–76, 1998.

Gunnar Bratbak, Jorun K Egge, and Mikal Heldal. Viral mortality of the marine algae *Emiliania huxleyi* (haptophyceae) and termination of algal blooms. *Marine Ecology Progress Series*, 93:39–48, 1993.

Gunnar Bratbak, Anita Jacobsen, Mikal Heldal, Keizo Nagasaki, and Frede Thingstad. Virus production in *Phaeocystis pouchetii* and its relation to host cell growth and nutrition. *Aquatic Microbial Ecology*, 16:1–9, 1998.

CPD Brussaard, KR Timmermans, J Uitz, and MJW Veldhuis. Virioplankton dynamics and virally induced phytoplankton lysis versus microzooplankton grazing southeast of the kerguelen (southern ocean). *Deep-Sea Research II*, 55:752–765, 2008.

Corina P.D. Brussard, Rob S Kempers, Arjen J Kop, Roel Riegman, and Mikal Heldal. Virus-like particles in a summer bloom of *Emiliania huxleyi* in the north sea. *Aquatic Microbial Ecology*, 10:105–113, 1996.

T Castberg, A Larsen, RA Sandaa, CPD Brussaard, JK Egge, M Heldal, R Thyrhaug, EJ van Hannen, and G Bratbak. Microbial population dynamics and diversity during a bloom of marine coccolithophorid *Emiliania huxleyi* (haptophyta). *Marine Ecology Progress Series*, 221:39–46, 2001.

J. Chattopadhyay and S. Pal. Viral infection on phytoplankton-zooplankton system - a mathematical model. *Ecological Modelling*, 151:15–28, 2002.

Roberto Danovaro, Cinzia Corinaldesi, Antoni Dell'Anno, Jed A. Fuhrman, Jack J. Middelburg, Rachel T. Nobel, and Curtis A. Suttle. Marine viruses and global climate change. *FEMS Microbiol Review*, 35:993–1034, 2011.

Claire Evans, Stephen D Archer, Stéphan Jacquet, and William H Wilson. Direct estimates of the contribution of viral lysis and microzooplankton grazing to the decline of a *Micromonas* spp. population. *Aquatic Microbial Ecology*, 30:207–219, 2003.

Geoffrey T. Evans and John S. Parlsow. A model of annual plankton cycles. *Biological Oceanography*, 3(3):327–347, 1985.

Peter J.S. Franks. Npz models of plankton dynamics: their construction, coupling to physics, and application. *Journal of Oceanography*, 58:379–387, 2002.

Jed A. Fuhrman. Marine viruses and their biogeochemical and ecological effects. *Nature Reviews: Microbiology*, 399:541–548, 1999.

Wendy Gentleman, Andrew Leising, Bruce Frost, Suzanne Strom, and James Murray. Funtional responses of zooplankton feeding on multiple resources: a review of assumptions and biological dynamics. *Deep-Sea Research II*, 50:2847–2875, 2003.

Roger P. Harris. Coccolithophorid dynamics: The european *Emiliania huxleyi* programme, ehux. *Journal of Marine Systems*, 9:1–11, 1996.

Stéphan Jacquet, Mikal Heldal, Debora Iglesias-Rodriguez, Aud Larsen, William Wilson, and Gunnar Bratbak. Flow cytometric analysis of an *Emiliana huxleyi* bloom terminated by viral infection. *Aquatic Microbial Ecology*, 27:111–124, 2002.

Bruce R. Levin, Frank M. Stewart, and Lin Chao. Resource-limited growth, competition, and predation: A model and experimental studies with bacteria and bacteriophage. *The American Naturalist*, 111(977):3–24, 1977.

Carole A Llewellyn, Glen A Tarran, Chris P Galliene, Denise G Cummings, Alex De Menezes, Andy P Rees, Jo L Dixon, Claire E Widdicombe, Elaine S Fileman, and Willie H Wilson. Microbial dynamics during the decline of spring diatom bloom in the northeast atlantic. *Journal of Plankton Research*, 30(3):261–272, 2008.

AJ Lotka. Fluctuations in the abundance of a species considered mathematically. *Nature*, 119(2983), 1927.

Joaquín Martínez Martínez, Svein Norland, Tron Frede Thingstad, Declan C. Schroeder, Gunnar Bratbrak, William H. Wilson, and Aud Larsen. Variability in microbial population dynamics between similarly perturbed mesocosms. *Journal of Plankton Research*, 28(8):783–791, 2006.

Joaquín Martínez Martínez, Declan C. Schroeder, Aud Larsen, Gunnar Bratbrak, and William H. Wilson. Molecular dynamics of *Emiliania huxleyi* and cooccurring viruses during two separate mesocosm studies. *Applied and Environmental Microbiology*, 73(2): 554–562, 2007.

Joaquín Martínez Martínez, Declan C. Schroeder, and William H. Wilson. Dynamics and genotypic composition of *Emiliania huxleyi* and their co-occurring viruses during a coccolithophore bloom in the north sea. *FEMS Microbiology Ecology*, pages 1–9, 2012.

Evangelia Michaloudi, Maria Moustaka-GOuni, Spyros Gkelis, and Kimon Pantelidakis. Plankton community structure during an ecosystem disruptive algal bloom of *Prynesium parvum*. *Journal of Plankton Research*, 31(3):301–309, 2009.

Yiwen Pan, Wei Fan, Ting-Hsuan Huang, Shu-Lun Wang, and Chen-Tung Arthur Chen. Evaluation of the sinks and sources of atmospheric co_2 by artificial upwelling. *Science of the Total Environment*, 511:692–702, 2015.

C.J. Rhodes and A.P. Martin. The influence of viral infection on a plankton ecosystem undergoing nutrient enrichment. *Journal of Theoretical Biology*, 265:225–237, 2010.

C.J. Rhodes, J.E. Truscott, and A.P. Martin. Viral infection as a regulator of oceanic phytoplankton populations. *Journal of Marine Systems*, 74:216–226, 2008.

Leslie Schwierzke, Daniel L Roelke, Bryan W Brooks, James P Grover, Theodore W Valenti, Mieke Lahousse, Carrie J Miller, and James L Pinckney‘. *Prymnesium parvum* population dynamics during bloom development: a role assessment of grazers and viruses. *Journal of the American Water Resources Assoviation*, 46(1):63–75, 2010.

Ivo Siekmann and Horst Malchow. An extension of the beretta-kuang model of viral disease. *Mathematical Biosciences and Engineering*, 5(3):549–565, 2008.

Olin K. Silander, Daniel M. Weinreich, Kevin M. Wright, Kara J. O'Keefe, Camilla U. Rang, Paul E. Turner, and Lin Chao. Widespread genetic exchange among terrestrial bacteriophages. *Proceedings of the Nation Academy of Sciences of the United States of America*, 102(52):19009–19014, 2005.

Brajendra K. Singh, Joydev Chattopadhyay, and Somdatta Sinha. The role of virus infection in a simple phytoplankton zooplankton system. *Journal of Theoretical Biology*, 231:153–166, 2004.

Curtis A Suttle. Mechanisms and rates of decay of marine viruses in seawater. *Applied and Environmental Microbiology*, 58(11):3721–3729, 1992.

Curtis A Suttle. Marine viruses - major players in the global ecosystem. *Nature Reviews: Microbiology*, 5:801–812, 2007.

T Frede Thingstad. Elements of a theory for the mecahnisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology and Oceanography*, 45(6):1320–1328, 2000.

Runar Thyrhaug, Aud Larsen, T Frede Thingstad, and Gunnar Bratbak. Stable coexistence in marine algal host-virus systems. *Marine Ecology Progress Series*, 254:27–35, 2003.

Rucheng C. Tian. Toward standard parameterizations in marine biological modeling. *Ecological Modelling*, 193:363–368, 2006.

Jefferson T. Turner. Planktonic marine copepods and harmful algae. *Harmful Algae*, 32: 81–93, 2014.

T Tyrrell and A.H. Taylor. A modelling study of *Emiliania huxleyi* in the ne atlantic. *Journal of Marine Systems*, 9:83–112, 1996.

V Volterra. Principes de biologie mathématique. *Acta Biotheoretica*, 1(3):1–36, 1937.

Carl J. Watras, Veronique C. Garçon, Robert J. Olson, Sallie W. Chisholm, and Donal M. Anderson. The effect of zooplankton grazing on estuarine blooms of the toxic dionflagellate *Gonyaulax tamarensis*. *Journal of Plankton Research*, 7(6):891–908, 1985.

Markus G. Weinbauer. Ecology of prokaryotic viruses. *FEMS Microbiol Review*, 28, 2004.

Markus G. Weinbauer and Fereidoun Rassoulzadegan. Are viruses driving microbial diversification and diversity? *Evironmental Microbiology*, 6(1):1–11, 2004.

Joshua S. Weitz and Jonathan Dushoff. Alternative stable states in host-phage dynamics. *Theoretical Ecology*, 1:13–19, 2008.

Steven W. Wilhelm, Wade H. Jefferey, Amanda L. Dean, Jarah Meador, J. Dean Pakulski, and David L. Mitchell. Uv radiation induced dna damage in marine viruses along a latitudinal gradient in the southereastern pacific ocean. *Aquatic Microbial Ecology*, 31: 1–8, 2003.

William H. Wilson, Glen Tarran, and Mikhail V Zubkov. Virus dynamics in a coccolithophore-dominated bloom in the north sea. *Deep-Sea Research II*, 49:2951–2963, 2002.