

Summary of Research

The focus of my research is to develop and analyze mathematical models that investigate the behavior of systems in biology and ecology. I am particularly interested in invasion dynamics and control strategies, species-environment interactions, and how population demographics can change over time in response to external drivers and community structure. My previous work includes modeling and analyzing the non-local spread of cheatgrass establishment in Rocky Mountain National Park, the time evolution of savanna tree composition in response to water resource availability and fire disturbance, and presently, the dispersal of parasitoid wasps from point release on the kilometer scale. The broad goal of my future work is to incorporate environmental conditions and dynamics, social network structure, and available field data into a probabilistic modeling framework that can inform risk management policy and aid in developing superior control strategies.

The approach I use to study each of these problems is highly motivated by the scientific setting and can involve a diverse range of mathematical topics including scientific computing, probabilistic modeling, network theory, dynamical systems, and Bayesian model selection. The modeling approach I use is based on principles of analytical tractability and broad scientific utility. My methods are then typically computational in nature, with the goal of providing results that admit a direct comparison to data. This last facet of my work has also led me to start pursuing some aspects of data science, including managing large data sets, converting between formats, and conducting data manipulation and analysis in software such as the Python pandas library or the 3D visualization software VisIt [16, 17]. Throughout the research process, I collaborate closely with scientists to ensure that my work is contextually relevant and provides new, actionable insight into the mechanistic properties of the system we are studying.

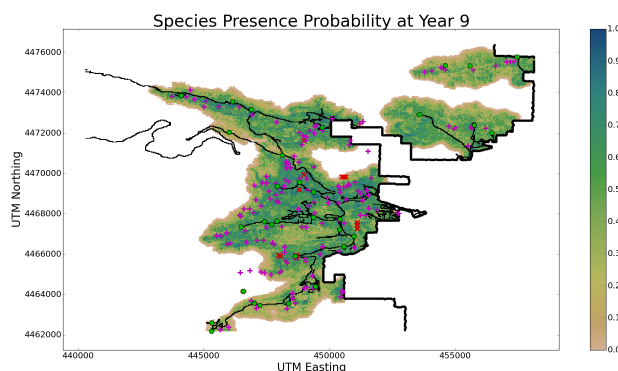


Figure 1: Presence probability of cheatgrass in Rocky Mountain National Park 9 years after initial presence data.

Interdisciplinary Collaboration

My research is primarily inspired by scientific questions that in turn drive mathematical and computational innovation. As a result, scientific collaboration is a critical part of my research [13]. For example, most of my graduate research addresses a problem in mathematical ecology involving the spread of a biological invader through nonhomogeneous domains with an embedded human transportation network. This project relied heavily on a collaboration with ecologists Tom Stohlgren and Sunil Kumar that produced not only results addressing data on the spread of cheatgrass in Rocky Mountain National Park [15] (Fig. 1), but also novel mathematics extending the work of Mollison [9] to heterogeneous domains and species presence probability [14]. In a current project, I am collaborating with Nadiah Kristensen, an ecologist, and Laura Miller, a biological fluid dynam-

icist, on a project that uses data driven mathematics to demonstrate the relative importance of air flow in modeling the kilometer-scale dispersal of tiny, wind-borne insects [16]. Future work includes incorporating these findings into modeling marine dispersal of small organisms with Virginia Pasour [17] and developing more accurate wind-based models.

Current and Previous Work

1. Modeling Invasive Plant Dispersal

Invasive species represent one of the major environmental threats of the 21st Century. Damage to native species, habitat, and agricultural lands leads to economic suppression, reduced food and water security, and direct threats to human health [3]. Control costs rise dramatically with population size, so mathematical models of potential spread are essential to inform management strategies. However, the traditional approach of modeling invasive plant dispersal using a Fickian diffusion system with logistic growth [8, 12] has been widely criticized because species dispersal in these models is assumed to be local, normally distributed, and uniform in all directions. In fact, dispersal mechanisms are often non-local with heavy-tailed distributions and dependent on environmental factors [2]. Current mathematical models also routinely ignore uncertainty in initial conditions and neglect heterogeneity in the landscape.

Our work addresses these and other concerns by deriving and analyzing a new model for invasive spread based on the stochastic contact birth process introduced by Mollison called the “simple epidemic” (Eqn. 1 in appendix) [9]. In forming our model, we note that the quantity of interest for ecologists is often not the expected population size, but rather the probability that an invader will be present at all. This fact is reflected in available data, which is often in the form of spatial presence/absence points with no information about population density. By first extending the model to a spatially continuous formulation and then using the Master equation generated by the simple epidemic, we derived an intergo-differential equation for the probability that the population is greater than zero in any specified area (Eqn. 2 in appendix). This new model includes landscape heterogeneity as well by incorporating data from ecological niche modeling software as a probability of establishment [14].

1.1. Function approximations and solving the inverse problem

Our model includes one term that must be cast as an unknown function of the presence probability and reconstructed. Rough approximations were found analytically and perform reasonably

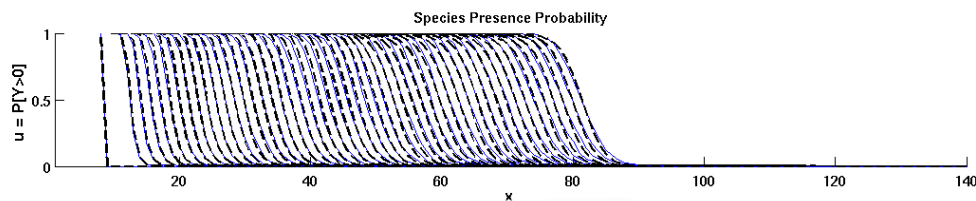


Figure 2: Average of pseudo-random realizations for $P[Y(x) > 0]$ (solid blue line) compared with homotopy model prediction through time 50. $K = 200$ with a standard normal dispersal kernel.

well under certain conditions. To obtain better approximations, we generated data for presence probability based on stochastic simulations and solved the ill-posed inverse problem using generalized Tikhonov regularization. We then matched long-term behavior with transients by applying a homotopy relation. Though the result is highly sensitive to parameter selection due to non-local effects, Fig. 2 demonstrates the reconstruction accuracy that is achievable via this method.

1.2. Network model

Modern landscapes rarely exist without some kind of transportation infrastructure, even in the case of preserves and natural parks. Roads and trails effectively connect otherwise distant locations and can facilitate the movement of an invader along their length. Assuming a general, spatially continuous spread model has been specified, we formulated a general method for attaching a network in the form of a system of coupled ordinary differential equations (Eqn. 3 in appendix) that describe dynamics for susceptible and disease carrying vectors traveling from node to node. For an herbaceous invader, there should also be some latency between network transportation and growth so we augment this model with an intermediate, spatially continuous latent step that realizes this dynamic. It is this formulation of the network model that is represented in Fig. 1, where we model the spread of cheatgrass in Rocky Mountain National Park including observed presence locations for validation [15]. We have also considered various control regimes and made the full Python code for this project available on GitHub.

2. Modeling Parasitoid Wasp Dispersal from Point Release

Biological invasions have movement at the core of their success. However, due to difficulties in collecting data, medium- and long-distance dispersal of small insects and seeds has long been poorly understood and likely underestimated. Data collected on the agricultural introduction of a previously absent species of parasitoid wasp, *Eretmocerus hayati*, in a hierarchical sampling design and over an area measured in kilometers [6] represents a special opportunity to reveal and study the key mechanisms behind tiny insect dispersal. To begin this process, we formulated a probabilistic model resulting in a special case of the Fokker-Planck equation (Eqn. 4 in appendix) which admits an analytic solution. Local, near-ground movement is modeled using simple diffusion while wind-based flight includes both an active and passive component: wasps may preferentially choose a take-off time based on environmental conditions but then float on the wind, which is assumed spatially but not temporally constant (Eqn. 5 in appendix).

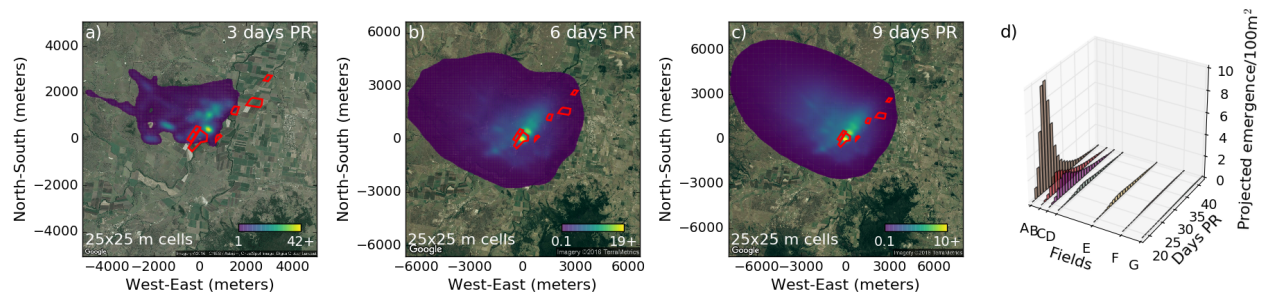


Figure 3: Model results for parasitoid numbers on data collection days (a-c) and predicted emergence (d).

2.1. Bayesian framework and scientific computation

To fit model parameters to field data in a robust way, it was necessary to superimpose a second, probabilistic framework onto the dispersal model that could describe multiple data collection methodologies at different scales while accounting for uncertainty. In particular, the statistical model had to correctly match geospatial data to locations within the model of different shapes and sizes while projecting estimated population density to subsequent adult emergences from the host species. After specifying some initial information about the expected value of parameters as prior distributions, we could then seek the parameters that maximize the probability of observing the data given the model and eventually quantify the uncertainty in those parameter choices as well. As this process involves a high-dimensional nonlinear optimization problem, methods typically involve Monte Carlo algorithms and necessitate fast evaluation of the model in order to obtain thousands of realizations from the parameter space.

The multiscale nature of this model presents significant challenges in terms of computational speed because spatial resolution must be fine enough to capture local movement of wasps < 1 mm in size while working on a domain over 100 km^2 in size. To address this problem, we utilized a two-pronged approach. The probability distribution for each day's movement was assumed to be independent and thus calculated in parallel. The distributions were then combined using an FFT convolution implemented on a GPU using CUDA Python libraries [5]. Numerical Python was used throughout the implementation with the goal of producing transparent and reusable code with high performance metrics. Fig. 3 shows results using the maximum a posteriori parameter estimates.

3. Savanna Modeling

Modeling is a vital part of savanna research as ecologists must rely on models to explore stand compositional changes across generations and the dynamics of tree and grass coexistence dictates how savanna will respond to various forms of disturbance (climate change, grazing livestock, fuel-wood harvest, etc.). However, there is currently no consensus as to why savanna occurs as a stable state between tropical grassland and forest. Numerous models have been suggested to address this problem, but often they take the form of process-based software simulations that are mathematically intractable and difficult to analyze. In collaboration with savanna ecologists Adam Liedloff and Garry Cook [7] of the Commonwealth Scientific and Industrial Research Organization (CSIRO) lab at Darwin, Australia, we developed an ODE-based model to explore the interaction between seasonal resource availability, fire disturbance, and stand structure in Australian savannas. The model shows how changes in climate seasonality and fire disturbance may effect woody biomass and seedling growth over centuries-long time scales [14].

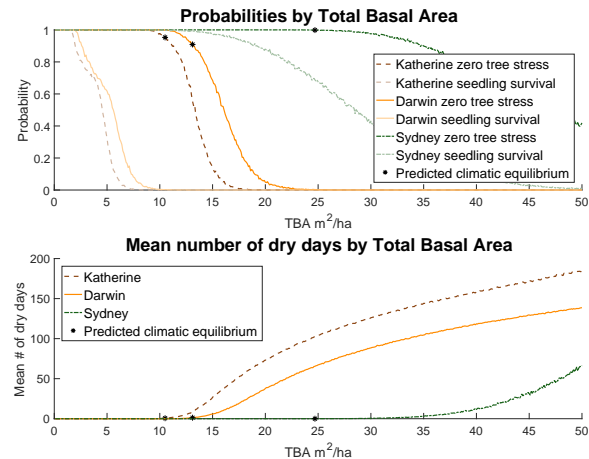


Figure 4: Climate effect on tree population by location and woody biomass.

3.1. Modeling approach and computational analysis

To correctly account for the stability of mesic Australian savannas, special attention was paid to modeling water resource availability based on soil properties, rooting depth, and daily rainfall history. The soil system was imagined as a cascading bucket model, where water flows downward through successive layers that grant different access to grasses and trees based on rooting depth or niche access. Tree population was modeled using an age-structured approach that allowed us examine structural dynamics and apply a non-uniform susceptibility to fire. The end result was a system of ordinary differential equations coupled with a discrete-time model for advancing monthly age classes (Eqn. 6 in appendix).

For analyzing the behavior of our system over long time scales, we utilized a Markov chain model for stochastic daily rainfall generation that was parameterized by location specific precipitation records. Our results revealed a long term oscillation for stand structure based on seedling cohorts which advance into maturity when woody biomass decreases and resources become available. However, we also showed that this oscillation is likely noisy and less well defined on the stand scale due to a spatial resource niche for seedlings that keeps total basal area roughly constant at a location specific climatic equilibrium. Using another discrete-time Markov process with absorbing boundary condition, we provided a method for approximating the probability distribution for drought stress and thus the basal area climatic equilibrium (see Fig. 4; this figure was generated using Lincoln Laboratory's MatlabMPI [4] to examine the parameter space in parallel on a cluster). We then examined the effects of fire as a stochastic perturbation away from underlying climatic equilibrium states [14].

Future work

An area that I am particularly targeting for future research are fluid dynamics extensions to our work on parasitoid wasp dispersal and modeling network dispersal within the context of social traits and/or demographics. Results (such as Fig. 3) comparing maximum a posteriori model results to field data make a clear case that while a spatially homogenous wind approximation may be sufficient to model parasitoid dispersal on local scales, it is insufficient to reproduce observed results at distances over 1 km away from the release point, at least in directions perpendicular to the average wind direction. Using techniques from computational fluid dynamics [1], we propose to extend the model by considering nonlinear and spatially inhomogenous wind patterns and variable flight time. We also have plans to extend the model to marine environments and to examine parasitoid flight characteristics in the lab using a wind tunnel.

Network dynamics is an area of great interest to me, and one that I see myself devoting more time to in my future work. In particular, I have recently started a project that uses a novel method for generating network models to mechanistically examine the spread of ideas within a social community. There are also plans to consider in more detail the network spread of a pathogen in an epidemiological context, based on my work on the spread of cheatgrass [15]. Finally, I continue to work with my Ph.D. advisor Patrick Shipman on various projects that utilize tools from mathematical ecology to numerically explore biological systems known to produce patterns [11, 10].

References

- [1] BATTISTA, N., STRICKLAND, C., AND MILLER, L. A. An easy to use immersed boundary method in 2d, with multiple options for fiber-structure models with possible porosity, advection-diffusion, and/or artificial forcing. Submitted.
- [2] FURTER, J., AND GRINFELD, M. Local vs. non-local interactions in population dynamics. *J. Math. Biol.* 27 (1989), 65–80.
- [3] KELLER, R. P., LODGE, D. M., LEWIS, M. A., AND SHOGREN, J. F., Eds. *Bioeconomics of Invasive Species: Integrating Ecology, Economics, Policy, and Management*. Oxford UP, 2009.
- [4] KEPNER, J. Parallel programming with matlabmpi. In *Proceedings of the High Performance Embedded Computing (HPEC 2001) workshop* (Lexington, MA, September 2001), MIT Lincoln Laboratory.
- [5] KLÖCKNER, A., PINTO, N., LEE, Y., CATANZARO, B., IVANOV, P., AND FASIH, A. PyCUDA and PyOpenCL: A Scripting-Based Approach to GPU Run-Time Code Generation. *Parallel Computing* 38, 3 (2012), 157–174.
- [6] KRISTENSEN, N. P., DE BARRO, P. J., AND SCHELLHORN, N. A. The initial dispersal and spread of an intentional invader at three spatial scales. *PLOS ONE* 8, 5 (2013), e62407.
- [7] LIEDLOFF, A. C., AND COOK, G. D. Modelling the effects of rainfall variability and fire on tree populations in an australian tropical savanna with the flames simulation model. *Ecol. Model.* 201 (2007), 269–282.
- [8] MARUVKA, Y. E., AND SHNERB, N. M. Nonlocal competition and logistic growth: patterns, defects, and fronts. *Phys. Rev. E* 73 (2006), 1–12.
- [9] MOLLISON, D. Spatial contact models for ecological and epidemic spread. *J. R. Statist. Soc.* 39, 3 (1977), 283–326.
- [10] PEARSON, D. A., STRICKLAND, C., AND SHIPMAN, P. D. Formation of square lattices in pattern-forming systems. Submitted.
- [11] SHIPMAN, P. D., FARIA, S. H., AND STRICKLAND, C. Towards a continuous population model for natural language vowel shift. *Journal of Theoretical Biology* 332 (2013), 123–135.
- [12] SKELLAM, J. G. Random dispersal in theoretical populations. *Biom.* 38, 1 (1951), 196–218.
- [13] STRICKLAND, C., BOURBOUBA, L., BROWN, C. S., CHILDS, L. M., HO, A. M., JENKINS, E. W., KRISTENSEN, N. P., KUMAR, S., PASOUR, V., SAMSON, J. E., SHIPMAN, P. D., TAYLOR, D., WALDROP, L., ZHU, L., AND MILLER, L. A. The fluid physics of long distance dispersal and its significance to ecology and epidemiology. Under revision for resubmission.
- [14] STRICKLAND, C., DANGELMAYR, G., AND SHIPMAN, P. Modeling the presence probability of invasive plant species with nonlocal dispersal. *J. Math. Biol.* 69, 2 (2014), 267–294.
- [15] STRICKLAND, C., DANGELMAYR, G., SHIPMAN, P. D., KUMAR, S., AND STOHLGREN, T. J. Network spread of invasive species and infectious diseases. *Ecological Modelling* 309–310 (2015), 1–9.
- [16] STRICKLAND, C., KRISTENSEN, N. P., AND MILLER, L. A. Modelling parasitoid wasp dispersal from point release. In preparation.
- [17] STRICKLAND, C., PASOUR, V., AND MILLER, L. A. Low reynolds number flow near leaves, branches, and trichomes. In preparation.

Mollison's Simple Epidemic. Consider a set of nonnegative integer-valued populations $Y_i(t)$ occupying a finite or countably infinite set of cells labeled by indices $i \in \mathcal{L}$, \mathcal{L} an index set. Assume that the Y_i evolve stochastically in time t according to a contact-birth process [9] defined by the transition rates

$$\lim_{\tau \rightarrow 0} \left(\frac{1}{\tau} \Pr[Y_i \rightarrow Y_i + 1 \text{ in } (t, t + \tau)] \right) = r \bar{Y}_i \left(1 - \frac{Y_i}{K}\right), \quad (1)$$

where $r > 0$, $\bar{Y}_i = \sum_{j \in \mathcal{L}} W_{ij} Y_j$, and W_{ij} is a stochastic matrix representing the probability that an individual located in cell j gives birth to a new individual in cell i . This process is called the “simple epidemic.”

Presence Probability Model. Let $Y(x, t)$ be the population within a defined area around location x at time t , and let $u(x, t) = \Pr[Y(x, t) > 0]$. Our derived model for presence probability is

$$\frac{\partial u(x, t)}{\partial t} = rs(x)(1 - u(x, t)) \int_{\Omega} E[Y(s, t) | Y(x, t) = 0] w(x - s) ds, \quad (2)$$

where $s(x)$ is the suitability of the environment at location x for the growth of Y and E denotes the expected value. $w(x)$ is the seed dispersal kernel. We look for approximations of $E[Y | Y(x, t) > 0]$ as a function of u .

Network Invasion Model. Let $Y(x, t)$ be the infected (non-network) target population over a location space Ω at time t , which obeys a spread model $\partial Y / \partial t = F(Y, x)$. Then we attach a carrier network via

$$\begin{aligned} \frac{d\mathbf{s}}{dt} &= [H - \text{diag}(\boldsymbol{\beta}(Y))] \mathbf{s} + (\boldsymbol{\nu} \boldsymbol{\mu}^T) \mathbf{c} \\ \frac{d\mathbf{c}}{dt} &= \text{diag}(\boldsymbol{\beta}(Y)) \mathbf{s} + (G - \text{diag}(\boldsymbol{\mu})) \mathbf{c} \\ \frac{\partial Y}{\partial t} &= F(Y, x, r \mathbf{w}(x) \cdot \mathbf{c}) \\ \boldsymbol{\beta}(Y) &= \gamma \int_{\Omega} Y(x, t) \mathbf{w}(x) dx \\ \mathbf{s}(0) &= \mathbf{s}_0, \quad \mathbf{c}(0) = \mathbf{c}_0, \quad Y(x, 0) = Y_0(x), \end{aligned} \quad (3)$$

where G is a matrix of network transition rates, \mathbf{s} and \mathbf{c} are susceptible and carrying disease vectors respectively, $\boldsymbol{\beta}$ is a vector of infection rates at each network node, $\boldsymbol{\mu}$ is the rate carriers leave the network at each node (balanced by susceptibles entering), and $\boldsymbol{\nu}$ is the probability that new susceptibles enter at each node (sums to 1). $H = G - \text{diag}(\boldsymbol{\mu}) + \boldsymbol{\nu} \boldsymbol{\mu}^T$.

Parasitoid Wasp Dispersal. We assume that the probability density $p(\mathbf{x}, t)$ describing the position \mathbf{x} of a wasp at time t (in hours) which begins at the origin and takes a wind-based flight from time t_0 to time $t_0 + \Delta t$ obeys a special form of the Fokker-Planck equation

$$\left. \frac{\partial p_w(\mathbf{x}, t)}{\partial t} \right|_{t=t_0}^{t=t_0+\Delta t} = - \sum_{i=1}^2 \frac{\partial}{\partial x_i} [\mu_i(\mathbf{w}(t)) p_w(\mathbf{x}, t)] + \frac{1}{2} \sum_{i=1}^2 \sum_{j=1}^2 \frac{\partial^2}{\partial x_i \partial x_j} [D_{ij} p_w(\mathbf{x}, t)], \quad (4)$$

where $\mathbf{w}(t)$ is the wind velocity at time t , assumed to be spatially homogenous, $\boldsymbol{\mu}$ relates wind velocity to flight drift, and D is a diffusion tensor. The position density for a wasp at the end of a full day is a weighted average between wind-based flight and local flight based on conditions measured by $h(t_0, \mathbf{w}(t_0))$ at each t_0 ,

$$\begin{aligned} p(\mathbf{x}) &= \int_0^{24} h(t_0, \mathbf{w}(t_0)) dt_0 p_w(\mathbf{x}, t) + \left(1 - \int_0^{24} h(t_0, \mathbf{w}(t_0)) dt_0 \right) p_l(\mathbf{x}, t) \\ \text{where } \frac{\partial p_l(\mathbf{x}, t)}{\partial t} &= \frac{1}{2} \sum_{i=1}^2 \sum_{j=1}^2 \frac{\partial^2}{\partial x_i \partial x_j} [\tilde{D}_{ij} p(\mathbf{x}, t)]. \end{aligned} \quad (5)$$

Savanna Model.

$$\begin{aligned}
\frac{d\Gamma}{dt} &= f(t)(1 - \mathcal{H}(\Gamma - V_\Gamma)) - \delta\mathcal{H}(\Gamma - F_\Gamma)(1 - \mathcal{H}(R - V_R)) - \epsilon\mathcal{H}(\Gamma - E_\Gamma) - (\gamma G + \boldsymbol{\omega} \cdot \mathbf{T})\mathcal{H}(\Gamma) \\
\frac{dR}{dt} &= \delta\mathcal{H}(\Gamma - F_\Gamma)(1 - \mathcal{H}(R - V_R)) - \boldsymbol{\omega} \cdot \mathbf{T}_m(1 - \mathcal{H}(\Gamma))\mathcal{H}(R) \\
\frac{dG}{dt} &= g\mathcal{H}(\Gamma) - cG(1 - \mathcal{H}(\Gamma)) \\
T_{i,m+1} &= \begin{cases} T_{i-1,m}(1 - \mu_{i-1})^{\Delta_m}(1 - \nu_{i-1}) & \text{if } i > \sigma \\ T_{i-1,m}(1 - \mu_{i-1})^{\Theta_m}(1 - \nu_{i-1}) & \text{if } 0 < i \leq \sigma \\ s(m) & \text{if } i = 0 \end{cases}
\end{aligned} \tag{6}$$

where:

- Γ = topsoil water with F_Γ = field capacity, E_Γ = evaporation depth, V_Γ = saturation capacity
- R = subsoil water with V_R = subsoil capacity, V_S = seedling rooting depth
- G = grass biomass in tonnes with g = grass growth rate, c = grass death rate
- \mathbf{T} = finite vector of trees belonging to uniform basal area classes, $f(t)$ = rainfall
- $\gamma, \boldsymbol{\omega}, \delta, \epsilon$ = water usage and rate constants
- $\mathcal{H}(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases}$
- m is an index of monthly time-steps, Δ_m = days in month m that $\Gamma = R = 0$, Θ_m = days in month m that $\Gamma = 0$ and $R < V_R - V_S$
- μ, ν = death rates due to drought and other causes, respectively
- σ = largest “seedling” size class, $s(m)$ seedling recruitment in month m