A Scalable Algorithm for Dispersing Population *

Sathish Govindarajan † Mike Dietze ‡ Pankaj K. Agarwal §
James S. Clark ¶

Abstract

Models of forest ecosystems are needed to understand how climate and land-use change can impact biodiversity. In this paper we describe an ecological dispersal model developed for the specific case of predicting seed dispersal by trees on a landscape for use in a forest simulation model. Unfortunately, performing realistic forest simulations of such models has proven to be computationally infeasible. We present efficient algorithms for computing seed dispersal. These algorithms allow us to simulate large landscapes for long periods of time.

* ADD GRANT DETAILS HERE.

† Department of Computer Science, Box 90129, Duke University, Durham, NC 27708, USA. E-mail: gsat@cs.duke.edu

‡ Department of Biology, Duke University, Durham, NC 27708, USA. E-mail: mcd7@duke.edu

§ Department of Computer Science, Box 90129, Duke University, Durham, NC 27708, USA. E-mail: pankaj@cs.duke.edu

¶ Department of Biology, Duke University, Durham, NC 27708, USA. E-mail: jimclark@duke.edu

1 Introduction

1.1 Motivation

Ecologists have long been interested in how organisms move about their environment. While the movement of motile animals is quite conspicuous, the dispersal of sessile organisms, such as by seed in vascular plants, is no less important ecologically. Recent ecological theory has highlighted the potential importance of dispersal in a number of contexts, such as maintaining biodiversity and predicting how species will respond to climate change and habitat fragmentation. However, predicting dispersal can be challenging. On one hand, dispersal is often a very stochastic process, with a high degree of spatial and temporal variability. On the other, dispersal is often dominated by spatially small-scale interactions but can be very sensitive to rare long-distance dispersal events (citation). In addition, predicting dispersal can be computationally challenging. Consider the common case where you need to produce a map of where organisms are dispersing to. This involves calculating the contribution by every reproductive individual on a landscape to seed rain at the every location on the landscape in all pair-wise combinations.

In this paper we will present an ecological dispersal model developed for the specifi c case of predicting seed dispersal by trees on a landscape for use in a forest simulation model (citation-SCG paper). However, it is important to keep in mind that the principles developed here remain general to many dispersal problems faced by ecologists.

An increased recognition of the impacts of dispersal on ecological processes has occured as ecologists have shifted from early mean-fi eld population models to more complicated spatial models. The use of spatial models in ecology has increased rapidly as we have come to recognize a number of ways in which spatial heterogeneity, both intrinsic in the landscape and generated by ecological dynamics, can strongly affect ecological dynamics [3].

Spatial and temporal variability in dispersal arises from a number of different sources. Considering the case of seed dispersal by trees, one must consider the densities of different species of tree, the heterogeneous arrangement of trees within a species, the variability among trees in seed production, and the stochastic nature of the dispersal process itself. Spatial and temporal variability in seed dispersal can have a number of interesting ecological consequences. For example, some species may buffer themselves against variability, referred to as the ”storage effect” (Chesson cite), while other species may exploit transient climatic conditions conducive to establishment or rare dispersal to good habits. Through periodic variation in dispersal, a phenomenon known as masting in trees, it is though that trees can temporarily satiate their predators, resulting in greater survival than would occur from

constant seed production.

Seed dispersal is critical for the generation of ecological pattern. Dispersal is the only means by which trees are able to move on the landscape, however most seed drops very close to the parent tree [11, 4] and thus tends to promote spatial aggregation. However, the small fraction that disperses long distance is essential for species migration in the face of global change [5] (Clark et al 2001b, 1998). In addition, recent theory [9, 10] (Hasting 1980, Tilman 1994, Hurtt and Pacala 1995) suggests that species composition may be determined largely by a trade-off between long-distance dispersal ability and competitive ability.

To be useful to scientists and managers, forest models must be sufficiently detailed to capture processes that affect the establishment of trees and yet sufficiently broad to admit landscape and atmospheric processes. This is particularly true for dispersal, where we believe that the fine scale temporal and spatial variability in seed rain is important to understanding key ecological processes like the maintenance of biodiversity and migration. However, the important effects of dispersal that we want to predict unfold over large spatial scales and over a long time.

Related work Within the context of forest models, concern about dispersal ability and seed availability is a relatively recent phenomenon. Early forest models assumed that there was a constant influx of seed of each species irregardless of what species are actually present in a stand or the surrounding area (CITE). In this way species never became locally extinct in a model because there was always new seedlings even if there were no adults to produce them. Later models accounted for the fact that the amount and species of seed arriving in a stand was dependent upon the reproductive adult trees present, but did so ignoring their spatial arrangement or dispersal ability. Clearly, neither type of model is suitable to answer questions about the effects of spatial or temporal heterogenity in dispersal, the impact of long-distance dispersal, or the ability of forests to shift spatially in response to climate change or anthropogenic disturbance. The SORTIE forest model (Pacala et al 1993, 1996, Rees?) was the first forest model to include dispersal explicitly. The algorithm used by the SORTIE model involves drawing for each propagule a dispersal distance and direction from a statistical distribution. If there are n individuals each producing an average of p propagules, this requires $O(np)$ calculations. If the number of propagules is small then this approach can be much faster than the alternative of computing the pair-wise interactions between n individuals to all locations on a landscape of area A, which requires $O(nA)$ calculations. The SORTIE model keeps p small by looking at the "dispersal" of established saplings, for which there are only a few per tree, rather than the total available seed, which can be many orders of magnitude greater.

Outside the context of forest models, most theoretical models treat dispersal as a global process, a lattice "neighboor" process, or consider all pair-wise dispersal events between a small number of "patches". In global dispersal models, propagules all enter a common pool which is then distributed amoung all locations in the model (e.g. Tilman 1994). In lattice models and cellular automata, dispersal is usually treated as just occurring between a cell and it's local neighboorhood, which is usually defined as the 4 or 8 adjacent cells on a square lattice or the 6 adjacent cells on a hex lattice e.g. (Tilman 1997?). In patch and metapopulation models, either all pair-wise dispersal events are considered, but the number of patches is generally small and doesn't present a computational challenge, or dispersal is assumed to be global (e.g Hanski). All of this is not to say that dispersal is simpler outside of forest models, but rather than dispersal can be just as complicated but is still treated in a very simplified manner in other ecological models.

Our approach. The dispersal process is computationally intensive – it requires quadratic number of calculations. Moreover, our forest model allows for many sources of variability and uncertainty that characterize forest processes and parameter data. For example, inherent stochasticity in dispersal calculation is quite high and can be quantified. This means that we can approximate the dispersal calculations, wherein the accuracy of the calculation is fixed based on the quantified stochasticity. Detailed description of statistical computation is presented in Clark *e*t al. [6, 7]. In general, knowledge of uncertainty and error associated with parameterization guides the development of efficient algorithms.

We exploit spatial coherence to design efficient algorithms for dispersal calculations. We use a hierarchical data structure to represent the forest at various spatial scales. Using the multi-resolution nature of the quad tree [8, 12], we make spatial approximations, depending on the required accuracy. To compute dispersal, we use the Monopole approximation [1] to aggregate seed dispersal from distant trees. This yields an efficiency-accuracy tradeoff scheme to compute dispersal.

Our results. We have developed efficient approximation algorithms for computing dispersal. For reasonable error, our approximation algorithm acheives a speedup of an order of magnitude. We have performed a series of experiments to quantify the inherent variability in the dispersal process. We have also performed a series of experiemnts to evaluate the dependence of approximation error and running time on approximation parameter, species parameters, number and distribution of trees in the forest. Based on these experiments, we determine the appropriate approximation for our forest simulations.

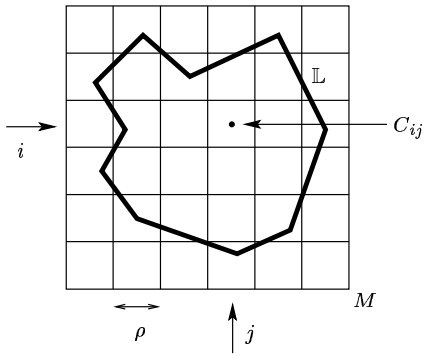

Figure 1: Landscape $\mathbb{L}$ and the underlying mesh M.

2 Overview of Our Model

We first give a brief overview of our forest model. The forest consists of a landscape $\mathbb{L}$ and a population of trees, which is modeled as family of densities and a family of individuals. The landscape remains fixed over time but the population evolves with time.

Landscape. Our model considers the landscape $\mathbb{L}$ of the forest as a planar region whose boundary is a closed polygonal curve. The area of the landscape varies from a few hectares to few hundreds of hectares. We discretize $\mathbb{L}$ by enclosing it with a square and laying a uniform grid(mesh) M on it. Each grid cell M_{ij} of M is a square of size ρ; we refer to ρ as the resolution of M. We use C_{ij} to denote the center of M_{ij}, and we associate an elavation (height) z_{ij} with C_{ij}. By interpolating the heights at other points $\mathbb{L}$, we can view $\mathbb{L}$ as a terrain. We could also associate various geological and urban features such as rivers, lakes, roads, etc. with $\mathbb{L}$. Figure 1 shows an example of a landscape alongwith the underlying mesh.

Population. Our model is hybrid in the sense that we use both densities and individuals to model the population of the forest. The early stages of trees are modelled as densities, and after some growth, they are modelled as individulas with unique physical attributes. More precisely, we classify the population into five stages: seed, yearling, seedling, sapling and adult (Figure 2). We further refine seeds into seed rains and seed banks - the former representing the seeds that are dispersed by trees and the latter representing the ones that are on the ground. The seeds that have germinated are called yearlings. We model seed rain, seed bank

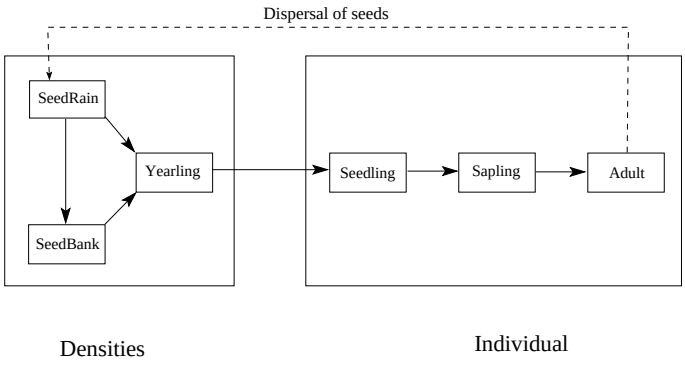

Figure 2: Evolution of densities of stages seed and yearling and growth of an individual from a seedling to an adult.

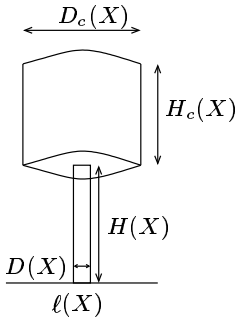

Figure 3: Geometric model of an individual.

and yearling as densities, as they do not have any geometric attributes and all of them with the same species are identical. We assume that the density is uniform within each grid cell.

We model the next three stages — seedling, sapling, adult — as families of individuals. Each individual X has a physical location $\ell(X) \in \mathbb{R}^2$ and various physical attributes. Currently, we model each individual as a cylindrical trunk and a cylindrical crown sitting on top of the trunk; see Figure 3. Let $D(X), H(X)$ (resp. $D_c(X), H_c(X)$) denote the diameter and height of the trunk (resp. crown) of X.

We have two threshold parameters τ_D and τ_P. An individual X is a seedling if $H(X) \leq \tau_D$, a sapling if $\tau_D \leq H(X) \leq \tau_P$, and an adult if $H(X) \geq \tau_P$.

Dynamics. The dynamics of our forest model consists of three parts - establishment of individuals, its growth and its mortality. Individuals are established by dispersal of seeds. The adult trees produce seeds depending on its DBH(diameter at breast height) and these seeds are dispersed based on a dispersal kernel. The dispersal kernel accounts for both short and long distance dispersal. Growth of each of the stages is calculated based on resource availability and local density. Individuals are promoted from one stage to next based on the growth thresholds. Mortality of individuals is done by calculating a mortality probability based on the growth suppression and natural disturbances.

Resources. The forest contains several resources like light, moisture, nitrogen, etc which are vital for the growth of the individuals in the forest. We model each resource as a separate submodel. Light is considered as one of the main resources in our model. We develop a sophisticated light model based on Casetti's [2] light model to calculate the availability of understory light at each gridcell.

3 Dispersal Model

The dispersal model determines the number of seeds that disperse into each gridcell of M. This quantity depends on:

- the number of seeds produced by each tree, denoted as *fecundity*
- spatial distribution of seeds, which is defi ned by the dispersal kernel.

Fecundity. The *fecundity*, $\beta_t(k)$, of an individual k at time t is based on the idea of decomposing the variability in reproductive output appropriately between species-level effects and individual-level effects that are constant, temporally varying, and size-dependent. It is calculated as follows:

$$\beta_t(k) = \chi(k) \cdot \Delta(k) \cdot 10^{a_0 + b(k) + \epsilon_t(k)} \cdot (D_t(k))^{a_1} \tag{1}$$

where

$$\chi(k) = \begin{cases} 1 & \text{if individual } k \text{ is female,} \\ 0 & \text{if individual } k \text{ is male.} \end{cases}$$

$$\Delta(k) = \begin{cases} 1 & \text{if } D_t(k) > \gamma(k), \\ 0 & \text{if } D_t(k) \leq \gamma(k). \end{cases}$$

$$\gamma(k) \sim \Gamma(m_0, m_1)$$

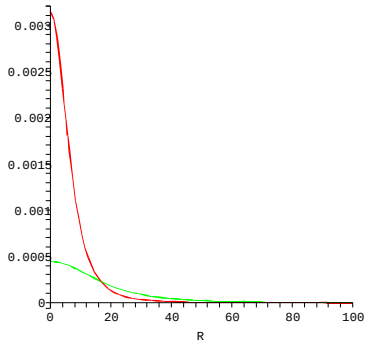

Figure 4: Dispersal kernel for parameters of species *Acer rubrum* p=1; u=101.3(steep curve) and species *Liriodendron tulipifera* p=1; u=719.8.

where a_0 and a_1 are species-specific scaling parameters, $b(k)$ is an individual-specific scaling parameter, $D_t(k)$ is the diameter of the trunk of individual k at time t, $\Gamma(m_0, m_1)$ is a Gamma distribution with species-specific maturity parameters m_0, m_1 and $\epsilon_t(k)$ is a temporally autocorrelated Gaussian stochastic process, defined as:

$$\epsilon_t(k) \sim N(\nu \cdot \epsilon_{t-1}(k), \sigma^2) \tag{2}$$

where ν and σ are species-specific parameters.

The dispersal kernel describes the spatial distribution of the scattering of seeds in the vicinity of the parent plant, as a function of distance r. We use a bivariate Student's t-distribution for the dispersal kernel, which has the following form:

$$f(r; u, p) = \frac{p}{\pi u \left[1 + \frac{r^2}{u}\right]^{p+1}},$$

where p and u are species specific parameters. Figure 4 shows the graph of the dispersal kernel for the parameters of two species: *Acer rubrum* and *Liriodendron tulipifera* (Clark et al. [7]).

The actual number of seeds dispersed into the grid cell M_{ij}, dnoted as s_{ijt} is drawn from a Poisson distribution

$$s_{ijt} \sim \Pi(q_t(C_{ij}) \cdot \rho^2),$$

where ρ is the side length of a gridcell M_{ij} and $q_t(x)$ is the expected seed density in location x at time t,

$$q_t(x) = \sum_k \beta_t(k) \cdot f(\|x - l(k)\|; u, p).$$

where the sum is taken over all the trees in the forest.

The functional form of the dispersal kernel was chosen from a number of potential functions using formal statistical techniques for model comparison (Clark *et* al [4]). Model parameterization is based on extensive fi eld data (Clark et al. [7]).

4 Computing Dispersal

We developed a forest simulator based on relationships outlined in Section 2. Submodels include dispersal, light, germination, mortality, disturbance, and climate. The simulator takes as input an initial confi guration of the forest and landscape. It simulates dynamics at annual time steps. Figure 5 shows the flowchart of operations.

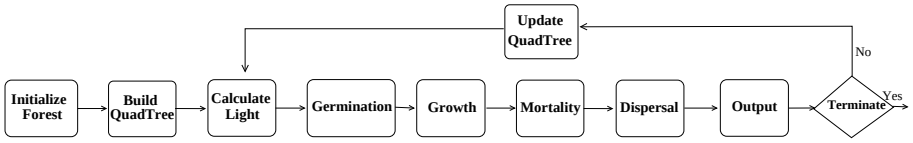

Figure 5: Flow chart of the sequence of operations performed by the simulator

Dispersal and light calculations are computationally intensive. Moreover ecological experiments require large landscapes (at least 1 sq. km.) and up to several thousand years. Using exact calculations, these simulations would take months. Our algorithms included a quad-tree like data structure that is hierarchical. Using the multi-resolution quad tree, we calculate dispersal using approximations at spatial resolutions that depend on the required accuracy. We describe the quad tree data structure below:

Quad tree data structure. We developed a 2D quad tree T on the coordinates of all the individuals(adults, saplings) in the forest. Let P denote the set of points representing the locations of the trees. All spatial processes (esp. dispersal) operate between and not within cells. Thus, we stop the recursion on the quad-tree when

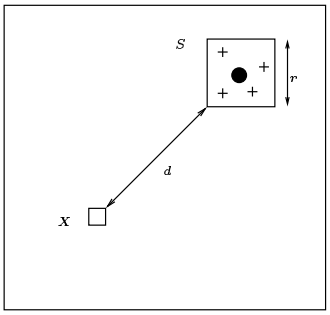

Figure 6: Monopole approximation for dispersal. Trees, represented by crosses, disperse seed to destination x.

the granularity of the region reaches the size of a grid-cell. This property of our model allows us to obtain a bound on the depth of T. In a general setting, the depth of the quad-tree depends on the distance of the closest pair of points in P, because recursion continues until we separate the closest pair.

For any node v in T, let R_v denote the region represented by v and P_v represent the set of all the trees contained in R_v. In each interior node of the quad-tree T, we store summary statistics for trees contained in the node. The summaries consist of number, total fecundity, basal area, etc. of trees $t \in P_v$. These summaries are used to develop approximation schemes for dispersal calculations.

Exact calculation of dispersal requires $O(nA)$ time, because each grid-cell receives seeds from n trees. However, when trees are far away from the destination grid cell, their individual distances from x are roughly the same(Figure 6). Therefore, we can replace the trees by a single large tree (represented by shaded circle) of appropriate fecundity to obtain an approximate solution.

Approximation scheme. We formally describe the algorithm as follows: Let r be the side-length of square S and d be the distance of S from X, as shown in Figure 6. We can perform the monopole approximation if $r/d < \mu$, where μ is the monopole coeffi cient. The approximation algorithm is a top-down traversal of the quad-tree starting at the root. The procedure at any node v is shown in Figure 1.

The quad-tree provides a multi-resolution mechanism to monopole approximate regions of varying dimensions depending on distance from X. Moreover, since the total fecundity of trees in P_v is stored at v, the monopole approximation can be computed in constant time.

We now analyze the running time of our algorithm. Let C be a circle of radius $1/\mu$, centered at the grid-cell X and let c_X is the number of trees that lie inside

Algorithm 1 `Monopole Approximation`(node v, cell X)

$r =$ side-length of square R_v
$d =$ distance between R_v and X
if $(r/d \leq \mu)$ **then**
  replace trees in P_v by a tree t of appropriate fecundity
  calculate dispersal in X due to tree t
else
  for each child w of v **do**
    `Monopole Approximation`(w, X)

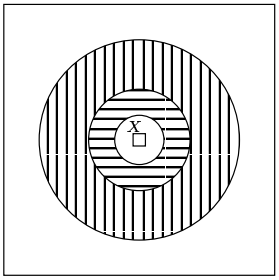

Figure 7: Concentric circles C_0, C_1, C_2 around gridcell X

circle C. The dispersal due to all plants in C is calculated by going over each tree individually. This is because $d < 1/\mu$ and thus $r/d > \mu$. The complexity of the dispersal calculation is $O(c_X)$.

Let $C_i, 1 \leq i \leq \log A$ be an annular ring, centered at the center of X and having inner radius $2^i/\mu$ and outer radius $2^{i+1}/\mu$. Figure 7 shows concentric rings $C = C_0, C_1, C_2$ The monopole condition is satisfi ed for all quad-tree regions R_v with side length 2^i, that are contained in C_i. Using a packing argument, the number of such regions R_v is at most $3/\mu^2$.

For any gridcell X, the forest with A gridcells is covered by atmost $\log A$ annular rings C_i. The number of monopole approximations performed in any such C_i is $3/\mu^2$. Thus the total complexity of the Monopole procedure is $O(\log A/\mu^2)$. Thus the total cost is $O(\log A/\mu^2 + c_X)$

Performing the Monopole procedure for each gridcell X, the time to compute dispersal for the entire forest is $O(A \log A + \sum_X c_X)$. Note that $\sum_X c_X$ denotes the total number of dispersal calculations done without monopole approximation. Dispersal is calculated for each tree in the forest located in grid cell X for all grid cells that are distance $1/\mu$. Thus each tree contributes to $1/\mu^2$ dispersal calcula-

tions. The total number of individual dispersal calculations is thus n/μ^2. The total time to calculate dispersal is thus $O((A \log A + n)/\mu^2)$.

Experimental results. We have performed a set of experiments to evaluate the computational performance of the approximation algorithm as a function of the area of the forest and the monopole error threshold. All our experiments are performed on a 2.2 GHz Intel PC with 4 GB memory, nVidia Quadro4 XGL 900 graphic card running Linux OS.

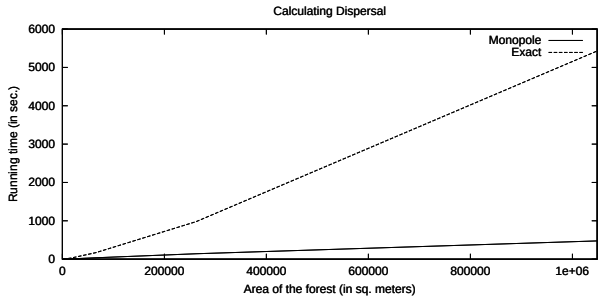

Figure 8: Running time of dispersal algorithm with monopole 0.1 for varying forest area sizes

We performed experiments on a forest initialized with the output of a 100 year simulation involving a single species. In our experiment, we varied the side length of the forest from 32 meters to 1024 meters. Figure 8 compares the running time of the exact algorithm with the monopole algorithm (for monopole threshold 0.1). The exact algorithm is the monopole algorithm with monopole factor set to 0. For a 1024×1024 sq. m forest, the monopole acheievs speedup of two orders of magnitude. Figure 9 plots the RMS error of seeds dispersed for monopole factor 0.1. Note that the error in seeds dispersed is less than 2% for the landscapes simulated.

Finally, we performed an experiemnt to evaluate the effect of monopole factor(μ) on the error incurred. We initialized the forest with trees of species Acer rubium from Duke forest site. Figure 10 plots the RMS error of seeds dispersed for monopole factors from 0 to 100. The RMS error increases rapidly for $\mu \leq 1$ and does not change for $\mu \geq 20$. The shape of the dispersal kernel for *Acer rubium*(see figure 4) explains this behaviour. Since the dispersal kernel is almost a constant after a 20 meter radius, any amount of monopole approximation involving distances above 20 meters does not affect the RMS error.

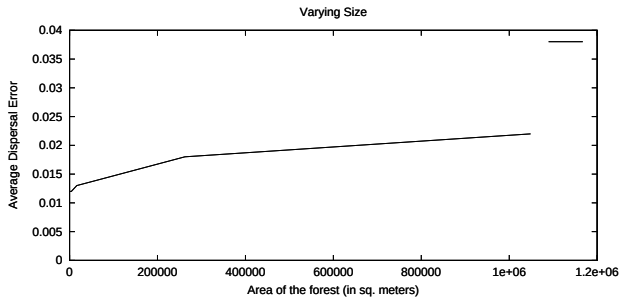

Figure 9: Relative error of dispersal algorithm with monopole 0.1 for varying forest area sizes

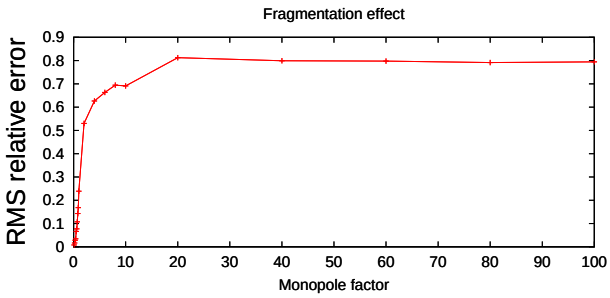

Figure 10: Relative error of dispersal algorithm for varying monopole factor. The forest is initialized with trees of species Acer rubium from Duke forest site.

Species	σ^2	ρ	τ^2	u	α_0	α_1	a_θ	b_θ
ACru	1.18	0.036	0.024	62	2.77	0.406	1.15	0.05
CAca	1.36	-0.025	0.020	50.5	1.53	0.923	16.11	2.12
CEca	0.37	0.272	0.040	163.9	1.62	0.757	2.84	0.7
FRam	1.70	0.105	0.015	34.7	2.28	0.425	3.10	0.14
LIst	0.57	-0.790	0.001	518	2.31	0.527	5.30	0.15
Litu	0.55	0.371	0.005	719.8	3.37	0.577	3.09	0.12
NYsy	1.16	-0.001	0.434	27.7	0.60	0.329	3.07	0.06
PIta	0.72	-0.349	0.145	1706.1	2.06	0.700	1.65	0.03
ULsp	0.31	0.189	0.315	391.2	2.27	0.888	6.25	0.42

Table 1: Species dispersal parameters derived from field experiments (Clark et al 2004)

5 Inherent variability

To test the ecological behavior of the dispersal model, we ran the simulation on the map of a real forest using the parameters derived empirically. The forest stand is located in the Blackwood Division of the Duke Forest in Chapel Hill, NC. Over an area covering 4 ha every tree over 2m tall was identified to species, mapped, and its diameter was measured at 1.45m high, a common metric in forestry and ecology referred to as Diameter Breast Height (DBH). In total there were 52 species observed in this stand, but for this experiment we will focus on nine species: *A*cer rubrum (ACru), *Carpinus caroliniana* (CAca), *Cercis canadensis* (CEca), *Fraxinus americana* (FRam), *Liquidambar styraciflua* (LIst), *Liriodendron tulipifera* (LItu), *Nyssa sylvatica* (NYsy), *Pinus taeda* (PIta), *Ulmus* species (ULsp). In the case of *Ulmus*, we have chosen to lump three species (*Ulmus alata*, *Ulmus americana*, *Ulmus rubrum*) which are ecologically very similar.

We focussed our experiment on species for which we have good parameter estimates and which show a number of contrasting ecological patterns. The criteria of having good parameter estimates results in common species being overrepresented in this sample. Table 1 lists the parameter values used for each of these species.

Dispersal calculations were conducted 1000 times for each species, using the same tree map in all cases, and allowing for temporal autocorrelation as defined in equation 1. Simulations were run on a 512x512m landscape at a 1x1m resolution with a monopole threshold of 0.125 and with the mapped forest approximately centered in the landscape. From the 1000 replicate dispersal maps, we calculated the mean and variance of seed rain for each species at each grid cell. ArcGIS (ESRI citation) was then used to visualize the data(Figure 13), generate a coefficient of variation map for each species(Figure 14), and to sample the mean and CV maps

at 124 known points Figures 11 and 12). These sample points, which coincide with the seed rain sampling points used in fi eld studies, were then used to compare dispersal ability among species and estimate what an acceptable level of approximation error is for the dispersal algorithm. The mean dispersal maps illustrate the spatial variability in seed rain and it's relation to parent trees, which are depicted as circles with diameter equal to their mean canopy diameter (Figure 13). The CV maps illustrate the spatial pattern of the temporal variability in seed rain (Figure 14.

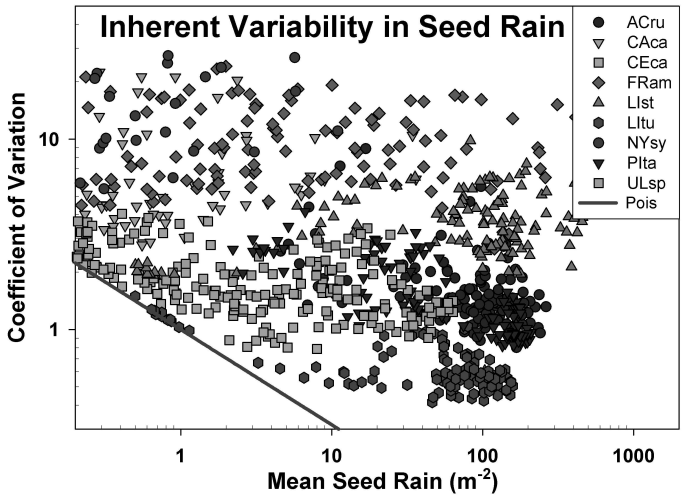

Figure 11: Plot of coeffi cient of variation against mean seed rain for 9 different species at 124 sample points in the Duke foest site.

There are a number of factors that control the spatial and temporal patterns of seed rain for each species: the abundance of parent trees, the spatial arrangement of the trees, the fecundity of the trees (α_0,α_1,a_θ, b_θ), the variability in fecundity (σ,ρ,τ), the dispersal ability of seed (u), and the stochasticity in dispersal ($Poisson(s)$). For Poisson process error the CV scales as $\lambda^{-1/2}$, so all else being equal we expect low seed density to be more variable. We plot this expectation in Figure 11 and fi nd, as we would expect, that all sample points are above this line, though at low seed rain some species defi nitely approach this limit. The variability above this line can be attributed to other factors such as variability in fecundity. Table 2 lists the mean and standard deviation of both mean dispersal and CV of dispersal, as well as the relative abundance of each species and the expected CV at the mean abundance attributable to the Poisson dispersal stochasticity.

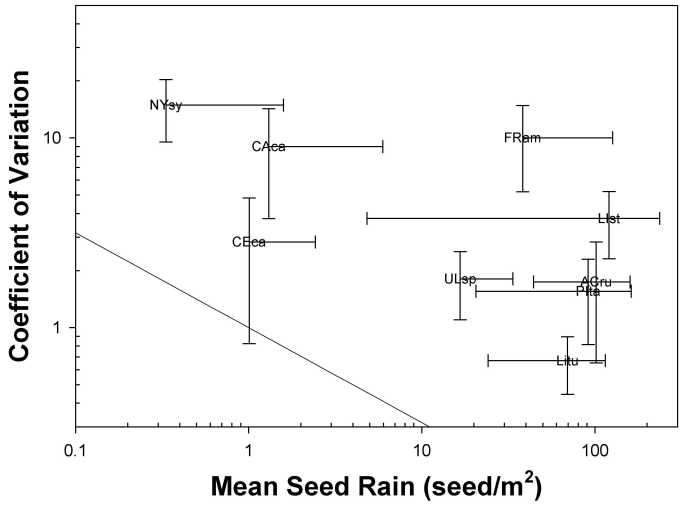

Figure 12:

Spp	$seed_{Mean}$	$seed_{sd}$	CV_{mean}	CV_{sd}	CV_{Pois}	RelAbund
ACru	101.76	57.64	1.74	1.08	0.09	0.23
CAca	1.30	4.63	9.01	5.25	0.87	0.01
CEca	1.01	1.42	2.82	2.00	0.99	0.03
FRam	38.24	88.37	10.02	4.82	0.16	0.06
LIst	120.63	115.81	3.76	1.46	0.09	0.11
Litu	69.46	45.34	0.67	0.22	0.12	0.03
NYsy	0.33	1.26	14.90	5.39	1.73	0.03
PIta	91.45	70.88	1.55	0.74	0.10	0.04
ULsp	16.67	16.87	1.81	0.71	0.24	0.08

Table 2: Summary statistics for sampled dispersal maps by species. CV_{Pois} is the expected CV due to the Poisson dispersal stochasticity alone at the mean seed rain. In all cases CV_{Pois} is substantially below the observed CV. $Rel\,Abund$ is the relative abundance of each species on the landscape based on stem density but not taking into account tree size. $Rel\,Abund$ does not sum to one because there are 43 other species in the forest which are not presented here.

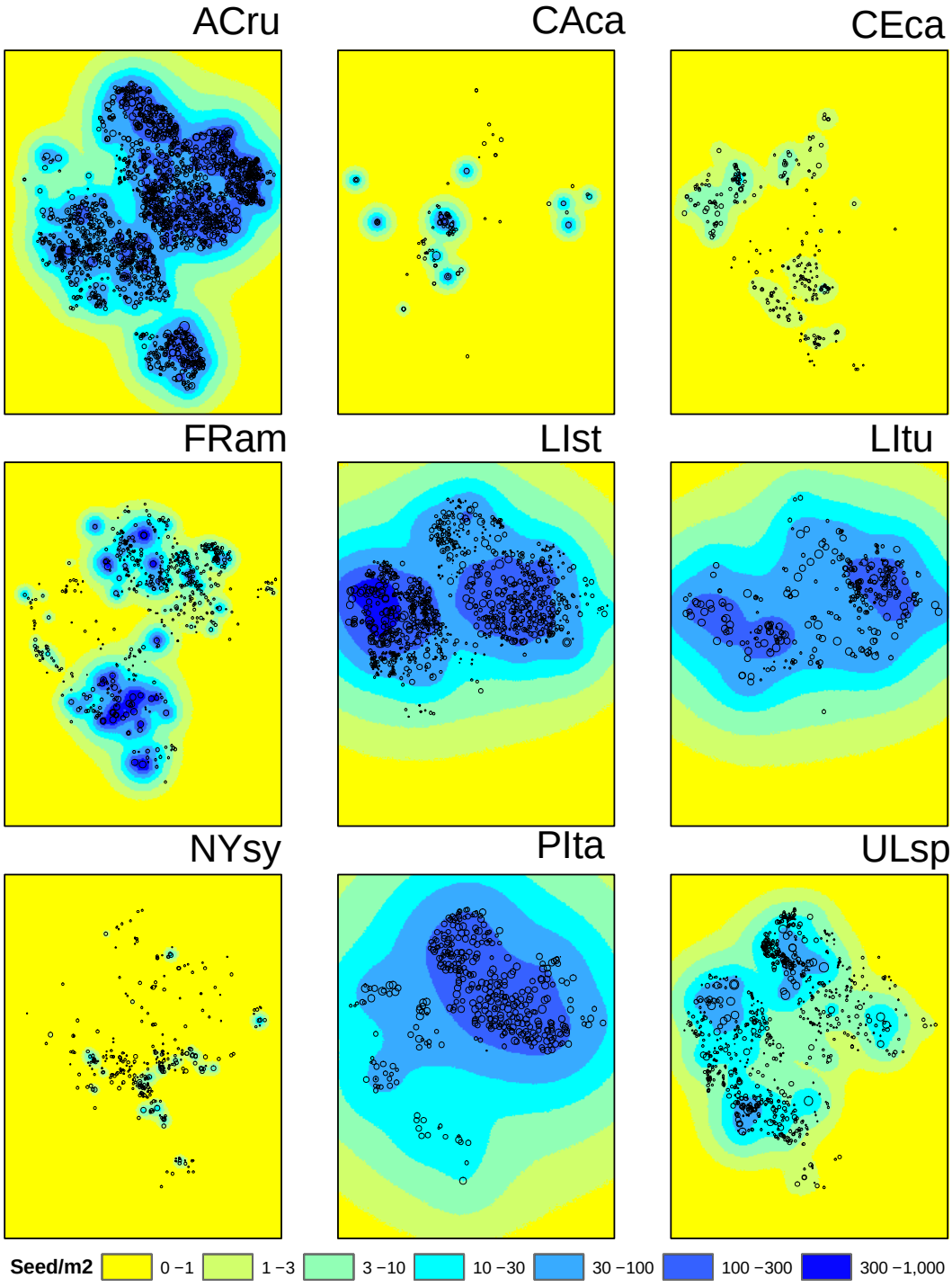

Figure 13: Spatial map of the mean number of seeds dispersed per sq. m for different species

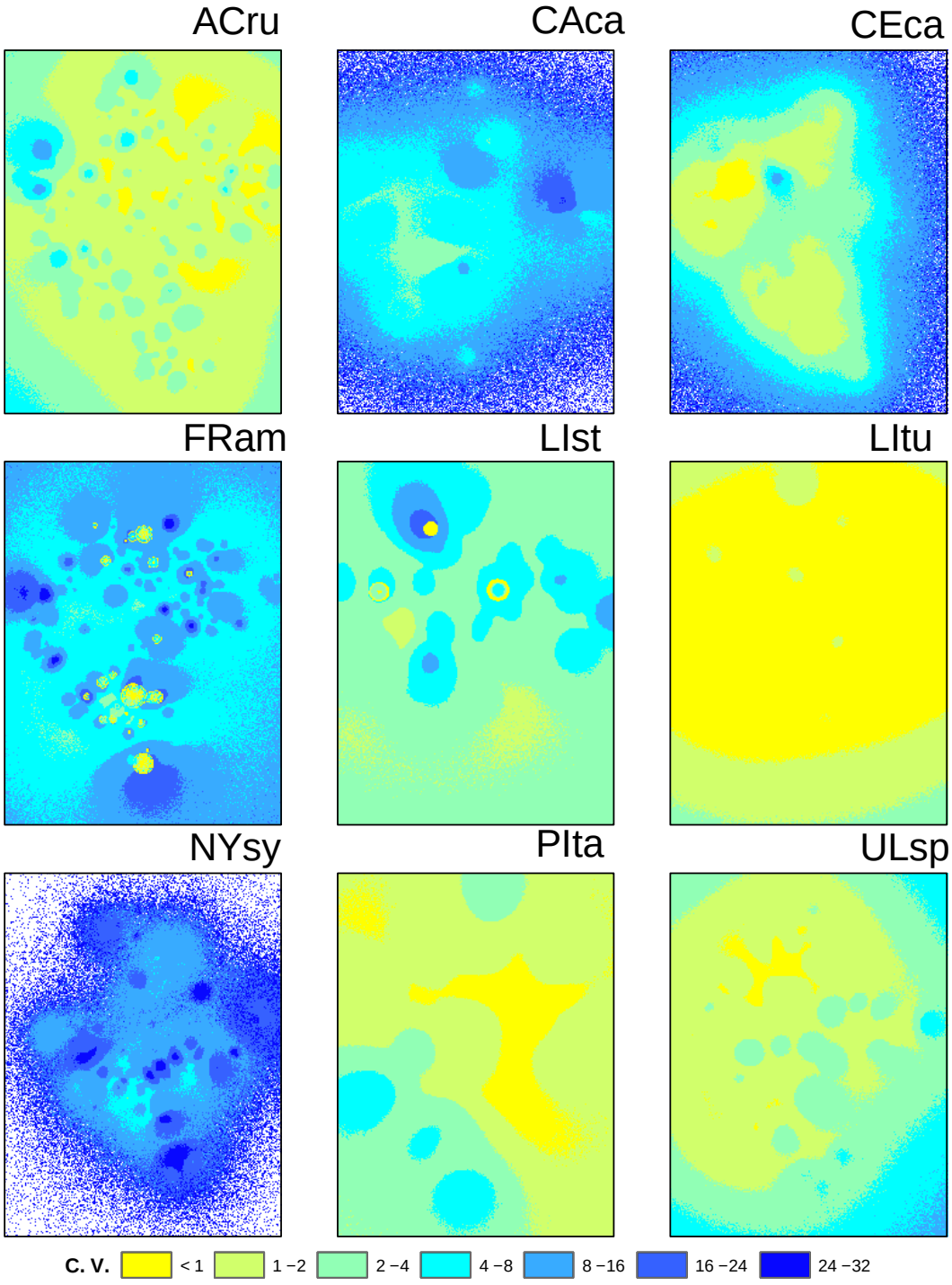

Figure 14: Spatial map of the coeffi cient of variation in the number of seeds dispersed for different species

Three of the species clearly appear to be dispersal limited (CAca, CEca, and NYsy) while the other six species appear to be reasonably abundant at most, though not necessarily all, places on the landscape (Figure 11). The seed limited species all appear to be relatively less abundant (column RelAbund in Table 2) and have short dispersal distances (column u in Table 1), but they also span a spectrum of temporal variability due largely to their different levels of variability in fecundity. Of the remaining species with relatively high mean seed rain, they also span a range of variability from LItu, which is widely dispersed, highly fecund, and has trees that are well distributed across the landscape, to FRam, which has high levels of variability in fecundity and a short dispersal distance, leading to high spatial and temporal variability in seed rain. In between are species like ACru and PIta, which have similar mean seed rain and CV, but reach this in very different ways. PIta has a very long dispersal distance but is highly aggregated to one part of the landscape, leading to very smooth dispersal contours. ACru on the other hand has a much smaller dispersal distance, but is very abundant and distributed across the landscape, which leads to a much more complicated pattern of both mean dispersal and CV (Figures 13, 14).

In setting the level of "acceptable" error it is important to understand the ecology underlying the dispersal process. For example, the lowest values for the CV we observed in our sampling were for LItu, which were around 40%. However, these low values also occur at high seed rains. At high seed density, density dependent mortality exerts a negative feedback, reducing the number of seeds available for recruitment and thus reducing the model's sensitive to variation at high seed densities. In the face of such high levels of inherent stochasticity in the process, it is tempting to ask why do we need to worry about calculating dispersal at all? Couldn't we do just as well just drawing numbers randomly? Here it is important to remember what types of inherent variability we consider important to capture. We do care about the overall spatial pattern of dispersal and it is important that we generate the correct pattern of spatial aggregation for seed rain. It is also important that we capture the temporal and spatial coherence within the temporal variability. In other words, we need to capture the bust-and-boom masting cycles exhibited by some species. The recruitment response by each species is nonlinear so actual recruitment can be very different than average seed rain times average survival. Where we have the greatest room to accomodate approximation error is in comparison to the dispersal stochasticity, in particular at low seed densities, and to the fecundity stochasticity at high seed rain that is being buffered out. Further, high seed rain most often occurs in close proximity to adult trees, which will often be calculated exactly by the monopole anyways.

6 Determining the appropriate monopole factor

We performed a set of experiments that illustrate the variation of RMS error of seeds dispersed with respect to the following factors – monopole factor, number of trees in the forest, spatial distribution of trees and species parameters. Using the data from these experiments, we can choose the appropriate monopole factor for a given simulation year, such that the RMS error incurred at that year is quantitatively similar to the inherent variability of the dispersal process.

The first experiemnt describe the variation of RMS error and running time of dispersal model with respect to monopole factor. The experiment was performed on the 9 species used in the previous section. We initialized the forest with tree data from a real forest. The side length of the forest is 512 meters and we varied the monopole thresholds from 0.1 to 1.0 As anticipated, the run time decreases with the increase in the monopole threshold and the RMS error increases with increase inthe monopole factor Figure 15 shows the running time and RMS error of seeds dispersed for the 9 different species as the monopole threshold is varied. Figure 16 shows the spatial distribution of relative error with monopole factor 0.5 for all the 9 species.

For a given monopole factor μ and species i, we quantify the tradeoff between RMS error and running time by calculating the "tradeoff index"($\Phi_i(\mu)$), which is defined as follows:

$$\Phi_i(\mu) = e^{c \cdot E_i(\mu)} \cdot \frac{T_i(\mu)}{T_i(0)} \tag{3}$$

where $E_i(\mu)$ is the RMS error and $T_i(\mu)$ is the running time of dispersal algorithm for monopole fator μ and c is a constant.

We also calculate the cumulative index by summing the index of all the 9 species. Figure 17 shows the variation of the cumulative index with monopole factor μ. It also shows the tradeoff index for species *Acer rubium* and *caca*.

The next experiemnt describe the variation of RMS error with respect to monopole factor and number of trees. The experiment was performed on *Acer rubium* species. We initialized the forest with synthetic tree data. The side length of the forest was fixed as 512 meters. We varied the monopole thresholds from 0.1 to 1.0 and the number of trees from 100 to 1000. Figure 18 shows the RMS error as a triangulated surface. As anticipated, the RMS error increases with the increase in the monopole factor and decreases with increase in the number of trees.

The next experiemnt describe the variation of RMS error and running time with respect to the percentage of forest area that contains trees(occupancy). The occupancy was varied from 5% to 100%. The experiment was performed on *Acer rubium* species. The side length of the forest was fixed as 512 meters. For a

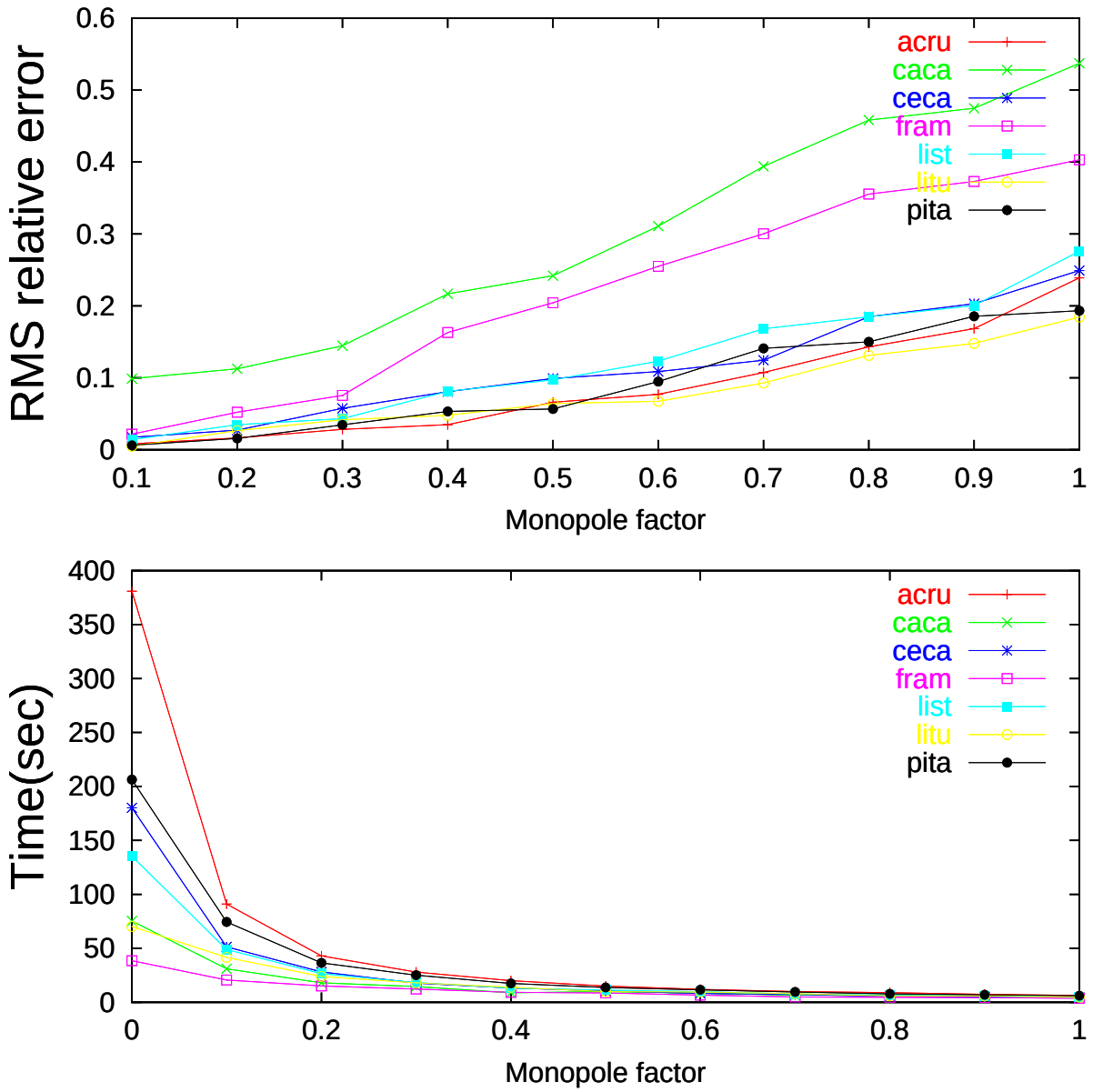

Figure 15: Tradeoff between running time and error for nine different species. The right figure shows the running time and the left figures show the RMS error incurred as the monopole factor is varied from from 0 to 1

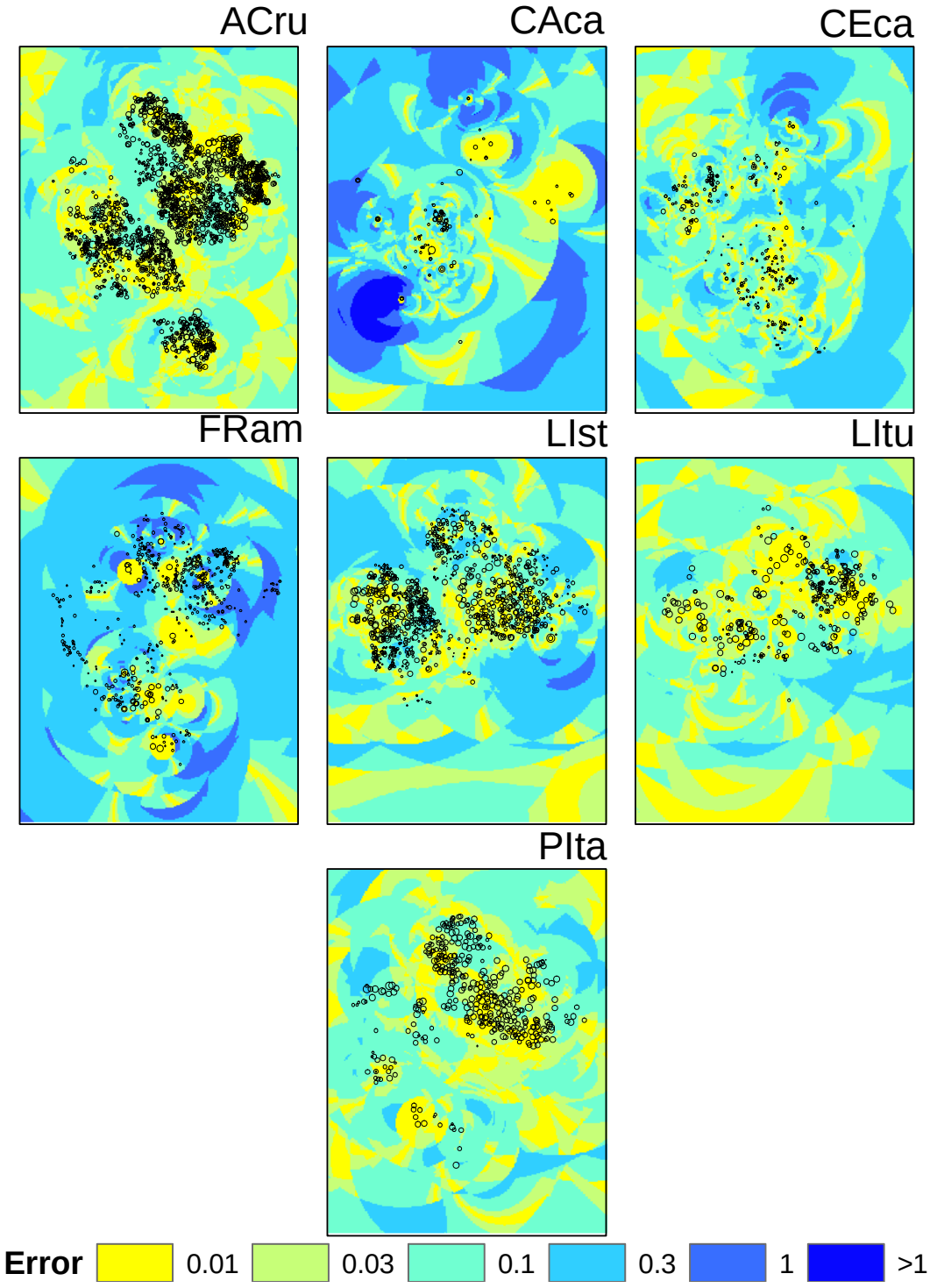

Figure 16: Spatial distribution of relative error for 9 species. Monopole factor is set to 0.5.

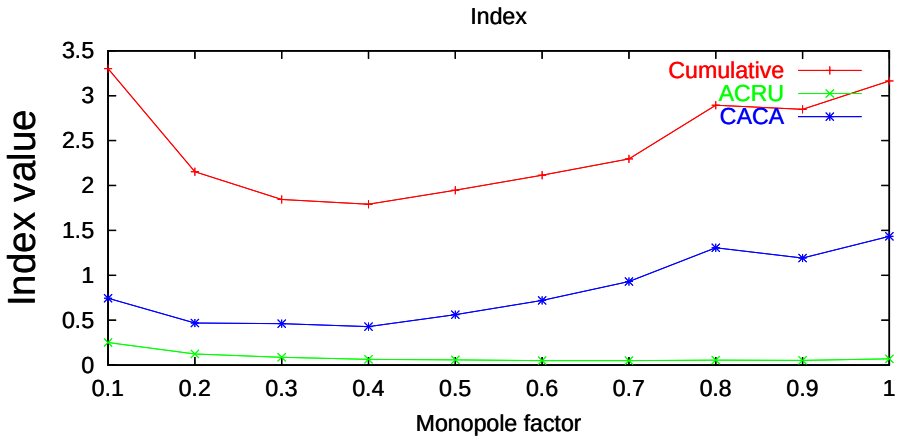

Figure 17: Plot of Φ for various monopole factors. Figure shows the cumulative index and the index for species *Acer rubium* and *caca*.

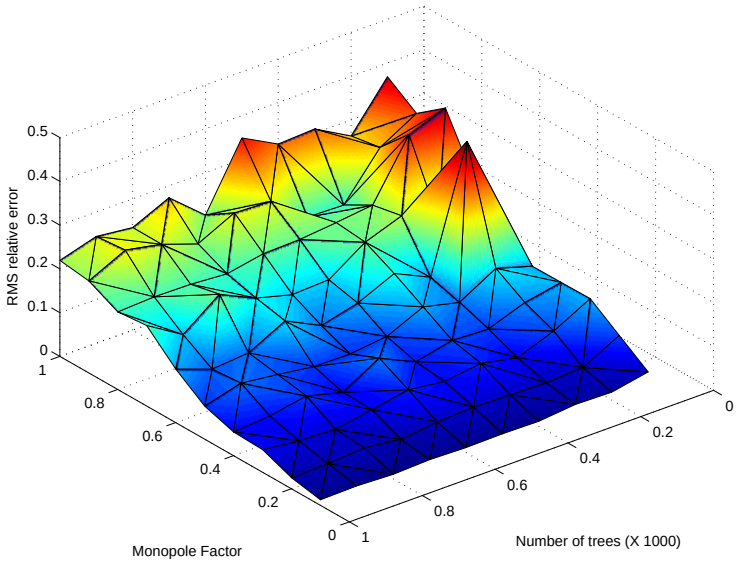

Figure 18: RMS error variation plotted as a surface. Monopole factor is varied from from 0 to 1 and number of trees is varied from 100 to 1000.

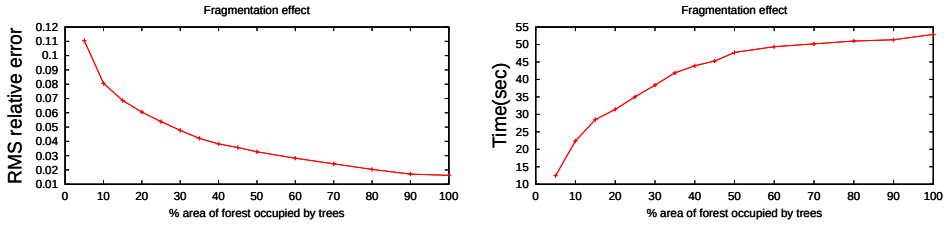

Figure 19: Variation of RMS error and running time for different occupancy values. The left fi gure plots the RMS error and the right fi gure plots the running time.

given occupancy, we obtain the initial tree distribution as follows: randomly select $16X16$ regions of appropriate number in the forest and fi ll each $16X16$ region with constant(in this experiment 20 trees) number of trees placed randoml inside the region. Figure 19 plots the RMS error as the occupancy is varied. As anticipated, the RMS error decreases and running time increases with increase in occupancy. Two factors contribute towards this behaviour: (i) the number of trees in the forest increases as occupancy increases; (ii) the distribution of trees becomes more uniform as occpancy increases.

A Appendix: Notation Index

Landscape

Mesh bounding the landscape	M
Grid cells	M_{ij}
Center of grid cell M_{ij}	C_{ij}
Resolution of each gridcell	ρ

Spatial Attributes of Individual

Position of tree T	$\ell(T)$
DBH of tree T	$D(T)$
Trunk height of tree T	$H(T)$
Crown height of tree T	$H_c(T)$
Crown diameter of tree T	$D_c(T)$
Fecundity of tree T	$\beta(T)$

Dispersal

Expected number of seeds dispersed in gridcell C_{ij}	$\bar{r}_{ij}$
Number of seeds dispersed in gridcell C_{ij}	r_{ij}
Indicator variable for sex of tree T	$\chi(T)$
Indicator variable for reproductivity of tree T	$\Delta(T)$
Dispersal Kernel - Student t-distribution with parameters p, u	$f(r; u, p)$

References

[1] J. Barnes and P. Hut, A hierarchical $O(n \log n)$ force-calculation algorithm, *Nature*, 324 (1986), 446–449.

[2] A. Cescatti, Modelling the radiative transfer in discontinuous canopies of asymetric crowns. i. model structure and algorithms, *Ecol. Modeling*, 101 (1997), 263–274.

[3] P. Chesson, General theory of competative coexistance in spatially-varying environments, *Theoretical Population Biology*, 58 (2000), 211–237.

[4] J. Clark, M. Silman, R. Kern, E. Maclin, and J. HilleRisLambers, Seed dispersal near and far: patterns across temparate and tropical forests, *Ecology*, 80 (1999), 1475–1494.

[5] J. S. Clark, R. Carpenter, M. Barber, S. Collins, A. Dobson, J. Foley, D. Lodge, M. Pascual, R. Pielke, W. Pizer, C. Pringle, W. V. Reid, K. A. Rose, O. Sall, W. H. Schlesinger, D. Wall, and D. Wear, Ecological forecasts: an emerging imperative, *Science*, 293 (2001), 657–660.

[6] J. S. Clark, M. Dietze, I. Ibanez, and J. Mohan, Coexistence: how to identify trophic tradeoffs, *Ecology*, 84 (2003), 17–31.

[7] J. S. Clark, S. LaDeau, and I. Ibanez, Fecundity of trees and the colonization-competition hypothesis, in press (2004).

[8] R. A. Finkel and J. L. Bentley, Quad trees: a data structure for retrieval on composite keys, *Acta Inform.*, 4 (1974), 1–9.

[9] A. Hastings, Disturbance, coexistance, history, and competition for space., *Theoretical Population Biology*, 18 (1980), 363–373.

[10] C. G. Hurtt and S. W. Pacala, The consequences of recruitment limitation: reconciling chance, history, and competative differences between plants., *J. theor. Biol*, (1995), 1–12.

[11] E. Ribbens, J. A. SIlander, and S. W. Pacala, Seedling recruitment in forests: calibrating models to predict patterns of tree seedling dispersal, *Ecology*, 75 (1994), 1794–1806.

[12] H. Samet, *Applications of Spatial Data Structures: Computer Graphics, Ima ge Processing, and GIS*, Addison-Wesley, Reading, MA, 1990.