

**EPCC-SS99-11**

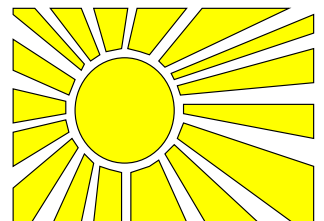
**Modelling the Impact of Genetically Modified crop  
on the Flora Associated with Agriculture**

**Wing-yun A. Yau**

**Abstract**

The Scottish Agricultural College (SAC) is currently undertaking a series of research projects commissioned by the Scottish Office to examine and model the impact of Genetically Modified (GM) oilseed rape on the weed flora at sites where GM oilseed rape is being grown on a trial basis. Since any trials conducted will be relatively short-term, and since the public require predictions of the likely long-term impact of GM crops on the environment, a computer modelling approach appears to be the only option for providing the “answers” that are required now.

Two modelling methods have been suggested for this project — a Markov model using transition matrices and a difference equation approach suggested by Cousens & Mortimer (1989). As the latter approach is closer to the natural process and more computationally complex it will be used in this project.



## Contents

<b>1</b>	<b>Acknowledgements</b>	<b>3</b>
<b>2</b>	<b>Introduction</b>	<b>4</b>
<b>3</b>	<b>Domain Decomposition — use of the MPI library</b>	<b>4</b>
<b>4</b>	<b>Simulating the Annual Agricultural Cycle</b>	<b>4</b>
4.1	Initialising plants per patch . . . . .	5
4.2	Application of herbicide . . . . .	6
4.3	Growth and competition between plants . . . . .	6
4.4	Harvesting . . . . .	6
4.5	Modelling dispersal . . . . .	8
<b>5</b>	<b>Program algorithm</b>	<b>8</b>
<b>6</b>	<b>File Input and Output</b>	<b>8</b>
<b>7</b>	<b>Simulation Case Study</b>	<b>10</b>
7.1	Weed-crop distribution . . . . .	10
7.2	Parameter Estimation . . . . .	11
7.3	Statistical Analysis . . . . .	12
<b>8</b>	<b>Conclusion</b>	<b>13</b>
<b>9</b>	<b>Appendix</b>	<b>15</b>

## **1 Acknowledgements**

Many thanks and appreciative gestures towards my fellow EPCC SSP colleagues for helping out especially where vast quantities of knowledge of the UNIX operating system was required. I would also like to thank the EPCC staff for their patience and generosity with their time, I won't name names as I think they know who they are. Finally many thanks to my supervisors for their guidance and support.

## 2 Introduction

The sensitive topic of genetic modification of crops is of current interest. Research conducted to explore the impact of Genetically-Modified crops on the environment has faced obstacles. For example, these trials have been thought of as unethical and this has encouraged damage of experiments by opponents of these crops. In addition, trials will be, in general, relatively short-term in duration. This project hopes to develop a robust code to model the likely long-term impact of GM oilseed rape on the environment. The interaction between a selection of crops and weeds over a decade will be simulated. This is of interest as one extreme outcome could involve the extinction of one or more weed types. Since the model will be required to simulate a realistically large number of fields to make predictions at a landscape scale, it will be computationally intensive, and a parallel processing environment may have real benefits. The initial phase of the work, with which this project is concerned will, however, concentrate on modelling the population dynamics in one field.

## 3 Domain Decomposition — use of the MPI library

The field will be decomposed into a collection of patches of which the size is dependent on the minimum dispersal measure as follows:

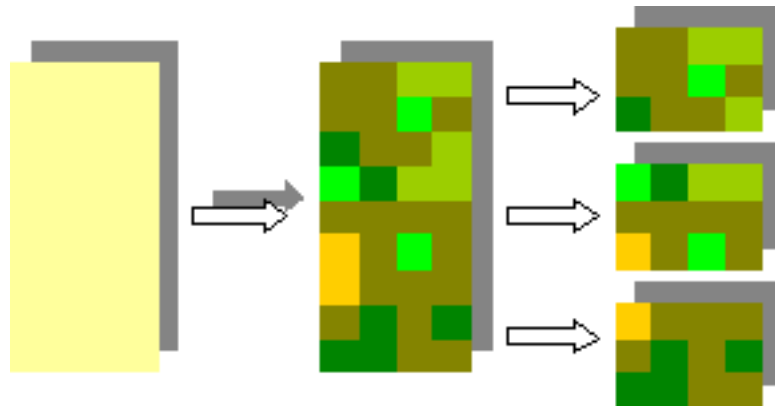


Figure 1: Decomposition of field into patches then sharing between processors

Once this has been completed then parallelisation of the code is performed by splitting the field into chunks and sharing chunks between processors. The main reason for parallelisation of code is that the size of patch dictates to a great extent the minimum amount of memory required to run the simulation. If a field of dimensions 200 by 300 metres is considered, then decomposition into patches of size 0.5 metres will need a minimum of 5Mbytes. As we successively halve the patch size the minimum amount of memory required increases exponentially, see Figure 3.

## 4 Simulating the Annual Agricultural Cycle

As Figure 3 shows the main stages which the model will be considering are:

- The sowing of crops at a set density.

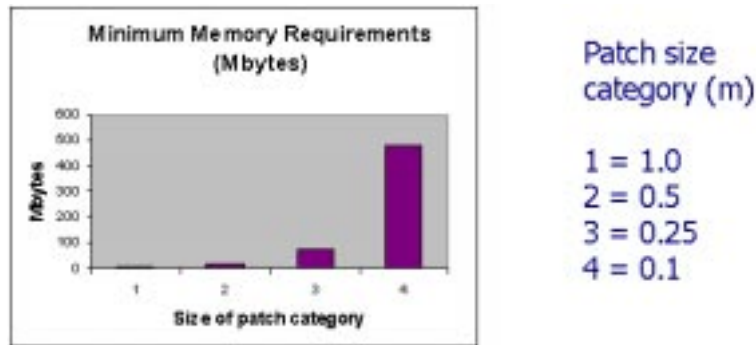


Figure 2: Minimum memory requirement increases exponentially with decreasing patch size

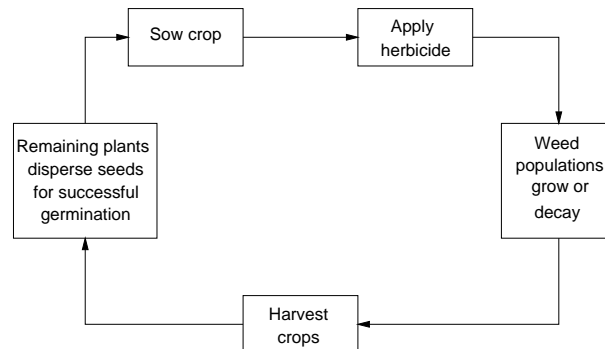


Figure 3: The annual agricultural cycle

- The application of herbicide thereby removing a large proportion of weeds.
- The change in population of each plant species due to competition and growth of plants.
- The harvest of crops. Any remaining crop plants are then treated as weeds, these are termed volunteers. A proportion of weeds are removed at this point.
- The dispersal of seeds from the remaining weeds for successful germination.

These stages are explained in further detail below.

#### 4.1 Initialising plants per patch

Following decomposition of a field into patches, each patch will be initialised with a average density of crops. For the weed species, two cases will be run:

- Random distribution of the weed types (Poisson distribution)
- Clusters of weeds in the field (Negative binomial distribution)

For the first case, a patch will be chosen and a number of weeds will be added to this patch. The distribution of weeds will follow a Poisson distribution of mean 10 plants per patch.

For the second case, the over-dispersal factor of the distribution will dictate the patchiness of weeds.

## 4.2 Application of herbicide

To simulate the application of herbicide to the field, a two-dimensional array has been set up consisting of the proportions of a weed type removed by a particular herbicide given that a certain type of crop has been sown.

## 4.3 Growth and competition between plants

Following the example of Mortimer, Sutton and Gould (1989) [1], the change in the population of a weed type is modelled by a difference equation. This is because of the discrete population behaviour of weeds in agricultural environment. In general, the population  $N$  at time  $t + 1$  is given as a function of  $N$  at time  $t$  by

$$N_{t+1} = F(N_t)N_t \quad (1)$$

where  $t \geq 0$  denotes the timestep and  $F$  is a non-linear function of  $N_t$ . The density of crop sown per patch will be, in general, non-constant. From Mortimer, Sutton and Gould (1989)

$$F(N_t) = \frac{R}{(1 + aN_t)^b} \quad (2)$$

where  $R$  is the asymptotic per capita growth rate of the population,  $a$  and  $b$  are shape and peak parameters respectively. The difference equation in full is,

$$N_{t+1} = \frac{RN_t}{(1 + aN_t)^b}. \quad (3)$$

For two species, the dynamics of one species is likely to be affected by the population level of another species. Equation 3 can be expanded to represent a two-species community:

$$\begin{aligned} N_{1,t+1} &= \frac{R_1 N_{1,t}}{[1 + a_1(N_{1,t} + \alpha N_{2,t})]^{b_1}} \\ N_{2,t+1} &= \frac{R_2 N_{2,t}}{[1 + a_2(N_{2,t} + \beta N_{1,t})]^{b_2}} \end{aligned} \quad (4)$$

from Firbank and Watkinson (1986) [2].  $\alpha$  and  $\beta$  are equivalence coefficients known as inter-specific competition effects.

For a mixture of more than two cultures, Equation 4 may be extended so that the dynamics of a particular species  $i$  are given by

$$N_{i,t+1} = \frac{R_i N_{i,t}}{[1 + a(N_{i,t} + \sum_{j=1, j \neq i}^{n_{species}} \alpha_j N_j)]^{b_i}} \quad (5)$$

where  $\alpha$  is the inter-specific effect between species  $i$  and  $j$ .

## 4.4 Harvesting

The idea used to model this phase is similar to the that used in the Herbicide phase. A large proportion of all plant types will be removed from each patch to simulate this action.

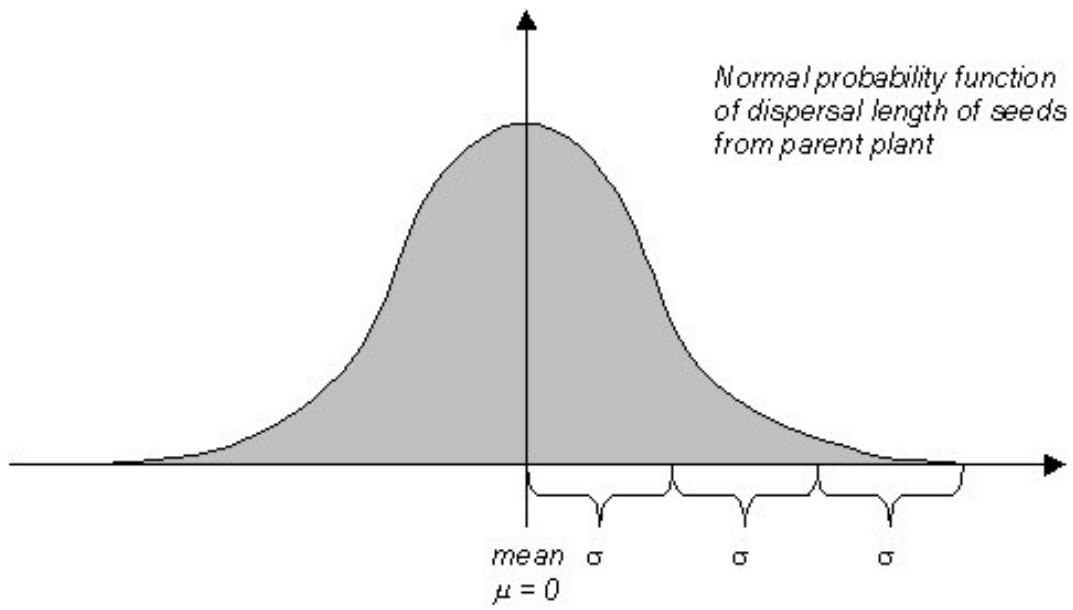


Figure 4: Probability distribution of seed dispersal distances from parent plant

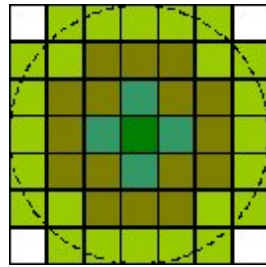


Figure 5: Seed dispersal region centred about parent plant

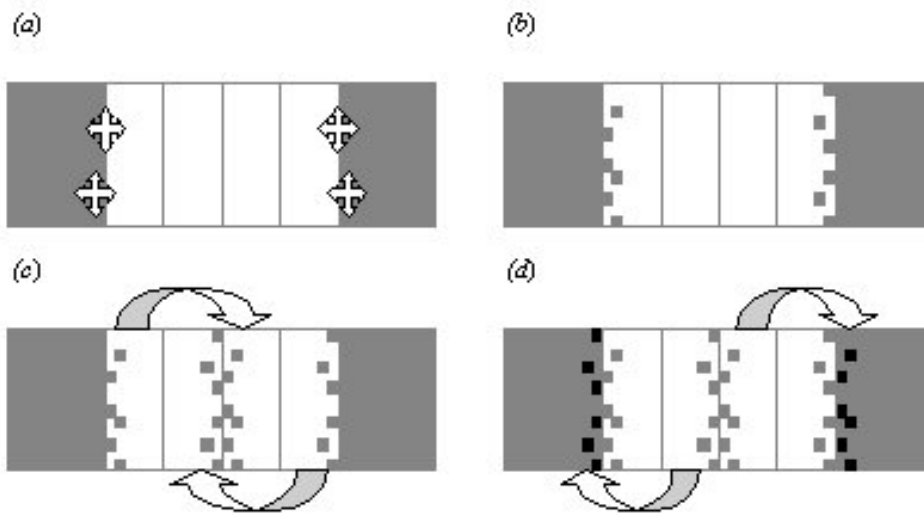


Figure 6: Dispersal over many processors

## 4.5 Modelling dispersal

Consider one patch and one type of plant. The number of seeds for successful germination the following year will be assumed to be constant for each individual plant. We assume the dispersal length of seeds from an individual plant follows a Gaussian distribution of mean  $\mu = 0$  and variance  $\sigma^2 = 2 \times (\text{DispersalMeasure})$ . The maximum length of seed dispersal from a parent plant which will be considered is three standard deviations from the mean so that 99% of seeds are expected to fall within this region — see Figure 4. The pattern of seeds distributed around a plant will be circular with seeds concentrated closest to the plant, as in Figure 5. To model this behaviour, a normal random number generator is used to choose a point within the dispersal region. The plant population of the patch corresponding to this position is then incremented by one. Since the field is divided into chunks, *halo regions* will be needed to receive seeds dispersed by plants at the edge of the fields, illustrated in Figure 6(a) and (b). The halo region is then sent back to the appropriate processor (c) and the seeds are then added back into the field at the required location (d).

Now the dispersal strengths for a selection of plants will not, in general, be the same, therefore an appropriate rescaling is required to determine the maximum dispersal length from the parent plant in terms of patches.

We have dispersal strengths for  $n$  plants,

$$D_1, D_2, \dots, D_n \quad (6)$$

and these are ordered such that

$$D_1 < D_2 \leq \dots < D_n. \quad (7)$$

Then  $D_1$  and  $D_n$  are the minimum and maximum dispersal measures respectively. For rescaling, we calculate the standard deviations  $\sigma_i = \sqrt{2D_i}$  for the  $i$ th plant. The patch size will be equal to the minimum standard deviation value  $\sigma_1$ , call this value *MinSD*. Since we set the patch size to be equal to the value  $\sigma_1$  then we can compute the ratios  $\frac{\sigma_i}{\text{MinSD}}$  and hence determine the maximum number of dispersal patches required for each weed type.

$$\frac{\sigma_i}{\text{MinSD}} \quad (8)$$

for the  $i$ th plant.

## 5 Program algorithm

Figure 7 shows the generic program algorithm.

## 6 File Input and Output

The input file format is illustrated in Figure 8. Here the field dimensions, dispersal strengths, and so forth, are fed into the program. After each time-step, the plant populations for each patch of every plant type is written to file. These files have been written in *portable grey map (pgm)* format so that they may be displayed using graphics package XV.



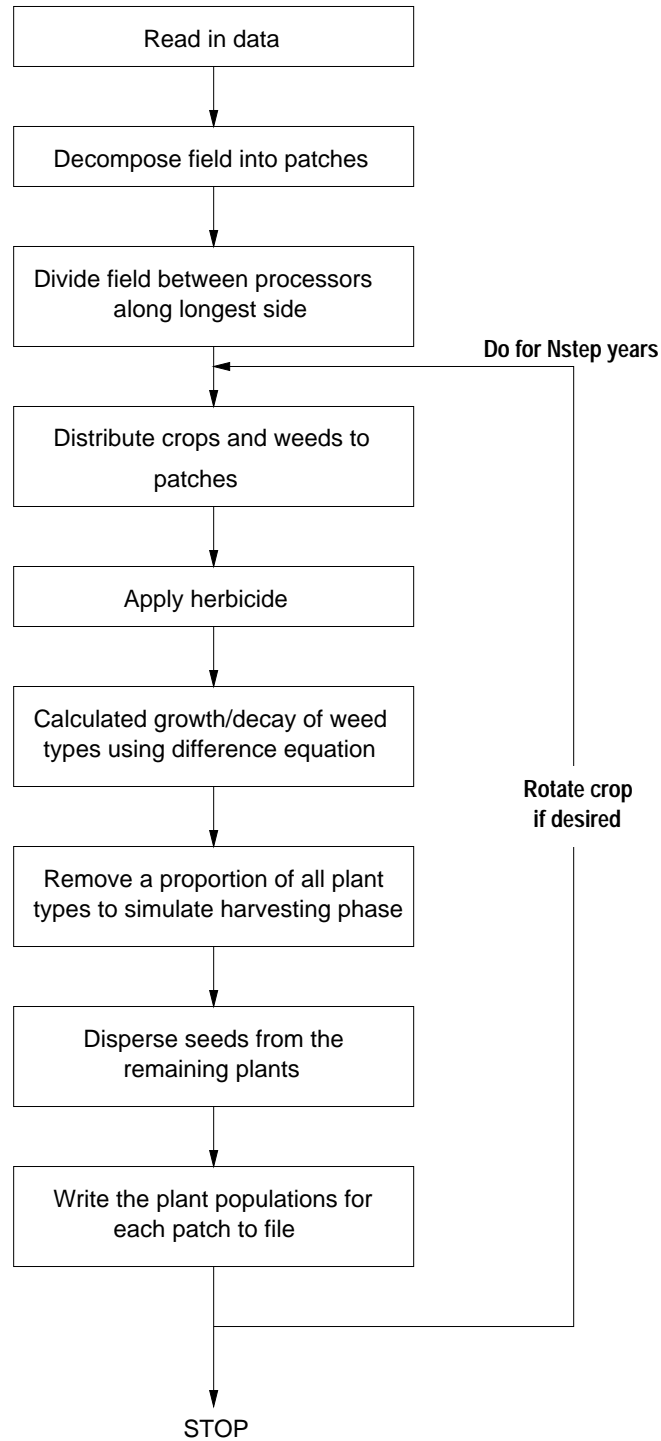


Figure 7: Flowchart of program algorithm

```

Field size
10.0 10.0
Number of plant species
3
Crop error (A parameter for making distribution of crops per patch more random)
0.0
Weed density (Plants per metre squared)
10.0
Total number of time steps to run simulation for (in years)
2
Dispersal measures
1.0 2.0 3.0
Growth rates
1.25 1.5 1.75
Herbicide plant removal proportions
1.00 0.90 0.90
0.90 0.90 0.90
0.90 0.90 0.90
Competition parameters
1.0 0.38 0.76
0.59 1.0 2.45
0.89 0.93 1.0
Shape parameters, a, for difference equation
21.14 33.2 43.1
Peak parameters, b, for difference equation
0.39 0.59 0.86
Total number of seed for dispersal (only consider for successful germination)
8.0 8.0 8.0
Crop density (plants per metre squared) for each timestep
100.0 150.0 125.0
Crop index
1 1 1

```

Figure 8: Sample input file

## 7 Simulation Case Study

A total of thirteen plant species were considered for the simulation of which three were crops and the rest a variety of weeds. The crop plants included Barley, Genetically Modified Oilseed Rape and conventional Oilseed Rape. Crop rotation was incorporated into the model and the simulation was allowed to run for ten years of simulation time.

### 7.1 Weed-crop distribution

One crucial step before interpretation of results can begin is to decide how to distribute weeds between patches. Depending on the dispersal strength and survival of plants, the frequency distribution of plants per patch should vary from random (Poisson) to over-dispersed (clusters of patches with an abundance of weeds, Negative Binomial distribution). From field studies performed by the SAC, two of the weed species were deemed to follow the Poisson distribution — the Sow Thistle and the Nettle. Weeds Mayweed, Shepherd's Purse, Annual Meadow Grass, Chickweed, Fathen, Knotgrass, Fumitory and Spurge seem to follow the Negative Binomial distribution (Table 1). The means and variances for plant populations per patch of area  $0.25m^2$  were obtained from field sample data collected in December 1998 and provided by the SAC consultant, Dr Neil McRoberts. The parameter estimates for the frequency distribution of each plant type were calculated from the given data. Random numbers were then generated for each frequency distribution of a plant type and the field patches initialised with these random

Weed	mean	variance	Frequency Distribution	p	k
Mayweed	0.6	3.05	Negative Binomial	0.2	1
Shepherd's Purse	12.0	81.1	Negative Binomial	0.15	2
Annual Meadow grass	10.9	53.7	Negative Binomial	0.2	3
Chickweed	7.8	27.2	Negative Binomial	0.29	3
Fathen	0.5	2.1	Negative Binomial	0.24	1
Knotgrass	0.3	0.6	Negative Binomial	0.5	1
Sow Thistle	0.05	0.05	Poisson	n/a	n/a
Fumitory	1.1	2.3	Negative Binomial	0.47	1
Spurge	0.3	4.0	Negative Binomial	0.08	1
Nettle	0.01	0.01	Poisson	n/a	n/a

Table 1: Mean and variance data for plant populations per patch donated by the SAC contact Dr. Neil McRoberts

numbers. For the crops, these were sown at a constant density of 300 plants per square metre.

For the Poisson distribution, the probability density function is given by,

$$P_r(N = n) = \frac{\mu e^{-\mu}}{n!}, \quad n \geq 0 \quad (9)$$

where  $n$  is the number of plants per patch and  $\mu$  is estimated by  $\bar{n}$ , the mean of  $x$  observations.

For the negative-binomial distribution,

$$P_r(N = n) = \binom{k+n-1}{n} \left(\frac{\mu}{k}\right)^n \left(1 + \frac{\mu}{k}\right)^{-(k+n)}, \quad n \geq 0 \quad (10)$$

where  $k > 0$  is the the over-dispersion parameter and the mean  $\mu$  is estimated by  $\bar{n}$  as above.

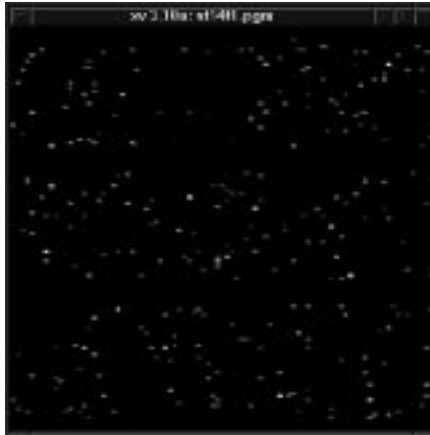


Figure 9: Example of weed population following a Poisson distribution

## 7.2 Parameter Estimation

Parameter estimations for dispersal strengths etc. were obtained by observed field analysis. Estimating these is also highly significant to the model's performance. Again the data was provided

by the SAC consultant Dr. Neil McRoberts.

### 7.3 Statistical Analysis

During the ten-year simulation the population for each plant for each patch was collected after each agricultural phase:

- Sowing of crop
- Application of herbicide
- Growth and competition
- Cropping
- Dispersal of seeds

The plant populations for each patch were then grouped into quadrants of four patches, two across and two down, as Figure 10 illustrates below.

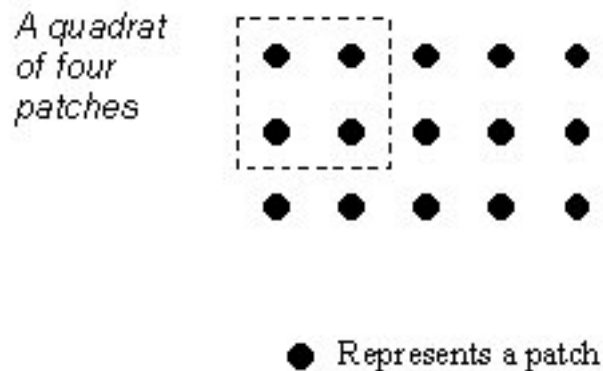


Figure 10: Grouping of plant populations by quadrants.

Boxplots for each plant type at each agricultural phase over the full ten years were constructed and are provided for perusal in the appendix.

For the crops, a strong seasonal component is present though the crop rotation amplified some of the results at times. The variance of plant populations fluctuates wildly, especially for the GM Oilseed Rape crop. It is difficult to see a general trend in the data for all crops, as the length of time for the run does not yield a strong upward or downward trend. A longer time period for simulation may be beneficial in establishing trends. The GM crop shows more activity in population change than the other crops; the average population of the conventional Oilseed Rape is more sedate by comparison. The Barley crop population seems the most stable over the time period with a growth spurt just after midway through the time period. For further analysis into this case, the plant populations of the remaining crop and weed types at this time point should be considered.

The weed types show a general downward trend towards extinction. This is probably due to the method in which weeds were distributed initially. Patches were chosen at random and a random number following the frequency distribution of plant populations per patch was then

assigned to this patch. Since a small proportion of patches (one percent of total number of patches) were assigned at random with plants, then there was little scope for clusters of patches with an abundance of weeds to appear. Further simulations with groups of patches with a large population of weeds would be an interesting route to follow.

## 8 Conclusion

The success of the model for predicting the long-term prospects for weed and crop populations is highly dependent on the parameter estimates and the initial weed-crop distribution. It is important to bear both the above in mind when interpreting the results. A large number of parameter estimates were needed to set up the simulation, a situation which is undesirable. As a rule, extrapolation outwith the bounds of any given data should be viewed as unreliable; in this light the model is to be thought of as a qualitative tool whereby the parameters may be adjusted and initial plant distribution changed. The program has been written with a modular approach in mind and this makes it simple for the user to build their agricultural cycle as they see fit. One direction for modification of the model would be to incorporate other agricultural aspects, the inclusion of seed banks for example. Another route is to check for the sensitive dependence on initial the parameter estimates. This would involve identifying parameters which greatly affect the outcome of the plant population change.

## References

- [1] Cousens, Mortimer (1995)*Dynamics of Weed Populations*, Cambridge University Press
- [2] Firbank & Watkinson (1985)*On the analysis of competition within two-species mixtures of plants*, Journal of Applied Ecology, Volume 23, pp147-159
- [3] Hughes, McRoberts, Madden & Nelson (1997)*Validating mathematical models of plant-disease progress in space and time*, IMA Journal of Mathematic Applied in Medicine & Biology, Volume 14, pp85-112
- [4] Knuth (1938)*Semi-numerical algorithms*, The Art of Computer Programming , Volume 2 (Second Edition)
- [5] Mortimer, Sutton & Gould (1989) *On robust weed population models*(1989), Weed Research, Volume 29, pp229-238
- [6] Press, Teukolsky, Vetterling & Flannery *Numerical Recipes in Fortran 77*, The Art of Scientific Computing, 2nd Edition

## Personal



Formally of Napier University, I have graduated with an Honours degree in Mathematics, 1999. I like to indulge myself with Belgian truffles.

My supervisors include Dr. Mark Parsons (EPCC) and Dr. Neil McRoberts (Scottish Agricultural College).

### 9 Appendix

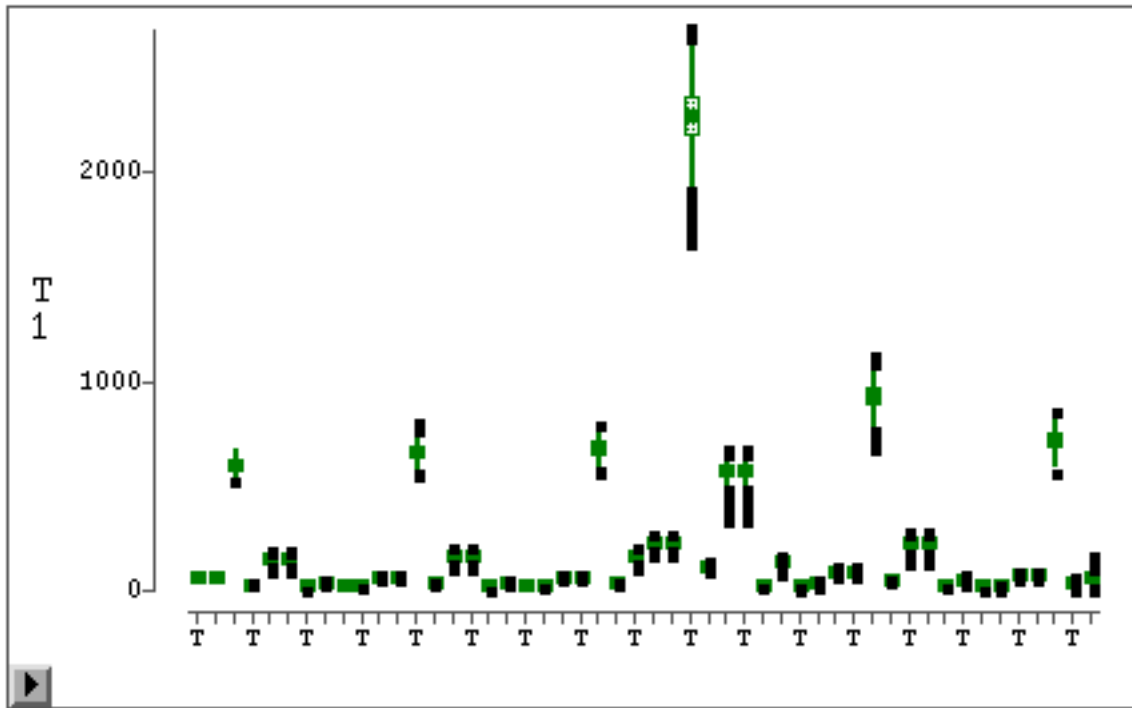


Figure 11: Barley

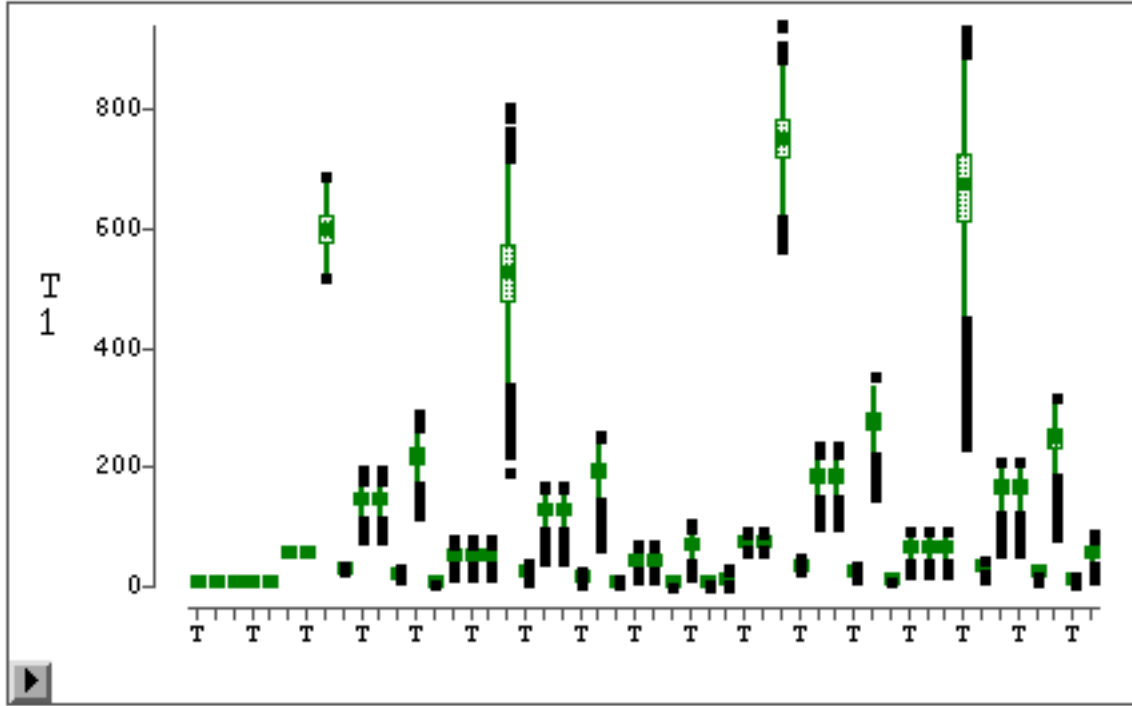


Figure 12: Genetically-modified Oilseed Rape

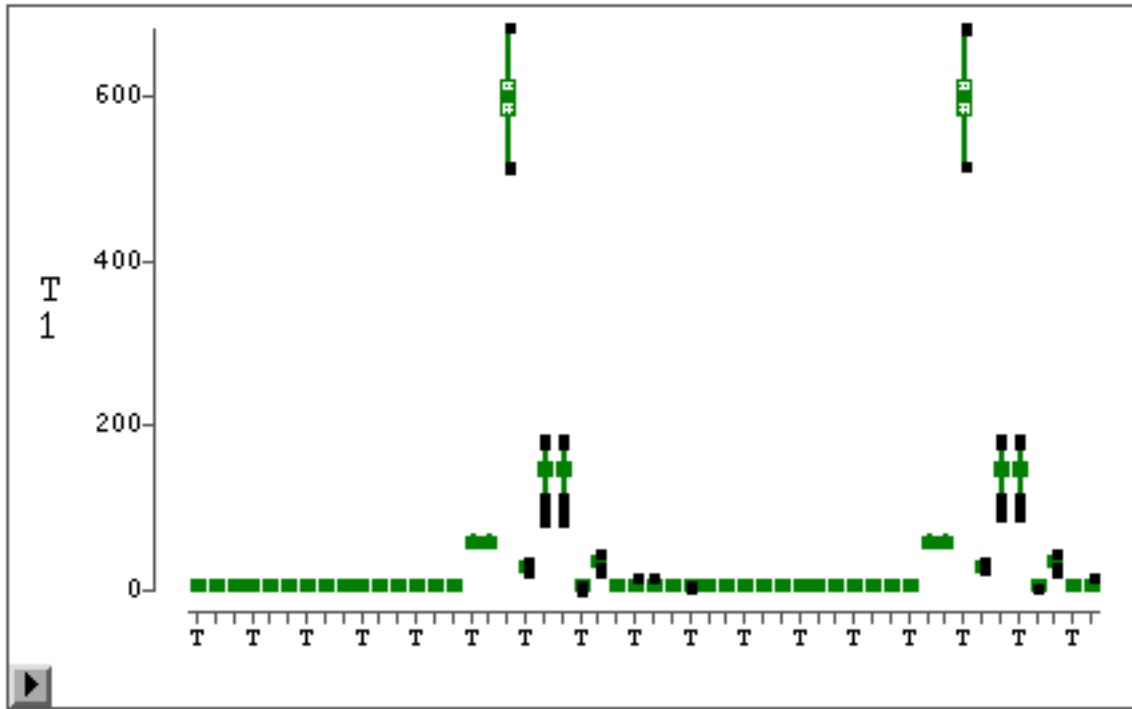


Figure 13: Conventional Oilseed Rape



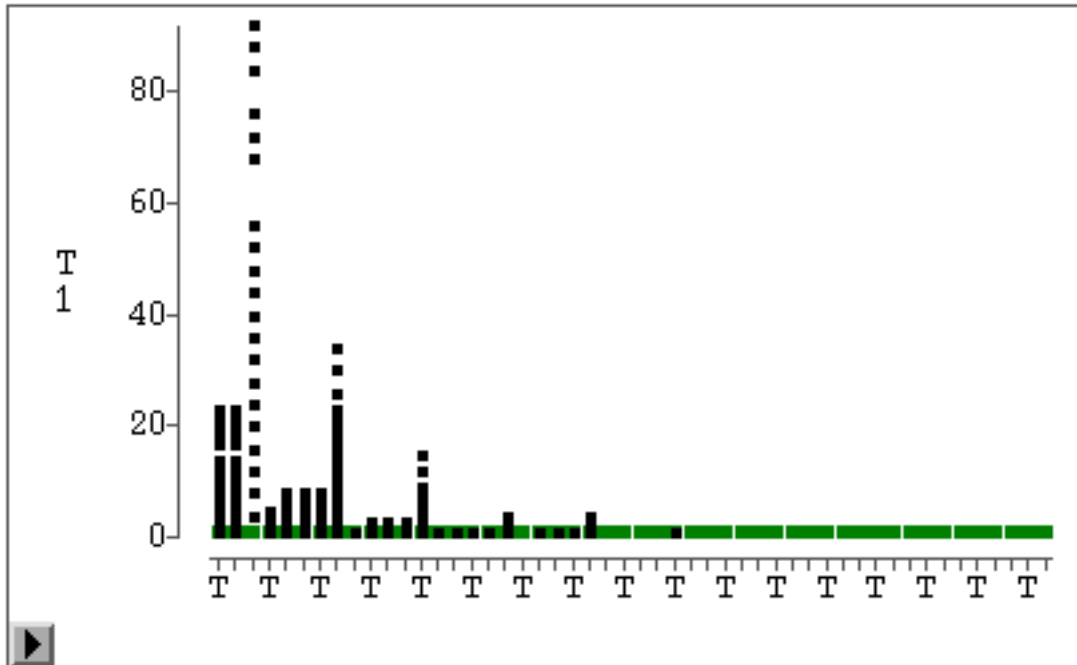


Figure 14: Mayweed

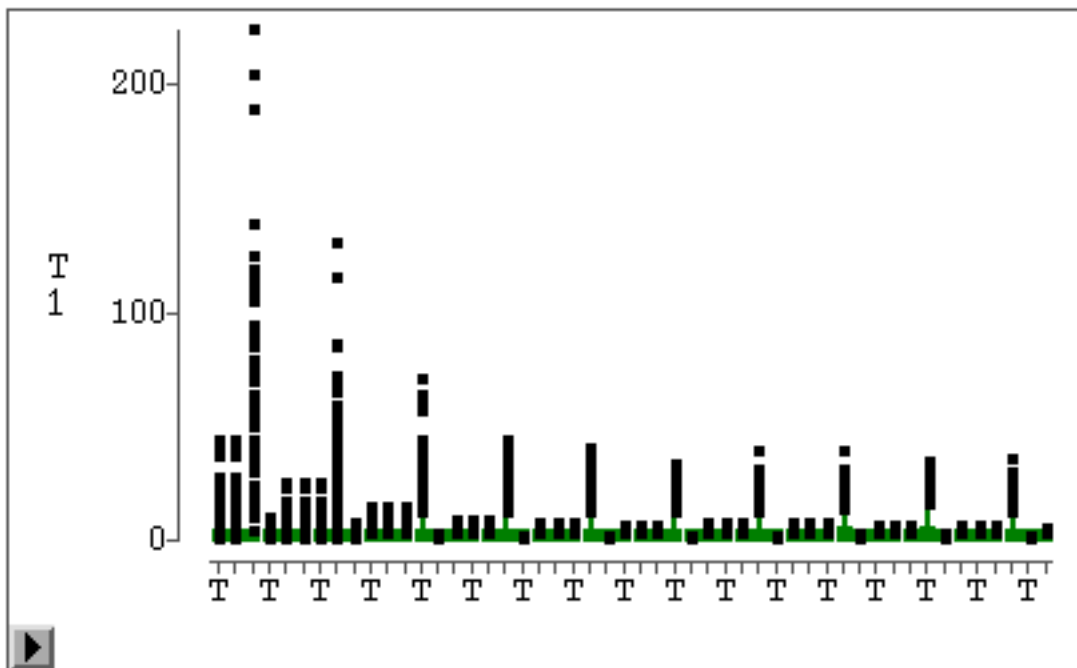


Figure 15: Shepherd's Purse

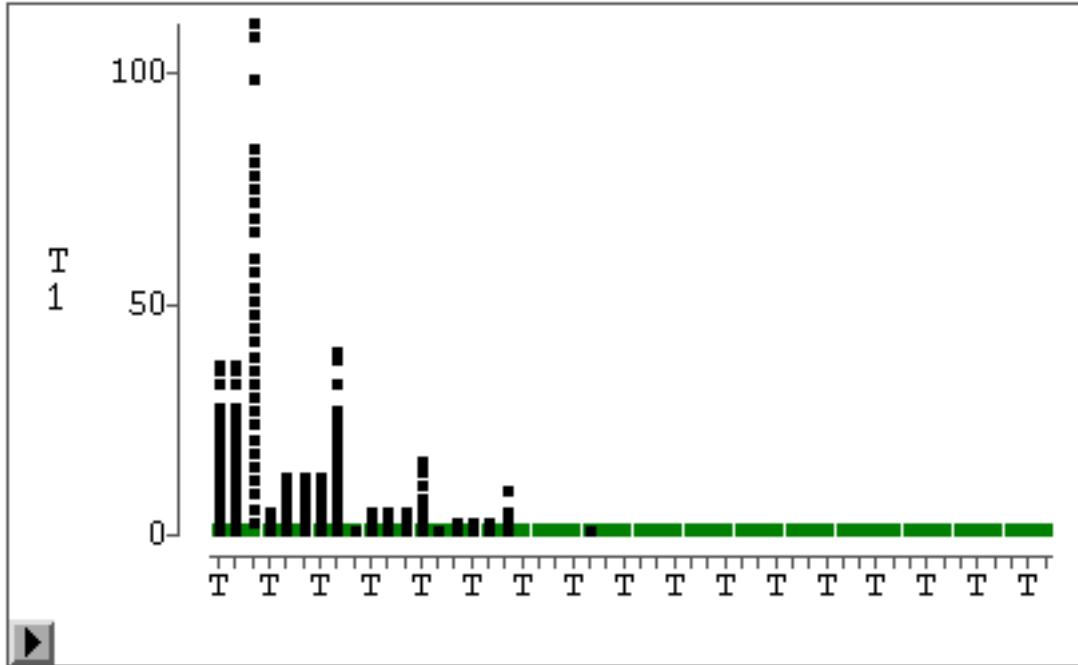


Figure 16: Annual Meadow Grass

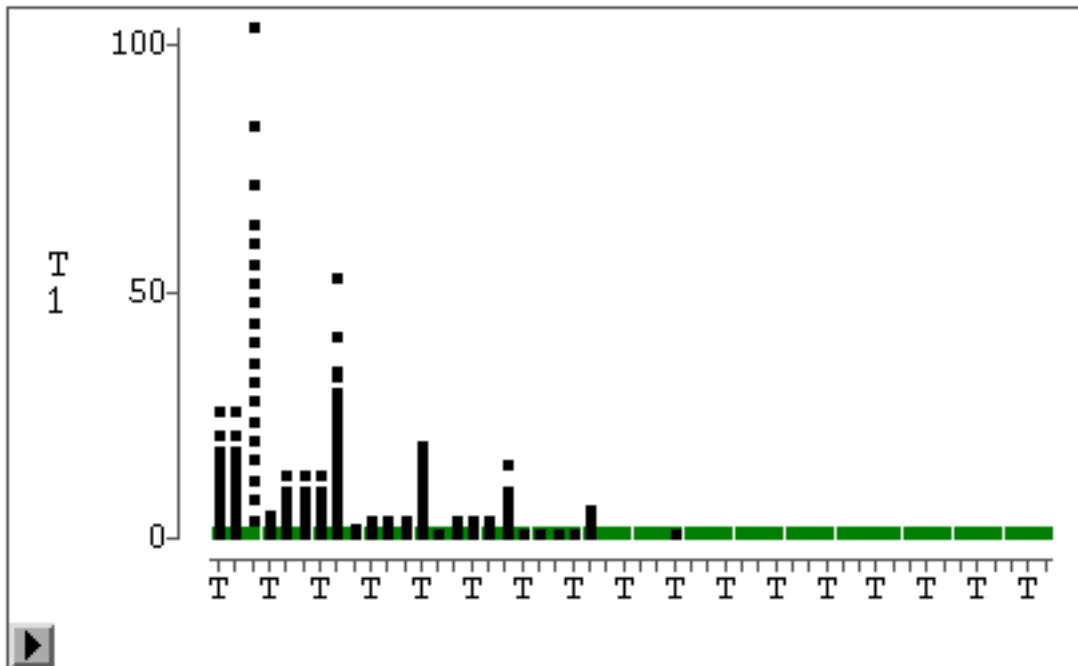


Figure 17: Chick weed

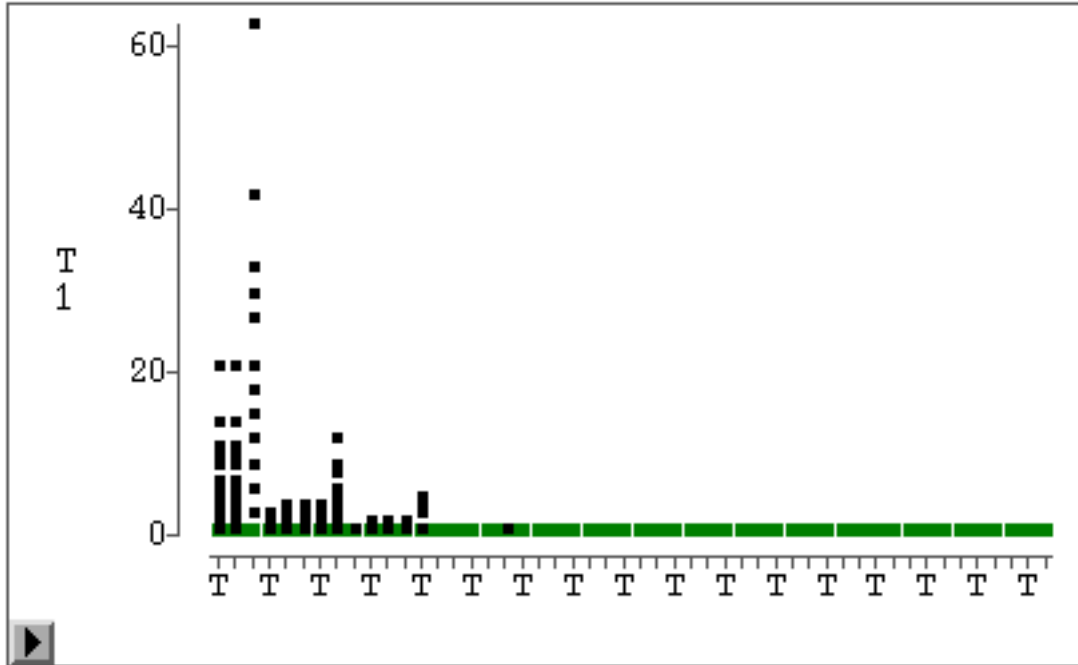


Figure 18: Fathen

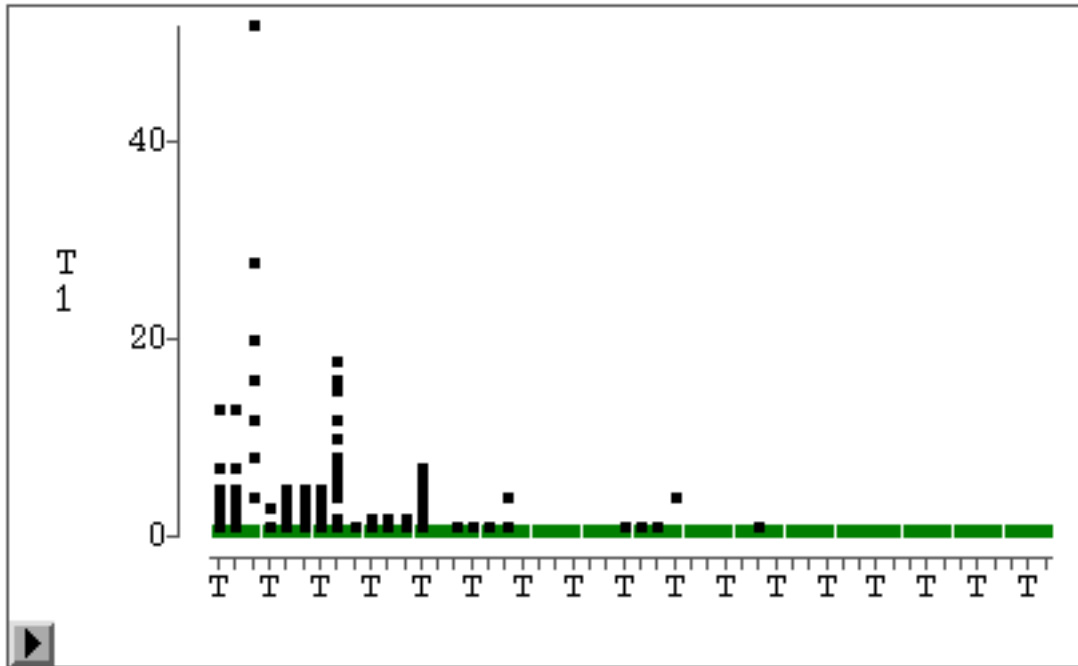


Figure 19: Knot Grass

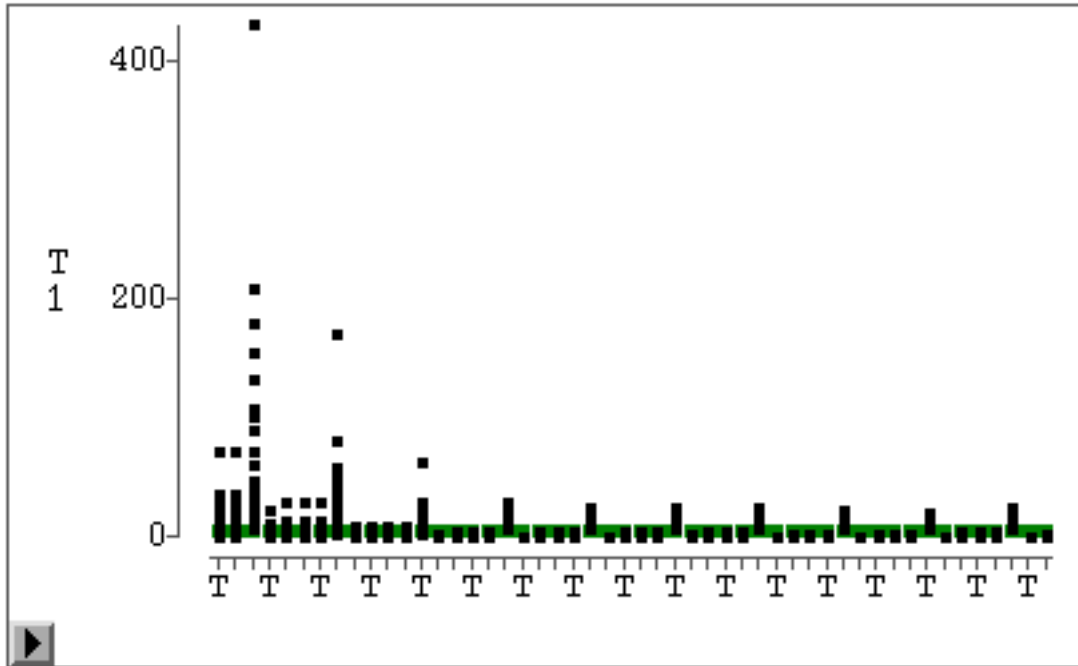


Figure 20: Sow Thistle

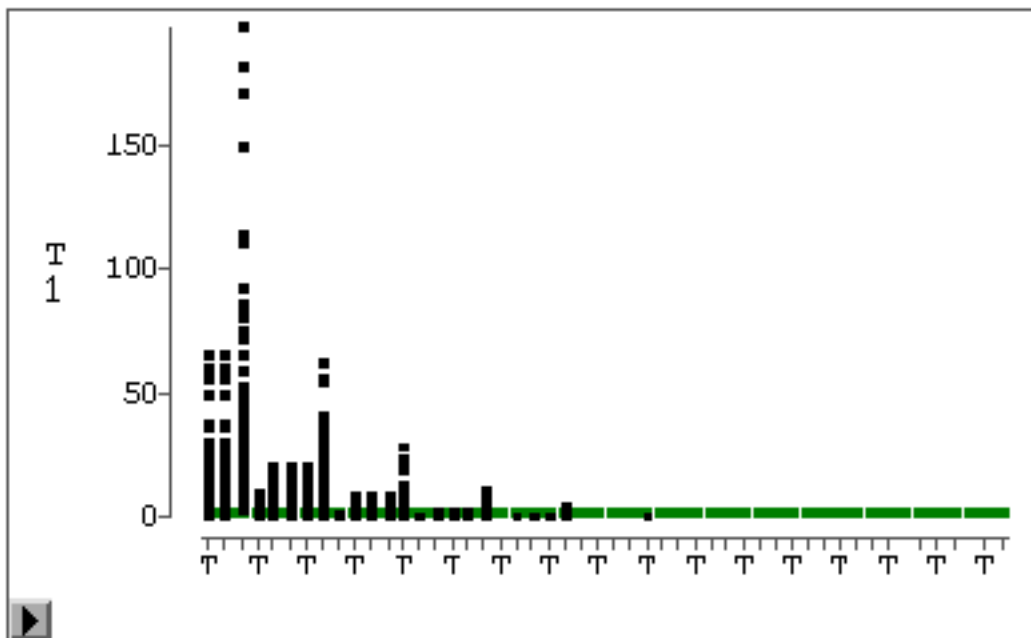


Figure 21: Fumitory

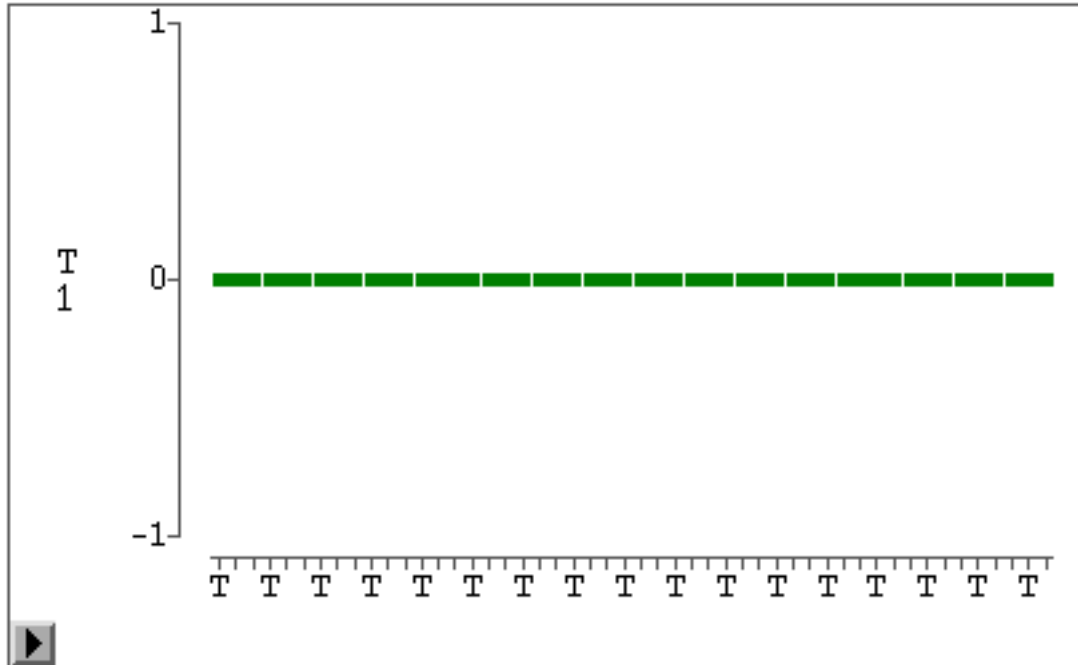


Figure 22: Spurge

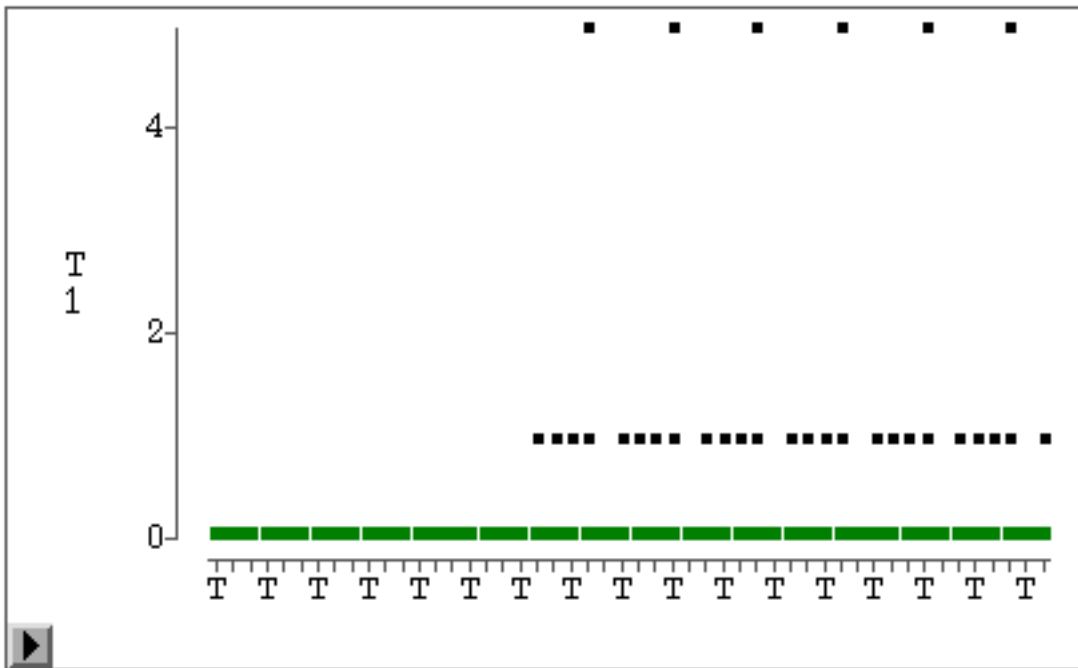


Figure 23: Nettle