

Black grass IPM

Shaun R. Coutts[†], Dylan Childs[†], Rob Freckleton[‡]

September 11, 2015

1 Model

The goal of this model is to predict the distribution of black grass and the evolution of herbicide resistance at the field scale (roughly a 100m x 100m). The model is spatially explicit and includes a seed-bank, density dependent germination out of the seed-bank, and an individual quality called herbicide resistance (r), which has a genetic and non-genetic component.

The model tracks the number of individuals of size k at location x with genotype g through time as a function of survival, fecundity and genetic mixing.

$$n(x, g, t + 1) \sim B(z(x, g, t), \phi_s(x, g, t)) \quad (1)$$

Where $n(x, g, t)$ is the number of individuals at location x with genotype g at time t . We use a yearly time-step and take t to be the point just before plants flower, set seed and die. Eq. 1 gives the number of new individuals establishing at location x and surviving until seed set. We assume the number of individuals with genotype g surviving until seed set is a random variable drawn from a binomial distribution where the number of trials is the number of individuals that germinate from the seed bank, $z(x, g, t)$, and the probability of survival for genotype g at location x is $\phi_s(x, g, t)$.

Genetic model

Herbicide resistance is an evolved trait that may be the result of a single mutated gene, or several different genes that control different metabolic pathways or toxin transport. Plants in our model are assigned a resistance score, $R(g)$ based on the genotype they inherit from their parents, g , and a random non-genetic component.

Each genotype consist of a set genes, each gene, j having two copies, that can either be a susceptible version, $v_{j,0}$, or a resistant version, $v_{j,r}$. The genotype is the set of J genes, $g = \{l_{g,1}, l_{g,2}, \dots, l_{g,J}\}$, with each gene coded as a tuple $l_{g,j} = \{\kappa_{1,g,j}, \kappa_{2,g,j}\}$, where $\kappa_{k,g,j}$ is the k^{th} copy of gene j in genotype g and can be either $v_{j,0}$ or $v_{j,r}$.

seed-bank and fecundity

Ultimately we are interested in how resistant genotypes spread through the population. The number of individuals with genotype g that establish at location x in time t is

$$z(x, g, t) \sim \text{B}(b(x, g, t), \phi_e) \quad (2a)$$

$$b(x, g, t) \sim \text{B}(b(x, g, t-1) - z(x, g, t-1), \phi_b) + \text{Pois} \left(\sum_{\forall y} f(y, g, t) K_s(x, y) \right) \quad (2b)$$

We model establishment of each genotype g from the seed bank at location x and time t ($z(x, g, t)$) as a random draw from a binomial distribution ($\text{B}(\cdot)$ Eq. 2a). The number of independent trials is $b(x, g, t)$, the number of seeds in the seed bank at location x of genotype g at time t , and the density independent probability that a seed germinates is ϕ_e . The number of seeds in the seed bank at time t (Eq. 2b) is made of the number of seeds in the seed bank in the previous time step that did not germinate in the last time step and survived to the current time step, plus the input of new seeds from the entire field (summation over all locations y). The number of seed bank seeds that don't germinate and survive is a random number drawn from a binomial distribution ($\text{B}(b(x, g, t-1) - z(x, g, t-1)$

1), ϕ_b), Eq. 2b), where ϕ_b is the probability that a seed survives one time step. The number of new seeds that arrive at location x is a random draw from a Poisson distribution. The expected arrival rate is $\sum_{y,g,t} f(y,g,t)K_s(x,y)$, where $f(y,g,t)$ is number of seeds produced at location y of genotype g at time t and $K_s(x,y)$ is a dispersal kernel which gives the probability of seeds dispersing between locations y and x . Drawing the number of arriving seeds from a Poisson distribution guarantees that the number of seeds arriving is integer, which is required by the binomial distribution. Over the whole landscape over a long time period the number of seeds that arrive at a location will be very close to the expected rate of arrival, but in any one time step the total number of seeds produced (summed over the whole landscape) might be larger or smaller than the expected number summed over the whole landscape. That is, using a random draw will 'create' or 'destroy' seeds in any given time step. This variation could be thought of as natural stochastic variation in seed production. There is no way to round the expected number of seeds to an integer number without creating or destroying some seeds.

The number of seeds produced at location y of genotype g at time t is determined by the gene frequencies in the population and how the population is distributed around the landscape. We assume that black grass is an obligate out-crossing species and so only model two sources of genes, the individual producing the seed and pollen. Because gene copies can spread through pollen individuals with a genotype other than g could still produce seeds with genotype g by receiving different versions of genes ($v_{j,w}$) from pollen. Thus, in Eq. 3a we sum the seed contribution across the set of all possible genotypes, G (first summation term).

$$f(y, g, t) = \sum_{\forall g' \in G} n(y, g', t) \psi \prod_{j=1}^{j=J} \left(\sum_{k=1}^{k=2} Q(\kappa_{k,g,j}, l_{g',j}) q(y, g, j, k, t) \right) \quad (3a)$$

$$Q(\kappa_{k,g,j}, l_{g',j}) = \begin{cases} 0.5 & \text{if } \kappa_{k,g,j} \in l_{g',j} \\ 0 & \text{otherwise} \end{cases} \quad (3b)$$

$$q(y, g, j, k, t) = \frac{\sum_{\forall y'} \sum_{\forall g^* \in G} n(y', g^*, t) \eta(k, g^*, g, j) K_p(y', y)}{\sum_{\forall y'} \sum_{\forall g^* \in G} n(y', g^*, t) K_p(y', y)} \quad (3c)$$

$$\eta(k, g^*, g, j) = \begin{cases} 0.5 & \text{if } \kappa_{h \neq k, g, j} \in l_{g^*, j} \\ 0 & \text{otherwise} \end{cases} \quad (3d)$$

The number of seeds contributed from each genotype g' to the target genotype g (Eq. 3) is the number of individuals of genotype g' ($n(y, g', t)$) multiplied by the number of seeds each individual produces, ψ , multiplied by the proportion of seeds with the target genotype (g) expected to be produced by the genotype being evaluated (g'). This proportion (the product term in Eq. 3a) is the probability all J tuples in genotype g ($l_{g,j}$) are produced in the same seed made by an individual of genotype g' . Genes copies have two sources, the individual producing the seed provides half the copies of a gene, while pollen provides the other half. In Eq. 3 these two processes are modelled with the functions $Q(\kappa_{k,g,j}, l_{g',j})$, which contributes gene versions ($v_{j,w}$) from the mother, and $q(y, g, j, k, t)$, which contributed gene versions from pollen. $Q(\kappa_{k,g,j}, l_{g',j})$ is a function that returns 0.5 if the value held by copy k of gene j in the target genotype g ($\kappa_{k,g,j}$) is in tuple j of the genotype being evaluated $l_{g',j}$, and returns 0 otherwise. This function returns 0.5 since the mother can only supply one out of two copies for each gene. The second copy of the gene is supplied from the pollen using function $q(y, g, j, k, t)$. We assume that the probability of getting the version of gene j required to complete the tuple $l_{g,j}$ is the proportion all pollen arriving at location y which contains that copy of gene j . We assume that the amount of pollen produced with genotype g is proportional to the number

of individuals with that genotype. Thus, the denominator of Eq. 3c is proportional to the total amount of pollen that arrives at location y from all other location (first summation in the denominator) for all genotypes (second summation in denominator). The numerator of Eq. 3c is proportional to the amount of pollen that contains the version of gene j in the target genotype g , that is not supplied by the mother. It is function $\eta(k, g^*, g, j)$ that determines if this gene version will be produced by the genotype being evaluated for pollen production, g^* (Eq. 3d). $\eta(k, g^*, g, j)$ returns 0.5 if the version of gene j in target genotype g not provided by copy $\kappa_{k,g,j}$ (i.e. copy provided by the mother) is a member of the tuple $l_{g^*,j}$ in the genotype being evaluated for pollen production.

survival-growth

We assume that probability of survival to flowering and seed set (ϕ_s) is influenced by genotype, herbicide application and density. It is at this point that herbicide application separates the different genotypes and drives the spread of herbicide resistant genotypes.

$$\phi_s(x, g, t) = \gamma(x, g, t)s(h_t, g) \quad (4)$$

Where $s(h_t, g)$ is the survival probability for an individual with genotype g given the herbicide application in time step t , h_t , and $\gamma(N_x)$ is a density dependent term that reduces establishment and survival above some density threshold.

Survival probability, $s(h_t, g)$, must be bound between 0 and 1, thus we model it as a generalised logistic function of genotype

$$s(h_t, g) = \begin{cases} \frac{\beta_0 e^{-C(g)}}{1 + \beta_0 (e^{-C(g)} - 1)} & : h_t = 0 \\ \frac{\beta_0 R(g) e^{-C(g)}}{1 + \beta_0 R(g) (e^{-C(g)} - 1)} & : h_t = 1 \end{cases} \quad (5)$$

where β_0 is the survival probability for an individual with no resistance gene versions in their genotype, e is Euler's constant, $C(g)$ is function that describes the survival costs

of having genotype g and $R(g)$ is a function that describe the herbicide resistance of genotype g .

The cost and resistance functions ($C(g)$ and $R(g)$) are how the genotype interacts with the population model through the cost of having a given genotype and the resistance that genotype provides. The survival cost of having genotype g is assumed to be a linear function of the summed cost of each gene in g , with an intercept of 0 (implying that individuals with no expressed copies of a resistant gene have no cost)

$$C(g) = \sum_{\forall j} \mathbf{max}(\epsilon(\kappa_{1,g,j}), \epsilon(\kappa_{2,g,j})) \quad (6)$$

where $\epsilon(\kappa_{k,g,j})$ is the cost, in terms of reduced survival, of having the version of gene j held in copy $\kappa_{k,g,j}$ (recall that $\kappa_{k,g,j}$ can be $v_{j,0}$ or $v_{j,r}$). The resistance function, $R(g)$ is bound to be between 0 and 1, thus we use another generalised logistic function to meet these constraints.

$$R(g) = \frac{\xi_0 e^{r(g)}}{1 + \xi_0 (e^{r(g)} - 1)} \quad (7a)$$

$$r(g) = \left(\sum_{\forall j} \mathbf{max}(\rho(\kappa_{1,g,j}), \rho(\kappa_{2,g,j})) \right) + N(0, \sigma_r) \quad (7b)$$

$$(7c)$$

where ξ_0 is the herbicide effectiveness for a completely susceptible individual and $\rho(\kappa_{k,g,j})$ is the amount of herbicide resistance conferred by having the version of gene j held in copy $\kappa_{k,g,j}$. $N(0, \sigma_r)$ is a random variable drawn from a normal distribution that gives the non-inheritable part of herbicide resistance, with σ_r controlling the strength of non-inheritable resistance. We assume $\rho(v_0) = \epsilon(v_0) = 0$ and for versions that confer resistance we assume $\rho(v_r)$ and $\epsilon(v_r)$ are greater than 0. Thus, the $\mathbf{max}(\cdot)$ function assumes that the resistance version of the gene is the dominant one for all j genes. We could use other cost and resistance functions to change this behaviour. For example we could use $r(g) = \sum_{\forall j} \mathbf{min}(\rho(\kappa_{1,g,j}), \rho(\kappa_{2,g,j}))$ to make resistant genes receive, or $\rho(\kappa_{1,g,j}) + \rho(\kappa_{2,g,j})$

to make the copies additive in their effect. If we want some genes to be dominant and others receive we need to replace the summation over all j genes with a function that has a separate aggregator for each gene.

Without a density dependent term any increasing population would continue growing until infinitely large. We model this density dependence with the term

$$\gamma(x, g, t) = \begin{cases} \frac{M}{\sum_{\forall g} s(h_t, g)z(x, g, t)} & \text{if } \sum_{\forall g} s(h_t, g)z(x, g, t) > M \\ 1 & \text{if } \sum_{\forall g} s(h_t, g)z(x, g, t) \leq M \end{cases} \quad (8)$$

where $\gamma(x, g, t)$ reduces establishment and survival probability at each location x in response to the total number of individuals at location x which are expected to establish at time t (summation term in Eq. 8). $s(h_t, g)$ and $z(x, g, t)$ are defined in Eq.'s 5 and 2 respectively. M is the average number of individuals that can establish at a given location.