Inferring monocotyledon crown root trajectories from limited data

Jesse Windle

jesse@bayesfactor.net

July 25, 2023

Abstract

Quantifying root growth over time is challenging. Here we consider models of root growth that can be calibrated using data from a novel root sensing device called RootTracker, which non-destructively detects roots over time using impedance sensing. The resulting models capture spatial and temporal patterns in root growth and can be used to recapitulate realistic looking root trajectories.

1 Introduction

Herein we consider modeling root trajectories; in particular the task of inferring a crown root trajectory as well as the underlying parameters governing its path, knowing only the root's origin and a single point along its trajectory. This problem is inspired by data produced by a novel root sensing device called RootTracker.

Since the data available about a crown root's trajectory are limited, the complexity of the models we can consider are limited as well. Ultimately, we reduce crown root growth to two parts: 1) the angle at which the root is growing relative the the soil surface with no lateral movement allowed, and 2) the distance the root travels before changing the direction in which it is traveling.

Our purpose of modeling root growth is twofold. First, we want to quantify differences in root growth, specifically differences in the depth of root growth by variety or time. Second, we want to produce visualizations that recapitulate typical root growth.

2 Background

Roots are used to acquire nutrients and water as well as to anchor a plant. Within a species, genetic diversity can lead to a wide range of root morphologies. The substrate in which roots grow also has an impact on morphology as roots actively seek out water and nutrients. These morphological differences can, in turn, convey an advantage or disadvantage to, e.g. drought stress, flooding stress, nutrient stress, etc. Within a commercial context, stress reduces yield and hence developing or finding varieties that are more resilient, including through
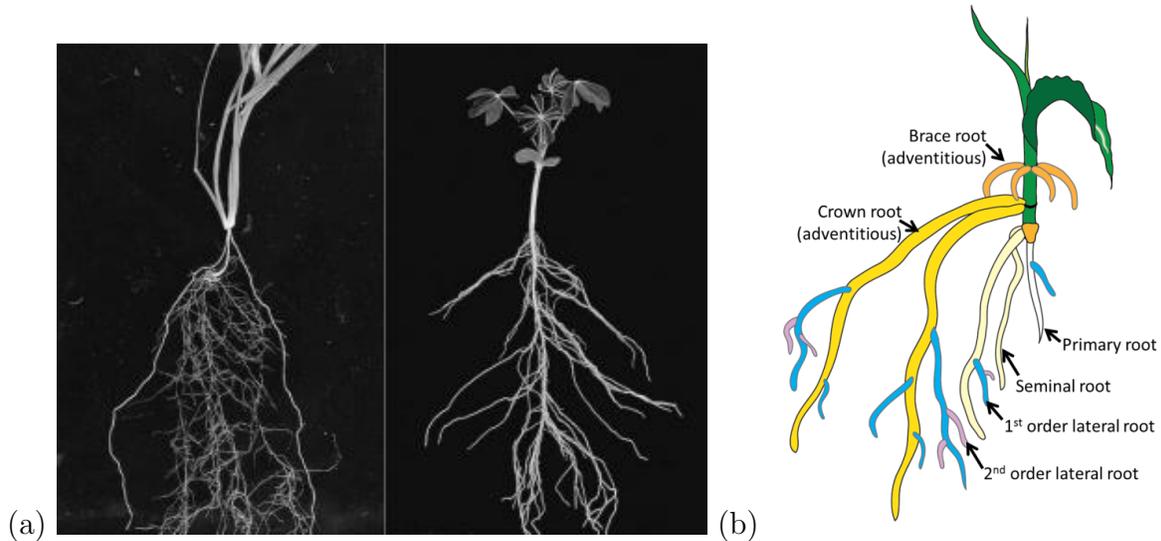
Figure 1: (a) Root systems of young wheat (monocot) and lupin (dicot) plants, left and right respectively. The wheat roots on the periphery are early nodal roots. Successive groups, called whorls, of nodal or crown roots grow above one another as the root system fills out. In contrast, lupin has a taproot which forms an axis along which lateral roots emerge. (b) Example root system architecture of a monocot, like maize. Crown roots emerge in successive groups called whorls, one on top of another. Eventually, these whorls form brace roots which emerge from the plant above ground. Lateral roots are roots that branch off of earlier formed roots. Reproduced from Plants in Action, `http://plantsinaction.science.uq.edu.au`, published by the Australian Society of Plant Scientists.

the genetic control of their root system, is of major importance. Thus, understanding root systems' structure and function is not only an important scientific endeavor but one with major implications for goods we consume on a daily basis.

2.1 Monocotyledon root systems

We are interested in modeling maize root systems or root systems that are similar to maize, like wheat or rice, that have a fibrous root system whose major elements, called crown roots or nodal roots, originate from a region near the soil surface and, roughly speaking, grow out and down. In contrast, crops like soybean or cotton are distinguished by having a taproot that forms the axis off of which other major elements grow. In either case, smaller lateral roots can branch off of a parent root. These two groups are classified respectively as monocotyledons or "monocots" and dicotyledons or "dicots" in reference to the number of embryonic leaves in a seed. Figure 1 provides an example of monocot and dicot root systems and a cartoon of monocot root system architecture (RSA). Herein we are most interested in (monocot) crown root and lateral root growth.

Hochholdinger [2009] is a common reference for an overview of maize RSA, but the book chapter is not freely available. To make it easier for readers to find sources, we focus on sources that are openly available as of this writing. For those looking for a basic introduction

to roots, Munns et al. [2018, ch. 4] is a undergraduate level overview. Nielsen [2020] provides a similarly leveled, brief description of early root growth in maize. Pages [2011] provides a very brief, but more technical, general overview of RSA. The review by Hodge et al. [2009] covers the regulators of RSA, its genetic components, different morphologies, as well as factors, like stressors, that influence root growth. References in the next section also generally contain some description of RSA.

2.2 Root phenotyping

Our data comes from a device called RootTracker, discussed in greater detail below, which uses impedance sensing to detect root growth [Aguilar et al., 2021]. Typical laboratory experiments involve gel imaging, i.e. growing roots in agar, or x-ray images [Mairhofer et al., 2013, Rogers et al., 2016]. However, these methods are not feasible in the field or when using substrates like soil or potting mix. Alternative phenotyping techniques for such substrates are shovelomics, soil cores, and minirhizotrons. Takahashi and Pradal [2021] and Paez-Garcia et al. [2015] provide an overview of root phenotyping techniques, along with further details of root system architecture and its relevance. We provide a description of phenotyping techniques similar to RootTracker which are used to measure aspects of the root crown and root growth at relatively shallow depths, e.g. less than 30cm in depth.

Shovelomics involves excavating a hemisphere of roots, cleaning and imaging the sample, and then extracting root characteristics with software [Trachsel et al., 2011, York and Lynch, 2015, Das et al., 2015, Liu et al., 2021]. (A brief overview can be found on the Jonathan Lynch's website.) A major disadvantage of shovelomics is that it is destructive. Soil coring, which involves taking a soil core and then measuring the amount of organic matter [Böhm, 2012, Prior and Rogers, 1992], is also destructive.

A minirhizotron is essentially a camera in a clear tube that is buried in the soil [Johnson et al., 2001, Maeght et al., 2013]. Like shovelomics, minirhizotrons capture images that must be processed to extract root characteristics. Minirhizotrons introduce some confounding, since roots respond to impediments. The installation process is also problematic, since one must auger a hole to create space for the instrument.
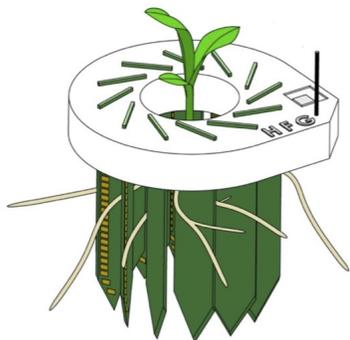
Figure 2: A graphic of RootTracker

RootTracker tries to overcome some of these challenges by non-destructively monitoring root growth with an array of sensors that are spaced in a cylindrical symmetry around a plant's root system. Figure 2 shows the appearance of RootTracker. The green "paddles" are printed circuit board (PCB). Along the outer edge of each paddle is a column of 22 electrodes. The centers of adjacent electrodes are about a half centimeter (cm) apart and have a range of 0.5 to 16.7 cm in depth. There are a total of 12 paddles, which are equally spaced at angles of 30 degrees.

RootTracker monitors voltages at each electrode that are passed through a detection algorithm to identify when a root passes close to a sensor. The algorithm also creates quality control metrics to assess when it is receiving valid signals. Noise

is introduced in at least two ways. First, it is possible to get false positive detections, as soil may shift near a sensor for reasons other than a root passing nearby. Second, roots may grow between paddles without growing near enough to sensors to trigger a detection. For the sake of tractability, we will assume that the false positive rate is low, and we will not try to distinguish between types of detections.

RootTracker can detect both crown roots and lateral roots, which are roots that branch off of crown roots. For the sake of developing a tractable model below, we will assume that all roots detected are crown roots, since as of this writing our detection algorithm does not distinguish between the types of roots detected. While the spatio-temporal pattern of the root detection may impart some information as to the type of root, introducing a hidden state characterizing the root type seriously complicates the task at hand. We leave that problem for later work.

2.3 Root models

There has been much work developing simulations of root growth, which is to say models that can realistically replicate root growth *in silico* [Lynch et al., 1997, Dupuy et al., 2010, Postma et al., 2017, Schäfer et al., 2022, Schnepf et al., 2018, Holzworth et al., 2014]. Such models contrast with statistical models of inference, whose purpose is to make inferences given limited data. While one could take a model designed for simulation and restrict its output to observed data to identify likely model parameters, this approach will carry a high computational cost, if it is even possible. (There are a few attempts along these lines using approximate Bayesian computation [Morandage et al., 2021]).

In contrast to the simulation-centric approaches, we start with the goal of inference and then try to find a minimal model that supports inferring model parameters and capturing important elements of root growth. To do this we reduce root growth to two elements: 1) the direction in which the root tip is growing and 2) the distance the root travels before changing direction. Under this setup, the we can capture and recapitulate gravitropism, one of the most prominent and longest recognized features of root growth [Rich and Watt, 2013]. Further, we can estimate the model parameters using observations of only a single point along a root trajectory.

3 Data

3.1 Methods

Data for this experiment was collected in a greenhouse at Alamance Community College, in Alamance County, NC. The aim was to compare early root growth for different species of plants. The treatments were maize, wheat, soybean, cotton, tomato, and a control (an empty pot). The fieldbook refers to maize as "corn" so we will use the terms maize and corn interchangeably. Five gallon pots were filled with soil. A RootTracker device was placed into the soil and then a seed was placed in the soil at the center of the RootTracker (see Figure 2). For a single treatment, 12 pots were arranged in a 4x3 grid on a bench to form what is akin to one replicate — though we do not pool data from that group into a single phenotype.

The locations of those replicates was placed in a randomized complete block design with 2 reps per treatment for a total of 24 RootTrackers per species. All pots were hand watered at the same time. RootTracker data was collected for 4 weeks.

3.2   RootTracker data structure

The most obvious representation of the data is the triplet of time, paddle, and depth at which a root is detected. Unfortunately, we also have to account for missing data, which can occur when quality control metrics are poor.

To that end, for a given device, we will consider the trajectory of matrices of root detections, $B_h \in \{0,1\}^{22 \times 12}, h = 0, \ldots, \Omega$, where h is indexed in hours and the response is an indicator if there was a detection or not at a given electrode-paddle combination within the hour. If, at any sensor, the quality control metric is poor for more than 10 minutes of an hour, then we assume the whole matrix of detections is missing — let $m_h \in \{0,1\}, t = 1, \ldots, \Omega$ indicate if at time h there is missing data.

We simplify this initial data further on the grounds that roots grow relatively slowly, hence we can still capture the most interesting root dynamics by aggregating to the daily level. Further, for data exploration and visualization purposes, it can be helpful to aggregate over adjacent electrodes as well. For the sake of generality, let ω be the period in hours, e.g. 24, over which to aggregate, let ν be the number of electrodes over which to aggregate, and let $H_t = \{h : \lfloor h/\omega \rfloor = t\}, t = 0, \ldots, T_\omega := \lfloor H/\omega \rfloor$ and $S_k = \{i : \lfloor i/\nu \rfloor = k\}, k = 0, \ldots, \lfloor 22/\nu \rfloor$. Then we can define the aggregated number of detections as

$$A_{tkj} = \max_{h \in H_t,\ i \in S_k} B_{hij},$$

so that $A_t \in \{0,1\}^{\lfloor 22/\nu \rfloor \times 12}$ is an indicator of if there was a detection on a given electrode-paddle combination. We use the max, instead of the sum, since when there are multiple detections at similar locations and times the detections are likely from the same root. This is usually not an issue and we only have a single detection within a space-time group, in which case the sum and the max yield the same aggregate value.

We can also aggregate the amount of non-missing data by letting

$$u_t = \sum_{h \in H_t} (1 - m_h).$$

It will be beneficial to partition our data into three parts: the number of detections in a day, the electrode (or depth) of detection, and the paddle on (or angle in cylindrical coordinates at) which the detection occurs. Let

$$c_t = \sum_{kj} A_{tkj}$$

be the number of detections within the day. Then let

$$F_t = \{(k, j, t) :\ A_{tkj} = 1\}$$

be the collection of electrode group-paddle-days where there are detections. We will conflate the set F_t as a matrix F_t of dimension $c_t \times 3$. We include the day so that we can stack the F_t to get a further matrix capturing the depth and angle information over time $F = (F_1, \cdots, F_T)$. We will model the number of root detected separately from where the detections occur.

3.3   Data exploration

We want to get a handle on what root growth looks like for this experiment — where does it occur and how does root growth change over time. To begin, we bin time and depth on a coarse grid and then compute average rates of root growth. In this case, average "rate" refers to the number of roots detected per unit of time (of non-missing data) and electrode area, averaged over all devices within a group.
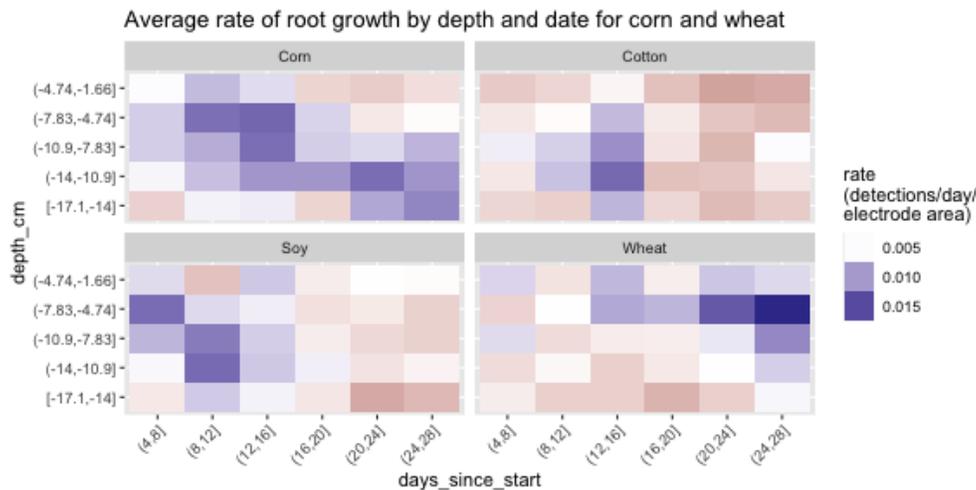
Figure 3: Heatmap showing the rates of root growth by depth and time for both corn and wheat.

Figure 3 shows the resulting rates of root growth by time and depth. For both corn and wheat, one sees greater rates at shallower depths initially, which then move lower over time. In contrast, for soybean and cotton, which have taproots (i.e. dicots not monocots), we see a different pattern of root growth in which there is a spike of growth in days 8-12 or 12-16, but little subsequent root growth. We will not dwell on these differences, since we are concerned with modeling monocot crown root growth, but they do show that the RootTracker is successfully distinguishing between species.

Restricting our attention to corn, we can plot the rates of root growth as a function of depth stratified by time, as seen in Figure 4. (We have flipped the axes to provide a visual sense of the distribution of root growth across depths.) The distributions are unimodal with the mode moving deeper over time.

We can refine these results using more sophisticated smoothing techniques. Figure 5 shows the typical rate of root growth by depth and time when we assume that the rate comes from a Gaussian process. Again, we see early root growth at shallower depths, which then moves deeper over time.
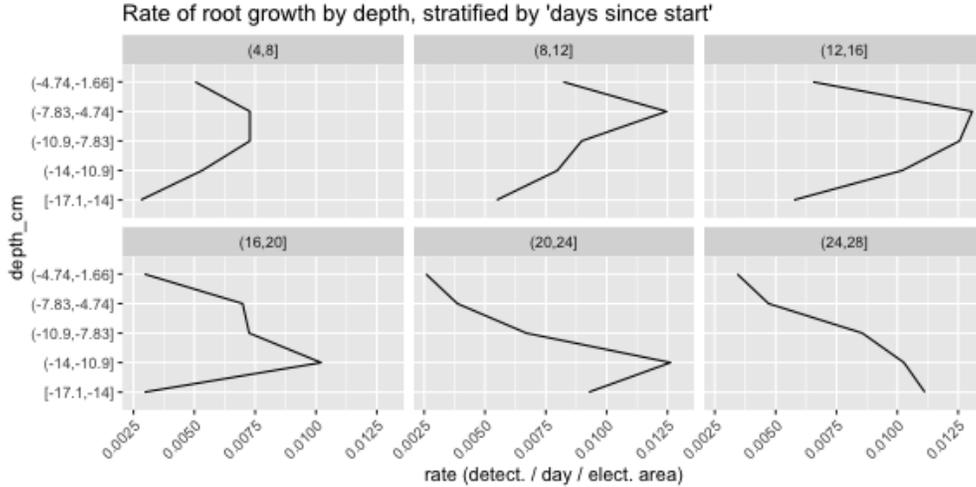
Figure 4: Rates of corn root growth as a function of depth and stratified by time group. The axes have been flipped so that depth has a more intuitive, vertical interpretation. These plots are analogous to unnormalized distributions of root growth by depth over time.
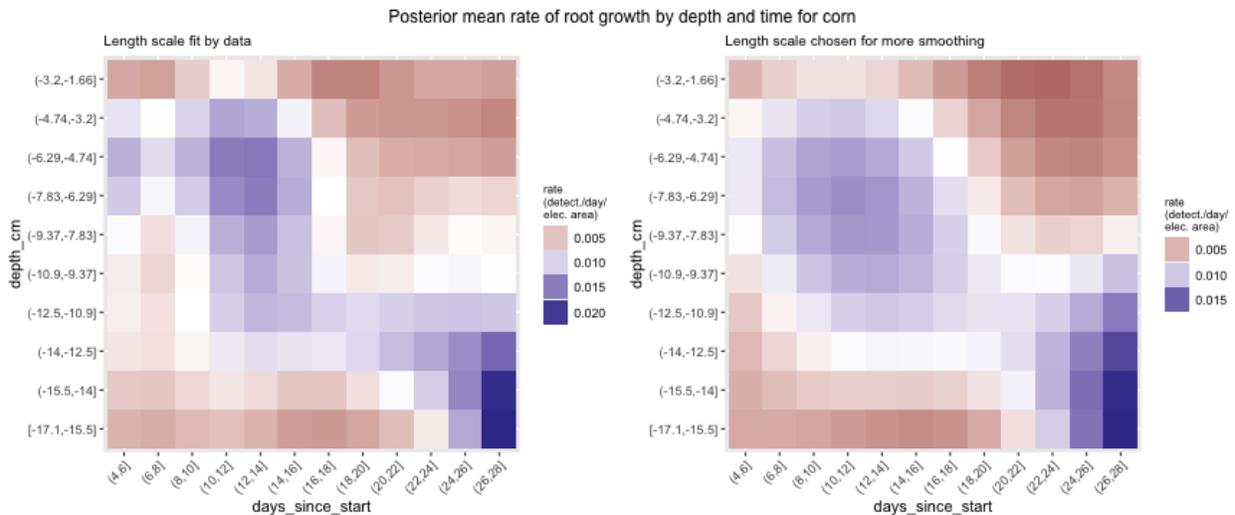
Figure 5: Rates of root growth fitted by a Gaussian process whose covariates are depth and time. We use the isotropic Euclidean distance (exponential squared distance kernel) for simplicity. The plot on the left corresponds to hyperparameters that are fit via cross validation. The plot on the right has hyperparameters that impose more smoothing.

4 Model

4.1 Model of root growth

We will consider growth in a vertical plane, to eliminate questions about growth side to side. In this simplified model, the root tip determines the direction of root growth, and the root grows linearly in that direction for a length (or time), before acquiring a new direction.

Thus, we can characterize root growth by a sequence of directions and lengths that piece together to recapitulate its trajectory.

In particular, let $\theta_i \in (-\pi/2, \pi/2)$ represent the angle (with respect to the horizontal) on piece i, and $\Delta\ell_i$ be the length traveled for that piece. In terms of changes in the xy-coordinates of the root, we have

$$(\Delta x_i, \Delta y_i) = \Delta\ell_i(\cos\theta_i, \sin\theta_i).$$
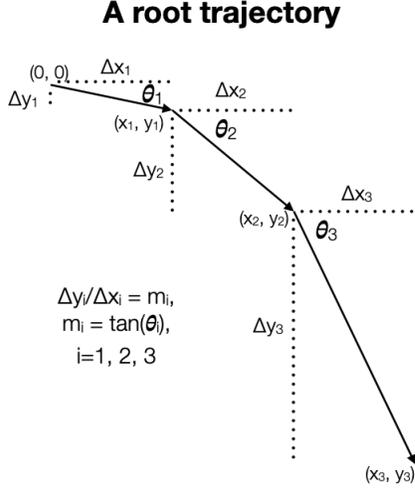
A root trajectory

Figure 6: A root trajectory. Under our model setup, the root travels in a piecewise linear fashion. The movement of each piece, $(\Delta x_i, \Delta y_i)$ is determined by the slope m_i or angle θ_i and distance traveld at each step $i = 1, 2, 3$.

While this parameterization makes sense in terms of our physical model of root growth, it will be convenient to construct a framework that accommodates our data. In particular, our data is such that we observe the depth of a root at a predetermined width (or radius). Thus, it will be helpful to derive a single expression for the observable depth.

To derive an expression for the observable depth a root reaches, we will employ Δx_i and the changes in slope Δm_i so that the slope on piece i is $m_i = \sum_{j=1}^{i} \Delta m_i$ and $\Delta y_i = m_i \Delta x_i$. In this case,

$$(\Delta x_i, \Delta y_i) = \Delta x_i(1, m_i).$$

Note, as a matter of modeling our root growth, we can now consider modeling the distributions of $\Delta x_i > 0$ and Δm_i, instead of θ_i and $\Delta\ell_i$. Connecting these quantities via our expression for Δx_i and Δy_i, we have that

$$m_i = \tan(\theta_i),$$
$$\Delta x_i = \Delta\ell_i \cos\theta_i,$$

and

$$\Delta m_i = \sum_{j=1}^{i} (-1)^{j+i} \tan\theta_j.$$

Let us assume that the root will take on K changes before continuing indefinitely in the direction of the $(K+1)$th piece. For $\Delta x_i, > 0$, $i = 1, \ldots, K$, the x-coordinate of the root at step i is $x_i = \sum_{j=1}^{i} \Delta x_j$, the change in depth is $\Delta y_i = m_i \Delta x_i$, and the y-coordinate is $y_i = \sum_{j=1}^{i} \Delta y_j$. By convention we will say that $(x_0, y_0) = 0$, so that $(x_i, y_i)_{i=0}^{n}$ define a piecewise linear path from the origin, which we can then extend outwards with slope m_{K+1} indefinitely.

To find the value of y, given a certain horizontal coordinate $x = r$, we can interpolate. There are $K + 1$ possible cases:

1. $x_{i-1} < r \leq x_i$: the root goes past (or up to) r on its ith step, $i = 1, \ldots, K$;

2. $x_K < r$: the root goes past r after its Kth step — assuming the root travels indefinitely from that point on at a slope of m_{K+1}.

For the kth case, the value of $y(r)$ will be

$$y = y_{k-1} + (r - x_{k-1})m_k. \tag{1}$$

We can modify (1) to find that

$$y = \left[\sum_{i=1}^{k-1} m_i \Delta x_i\right] + (r - x_{k-1})m_k = \sum_{i=1}^{k-1}\sum_{j=1}^{k-1} \Delta m_j \Delta x_i \mathbb{I}\{j \le i\} + \left(r - \sum_{i=1}^{k-1} \Delta x_i\right)\sum_{j=1}^{k} \Delta m_j$$

$$= \sum_{j=1}^{k-1} \Delta m_j \left[r - x_{j-1} + \sum_{i=j}^{k-1}\left(\Delta x_i - \Delta x_i\right)\right] + (r - x_{k-1})\Delta m_k$$

$$= \sum_{j=1}^{k} \Delta m_j (r - x_{j-1})$$

Thus, regardless of the specific case, we can write

$$y = \sum_{j=1}^{K+1} \Delta m_j (r - x_{j-1})^+. \tag{2}$$

The expression

$$\lambda(\Delta x)_j = (r - x_{j-1})^+, j = 1, \dots, K + 1, \tag{3}$$

comes up often, so we define it here. Now we can succinctly say, for instance, that $y = \lambda' \Delta m$, where λ is implicitly evaluated at Δx.

Alternatively, if we want to find an inner product in terms of the slopes, we can construct

$$\delta(\Delta_x)_j = \begin{cases} \Delta x_j, & x_j < r \\ (r - x_{j-1})^+, & x_j \ge r, \end{cases} \quad j = 1, \dots, K + 1 \tag{4}$$

so that $y = \delta' m$ where δ is implicitly evaluated at Δx.

4.2   Log depths

Because we are modeling depths, which are always below zero, we may also want to consider the same model above on the log scale, like $\log(-y)$. However, now we also need to consider where the root starts. Previously, we have assumed the root starts at zero, but this does not make sense from the standpoint of the log transform, since that corresponds to $-\infty$. Thus, it makes more sense to consider the origin of the root to be slightly below zero at y_0. In that case, to apply the same model as above, but on a log-like scale we would consider the transformed response $\tilde{y} = \log(-y) - \log(-y_0)$. We can invert this transform via $y_0 \exp(\tilde{y})$, or if $\tilde{y}_0 = \log(-y_0)$, then $-\exp(\tilde{y} + \tilde{y}_0)$.

If we consider the paths to be straight lines on the log scale, then they will be curved when returning to the original scale.

5 Identifiability

We will need to restrict the model above in order to have efficient Markov Chain Monte Carlo (MCMC) sampling. If we do not, then we run into issues of identifiability and flat regions in the posterior.

To provide a brief sketch of the issue, before we explore it more fully, let us approximate the trajectory of a root. To that end, suppose that neither $\Delta x_i, i = 1, \ldots K$ nor $\Delta m_i, i = 1, \ldots, n = K + 1$ varies much and $\Delta x_i \simeq \mu_{dx} = r/(K+1)$ and $\Delta m_i \simeq \mu_{dm}$. Then, using the fact that $x_n \simeq n\mu_{dx}$, we find that the depth of the root at x_n is:

$$y_n \simeq \sum_{i=1}^{n} \mu_{dx}(i\mu_{dm}) \simeq \frac{1}{2}x_n^2 \frac{\mu_{dm}}{\mu_{dx}}.$$

Hence, the curvature of the root trajectory is determined by the ratio μ_{dm}/μ_{dx} and is invariant along these level sets. Even if we change the number of kinks along the root prior to r this relationship holds at r, so that the quantify of interest is the ratio, not either mean parameter individually. This phenomenon can also be seen with numerical experiments, even when we allow Δx_i and Δm_i to vary, as seen in the next section.

5.1 Parameter exploration

For the sake of our numerical experiments, we need to place data generating distributions on Δx_i and Δm_i, which are easier to work with than θ_i and $\Delta \ell_i$.

In particular, we will assume that $\Delta x_i, i = 1, \ldots, K$ and $\Delta m_i, i = 1, \ldots, K + 1$ are independent and (truncated) Gaussian:

$$\Delta x_i \sim N_{(0,r)}(\mu_x, \sigma_x^2), \ i = 1, \ldots, K,$$

and

$$\Delta m_i \sim N(\mu_m, \sigma_m^2), \ i = 1, \ldots, K + 1.$$

Here we enforce having at least one kink in a root by insisting that $\Delta x_i \in (0, r)$. Letting Δm_i have some chance of taking on positive values allows the root to turn shallower at times.

Figure 7 shows how the parameter valus impact the mean and standard distribution of the depth. For both the mean and the standard deviation, one finds that there are level sets within the parameter space that leave the mean or standard deviation nearly invariant. In other words, one can arrive at similar looking distributions for different parameter values. Thus, the posterior will not concentrate around a single point, but rather along a surface, which makes posterior computations more difficult [Betancourt, 2016].

5.2 Non-identifiability

The numerical experiments above, suggest that our posterior can have flat regions. Here we show, specifically, how that comes about. For this section, we will drop the assumption that Δx is truncated, which simplifies the argument. If we assume that Δx is sufficiently unlikely to be below zero or above r, then the same argument holds, approximately, in the truncated case.
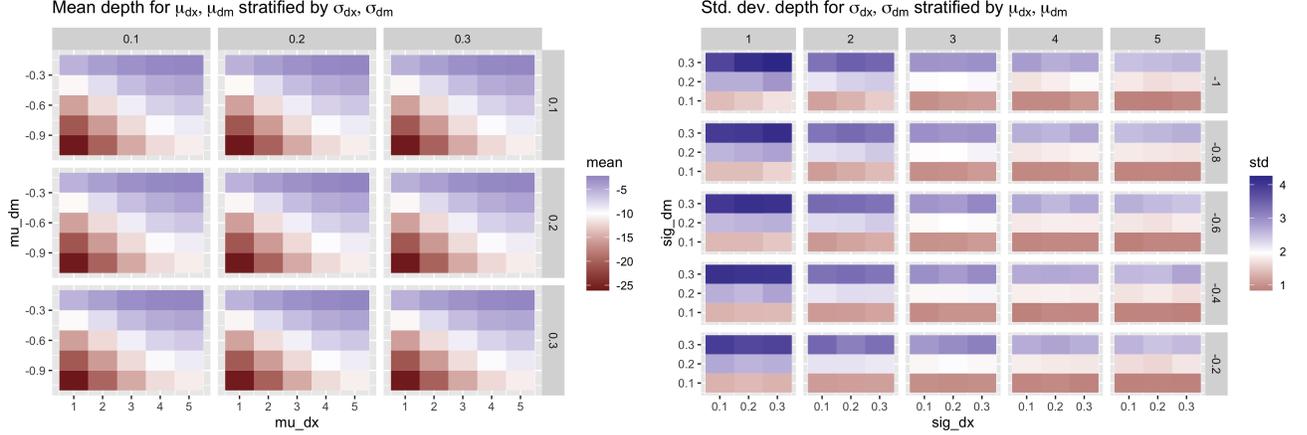
Figure 7: Left: a heatmap of the mean depth by μ_{dx} and μ_{dm}, stratified by σ_{dx} and σ_{dm}. There are level sets along a roughly linear, inverse relationship between μ_{dx} and μ_{dm} and this holds across the various standard deviations. Thus, in terms of the mean, μ_{dx} and μ_{dm} are not well identified. Right: a heatmap of the standard deviation by σ_{dx} and σ_{dm} stratified by μ_{dx} and μ_{dm}. The strongest changes are due to σ_{dm}. As it increases, the variation in the depth distribution markedly increases. On a secondary basis, μ_{dx} impacts the variability of the distribution, with more variation occurring when μ_{dx} is smaller, and hence there are more kinks in the root. Thus, one can trade some σ_{dm} for μ_{dx} and keep the standard deviation of the distribution similar. Given, the interplay of μ_{dx}, μ_{dm}, and σ_{dm}, it is possible to arrive at similar looking distributions for different parameter values.

Let us factor the distribution for depth as

$$\underbrace{p(y|\Delta x, \phi)}_{\text{term 1}} \underbrace{p(\Delta x|\mu_{dx}, \sigma_{dx})}_{\text{term 2}} \underbrace{p(\phi)}_{\text{prior}}.$$

For term 1, from (2) and (3), we have

$$y = \Delta m' \lambda(\Delta x).$$

Hence

$$(y|\Delta x, \phi) \sim N(\mu_{dm}\lambda'\mathbf{1}, \sigma_{dm}^2\lambda'\lambda).$$

For two different points, $\Delta x^{(1)}$ and $\Delta x^{(2)}$, and two different parameter sets, $(\mu_{dm}^{(1)}, \sigma_{dm}^{(1)})$ and $(\mu_{dm}^{(2)}, \sigma_{dm}^{(2)})$, the probability densities are identical if their respective parameters satisfy

$$\mu_{dm}^{(1)} \lambda(\Delta x^{(1)})'\mathbf{1} = \mu_{dm}^{(2)} \lambda(\Delta x^{(2)})'\mathbf{1}$$

and

$$\sigma_{dm}^{(1)} \|\lambda(\Delta x^{(1)})\| = \sigma_{dm}^{(2)} \|\lambda(\Delta x^{(2)})\|.$$

For term 2, fix σ_{dx} to any value. The respectively likelihoods for $\Delta x^{(1)}$ and $\Delta x^{(2)}$ are then identical when

$$\|(\Delta x^{(1)} - \mathbf{1}\mu_{dx}^{(1)})\| = \|(\Delta x^{(2)} - \mathbf{1}\mu_{dx}^{(2)})\|.$$

11

Completing the square for the respective parameters, we see that the level set is a hyperbola defined by

$$\left(\mu_{dx}^{(1)} - \overline{\Delta x}^{(1)}\right)^2 - \left(\mu_{dx}^{(2)} - \overline{\Delta x}^{(2)}\right)^2 = \text{var}(\Delta x^{(2)}) - \text{var}(\Delta x^{(1)})$$

where the overline denotes the empirical mean and var denotes the empirical variation across the K entries of the vector.

For an arbitrary pair of points, we have constructed levels sets in the parameters space for the product of terms 1 and 2. Thus, if the prior term is uninformative, the corresponding posterior will not concentrate around a single point in the parameter space, but rather have high probability on a surface within the parameter space.

5.3   Further constraints needed

The implication of all of this is that we cannot learn μ_{dx}, μ_{dm}, σ_{dx}, and σ_{dm} by only observing the depth the root reaches at radius r. This might be possible, if we had some additional information, regarding, for instance Δx, but we do not.

This is in principle acceptable. The posterior distribution, instead of concentrating around a point, will concentrate on a surface. If we were to think about simulating root trajectories from the posterior predictive distribution, we would get trajectories that might be different in the sense that they have differing numbers of kinks prior to being detected, but they would all result in similar distributions of root detections by depth at a given radius. The problem is that having an entire surface of likely parameter values slows the convergence of our HMC sampler and leads to poor convergence diagnostics.

To get around this problem, we will simply fix Δx_i at a constant value. In other words, we will make a very strong assumption about how often the root tip changes direction. Keep in mind, our primary goal is to recapitulate plausible root growth. To that end, the exact distance the root travels before making a change in direction is not critical, even though it would be nice not to have to make such a strong assumption.

6   Root trajectory process

Now that we have accepted that we need to fix Δx_i (or $\Delta \ell_i$), we can try to find a reasonable choice for the distribution of $\Delta m_i, i = 1, \ldots, K + 1$. There are several competing desires when trying to find a good data generating process for Δm. We want to

- recapitulate the distribution of root growth by depth;
- prevent unrealistic root growth, i.e. prevent roots from growing above the ground or too steeply; and
- produce realistic looking root trajectories.

Bounding the growth of roots is somewhat problematic. While we can only measure root detections to a certain depth (at a certain radius) that is not to say that there are no roots growing below that point. Conversely, the radius at which roots are detected is not that great, so perhaps it is reasonable to assume that there is little root growth below the deepest sensor. Truncating a sum of random variables is challenging since this is not necessarily

easily encoded in Stan, our software of choice for posterior inference [Stan Development Team, 2023].

We consider several options for modeling Δm_i or θ_i, $i = 1, \ldots, K + 1$. The details for these models can be found in Appendix A.

M08 (Normal, truncated) For all roots, each $\Delta m_i, i = 1, \ldots K + 1$ is independent and identically normally distributed with the restriction that $\ell < \Delta m' \lambda < u$ where l is a lower bound and u is an upper bound to root growth at radius r.

M09 (Correlated normal, truncated) For a given root, $\Delta m_i, i = 1, \ldots K + 1$ is independent and identically normally distributed, but the mean of that distribution is randomly generated for each root. The subsequent observed root depth is restricted so that $\ell < \Delta m' \lambda < u$ as above.

M10 (Skew normal) For all roots, each $\Delta m_i, i = 1, \ldots K + 1$ is independent and identically skew normally distributed.

M14 (Normal, truncated on the log scale) Like the normal, truncated model, but modeling log depths.

M21 (Normal on a log-odds scale) We use a sigmoid transform s so that $\theta_i = s(z_i)$ where $z_i = z_0 + \sum_{j=1}^{i} \Delta z_i$ and Δz_i are independent and identically distributed over all roots.

As you can see by the model numbers, we have tried many models and different parameterizations for each. For instance, the skew-normal model has much better MCMC convergence properties when parameterized as mean, standard deviation, and shape, rather than a location, scale, shape parameterization. We keep the original model number to match our Git repository code [Windle, 2023].

6.1 Performance

To compare the models we broke the observed root data into four epochs by time. For each model and epoch, we computed the mean absolute error (MAE) between the posterior predictive distribution and the empirical probability — the posterior predictive distribution was discretized using the depth bins used to describe the electrode locations and any weight outside of those bins was discarded. To create a single numeric summary, we computed the mean MAE (MMAE) over epochs. Figure 8 shows the probability mass functions for corn. We had tried to compare the models by Bayes factors using bridge sampling [Gelman and Meng, 1998, Gronau et al., 2017], but seemingly due to the high number of latent variables, the computation could not complete in a timely manner for some comparisons. We also inspected the trajectories to see if roots traveled to unrealistic values, i.e. above ground, and if they could pass for real root trajectories.

M08, M14, and M21 have similar MMAE performance, followed by M10. M09 had variable performance. Though M10 had slightly worse performance that M08, M14, and M21, the lack of truncation forced the parameters to fit the mode of the empirical distribution well, which is a critical element to capture. Further, its root trajectories look reasonable. Thus, for the purposes of both matching the empirical distribution and producing compelling root trajectories, M21 and M10 seem to be the best choices. The numeric metrics across the
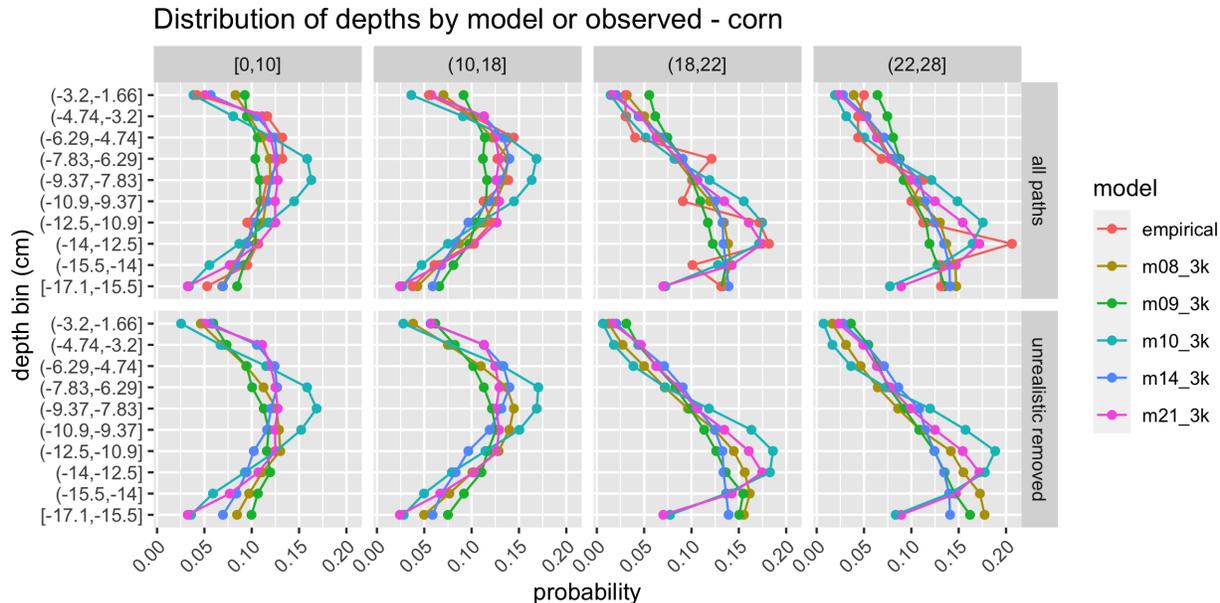
Figure 8: The empirical distribution of depth along with the discretized posterior predictive distributions of depth for each model. The top panels correspond the posterior predictive distribution when no unrealistic paths are removed. The bottom panels correspond to when unrealistic paths are removed. Removal of unrealistic paths can have and adverse impact. For instance, for the first epoch $[0, 10]$, M08 and M09 have much lower probability at shallower depths after the removal of unrealistic paths. For each model the x-step size is 2, which corresponds to 3 "kinks" before reaching the radius $r = 8$, hence the "3k" suffix for each model in the legend.

models for corn and wheat can be found in Appendix B, Table 1. A more detailed summary is as follows.

M08 M08 had good performance, but it has root trajectories that often go above zero. Removing those trajectories in the posterior predictive computation degraded the MMAE and qualitatively shifted the look of the posterior predictive depths when the depth distribution favored shallow roots. Another downside is that when simulating from the posterior predictive distribution without a lower bound, roots could travel very deep. The MCMC sampler was relatively quick with good performance diagnostics.

M09 More correlation in Δm_i had the intended effect of reducing the number of unrealistic paths. When the the parameter γ as described A.2 was set to 0.5, the performance was fair. However, M09 had poor performance when γ was high. Further, this seems to coincide with worse MCMC performance.

M10 M10 had fair performance. Because there was no truncation imposed here, fitting the model makes the implicit assumption that root growth drops off completely outside of the sensor range. As a consequence, the model tends to underestimate the amount of root growth at the edges of the paddle. However, an advantage of this is that the

model overshoots at the peaks of the empirical distribution, leading to a more decisive appearance in terms of where the roots are growing. Unlike M08 and M09 you cannot easily marginalize out the $K + 1$ latent variables.

M14 M14 had good performance. The results are like M08 with one exception. While one need not worry about unrealistic root growth above ground by modeling on the log scale, it does lead to what look like unrealistic root trajectories. Further, the problem of excessively deep root growth when removing the lower truncation, as seen in M08, is exacerbated.

M21 M21 had good performance. M21 tended to do a better job recapitulating the distribution of root growth when that root growth was concentrated at depth compared to M08 and M09. The trajectories look reasonable, though they can concentrate as a straight line of root growth owing to the sigmoid transformation. The model is more complicated than M08, M09, or M10 and the MCMC convergence is worse than M08 or M10.

7 Modeling root emergence

We know that crown roots emerge periodically in groups called whorls. One could consider a mixture model whose components represent whorls, but we avoid that approach for the following reason. While our focus is on crown roots, RootTracker detects roots from both crown roots and lateral roots, which are roots that branch off of crown roots. Thus, it may be unreasonable to assume that our detections will possess a very strong periodic pattern as suggested by a whorl model. That being said, if the lateral roots that are detected are relatively close to the crown root from which they emerge, it is still reasonable to use those detections to represent crown root paths — it is just that the timing of the detection will be perturbed. However, we still do want to leave open the possibility of capturing temporal patterns if they exist. For these reasons, we model the root growth using a Gaussian process with a periodic kernel. We adopt the following model.

From our description of the data structure in Section 3.2, for each device $i, i = 1, \ldots, D$, we aggregate $A_t^{(i)}$ to the daily level and then assume that the number of detections within day t for device i, which corresponds to the nth observation, is binomially distributed with probability p_n and trial size $2u_n$:

$$c_n \sim \text{Binom}(p_n, 2u_n).$$

The factor of 2 comes from the assumption that there we do not expect to have more than 2 detections within an hour period. When we have no missing data $u_n = 24$ so that the maximum trial size is 48. We model p_n on the log-odds scale z_n so that

$$z_n = \log\left(\frac{p_n}{1 - p_n}\right)$$

and for a single group, i.e. species, z_n follows the model

$$z_n = \mu + f_{t(n)} + \varepsilon_n, \ \varepsilon_n \sim N(0, \sigma_{obs}^2)$$

where $t(n)$ is the time, in days, of the nth observation and $f \in \mathbb{R}^T$ comes from a mean zero Gaussian process whose covariance function is

$$K(s,t) = \sigma_K^2 \exp\left[-\frac{2\sin^2\left(\pi\frac{|s-t|}{\tau}\right)}{\ell^2}\right] + \sigma_N^2 \delta_s(t).$$

Our prior choices are summarized in the following table.

Param	Prior
σ_{obs}	$N_{[0,\infty)}(0,1)$
μ	$N(-3, 1.5^2)$
σ_K	$N_{[0,\infty)}(0,1)$
ℓ	$\mathrm{InvGamma}(5,5)$
τ	$\mathrm{Unif}[7,21]$
σ_N	$N_{[10^{-4.5},\infty)}(0,1)$

The prior on μ was chosen based on our assumption that we are unlikely to observe more than two roots per day, and we strongly expect the probability of a root emerging to be less than 0.5 when the trial size is 48.

The posterior distribution of f and p is summarized in Figure 9 and discussed further therein. The posterior was generated using Stan with 3 chains that each ran for 7000 iterations with 4000 warmup iterations, leading to 9000 total samples. Of the underlying parameters (excluding the latent variables), σ_{obs} had the lowest effective sample size of just over 500, with the second lowest being σ_N at about 2100. The Stan program was run with the default parameters, except for `adapt_delta`, which we set to 0.99 to avoid divergent transitions that had a tendency to occur at lower settings.

8 Dynamic model parameters

Instead of breaking the root data into epochs and fitting each epoch separately, as we did in Section 6.1, we would like to make the model parameters dynamic and allow them to change smoothly in time. A common way to do this is to assume that the dynamic parameters come from a Gaussian process. Any of the models we have considered can accommodate this approach.

As an example, we introduce dynamic parameters to M10, which models the changes in slope as a skew normal distribution. The parameters of the skew normal innovations, as described in Section A.3 are μ, σ, and α or some transformed version thereof, e.g. $\tilde{\mu}$. Let us consider the mean parameter $\tilde{\mu}$ — we will drop the tilde for notational convenience. In Section 6.1 we effectively modeled that as

$$\mu_t = \mu_j^{(0)}, \quad \text{for } t \text{ is in epoch } E_j$$
$$\mu_j^{(0)} \sim N(d, (d/2)^2), \ j = 1, \dots, 4.$$

To allow this parameter to change on a daily basis we can modify the above to

$$(\mu_t)_{t=1}^T \sim N(\mu^{(0)}\mathbf{1}, \sigma^2 K)$$
$$\mu^{(0)} \sim N(d, (d/2)^2)$$

16
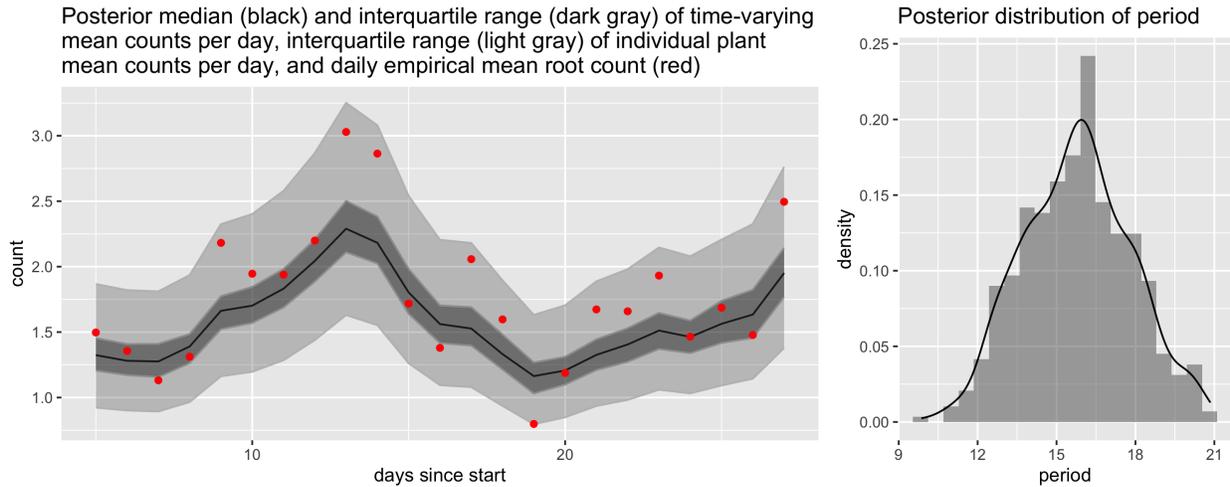
Figure 9: Left: posterior summary of mean count levels; right posterior distribution of the period parameter. The mean count level refers to the mean of the binomial distribution from which the counts are sampled and is equal to the the probability of "success" $\times 48$. There are two quantities plotted, both transformed from the log-odds scale back to the binomial mean scale: the posterior median and interquartile range of $\mu + f_t$ and the interquartile range of $\mu + f_t + \varepsilon$, $t = 1, \ldots, T$ where the residual ε is taken in a predictive sense, i.e. for a new plant and not one for which we have data. The empirical daily mean is plotted in red. The period is concentrated around 15 days.

17

where
$$K(s,t) = \sigma_K^2 \exp\left[-\left(\frac{s-t}{\ell}\right)^2\right], \; s,t = 1,\dots,T.$$

Similar accommodations can be made for σ and α. This leads us to:

M23 M10 with dynamic mean, standard deviation, and shape parameters that change smoothly in time (days).

We have implemented M23 in Stan using priors similar the model of counts described perviously and used for our other models, e.g M10. Details can be found in the Git repo [Windle, 2023].

For the purposes of fitting the model let t be time in days, c_t denote the number of detections on day t, d_{it} denote the ith detection depth on day t, and a_{it} denote the angle of the ith detection on day t. Assume the density takes the form

$$\prod_{t=1}^{T}\left[\prod_{i=1}^{c_t} p(d_{it}|\theta_d,t)p(a_{it}|\theta_a,t)\right]p(c_t|\theta_c,t)p(\theta_d)p(\theta_a)p(\theta_c)$$

where θ_d are the parameters related to depth, θ_a are the parameters related to the angle, and θ_c is the parameter related to the number of detections recorded. This structure lets us fit the time component, the depth component, and the angle component separately. Since our focus in this paper is modeling root trajectories and the distribution of detections by depth, we have simply assumed a_{it} comes from a uniform distribution. (In future work, we can explore how to take advantage of this information.)
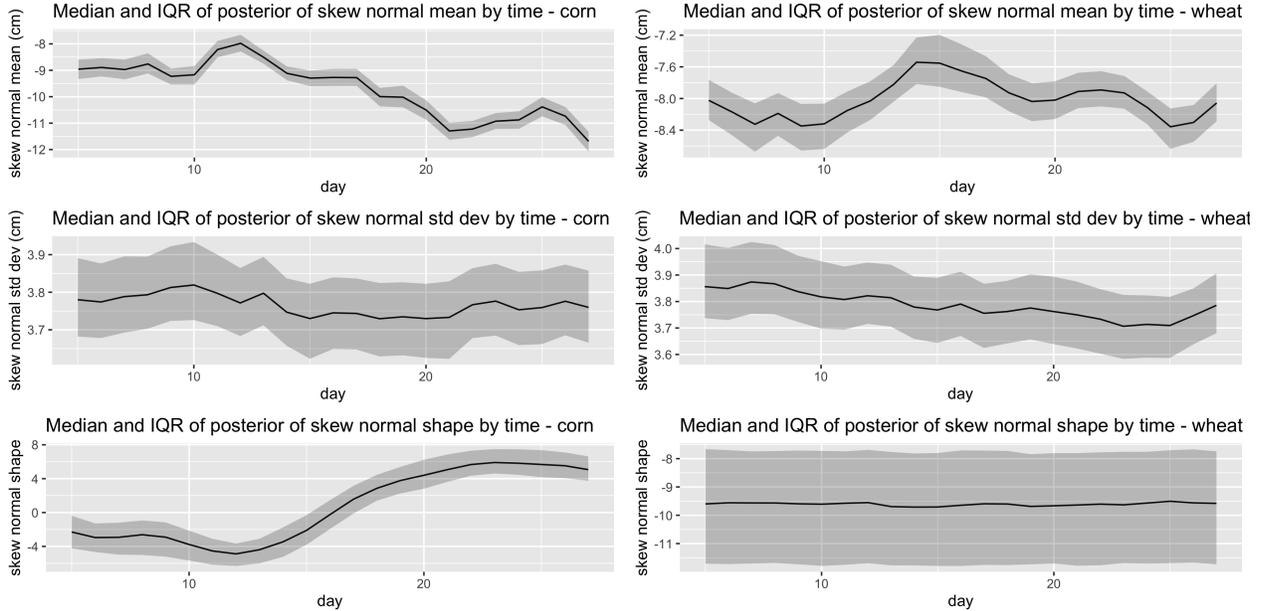
Figure 10: M23 parameters over time for corn and wheat.

Figure 10 shows the posterior distribution of the model parameters in time for corn and wheat. The mean, which has been transformed to reflect the mean observed depth, changes

noticeably in time for corn. This reflects what we observed in Figures 3 and 5 where the distribution appears to move deeper in time. We also see the shape of the distribution changing, which makes sense as the distribution moves from left skewed to right skewed over the interval of interest. We see a similar movement downward for wheat, but only from day 13 to 28. This is not totally surprising since there is very little wheat root growth initially, as seen in Figures 3 and 13. Under the model above, when there is little information about the distribution by depth, as is the case when there are few root detections, then a dynamic parameter will revert to the global average. We might be able to overcome this problem by introducing a trend into the model.
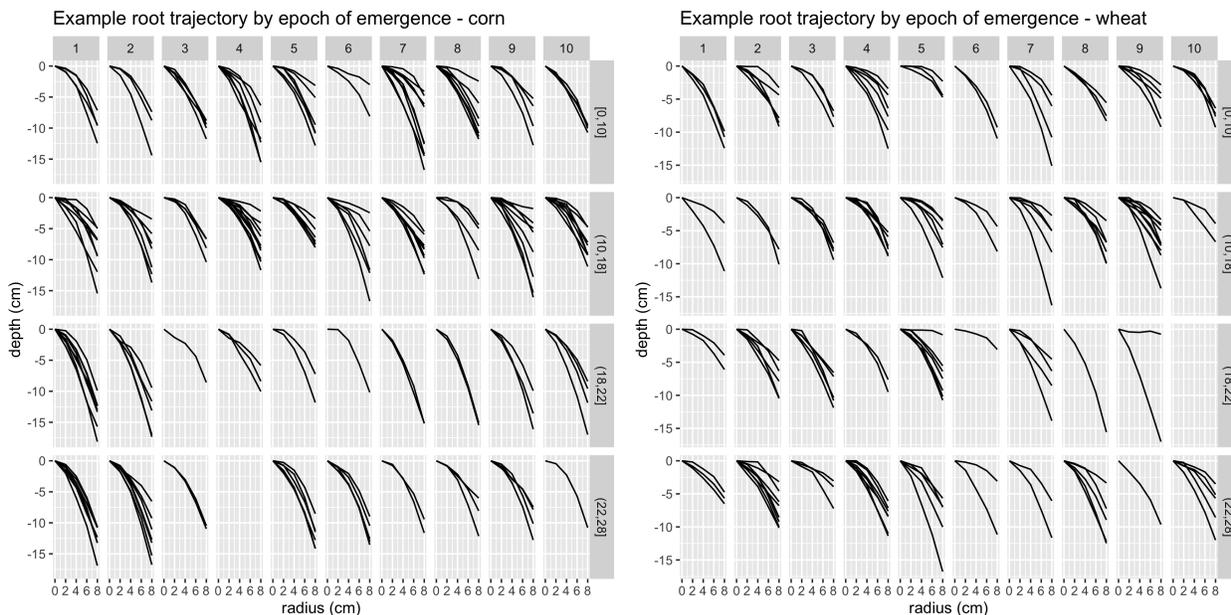
Figure 11:

To get a sense of what the root trajectories look like we simulate from the posterior predictive distribution and plot them in Figure 11. A few modifications were made to clean up the images. We excluded unrealistic trajectories, which is justified since we know that this has little impact on the subsequent distribution by Table 1. We also halved the number of trials in the binomial distribution from 48 to 24. As noted previously, the RootTracker device can detect both crown roots and lateral roots. We have assumed that all detections are crown roots to this point. However, it is likely some of these detections are lateral roots that grew off of a nearby crown root. Reducing the number of trials reduces the number of crown roots generated, reflecting the previous point, and making the images more clear in the process.

Examining the root trajectories in Figure 11, we can see similar features as suggested by Figures 10, 9, and 13. Corn tends to produced more root detections, except perhaps at the very end of the trial, and it seems to have deeper root distributions later in the experiment.

19

9 Conclusion

Herein we have considered a parsimonious model of monocot root growth. The models we considered were necessarily simple, since the data we have to fit these models is limited — we only observe one point along the trajectory of each root. Despite this simplicity, we were able to adequately capture the distribution of roots by depth while also producing plausible root trajectories.

There are many further avenues to explore, including the following.

- Fixed length of root growth. Herein we took advantage of a fixed horizontal step size when modeling root trajectories. That has the advantage of computational simplicity, but the disadvantage of being less realistic than perhaps having a fixed piecewise length of root growth. As shown in Section 4, we can accomodate this change, though it may adversely impact the sampling efficiency.

- Missing data. In some of our models, we have not acknowledged that there is missing data. And in other models where we have, simulations from the non-truncated distributions is not realistic. It would be preferable for the models to both reflect that there is missing data while also generating realistic posterior predictive distributions.

- Model both crown roots and lateral roots. Herein, we modeled all roots as crown roots. However, we know RootTracker also detects lateral roots. Given strong assumptions about when whorls of crown roots emerge and the likelihood of having detections in similar regions of space and time, it might be possible to infer what type of root is being detected.

- Angle of detection. We focused on recapitulating distributions by depth, but we have not conducted a detailed analysis of the angle of detection, which could be useful for flagging lateral roots.

- Capturing plant-to-plant variation or other partitions of variation. For the sake of simplicity, we have assumed that all e.g. corn plants in this experiment come from the same distribution (in a given epoch or on a given day). In agricultural experiments it is common to stratify by rep, location, or other factors, which would correspond to introducing a hierarchy for the model parameters. The results from M23 suggests that this is possible.

Code for replicating our results can be found at this paper's companion Git repository [Windle, 2023].

References

J. J. Aguilar, M. Moore, L. Johnson, R. F. Greenhut, E. Rogers, D. Walker, F. O'Neil, J. L. Edwards, J. Thystrup, and S. Farrow. Capturing in-field root system dynamics with RootTracker. *Plant physiology*, 187(3):1117–1130, 2021. ISSN 0032-0889.

M. Betancourt. Diagnosing suboptimal cotangent disintegrations in Hamiltonian Monte Carlo. *arXiv preprint arXiv:1604.00695*, 2016.

W. Böhm. *Methods of studying root systems*, volume 33. Springer Science & Business Media, 2012. ISBN 3642672825.

A. Das, H. Schneider, J. Burridge, A. K. M. Ascanio, T. Wojciechowski, C. N. Topp, J. P. Lynch, J. S. Weitz, and A. Bucksch. Digital imaging of root traits (DIRT): a high-throughput computing and collaboration platform for field-based root phenomics. *Plant methods*, 11(1):1–12, 2015. ISSN 1746-4811.

L. Dupuy, P. J. Gregory, and A. G. Bengough. Root growth models: towards a new generation of continuous approaches. *Journal of experimental botany*, 61(8):2131–2143, 2010. ISSN 1460-2431.

A. Gelman and X.-L. Meng. Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical science*, pages 163–185, 1998. ISSN 0883-4237.

Q. F. Gronau, A. Sarafoglou, D. Matzke, A. Ly, U. Boehm, M. Marsman, D. S. Leslie, J. J. Forster, E.-J. Wagenmakers, and H. Steingroever. A tutorial on bridge sampling. *Journal of Mathematical Psychology*, 81:80–97, 2017. ISSN 0022-2496. doi: https://doi.org/10.1016/j.jmp.2017.09.005. URL https://www.sciencedirect.com/science/article/pii/S0022249617300640.

F. Hochholdinger. The maize root system: morphology, anatomy, and genetics. *Handbook of maize: Its biology*, pages 145–160, 2009. ISSN 0387794174.

A. Hodge, G. Berta, C. Doussan, F. Merchan, and M. Crespi. Plant root growth, architecture and function, 2009.

D. P. Holzworth, N. I. Huth, P. G. DeVoil, E. J. Zurcher, N. I. Herrmann, G. McLean, K. Chenu, E. J. van Oosterom, V. Snow, and C. Murphy. APSIM–evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software*, 62:327–350, 2014. ISSN 1364-8152.

M. G. Johnson, D. T. Tingey, D. L. Phillips, and M. J. Storm. Advancing fine root research with minirhizotrons. *Environmental and Experimental Botany*, 45(3):263–289, 2001. ISSN 0098-8472.

S. Liu, C. S. Barrow, M. Hanlon, J. P. Lynch, and A. Bucksch. DIRT/3D: 3D root phenotyping for field-grown maize (Zea mays). *Plant Physiology*, 187(2):739–757, 2021. ISSN 0032-0889.

J. P. Lynch, K. L. Nielsen, R. D. Davis, and A. G. Jablokow. SimRoot: modelling and visualization of root systems. *Plant and Soil*, 188:139–151, 1997. ISSN 0032-079X.

J.-L. Maeght, B. Rewald, and A. Pierret. How to study deep roots—and why it matters. *Frontiers in plant science*, 4, 2013.

S. Mairhofer, S. Zappala, S. Tracy, C. Sturrock, M. J. Bennett, S. J. Mooney, and T. P. Pridmore. Recovering complete plant root system architectures from soil via X-ray μ-computed tomography. *Plant methods*, 9(1):1–7, 2013. ISSN 1746-4811.

S. Morandage, E. Laloy, A. Schnepf, H. Vereecken, and J. Vanderborght. Bayesian inference of root architectural model parameters from synthetic field data. *Plant and soil*, 467: 67–89, 2021. ISSN 0032-079X.

R. Munns, S. Schmidt, and C. Beveridge, editors. *Plants in Action*. Australian Society of Plant Scientists, 2 edition, 2018. URL `https://www.asps.org.au/plants-in-action-2nd-edition-pdf-files`.

S. Nadarajah and R. Li. The exact density of the sum of independent skew normal random variables. *Journal of Computational and Applied Mathematics*, 311:1–10, 2017. ISSN 0377-0427. doi: https://doi.org/10.1016/j.cam.2016.06.032. URL `https://www.sciencedirect.com/science/article/pii/S0377042716303041`.

R. B. Nielsen. Root Development in Young Corn, 2020. URL `https://www.agry.purdue.edu/ext/corn/news/timeless/roots.html`.

A. Paez-Garcia, C. M. Motes, W.-R. Scheible, R. Chen, E. B. Blancaflor, and M. J. Monteros. Root traits and phenotyping strategies for plant improvement. *Plants*, 4(2):334–355, 2015. ISSN 2223-7747.

L. Pages. Root system architecture: analysis from root systems to individual roots. *Encyclopedia of Agrophysics. Springe r Netherlands, Dordrecht*, 2011.

A. Pewsey. Problems of inference for Azzalini's skewnormal distribution. *Journal of applied statistics*, 27(7):859–870, 2000. ISSN 0266-4763.

J. A. Postma, C. Kuppe, M. R. Owen, N. Mellor, M. Griffiths, M. J. Bennett, J. P. Lynch, and M. Watt. OpenSimRoot: widening the scope and application of root architectural models. *New Phytologist*, 215(3):1274–1286, 2017. ISSN 0028-646X.

S. A. Prior and H. H. Rogers. Portable soil coring system that minimizes plot disturbance. *Agronomy Journal*, 84(6):1073–1077, 1992. ISSN 0002-1962.

S. M. Rich and M. Watt. Soil conditions and cereal root system architecture: review and considerations for linking Darwin and Weaver. *Journal of experimental botany*, 64(5): 1193–1208, 2013. ISSN 1460-2431.

E. D. Rogers, D. Monaenkova, M. Mijar, A. Nori, D. I. Goldman, and P. N. Benfey. X-ray computed tomography reveals the response of root system architecture to soil texture. *Plant physiology*, page pp. 00397.2016, 2016.

E. D. Schäfer, M. R. Owen, J. A. Postma, C. Kuppe, C. K. Black, and J. P. Lynch. Simulating crop root systems using OpenSimRoot. *Plant Systems Biology: Methods and Protocols*, pages 293–323, 2022. ISSN 1071618148.

A. Schnepf, D. Leitner, M. Landl, G. Lobet, T. H. Mai, S. Morandage, C. Sheng, M. Zörner, J. Vanderborght, and H. Vereecken. CRootBox: a structural–functional modelling framework for root systems. *Annals of botany*, 121(5):1033–1053, 2018. ISSN 0305-7364.

Stan Development Team. Stan Modeling Language Users Guide and Reference Manual, 2.31, 2023.

H. Takahashi and C. Pradal. Root phenotyping: important and minimum information required for root modeling in crop plants. *Breeding Science*, 71(1):109–116, 2021. ISSN 1344-7610.

S. Trachsel, S. M. Kaeppler, K. M. Brown, and J. P. Lynch. Shovelomics: high throughput phenotyping of maize (Zea mays L.) root architecture in the field. *Plant and soil*, 341 (1-2):75–87, 2011.

J. Windle. rootmodel, 2023. URL `https://github.com/jwindle/rootmodel`.

L. M. York and J. P. Lynch. Intensive field phenotyping of maize (Zea mays L.) root crowns identifies phenes and phene integration associated with plant growth and nitrogen acquisition. *Journal of experimental botany*, 66(18):5493–5505, 2015.

A   Root trajectory models

The depth of observed root growth is

$$y = \Delta m' \lambda$$

where λ is defined in (3). We are interested in choices of distribution for Δm. We can also express the depth as

$$y = m' \delta$$

where δ is as in (4), which is more useful when modeling changes in the angle directly, since $m = \tan \theta$.

A.1   Normal, truncated model

Assume that $\Delta m_j \sim N(\mu, \sigma^2)$, $j = 1, \ldots, K + 1$ are iid. We can marginalize this so that

$$y \sim N(\mu a, \sigma^2 b^2)$$

where

$$a = \lambda' \mathbf{1} \ \text{ and } b = \|\lambda\|.$$

By reparameterizing to $\tilde{\mu} = \mu a$ and $\tilde{\sigma} = \sigma b$ we can retrieve a model that is independent of the choice of λ, i.e.

$$y \sim N(\tilde{\mu}, \tilde{\sigma}^2).$$

This is useful, because we can place a prior on $\tilde{\mu}$ and $\tilde{\sigma}$ directly. Note that we can also impose the restriction that $y \in [\ell, u]$.

Let's go a step further, so that we can recover Δm, if we want. Let Q be an orthonormal matrix whose first column is proportional to λ and $w = bQ'\Delta m$. Note $y = w_1$ and has the same distribution as stated above. Because of the independence the components of w, when we truncate w_1 to $[\ell, u]$, we do not lose the Gaussian shape for the other components and the posterior of $w_{2:K+1}$ is independent of w_1 given μ and σ. Thus, for a given λ, and hence a and b, to recover μ, σ, and Δm, we can do $\mu = \tilde{\mu}/a$, $\sigma = \tilde{\sigma}/b$, $w_1 = y$,

$$w_{2:K+1} \sim N\Big(\tilde{\mu} \, \frac{b}{a} Q'_{2:K+1}\mathbf{1}, \tilde{\sigma}^2 I_K\Big),$$

and $\Delta m = Qw/b$.

As suggested above, we want to place a prior distribution on $\tilde{\mu}$ and $\tilde{\sigma}$, since these do not depend on λ. Because of the truncation, it is not necessarily the case that we want to limit $\tilde{\mu}$, so it is preferable to chose relatively uninformative priors. For instance, letting $d = (\ell + u)/2$, we could do

$$\tilde{\mu} \sim N(d, (d/2)^2)$$
$$\tilde{\sigma} \sim N_{[0,\infty]}(0, d^2).$$

A.2 Correlated normal, truncated model

The problem with the model above is that while it accommodates truncation of the observed depths, it does not take into account unrealistic trajectories, e.g. paths that go above the surface of the soil. The problem is that introducing these constraints into the model leads to a multivariate truncated normal distribution whose truncation is a convex space for which we cannot compute a normalizing constant. In other words, it messes up our ability to make inferences on μ and σ. However, for the purposes of generating realistic paths, we can expand the model to try and avoid such paths.

We can alter the normal, truncated model to avoid unlikely paths by introducing a hierarchy, which induces correlation in the trajectories of Δm_j. In particular, introducing notation for both root and piece:

$$\Delta m_{ij} \sim N(z_i, \sigma_2^2), j = 1, \ldots, K+1$$
$$z_i \sim N(\mu, \sigma_1^2)$$

for root i.

Let us ignore the issue of truncation for the moment. In terms of estimating the underlying parameters, it is a similar story as above, albeit with too many degrees of freedom. If we marginalize the z_i, then we have

$$\Delta m_{ij} \sim N(\mu\mathbf{1}, \mathbf{1}\mathbf{1}'\sigma_1^2 + I\sigma_2^2).$$

Using a reparameterization like above we have

$$y_i \sim N(\tilde{\mu}, \tilde{\sigma}^2)$$

24

where now $\tilde{\mu} = a\mu$ and

$$\tilde{\sigma}^2 = a^2\sigma_1^2 + b^2\sigma_2^2.$$

The latter makes it evident that we cannot learn both σ_1 and σ_2 in the non-truncated case and that we must chose their relative importances, which we can do by e.g. choosing γ for

$$\gamma = \frac{a^2\sigma_1^2}{a^2\sigma_1^2 + b^2\sigma_2^2}.$$

By choosing a larger σ_1 relative to σ_2, we introduce more correlation within the path of the trajectory. In other words, roots that start shallow will stay shallow and roots that start steep will stay steep. In the limiting case, we have a random start with a deterministic path. Within the context of truncation, the truncation would eliminate all paths that are too shallow, which are precisely the paths we are worried about — the ones that might go above the level of the soil.

However, we have to be careful about where we introduce truncation. The problem is that we cannot just truncate $\lambda'\Delta m$ for the distribution of Δm after marginalizing z_i, since that has a correlation structure that doesn't fit well with the transformation $w = bQ'\Delta m$. But, we can do something similar to above.

Just as above, we did not integrate out the mean parameters and we must avoid doing the same here. Following the same transformation $w = bQ'\Delta m$ as previously and letting $\tilde{z}_i = az_i$, we have

$$w_i \sim N(\tilde{z}_i\frac{b}{a}Q'\mathbf{1}, b^2\sigma_2^2 I)$$
$$\tilde{z}_i \sim N(\tilde{\mu}, a^2\sigma_1^2).$$

Given a choice of γ, we can rewrite this as

$$w_i \sim N\left(\tilde{z}_i\frac{b}{a}Q'\mathbf{1}, (1-\gamma)\tilde{\sigma}^2 I\right)$$
$$\tilde{z}_i \sim N(\tilde{\mu}, \gamma\tilde{\sigma}^2).$$

Conditional on $\tilde{z}_i$, we can truncate the first component of w and still write down a likelihood. Thus we learn about $\tilde{\mu}$ and $\tilde{\sigma}$ through $y = w_1$, which has the distribution (and being explicitly about the conditioning)

$$(y_i | \tilde{z}_i, \tilde{\mu}, \tilde{\sigma}) \sim N_{[\ell, u]}(\tilde{z}_i, (1-\gamma)\tilde{\sigma}^2),$$

which does not care about λ, a, or b! And then the other components of w follow

$$(w_{i,2:K_1} | \tilde{z}_i, \tilde{\mu}, \tilde{\sigma}) \sim N\left(\tilde{z}_i\frac{b}{a}Q'_{2:K+1}\mathbf{1}, (1-\gamma)\tilde{\sigma}^2 I_K\right).$$

Thus, after all is said and done, for the purposes of inference, we have

$$y_i \sim N_{[\ell, u]}(\tilde{z}_i, (1-\gamma)\tilde{\sigma}^2)$$
$$\tilde{z}_i \sim N(\tilde{\mu}, \gamma\tilde{\sigma}^2).$$

This makes it clear that we cannot just marginalize out $\tilde{z}_i$. The only external dependence that remains is the choice of γ. If we lose the truncation, then we can marginalize $\tilde{z}_i$, in which case it becomes clear that γ is free, which is a bit of a warning that this parameter may not be easy to learn. We at least know it must reside in $[0, 1]$. As before, the connection between y and $\tilde{\mu}$ and $\tilde{\sigma}$, suggest place priors on those parameters similar to above.

A.3 Skew normal

This is perhaps the most straightforward of all the models. A skew normal distribution is typically parameterized by location, shape, and scale (as it is in Stan). However, as explained by Pewsey [2000], there can be advantages to using the mean, standard deviation, and shape as the parameterization and that is what we found to be the case.

The skew normal density is

$$2\phi(x|\xi,\omega)\Phi\left(\alpha\,\frac{x-\xi}{\omega}\right)$$

where ϕ is a normal kernel with location ξ and scale ω, Φ is a standard normal cumulative distribution function, and α is the shape parameter. It is also common to replace the shape with

$$\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}.$$

The mean and variance are related to ξ, ω, δ by

$$\mu = \xi + \sqrt{\frac{2}{\pi}}\omega\delta \ \text{ and } \sigma^2 = \omega^2\left(1 - \frac{2}{\pi}\delta^2\right).$$

Further, the skewness (defined as $\mathbb{E}[(X-\mu)^3/\sigma^3]$) is

$$\gamma = \frac{4-\pi}{2}\left(\frac{\frac{2}{\pi}\delta^2}{1-\frac{2}{\pi}\delta^2}\right)^3.$$

We model the changing slopes as iid

$$\Delta m_i \sim \text{SN}(\mu, \sigma, \alpha), \ i = 1, \ldots K+1,$$

where SN refers to skew normal, μ and σ are the mean and standard deviation, not the location and scale, and α is the shape.

We would prefer to have priors that do not depend on K, the number of kinks before r, too much. To that end we use a non-informative Gaussian prior on α, like $\alpha \sim N(0, 5^2)$ and we modify μ and σ as in M08 so that

$$\tilde{\mu} = \|\lambda\|_1\,\mu \ \text{ and } \tilde{\sigma} = \|\lambda\|_2\,\sigma$$

We then place priors on $\tilde{\mu}$ and $\tilde{\sigma}$ like

$$\tilde{\mu} \sim N(d, (d/2)^2)$$
$$\tilde{\sigma} \sim N_{[0,\infty)}(0, d^2).$$

Aside: It is possible to write down the exact density of a sum of skew normal distributions [Nadarajah and Li, 2017]. However, this is not implemented by default in Stan nor is it a trivial expression.

A.4 Normal, truncated on the log scale

This model is almost identical to M08, but instead of modeling the depth, we model the log depth $\log(-y) - \log(1) = \log(-y)$ with the bounds adjusted accordingly.

A.5 Modeling angle, not slope

Now we consider modeling changes in angle directly, and not the slope. We use a shifted and scaled sigmoid transform s, where s has range $[0, 1]$ and is symmetric about $(0, 1/2)$, so that

$$\theta = s(\xi; u, \ell) = (u - \ell)s(\xi) + \ell.$$

The parameters u and ℓ have the effect of controlling the maximum and minimum angle so that $\theta \in [\ell, u]$. For each root, the random slopes are generated by

$$m_i = \tan(\theta_i), \ i = 1, \ldots, K + 1$$
$$\theta_i = s(\xi_i; u, \ell) \ i = 1, \ldots, K + 1$$
$$\xi_i = \xi_0 + \sum_{j=1}^{i} \Delta\xi_i \ i = 1, \ldots, K + 1$$
$$\Delta\xi_i \sim N(\mu, \sigma), \ i = 1, \ldots K + 1.$$

One of the advantages of M08 is that we defined its prior in terms of the observed depths. This is helpful since the number of kinks in the root before r can change, and we would like to have priors that are insensitive to that choice. To that end, we want to try to relate the parameters we have specified for Δx_i in a way that connects them to the distribution of depths. Let C be the matrix whose ith row sums the first i elements of the vector to which it is applied. We have

$$\tilde{\xi} = C\Delta\xi \sim N(\mu C\mathbf{1}, \sigma^2 CC').$$

The function $f(\xi) = \tan(s(\xi; \ell, u))$ will itself be sigmoidal when $\ell > -\pi/2$ and $u < \pi/2$ — it is a composition of increasing functions and has limits in both directions. Assume for the moment that we have an approximation to f that takes the form $\beta_0 + \beta_1 x$. Since $y = m'\delta$ using (4), we thus have

$$y \simeq \delta'[\beta_0 \mathbf{1} + \beta_1(\xi_0 \mathbf{1} + \tilde{\xi})]$$
$$= \beta_0(\delta'\mathbf{1}) + \beta_1\xi_0(\delta'\mathbf{1}) + \beta_1\delta'\tilde{\xi}.$$

Note that

$$\beta_1\delta'\tilde{\xi} \sim N(\beta_1\mu\delta'C\mathbf{1}, \sigma^2\beta_1^2\|C\delta\|^2).$$

We can simplify all of this by observing that $\delta'\mathbf{1} = r$ and $C'\delta = \text{rev } \tilde{x}$ where

$$\tilde{x}_i = \begin{cases} x_i, \ x_i < r \\ r, \ x_i \geq r, \end{cases} \ i = 1, \ldots, K + 1.$$

Hence y is approximately distributed as

$$N\left(r\beta_0 + r\beta_1\xi_0 + \beta_1\|\tilde{x}\|_1\mu, \sigma^2\beta_1^2\|\tilde{x}\|_2^2\right).$$

The transformed parameters of interest are then

$$\tilde{\mu} = r\beta_0 + r\beta_1\xi_0 + \beta_1\|\tilde{x}\|_1\mu$$

and
$$\tilde{\sigma} = \sigma \beta_1 \|\tilde{x}\|_2.$$

Before we get to the priors on $\tilde{\mu}$ and $\tilde{\sigma}$, we need some further definitions.

To add some further flexibility to the model, we allow u, ℓ, and ξ_0 to be free as well, but with fairly strong priors. We will describe the priors on these quantities, and then get to the priors for $\tilde{\mu}$ and $\tilde{\sigma}$. The prior set up for the bounds and for the starting point are

$$\xi_0 = \rho(u - \ell) + \ell$$
$$\rho \sim [0.7, 0.95]$$
$$\ell \sim [m_\ell, m_\ell + w_u]$$
$$u \sim [m_u - w_u, m_w]$$

In other words, the lower and upper bounds can vary a little from some minimum and maximum values and the starting point ξ_0 is randomly chosen so that it is near the top of the sigmoid curve. Of course, we can adjust those bounds so that instead of 70-95% up the curve we are 90-95% up the curve, or whatever seems reasonable. The important thing is that we want to move down the curve to angles that imply a steeper root growth.

To induce a prior on μ and σ, we consider the linear approximation to the sigmoid curve. In particular, we will use a secant line, which we hope is an approximation of the line of best fit over the ξ values that we expect to encounter.

Let θ_s and θ_e represent typical starting and ending angles and define $\xi_i = s^{-1}(\theta_i; \ell, u)$, $i = e, s$. The secant line is $\beta_0 + \beta_1 \xi$ where

$$\beta_1 = \frac{\tan(\theta_s) - \tan(\theta_e)}{\xi_s - \xi_e}$$

and

$$\beta_0 = \tan(\theta_e) - \beta_1 \xi_e.$$

Note that

$$\xi_i = s^{-1}\left(\frac{\theta_i - \ell}{u - \ell}\right), \ i = e, s.$$

Thus, if we specify not the angle, but the fraction of the angle, e.g. $\rho_i = \frac{\theta_i - \ell}{u - \ell}$, then

$$\xi_i = s^{-1}(\rho_i), \ i = e, s.$$

and

$$\theta_i = \ell + (u - \ell)\rho_i = \rho_i u + (1 - \rho_i)\ell, i = e, s.$$

Hence, if the fraction of the angle is fixed and used to determine the secant line, then ξ_e and ξ_s will be fixed as well and we just need to compute θ_e and θ_s to get β_1 and β_0. The one remaining point is what u and ℓ to use. We could choose m_u and m_ℓ, if we want the secant line to be independent of u and ℓ or we could chose to use the random values u and ℓ themselves.

Hence, we can place priors on $\tilde{\mu}$ and $\tilde{\sigma}$ as before, like $\tilde{\mu} \sim N(d, (d/2)^2)$ and $\tilde{\sigma} \sim N(0, d^2)$, and then retrive an implied prior on μ and σ, conditional on ξ_0 and perhaps u and ℓ via the secant line approximation.

B   Model Performance

Table 1 show the performance of the models for corn and wheat. Brief descriptions of the models and model names can be found in Section 6. More detailed descriptions are in Appendix A. The suffix of 1k, 2k, or, 3k indicates if the model was fit with 1, 2, or 3 kinks before the detection radius. The kinks were assumed to be equally spaced, which corresponds to a fixed μ_{dx} value of $r/(k+1), k = 1, 2, 3$. Each model was fit for corn and wheat in four different epochs: 1) $[0, 10]$ days, $(10, 18]$ days, $(18, 22]$ days, and $(22, 28]$ days. Posterior predictive distributions were generated for each and those posteriors were binned according to the depth levels of the electrodes. Data outside of those bins was excluded when compute the metrics. For each epoch, we computed the mean absolute error and the Kullback-Liebler divergence between the posterior predictive probability mass function versus the empirical mass function. Averages of those values are in the table below as "mmae" and "mkl". The column "thresh" refers to a threshold used to discard root sample paths. When the threshold was ∞ no roots trajectories were discarded. When it was 0, root trajectories that went above ground were discarded. You can see below that throwing away trajectories can adversely affect the fit for some models, especially M08 for corn.

C   Additional figures
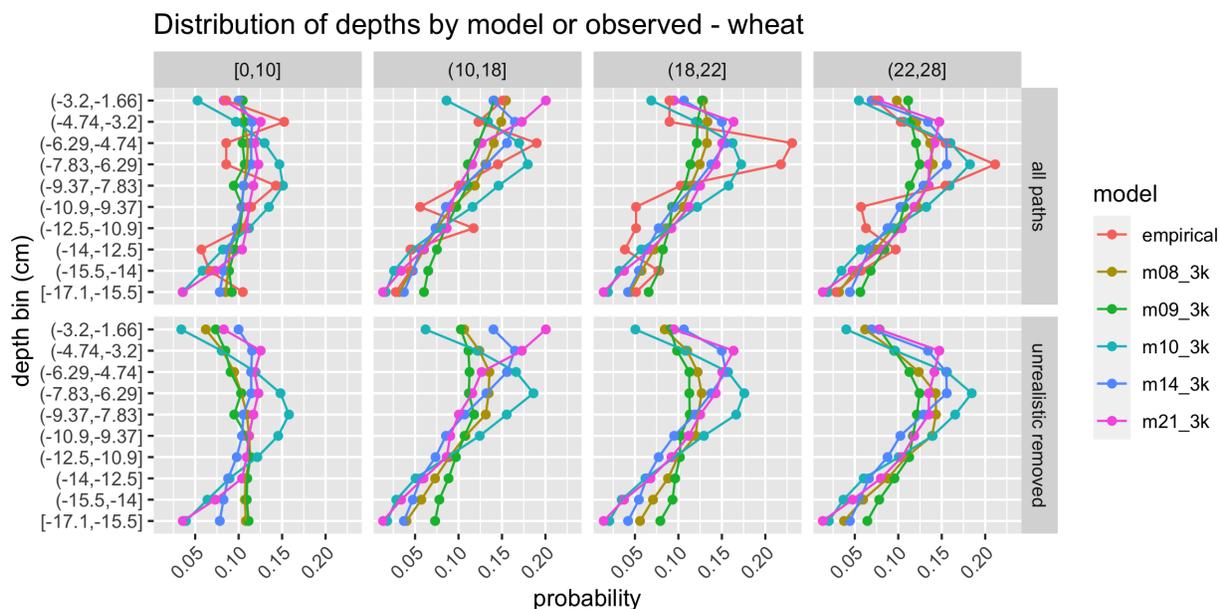
Figure 12: The posterior predictive probability mass functions for wheat, along with the empirical distribution as well.

Corn				Wheat			
model	thresh	mmae	mkl	model	thresh	mmae	mkl
empirical	Inf	0.000	0.000	empirical	Inf	0.000	0.000
m08_1k	Inf	0.169	0.023	m08_1k	Inf	0.286	0.057
m08_2k	Inf	0.172	0.023	m08_2k	Inf	0.293	0.059
m08_3k	Inf	0.173	0.024	m08_3k	Inf	0.292	0.059
m09_1k	Inf	0.222	0.043	m09_1k	Inf	0.327	0.076
m09_2k	Inf	0.214	0.042	m09_2k	Inf	0.330	0.075
m09_3k	Inf	0.218	0.043	m09_3k	Inf	0.326	0.075
m10_1k	Inf	0.226	0.042	m10_1k	Inf	0.323	0.081
m10_2k	Inf	0.229	0.043	m10_2k	Inf	0.330	0.085
m10_3k	Inf	0.246	0.047	m10_3k	Inf	0.326	0.085
m14_1k	Inf	0.169	0.024	m14_1k	Inf	0.272	0.050
m14_2k	Inf	0.171	0.025	m14_2k	Inf	0.272	0.048
m14_3k	Inf	0.167	0.023	m14_3k	Inf	0.263	0.046
m21_1k	Inf	0.187	0.026	m21_1k	Inf	0.333	0.079
m21_2k	Inf	0.184	0.030	m21_2k	Inf	0.328	0.084
m21_3k	Inf	0.177	0.029	m21_3k	Inf	0.330	0.084
m08_1k	0.000	0.183	0.026	m08_1k	0.000	0.281	0.061
m08_2k	0.000	0.209	0.035	m08_2k	0.000	0.300	0.071
m08_3k	0.000	0.220	0.039	m08_3k	0.000	0.309	0.075
m09_1k	0.000	0.196	0.032	m09_1k	0.000	0.328	0.082
m09_2k	0.000	0.206	0.033	m09_2k	0.000	0.345	0.090
m09_3k	0.000	0.225	0.037	m09_3k	0.000	0.359	0.097
m10_1k	0.000	0.239	0.053	m10_1k	0.000	0.329	0.083
m10_2k	0.000	0.250	0.059	m10_2k	0.000	0.350	0.095
m10_3k	0.000	0.278	0.071	m10_3k	0.000	0.361	0.106

Table 1: Model performance for corn and wheat. The performance is measured by discretizing the posterior predictive distribution and then comparing it to the empirical distribution. "mmae" refers to the mean mean absolute error across four epochs while "mkl" refers to the mean Kullback-Liebler divergence across the epochs. The column "thresh" refers to a threshold used to discard root trajectories. If a root trajectory goes above a threshold, then it is removed in the subsequent distribution. For instance, M08 in corn is adversely affected when trajectories go above ground. The results for M09 are for when the parameter $\gamma = 0.5$ as defined in A.2.
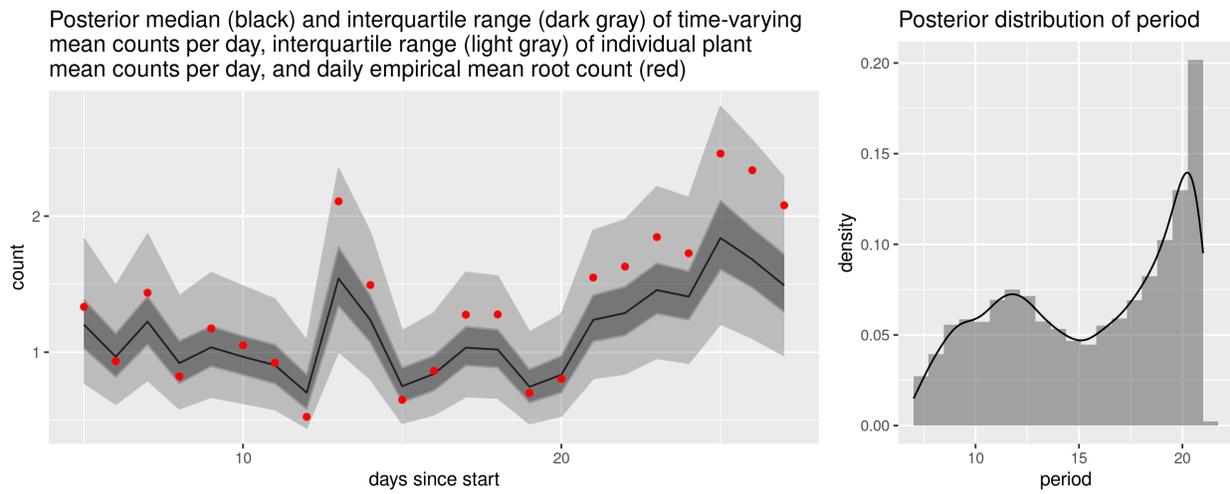
Figure 13: The posterior distribution for the mean number of roots expected day day for wheat. Unlike corn, estimating the periodicity of wheat emergence is more difficult. Pooling the data and using a common periodicity parameter could improve the estimation.