BioOne COMPLETE

An adapted Weibull function for agricultural applications

Authors: Reynolds, W. Daniel, Drury, Craig F., Phillips, Lori A., Yang, Xueming, and Agomoh, Ikechukwu V.

Source: Canadian Journal of Soil Science, 101(4) : 680-702

Published By: Canadian Science Publishing

URL: https://doi.org/10.1139/cjss-2021-0046

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

NRC Research Press

ARTICLE

An adapted Weibull function for agricultural applications

W. Daniel Reynolds, Craig F. Drury, Lori A. Phillips, Xueming Yang, and Ikechukwu V. Agomoh

Abstract: The Weibull function is applied extensively in the life sciences and engineering but underused in agriculture. The function was consequently adapted to include parameters and metrics that increase its utility for characterizing agricultural processes. The parameters included initial and final dependent variables (Y_0 and Y_F, respectively), initial independent variable (x_0), a scale constant (k), and a shape constant (c). The primary metrics included mode, integral average, domain, skewness, and kurtosis. Nested within the Weibull function are the Mitscherlich and Rayleigh functions where c is fixed at 1 and 2, respectively. At least one of the three models provided an excellent fit to six example agricultural datasets, as evidenced by large adjusted coefficient of determination ($R_A^2 \geq 0.9266$), small normalized mean bias error ($MBE_N \leq 1.49\%$), and small normalized standard error of regression ($SER_N \leq 8.08\%$). The Mitscherlich function provided the most probable (P_X) representation of corn (*Zea mays* L.) yield ($P_M = 87.2\%$); Rayleigh was most probable for soil organic carbon depth profile ($P_R = 96.4\%$); and Weibull was most probable for corn seedling emergence ($P_W = 100\%$), nitrous oxide emissions ($P_W = 100\%$), nitrogen mineralization ($P_W = 58.4\%$), and soil water desorption ($P_W = 100\%$). The Weibull fit to the desorption data was also equivalent to those of the well-established van Genuchten and Groenevelt–Grant desorption models. It was concluded that the adapted Weibull function has good potential for widespread and informative application to agricultural data and processes.

Key words: Weibull function, Mitscherlich function, Rayleigh function, least-squares regression, parsimony, goodness of model-data fit.

Résumé : Si on y recourt abondamment dans les sciences de la vie et en génie, en revanche, la fonction de Weibull est sous-utilisée en agriculture. Les auteurs ont adapté cette fonction en y incluant les paramètres et les mesures qui en accroîtront l'utilité lors de la caractérisation des processus agricoles, notamment les variables dépendantes initiale et finale (Y_0 et Y_F, respectivement), la variable indépendante initiale (x_0), une constante pour l'échelle (k) et une autre pour la forme (c). Parmi les principales mesures, mentionnons le mode, la moyenne intégrale, le domaine, le voile et l'aplatissement. Dans la fonction de Weibull figurent les fonctions de Mitscherlich et de Rayleigh, où c a respectivement pour valeur 1 et 2. Au moins un des trois modèles s'ajuste très bien aux six jeux de données agricoles employés comme illustration, ainsi que le prouve l'important coefficient de détermination après correction ($R_A^2 \geq 0{,}9266$), la faible erreur de justesse moyenne normalisée ($MBE_N \leq 1{,}49$ %) et le petit écart-type normalisé pour la régression ($SER_N \leq 8{,}08$ %). La fonction de Mitscherlich engendre la représentation la plus probable (P_X) du rendement du maïs (*Zea mays* L.) ($P_M = 87{,}2$ %), celle de Rayleigh, la représentation la plus probable du profil vertical du carbone organique dans le sol ($P_R = 96{,}4$ %) et celle de Weibull, la représentation la plus probable de la levée du maïs ($P_W = 100$ %), des émissions d'oxyde nitreux ($P_W = 100$ %), de la minéralisation de l'azote ($P_W = 58{,}4$ %) et de la désorption de l'eau dans le sol ($P_W = 100$ %). L'ajustement de la fonction de Weibull aux données sur la désorption équivaut à celui des modèles bien connus de van Genuchten et de Groenevelt–Grant. Les auteurs en concluent que la version adaptée de la fonction de Weibull pourrait largement s'appliquer aux données et aux processus agricoles, et s'avérer fort instructive. [Traduit par la Rédaction]

Mots-clés : fonction de Weibull, fonction de Mitscherlich, fonction de Rayleigh, régression des moindres carrés, parcimonie, adéquation de la justesse modèle-données.

Received 15 April 2021. Accepted 9 June 2021.

W.D. Reynolds, C.F. Drury, L.A. Phillips, X. Yang, and I.V. Agomoh. Harrow Research and Development Centre, Agriculture and Agri-Food Canada, Harrow, ON N0R 1G0, Canada.

Corresponding author: W. Daniel Reynolds (email: Dan.Reynolds@agr.gc.ca).

© Her Majesty the Queen in Right of Canada, as represented by the Minister of Agriculture and Agri-Food Canada 2021. This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

1. Introduction

Agricultural processes are amongst the most complex known, owing to their myriad of interacting chemical, physical, and biological factors. As a result, description and characterization of agricultural processes rely largely on fitted polynomial splines and empirical functions, rather than on theoretical models derived from first principles. Although this may be less than ideal, fitted splines and empirical functions can still be very useful. For example, splines and functions can often simulate the highly complex process of crop seedling emergence and can thereby provide objective estimates of important seed performance indicators such as time of first emergence, duration of emergence, maximum emergence rate, and average emergence rate. Fitted splines and empirical functions can also model strongly non-linear agricultural processes, such as nitrogen dynamics and soil water retention/release, which then serve as essential regulating functions in numerical crop growth and soil gas/water/solute balance models.

However, splines and empirical functions are frequently not equal when applied to agricultural data. Splines often perform unrealistically when applied to data that are irregularly spaced, change rapidly in value, exhibit outliers, and contain appreciable experimental and (or) random error, which are all "hallmark traits" of agricultural datasets. Empirical functions, in contrast, are not unduly perturbed by the above data traits; and when regression fitted, they are particularly adept at separating out the underlying data signal from random noise. As a result, empirical functions are often preferred over splines for describing and characterizing agricultural datasets.

One well-established empirical function that has potentially widespread agricultural applications is the so-called Weibull function. The function was first promoted in a publication by W. Weibull (Weibull 1951) and it has increased in popularity ever since, especially for characterizing extreme event processes. Some common life science and engineering applications include characterizing biological growth, ocean currents, particle size distributions, and failure of electrical insulation; determining breaking strength of natural fibers and building materials; assessing yield force, fatigue, and brittle failure of metals; estimating likelihood of infrequent events such as earthquakes, floods, gales, and tsunamis; and failure analysis of equipment such as lightbulbs, automotive parts, electronics, and industrial machinery (Weibull 1951; Monahan 2006; Abernethy 2010; Chu 2013; Brown and Mayer 1988; Mahanta and Borah 2014). Although the Weibull function is quintessentially empirical (Weibull 1951), it is nonetheless closely related to the power law distribution of flaw sizes, classical fracture mechanics, and weakest link theory (Zok 2017; Quinn and Quinn 2010).

The main popularizing features of the Weibull model are its relative simplicity, extensive versatility, and ability to perform well with small datasets and extreme data values (e.g., Abernethy 2010; Chu and Ke 2012; NCSS, Weibull 2020). More specifically, the standard Weibull function has only three fitting parameters, but it can still exhibit varying degrees of left skew, symmetry, right skew, leptokurtosis, mesokurtosis, and platykurtosis; and it can accommodate a variable inflection point, a variable *y*-axis intercept, and a variable *x*-axis threshold. In addition, Chu and Ke (2012) found that estimates of Weibull parameters can remain useably accurate even with datasets as small as 5–10 points. None of the other commonly used empirical functions (e.g., Gompertz, Logistic, Richards, Beta, Gumbel, Normal, Lognormal, etc.) can match the Weibull combination of simplicity, flexibility, and range of application.

Despite the above advantages, Weibull applications to agricultural processes appear limited to a relatively small number of studies using specialized (and restrictive) equation forms for characterizing crop yield and biomass production (e.g., Karadavut and Tozluca 2005; Karadavut et al. 2010), seed germination (e.g., Brown and Mayer 1988; Gardarin et al. 2011), and seedling emergence (e.g., Aboutalebian et al. 2017; Izquierdo et al. 2013; Navarro et al. 2013; Gan et al. 1996). The objectives of this study were consequently to (*i*) derive an adapted Weibull function that is applicable to a wide range of agricultural processes; (*ii*) describe Weibull function characteristics, parameters, and metrics that may be useful in agricultural applications; and (*iii*) present six illustrative Weibull characterizations of agricultural datasets.

We start by deriving an adapted Weibull function (Section 2.1.), then illustrate the primary Weibull characteristics (Section 2.2.), describe potentially useful function parameters and metrics (Section 2.3.), identify nested special case and related functions (Section 2.4.), describe a convenient but rigorous method for determining Weibull parameters (Section 2.5.), and finally, present six example applications to diverse agricultural datasets (Section 3.). Although the Weibull model may have theoretical justification for some processes, it is applied here as a purely empirical function for parameterizing and characterizing agricultural processes and not for making time/space/quantity predictions or extrapolations beyond fitted datasets.

2. Materials and Methods

2.1. Derivation of an adapted Weibull function

The Weibull relationship includes a "derivative function" and a "cumulative function". The derivative function can be written as a "diminishing response" relationship:

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

$$(1)\qquad \frac{dY}{dx} = c\alpha^{-c}(x - x_0)^{c-1}|Y_F - Y|; \quad \alpha, c > 0; x \geq x_0 \geq 0$$

where Y is the dependent variable, x is the independent variable, c is a shape parameter affecting function symmetry, α is a scale or proportionality parameter affecting function domain, x_0 is an initial/threshold/minimum allowable x-value, and Y_F is the final (end-point) Y-value as x becomes large. By "diminishing response", we mean that $dY/dx \rightarrow 0$ as $Y \rightarrow Y_F$, regardless of other parameter values.

Given eq. 1, an adapted Weibull cumulative function is obtained by separating variables and integrating between (x_0,Y_0) and (x,Y):

$$(2)\qquad \int_{Y}^{Y_0} \frac{dY}{(A + EY)} = \int_{x}^{x_0} (x - x_0)^{c-1} dx;$$
$$A = c\alpha^{-c}Y_F = \text{Constant}; \; E = -c\alpha^{-c} = \text{Constant}$$

which leads to

$$(3)\qquad \ln\left(\frac{A + EY}{A + EY_0}\right) = \frac{E}{c}(x - x_0)^c;$$

and ultimately simplifies to

$$(4)\qquad Y = Y_F - (Y_F - Y_0)e^{-k(x-x_0)^c}; \quad k, c > 0; \quad x \geq x_0$$

where $Y = Y_0$ at $x = x_0$, and $k = \alpha^{-c} = k_W^c$. Equation 4 is effectively a five-parameter Weibull cumulative function, as it contains five independent parameters (Y_F, Y_0, x_0, k, and c) that can be (in any combination) specified, independently measured, or determined by curve fitting to Y vs. x measurements (elaborated later). In addition, eq. 4 accommodates both $Y_F > Y_0$ and $Y_F < Y_0$ (Section 2.2.). The c parameter in eq. 4 is occasionally referred to as the "Weibull slope" (Abernethy 2010) because in traditional engineering analysis, Y vs. x data were plotted as $\ln[-\ln(Y)]$ vs. $\ln(x)$, c.f.

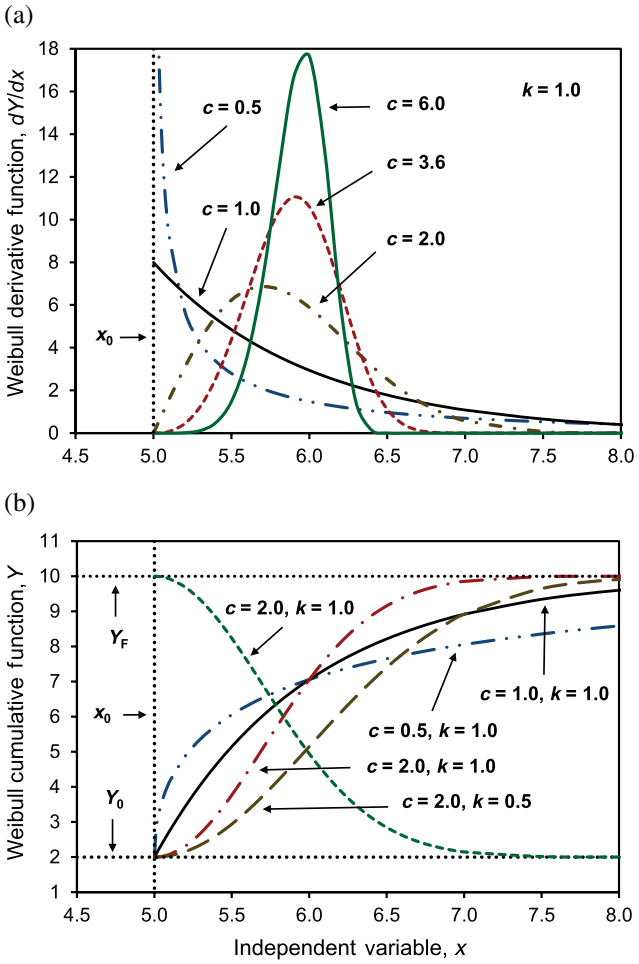
$$(5)\qquad \ln\left[-\ln\left(\frac{Y_F - Y}{Y_F - Y_0}\right)\right] = c\ln(x - x_0) + \ln(k); \quad x > x_0$$

which yields a straight line with slope = c and Y-axis intercept = $\ln(k) = c\ln(k_W)$ if the data follow a Weibull function. The k_W parameter is sometimes referred to as the "Weibull scale constant", or the "Weibull rate constant" if the x-axis represents time.

2.2. Essential characteristics of the five-parameter Weibull function

The essential characteristics of the five-parameter Weibull derivative and cumulative functions are illustrated in Fig. 1. The derivative function is negative monotone and unbounded ($dY/dx \rightarrow \infty$ as $x \rightarrow x_0$) when $c < 1$; negative monotone and bounded (dY/dx finite at $x = x_0$) when $c = 1$; bell shaped and right skewed when $1 < c < 3.6$; bell shaped and near symmetrical when $c = 3.6$; and bell shaped and left skewed when $c > 3.6$ (Fig. 1*a*). The corresponding cumulative function has an inflection when $c > 1$; no inflection when $c \leq 1$; and decreasing k causes increasing function domain but decreasing average slope (Fig. 1*b*). Note also that constant k, x_0, Y_F, and Y_0 causes all Y vs. x relationships to intersect at a common point regardless of c value; and that Y vs. x is increasing when $Y_F > Y_0$, but decreasing when $Y_F < Y_0$ (Fig. 1*b*).

Fig. 1. Essential characteristics of the five-parameter Weibull function: (*a*) effects of shape parameter, c, and independent variable offset, x_0, on derivative function, dY/dx vs. x (eq. 1, $k = 1.0$); (*b*) effects of initial and final dependent variables, Y_0 and Y_F, respectively, dependent variable offset, x_0, shape parameter, c, and scale parameter, k, on cumulative function, Y vs. x (eq. 4). For the two curves labelled $c = 2.0$, $k = 1.0$ in panel (*b*), the dash-dot line applies for $Y_0 = 2$, $Y_F = 10$, and the short-dash line applies for $Y_0 = 10$, $Y_F = 2$; both curves produce the derivative function labelled $c = 2.0$ in panel (*a*). [Colour online.]



An interesting feature of the Weibull function is its ability to indicate the likelihood (or chance) of an event or condition happening at any particular x value, given

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

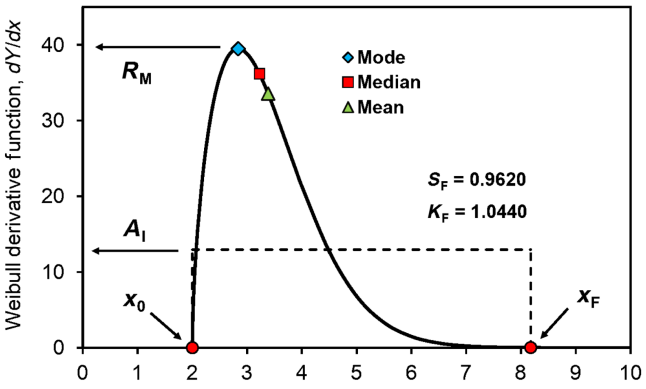
Table 1. Selected metrics of an adapted five-parameter Weibull function (see also Appendix A, Table 2, and Fig. 2).

Adapted Weibull function with five fitting parameters (Y_F, Y_0, k, x_0, and c)	
$Y = Y_F - (Y_F - Y_0)e^{-k(x-x_0)^c}$; Y = dependent variable Y_F = final (end-point) Y Y_0 = initial Y k = scale constant	$k,c > 0; x \geq x_0$ x = independent variable x_F = final (end point) x x_0 = initial x c = shape constant
Metric	**Comments**
Mode $x_M = x_0 + \left[\frac{1}{k}\left(\frac{c-1}{c}\right)\right]^{1/c}$ $Y_M = Y_F - (Y_F - Y_0)e^{[(1-c)/c]}$ $R_M = ck(x_M - x_0)^{c-1}\lvert Y_F - Y_M\rvert$	$c \geq 1$. Coordinates (x_M,Y_M,R_M) of derivative function peak and inflection of cumulative function. R_M = maximum dY/dx = maximum slope of Y vs. x = derivative function peak.
Median $x_N = x_0 + \left[-\frac{1}{k}\ln(0.5)\right]^{1/c}$ $Y_N = Y_F - \left(Y_F - Y_0\right)e^{-k(x_N - x_0)^c}$ $\left.\frac{dY}{dx}\right\rvert_N = ck(x_N - x_0)^{c-1}\lvert Y_F - Y_N\rvert$	Coordinates (x_N,Y_N,$dY/dx\rvert_N$) that bisect the area under the derivative function into two equal halves. $dY/dx\rvert_N$ = slope of Y vs. x at (x_N,Y_N).
Mean $\overline{x} = x_0 + \left[k^{(-1/c)}\right]\Gamma\left(1+\frac{1}{c}\right)$ $\overline{Y} = Y_F - (Y_F - Y_0)e^{-k(\overline{x} - x_0)^c}$	$\Gamma\left(1+\frac{1}{c}\right)$ is the Gamma function.
Integral average $A_I = \frac{(Y_F - Y_0)}{(x_F - x_0)}\left[1 - e^{-k(x_F - x_0)^c}\right]$	Mean rate of change of Y with respect to x (i.e., dY/dx) between $x = x_0$ and $x = x_F$.
Inflection $x_I = x_0 + \left[\frac{1}{k}\left(\frac{c-1}{c}\right)\right]^{1/c}$ $Y_I = Y_F - (Y_F - Y_0)e^{[(1-c)/c]}$	$c \geq 1$. Coordinates (x_I,Y_I) where cumulative function changes from convex to concave. Same coordinates as derivative function Mode.
Quantiles $x_Q = x_0 + \left[-\frac{1}{k}\ln(1-Q)\right]^{1/c}$ $Y_Q = Y_F - (Y_F - Y_0)e^{-k(x_Q - x_0)^c}$	Coordinates (x_Q,Y_Q) of specified quantiles, Q, where $0 \leq Q \leq 1$ The 0.5 quantile is equivalent to the Median.
Domain $x_D = x_F - x_0$	Difference between maximum (x_F) and minimum (x_0) x values. Defines the relevant range of Y, i.e., $Y_R = \lvert Y(x_F) - Y(x_0)\rvert$.
Fisher population skewness $S_F = \frac{\Gamma_3 - 3\Gamma_1\Gamma_2 + 2\Gamma_1^3}{(\Gamma_2 - \Gamma_1^2)^{3/2}}$	Quantifies derivative function symmetry. $\Gamma_i = \Gamma\left(1+\frac{i}{c}\right)$ is the Gamma distribution. See Table 2 for proposed skewness categories.
Fisher population kurtosis $K_F = \frac{\Gamma_4 - 4\Gamma_1\Gamma_3 + 6\Gamma_1^2\Gamma_2 - 3\left[\Gamma_1^4 + (\Gamma_2 - \Gamma_1^2)^2\right]}{(\Gamma_2 - \Gamma_1^2)^2}$	Quantifies derivative function tailing. $\Gamma_i = \Gamma\left(1+\frac{i}{c}\right)$ is the Gamma distribution. See Table 2 for proposed kurtosis categories.

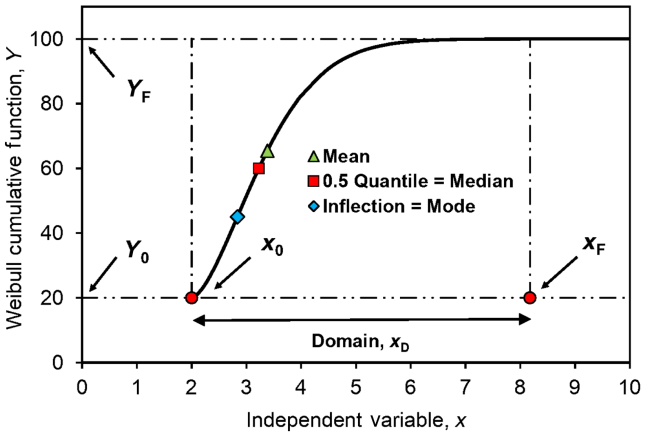
Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Fig. 2. Some metrics of the five-parameter Weibull function ($k = 0.5$, $c = 1.6$): (*a*) derivative function, dY/dx vs. x; (*b*) cumulative function, Y vs. x. Metrics listed include initial and final independent variables ($x_0 = 2.0$ and $x_F = 8.2$, respectively), initial and final dependent variables ($Y_0 = 20$ and $Y_F = 100$, respectively), mode (Mode), median (Median), mean (Mean), inflection (Inflection), integral average of derivative function (A_I), maximum value of derivative function (R_M), median quantile (0.5 Quantile), and domain (x_D) (see Appendix A for details). In panel (*a*), the area under the rectangle bounded by the dashed lines equals the area under the derivative function curve between x_0 and x_F. The derivative function is moderately right skewed ($S_F = 0.9620$) and slightly leptokurtic ($K_F = 1.0440$) (see Table 2). [Colour online.]

(a)



(b)



that the event or condition has not already happened (known in engineering as the Weibull "hazard" function). In processes where $c \neq 1$, the likelihood of an event or condition occurring changes with x (these processes are said to have "memory"), whereas in processes with $c = 1$, the likelihood does not change with x (these processes are said to be "memoryless") (Glen 2020*a*). Furthermore, when $c < 1$, likelihood of the event or condition happening decreases at a decreasing rate with increasing x; when $c = 1$, likelihood is constant (independent of x); when $1 < c < 2$, the likelihood increases at a decreasing rate with increasing x; when $c = 2$, likelihood increases linearly with increasing x; and when $c > 2$, likelihood increases at an increasing rate with increasing x (see e.g., figs. 2–8 in Abernethy 2010). Seedling emergence is an example of an agricultural process which may or may not exhibit memory. If a Weibull fit to measured seedling emergence (Y) vs. time (x) yielded $c > 1$, the likelihood of emergence increased as time increased, presumably due to viable seed and favourable seedbed conditions (e.g., amenable soil temperature, moisture, density, fertility, etc.). If the fitted function yielded $c < 1$, likelihood of emergence decreased with increasing time, suggesting low viability seed and (or) poor seedbed conditions. If $c = 1$ was obtained, likelihood of seedling emergence was constant through time and not affected by seed quality or seedbed conditions. The memory property is illustrated further in Section 3.

Table 2. Proposed skewness (S_F) and kurtosis (K_F) categories based on the inclusive graphic classifications of Folk (1980).

Skewness (S_F) categories	Numerical value
Extreme positive (right) skew	>4
Strong positive (right) skew	1.25 to 4
Moderate positive (right) skew	0.4 to 1.25
Slight positive (right) skew	0 to 0.4
Normal distribution (symmetrical)	**0**
Slight negative (left) skew	−0.4 to 0
Moderate negative (left) skew	−1.25 to −0.4
Strong negative (left) skew	−4 to −1.25
Extreme negative (left) skew	<−4
Kurtosis (K_F) categories	**Numerical value**
Extremely leptokurtic kurtosis	>20
Strongly leptokurtic kurtosis	6 to 20
Moderately leptokurtic kurtosis	3 to 6
Slightly leptokurtic kurtosis	0 to 3
Normal distribution	**0**
Slightly platykurtic kurtosis	−3 to 0
Moderately platykurtic kurtosis	−6 to −3
Strongly platykurtic kurtosis	−20 to −6
Extremely platykurtic kurtosis	<−20

2.3. Parameters and metrics obtainable from the adapted Weibull function

As mentioned above, the parameters of the adapted (five-parameter) Weibull derivative and cumulative functions (i.e., Y_F, Y_0, x_0, k, and c) may be obtained by any combination of specification, independent measurement, and estimation by curve fitting to Y vs. x data. Potentially useful derivative function metrics include mode (x_M, Y_M, R_M), median $(x_N, dY/dx|_N)$, skewness (S_F), kurtosis (K_F), and integral average (A_I). Cumulative function metrics of potential use include mean $(\overline{x}, \overline{Y})$, inflection (x_I, Y_I), quantiles (x_Q, Y_Q), and domain $(x_D = x_F - x_0)$. These metrics are briefly derived and discussed in Appendix A,

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

summarized in Table 1, and illustrated in Fig. 2. This study focuses on the Y_F, Y_0, x_0, and c parameters and on the x_D, x_M, R_M, A_I, S_F, and K_F metrics.

2.4. Special cases of the Weibull function and related functions

The five-parameter Weibull function collapses to four-parameter Mitscherlich and Rayleigh functions when $c = 1$ and 2, respectively (e.g., Lai and Xie 2006). The four-parameter Mitscherlich functions (which are effectively exponential functions) are given by

$$\frac{dY}{dx} = k_M|Y_F - Y|; \qquad k_M > 0,\ c = 1 \tag{6}$$

$$Y = Y_F - (Y_F - Y_0)e^{-k_M(x-x_0)}; \quad k_M > 0,\ c = 1; x \geq x_0 \tag{7}$$

where $k_M = \alpha^{-1} = k$ is the Mitscherlich scale constant, and the shape constant is fixed at $c = 1$. Distinguishing features of the Mitscherlich functions include no inflection, mode located at x_0, and fixed Fisher skewness and kurtosis values of $S_F = 2$ and $K_F = 6$, respectively (Weisstein 2004). Some traditional agricultural applications of Mitscherlich functions include description of nitrogen mineralization (e.g., Yang et al. 2020), characterization of nitrous oxide and carbon dioxide emissions from soil (e.g., Hoben et al. 2011; Gillis and Price 2011), depiction of crop and weed seed germination and (or) emergence (e.g., Brown and Mayer 1988; Gill et al. 1996), modeling crop growth and (or) yield responses to fertilizer rates (e.g., Srivastava et al. 2006), and simulating annual rainfall (e.g., Harmsen 2000) and water infiltration rate (e.g., Horton 1940).

It should also be noted that the four-parameter Mitscherlich cumulative function (eq. 7) is a generalization of a single-pool, first-order reaction kinetics model developed by Beauchamp et al. (1986) for describing mineralization of soil nitrogen (N) in laboratory incubation studies:

$$N_M = N_0 - (N_0 - N_E)e^{-\beta t}; \qquad \beta > 0,\ t \geq 0 \tag{8}$$

where N_M (mg N·kg^{-1} soil) is the amount of nitrogen mineralized at time t (d), N_0 (mg N·kg^{-1}) is the total amount of nitrogen available to mineralize (pool of mineralizable soil nitrogen), β (d^{-1}) is the mineralization rate constant, and N_E (mg N·kg^{-1}) is either the amount of mineralized N already present in the soil at $t = 0$, or the amount of "easily mineralizable" N that mineralized rapidly in the early stages of the incubation before measurements started. As eq. 8 assumes implicitly that $t_0 = 0$ (i.e., $x_0 = 0$ in eq. 7), then plots of N_M vs. t intersect the y-axis at $N_M = N_E$ (see e.g., Figure 4 in Beauchamp et al. 1986). Note also that the mineralization "half time" or "half life", $t_{1/2}$ (i.e., the time required for half of N_0 to be mineralized), is a special case of the median x (x_N, eq. A6) or the 0.5 quantile x ($x_{0.5}$, eq. A23):

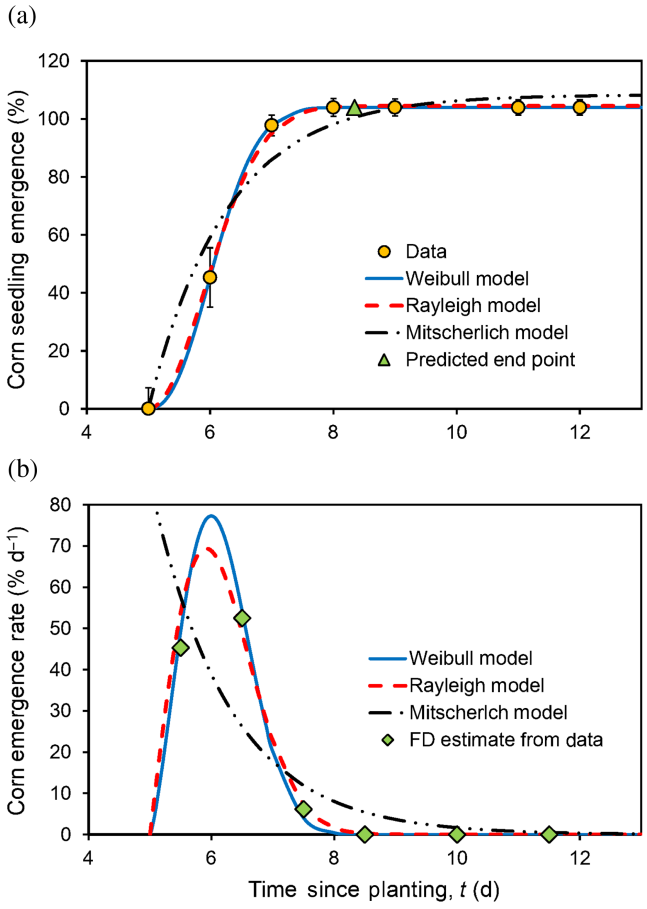


Fig. 3. Weibull, Mitscherlich, and Rayleigh functions fitted to corn seedling emergence data: (*a*) cumulative seedling emergence, Y vs. t; (*b*) seedling emergence rate, dY/dt vs. t. Circles are measured cumulative seedling emergence expressed as percentage of seed planting rate; triangles are finite difference (FD) estimates of emergence rate based on measured cumulative seedling emergence (eq. 21); predicted end point (triangle) is time when complete emergence was reached (x_F,Y_F). Vertical "T bars" are standard error ($n = 4$). Data from Agomoh et al. (2021). [Colour online.]

$$t_{1/2} = \frac{\ln(2)}{\beta} = x_N = x_0 + \left[-\frac{1}{k}\ln(0.5)\right]^{1/c} = x_{0.5} = x_0 + \left[\frac{1}{k}\ln(2)\right]^{1/c} \tag{9}$$

since $x_0 = 0$, $c = 1$ and $\beta = k$.

The four-parameter Rayleigh functions are

$$\frac{dY}{dx} = 2k_R(x - x_0)|Y_F - Y|; \qquad k_R > 0,\ c = 2 \tag{10}$$

$$Y = Y_F - (Y_F - Y_0)e^{-k_R(x-x_0)^2}; \qquad k_R > 0,\ c = 2; x \geq x_0 \tag{11}$$

where $k_R = \alpha^{-1} = k^{1/2}$ is the Rayleigh scale constant, and the shape constant is fixed at $c = 2$. The Rayleigh functions have variable mode, median, mean, and inflection (as with Weibull) but fixed Fisher skewness

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Table 3. Confidence limits (CL, 95%) on slope and *y*-axis intercept of measured dependent variable (Y_i^M) regressed against model-predicted dependent variable (Y_i^P) for the Weibull, Mitscherlich, Rayleigh, van Genuchten, and Groenevelt–Grant models.

Application	Weibull	Mitscherlich	Rayleigh
Corn seedling emergence (Fig. 3)			
95% CL on slope	0.9992 to 1.0008	0.7956 to 1.2618	0.9671 to 1.0448
95% CL on intercept (% planted)	−0.07 to 0.07	−23.3 to 17.8	−4.02 to 2.85
Nitrous oxide emissions (Fig. 4)			
95% CL on slope	0.9861 to 1.0190	0.9620 to 1.3220	0.9943 to 1.1264
95% CL on intercept (g N·ha^{-1})	−276.6 to 195.9	−4711.7 to 393.7	−1887.8 to 1.9
Corn grain yield (Fig. 5)			
95% CL on slope	0.7426 to 1.2528	0.7329 to 1.2494	0.6320 to 1.3520
95% CL on intercept (t·ha^{-1})	−2.4 to 2.4	−2.4 to 2.5	−3.3 to 3.5
Nitrogen mineralization (Fig. 6)			
95% CL on slope	0.9816 to 1.0184	0.9795 to 1.0322	0.8680 to 1.1320
95% CL on intercept (% of N_0)	−0.59 to 0.59	−1.05 to 0.65	−4.25 to 4.25
Soil organic carbon depth profile (Fig. 7)			
95% CL on slope	0.9434 to 1.0704	0.5672 to 1.8636	0.9044 to 1.1523
95% CL on intercept (wt. %)	−0.12 to 0.11	−1.42 to 0.82	−0.25 to 0.20
Soil water desorption curve (Fig. 8)			
95% CL on slope	0.9712 to 1.0322	0.6546 to 1.1364	0.49910 to 1.1592
95% CL on intercept (m^3·m^{-3})	−0.01 to 0.01	−0.06 to 0.11	−0.08 to 0.17
Soil water desorption curve (Fig. 9)	Weibull	van Genuchten	Groenevelt–Grant
95% CL on slope	0.9712 to 1.0322	0.9191 to 1.0645	0.9356 to 1.0535
95% CL on intercept (m^3·m^{-3})	−0.01 to 0.01	−0.02 to 0.03	−0.02 to 0.02

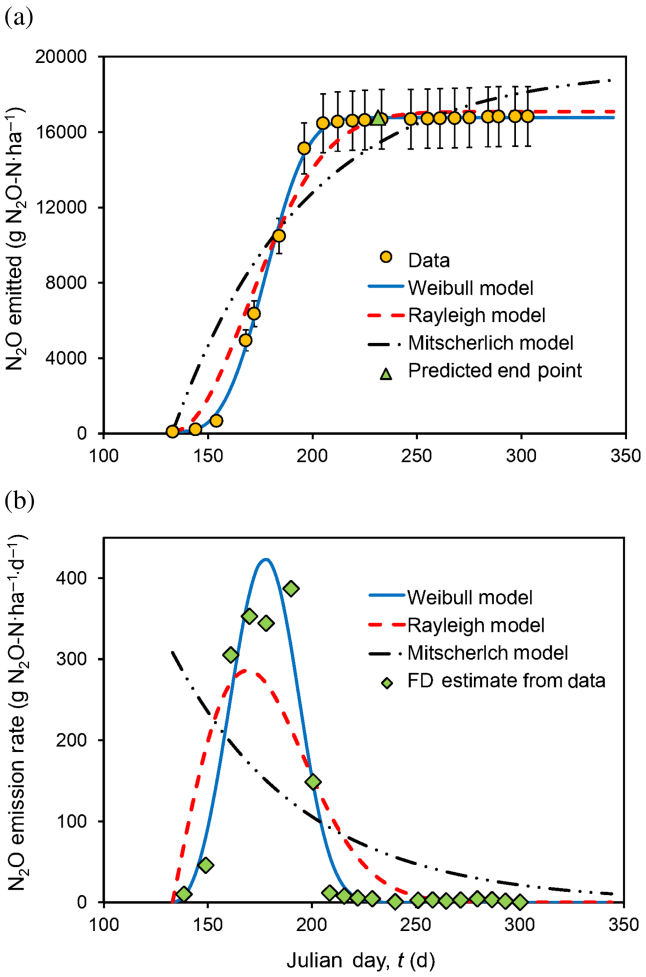
Table 4. Selected parameter values and associated metrics for Weibull, Mitscherlich, and Rayleigh models fitted to corn emergence data (Figs. 3*a*, 3*b*).

Parameter or metric	Weibull model	Mitscherlich model	Rayleigh model
Observed initial emergence, Y_0 (%)[a]	0	0	0
Fitted final emergence, Y_F (%)[a]	104.0	108.3	104.5
Measured first emergence, x_0 (d)[b]	5	5	5
Fitted final emergence, x_F (d)[b]	8.34	16.69	8.92
Time domain of emergence, x_D (d)	3.34	11.69	3.92
Scale constant: k_W, k_M, or k_R (d^{-1})	0.7845	0.7882	0.7743
Shape constant, c (—)	2.3022	1	2
Maximum emergence rate, R_M (% d^{-1})	77.27	85.39[c]	69.39
Time of maximum emergence rate, x_M (d)	6.0	5.0[c]	5.9
Integral aver. emergence rate, A_I (% d^{-1})	31.09	9.27	26.65
Fisher population skewness, S_F (—)	0.4536	2	0.6311
Fisher population kurtosis, K_F (—)	−0.0353	6	0.2451
Adjusted coefficient of determination, R_A^2 (—)	1.0000	0.9617	0.9988
Normalized mean bias error, MBE_N (%)	-3.56×10^{-5}	0.59	0.14
Normalized std. err. of regression, SER_N (%)	0.05	12.40	2.17
Significance level of Weibull fit: $F_{W\text{-}M}$ or $F_{W\text{-}R}$	—	<0.001	<0.001
Corrected Akaike inform. criterion, AIC_C (—)	−33.05	38.96	14.55
Prob. that best model: P_W, P_M, or P_R (%)	100.00	0.00	0.00

Note: Fitted parameters included Y_F, Weibull c, and k_W, k_M, or k_R.
[a]Seedling emergence calculated as percentage of seed planting rate (76 700 seeds·ha^{-1}).
[b]Time of seedling emergence (*x*-axis) is measured in days after planting.
[c](x_M,R_M) occur at (x_0,Y_0) for $c \leq 1$.

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Fig. 4. Weibull, Mitscherlich, and Rayleigh functions fitted to nitrous oxide emissions from soil: (*a*) nitrous oxide (N_2O) emitted vs. time (*t*); (*b*) N_2O emission rate, *dN*/*dt* vs. *t*. Circles are measured cumulative N_2O emitted; diamonds are finite difference (FD) estimates of *dN*/*dt* based on measured cumulative N_2O (eq. 21); predicted end point (triangle) is estimated time when emissions ceased (x_F,Y_F). Vertical "T bars" are standard error ($n = 4$). Data from Drury et al. (2021). [Colour online.]



and kurtosis of $S_F = 0.6311$ and $K_F = 0.2451$, respectively (Weisstein 2020). To the authors' knowledge, Rayleigh functions have not been applied to agricultural processes but are used extensively for characterizing communication signals, wind speed, wave height, sound and (or) light radiation, electronics longevity, and magnetic resonance images (e.g., Glen 2020*b*).

In addition to the above special cases, Weibull functions provide close approximations to the Lognormal distribution when $c = 2.5$ (NCSS, Weibull 2020), the Normal distribution when $c = 3.6023494$ (Cousineau 2011; Lai and Xie 2006; NCSS, Weibull 2020; Fig. 1*a*), and the Gumbel distribution when c is large (Cousineau 2011). This obviously serves to further extend the flexibility and applications of Weibull functions.

2.5. Determining Weibull parameters

2.5.1. Parameter estimation methods

The main approaches for estimating Weibull parameters from *Y* vs. *x* data include method of moments, maximum likelihood estimation, and model-data curve fitting using Weibull probability plots (eq. 5) or iterative non-linear least-squares regression (Cousineau 2009; Abernethy 2010; Lai and Xie 2006; Quinn and Quinn 2010; Evans et al. 2019; NCSS, Weibull 2020). Of these, maximum likelihood and curve fitting via regression provide objective (and often the most accurate) parameter estimates (NCSS, Weibull 2020). However, regression may provide more accurate estimates of x_0 than maximum likelihood (NCSS, Weibull 2020), as well as more accurate estimates of all Weibull parameters when the dataset is small (Chu and Ke 2012). Notwithstanding that commercial software packages (e.g., Relyence, Weibull 2020; ReliaSoft, Weibull 7.0 2020; Reliability Workbench, Weibull Analysis 2020; NCSS, Weibull 2020) provide a variety of methods for estimating Weibull parameters, model-data curve fitting via regression may be the most widely accessible because it resides as a built-in application in most computer spreadsheets. Iterative non-linear least-squares regression was used here by applying the Solver® algorithm in the Excel® spreadsheet to minimize the sum of squared errors between the adapted Weibull cumulative function (eq. 4) and *Y* vs. *x* data. One or more of Y_F, Y_0, x_0, *k*, and *c* were treated as curve-fitting parameters, and the Weibull scale parameter, k_W, was separated from the shape parameter, *c*, to improve parameter identifiability (i.e., *k* was used instead of $k_W{}^c$, Banks and Joyner 2017). Initial guess values for the curve-fitting parameters were obtained by manual "trial-and-error" matching of model prediction to *Y* vs. *x* data and by plotting the data according to eq. 5 (when possible) to estimate *c* from slope and *k* from intercept. Whenever possible, the number of parameters used for curve fitting was ≤ half the number of data points to increase the likelihood of obtaining unique parameter values. The options used in the Solver fitting algorithm included generalized reduced gradient non-linear numerical iteration, slope calculation using central derivatives, automatic scaling, and a 10^{-10} convergence criterion.

2.5.2. Assessing model-data fits

The accuracy and validity of model-data fits were assessed using various "goodness of fit" indicators, including adjusted coefficient of determination, normalized mean bias error, normalized standard error of regression (e.g., Archontoulis and Miguez 2015), and regression of measured *Y* (*y*-axis) against model-predicted *Y* (*x*-axis) (Piñeiro et al. 2008).

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Table 5. Selected parameter values and associated metrics for Weibull, Mitscherlich, and Rayleigh models fitted to nitrous oxide emissions from a clay loam soil (Figs. 4*a*, 4*b*).

Parameter or metric	Weibull model	Mitscherlich model	Rayleigh model
Fitted initial emission, Y_0 (g N·ha^{-1})[a]	101	101	101
Fitted final emission, Y_F (g N·ha^{-1})[a]	16 769	19 430	17 091
Fitted time of first emission, x_0 (d)[b]	133	133	133
Fitted time of final emission, x_F (d)[b]	231	711	288
Time domain of emissions, x_D (d)	98	578	155
Scale constant: k_W, k_M, or k_R (d^{-1})	0.0201	0.0159	0.0196
Shape constant, c (—)	3.2669	1	2
Maximum emission rate, R_M (g N·ha^{-1}·d^{-1})	423.8	308.2[c]	286.0
Time of maximum emission rate, x_M (d)	178	133[c]	169
Integral aver. emission rate, A_I (g N·ha^{-1}·d^{-1})	169.6	33.5	109.9
Fisher population skewness, S_F (—)	0.0872	2	0.6311
Fisher population kurtosis, K_F (—)	−0.2884	6	0.2451
Adjusted coefficient of determination, $R_A{}^2$ (—)	0.9986	0.8724	0.9764
Normalized mean bias error, MBE_N (%)	0.06	2.24	1.28
Normalized std. err. of regression, SER_N (%)	1.81	17.39	7.47
Significance level of Weibull fit: $F_{W\text{-}M}$ or $F_{W\text{-}R}$	—	<0.001	<0.001
Corrected Akaike inform. criterion, AIC_C (—)	235.04	382.24	292.76
Prob. that best model: P_W, P_M, or P_R (%)	100.00	0.00	0.00

Note: Fitted parameters included Y_0, Y_F, x_0, Weibull c, and k_W, k_M, or k_R.
[a]Cumulative nitrous oxide emissions calculated as g N_2O-N·ha^{-1}.
[b]Julian calendar days.
[c](x_M,R_M) occur at (x_0,Y_0) for $c \leq 1$.

Adjusted coefficient of determination for non-linear least-squares regression ($R_A{}^2$) is given by

$$R_A{}^2 = 1 - \frac{SSE}{SST}\left(\frac{B-1}{B-L-1}\right);\ 0 \leq R_A^2 \leq 1 \quad (12)$$

normalized mean bias error (MBE_N) by

$$MBE_N = \left[\frac{\sum_{i=1}^{B}(Y_i^P - Y_i^M)}{B}\right] / \overline{Y_i^M} \quad (13)$$

$$i = 1, 2, 3, \ldots, B;\ |MBE_N| \geq 0$$

and normalized standard error of regression (SER_N) by

$$SER_N = \left[\frac{SSE}{(B-L-1)}\right]^{1/2} / \overline{Y_i^M} \quad (14.1)$$

$$i = 1, 2, 3, \ldots, B;\ SER_N \geq 0$$

where $SSE = \sum_{i=1}^{B}(Y_i^P - Y_i^M)^2$ is the regression sum of squared errors, $SST = \sum_{i=1}^{B}\left(Y_i^M - \overline{Y_i^M}\right)^2$ is the regression total sum of squares, Y_i^P and Y_i^M are, respectively, the model-predicted and measured Y values at each x-value, $\overline{Y_i^M}$ is the mean of the measured Y values, B is the number of data points (measurements), and L is the number of model fitting parameters (excluding intercept). The $R_A{}^2$ metric indicates degree to which the fitted model predicts (or explains) systematic variation in the data, and accounts for the number of model fitting parameters (L) (e.g., Archontoulis and Miguez 2015). The MBE_N (also known as "normalized mean prediction error", MPE_N) indicates degree of systematic model bias, with positive values indicating net overestimate of the data by the fitted model, and negative values indicating net underestimate (e.g., Archontoulis and Miguez 2015; Schjønning et al. 2017). The SER_N indicates the "predictive power" of a fitted model with L fitting parameters, where $SER_N = 0$ indicates perfect prediction (i.e., no discrepancies between predicted and measured values). Given that SER_N is functionally related to normalized root mean square error, $RMSE_N$, i.e.,

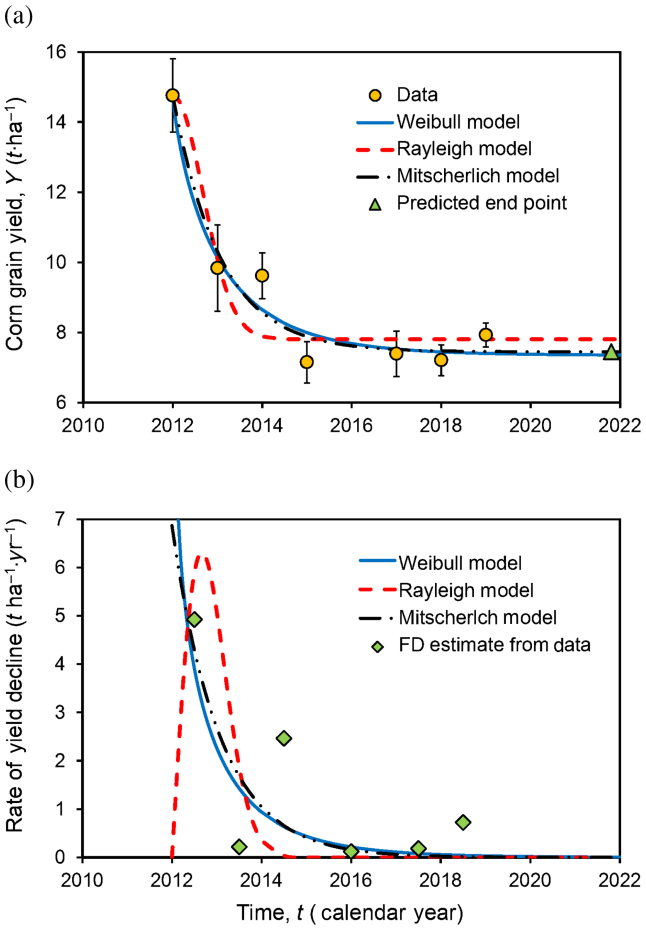
$$SER_N = RMSE_N\left(\frac{B}{B-L-1}\right)^{1/2} \quad (14.2)$$

the $RMSE_N$ predictive power categories of Jamieson et al. (1991) can be used as conservative SER_N categories, i.e., $0 \leq SER_N < 10\%$ indicates excellent model prediction of the data; $10\% \leq SER_N \leq 20\%$ signifies good prediction; $20\% < SER_N \leq 30\%$ indicates fair prediction; and $SER_N > 30\%$ signifies poor prediction.

According to the analysis of Piñeiro et al. (2008), model representations of data are viable only when the slope and intercept of measured Y (Y_i^M) regressed against model-predicted Y (Y_i^P) are not significantly different from unity and zero, respectively. When this occurs, the 95% confidence limits on slope fall on either side of unity, and the 95% confidence limits on intercept fall on either side of zero (at $P < 0.05$ significance level). If

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Fig. 5. Weibull, Mitscherlich, and Rayleigh functions fitted to decline in grain yield of unfertilized monoculture corn: (*a*) yield (*Y*) vs. time (*t*) after a one-time (spring 2012) application of compost; (*b*) rate of yield decline, $-dY/dt$ vs. *t*. Circles are measured corn grain yield at 15.5 wt. % moisture content; diamonds are finite difference (FD) estimates of $-dY/dt$ based on measured cumulative yield (eq. 21); predicted end point (triangle) is estimated time when stable yield is reached (x_F,Y_F). Vertical "T bars" are standard error ($n = 5$). [Colour online.]



the regression slope is significantly different from unity (i.e., 95% confidence interval excludes unity), the model may be compromised due to "inconsistency" (with the data); and if the regression intercept is significantly different from zero (i.e., 95% confidence interval excludes zero), the model may be compromised due to "bias". Generally speaking, the confidence limit "band" (i.e., upper confidence limit minus lower limit) becomes narrower and more symmetrical about the Y_i^M vs. Y_i^P regression line as fit between model and data improves.

2.5.3. Selecting the most suitable function or model

As the Mitscherlich and Rayleigh models are simpler special cases of the Weibull model (i.e., they are "nested" within the Weibull model because they have fixed *c* values), it is advisable to determine which of the three models provides the best balance between goodness of fit to the data and model simplicity, or "parsimony" (e.g., Johnson and Omland 2004; Vandekerckhove et al. 2015). This was achieved here using the two-model partial *F* test and the corrected Akaike information criterion (AIC_C).

The two-model partial *F* test (F_P) may be given as (e.g., Archontoulis and Miguez 2015):

$$F_P = \left(\frac{SSE_X - SSE_W}{SSE_W}\right)\left(\frac{B-K-1}{K-J}\right) \quad (15)$$

$$F_{W-X} = F.DIST.RT[F_P,(K-J),(B-K-1)] \quad (16)$$

where SSE_W and *K* are, respectively, the sum of squared errors and number of fitting parameters for the Weibull model, *B* is the number of *Y* vs. *x* measurements, $F_{W\text{-}X}$ is *F* test significance level, *F*.DIST.RT is the right-tailed *F* probability distribution (as represented in Excel®), and SSE_X and *J* are, respectively, the sum of squared errors and number of fitting parameters for the Mitscherlich or Rayleigh models. If $F_{W\text{-}M} \leq 0.05$, the Weibull fit is significantly better (at $P < 0.05$) than the Mitscherlich fit; and if $F_{W\text{-}R} \leq 0.05$, the Weibull fit is significantly better (at $P < 0.05$) than the Rayleigh fit. Note that the $F_{W\text{-}X}$ test cannot compare models with an equal number of fitting parameters ($K = J$).

For ordinary least-squares regression, the AIC_C is given by (Banks and Joyner 2017)

$$AIC_C = B\ln\left(\frac{SSE_X}{B}\right) + 2(L+1) + \left[\frac{2(L+1)(L+2)}{B-L}\right] \quad (17)$$

where SSE_X is the regression sum of squared errors for the fitted model (i.e., Weibull, Mitscherlich, or Rayleigh), *B* is the number of data points, and *L* is the number of model parameters used as least-squares fitting variables (as *c* is fixed in Mitscherlich and Rayleigh, these models will usually have one less fitting variable than the Weibull model). The first term on the right of eq. 17 represents the model's "goodness of fit", and the last two terms represent a "parsimony penalty" (Banks and Joyner 2017; Vandekerckhove et al. 2015). The parsimony penalty increases with increasing number of model fitting variables (increasing model complexity), and it is designed to compensate for model over-fitting; i.e., spurious improvements in model-data fit caused by adding fitting variables that fit largely to the data's random noise, rather than to the data's underlying signal (Johnson and Omland 2004; Vandekerckhove et al. 2015). Generally speaking, the model providing the lowest AIC_C value is considered the "best estimator" of the data (Banks and Joyner 2017; Vandekerckhove et al. 2015). Note also that AIC_C applies for both nested and un-nested models, and that two or more models can be compared simultaneously (see e.g., Banks and Joyner 2017).

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Table 6. Parameter values and associated metrics for Weibull, Mitscherlich, and Rayleigh models fitted to decline in grain yield of unfertilized monoculture corn after a one-time application of compost (Figs. 5*a*, 5*b*).

Parameter or metric	Weibull model	Mitscherlich model	Rayleigh model
Measured initial yield, Y_0 (t·ha^{-1} at 15.5%)[a]	14.8	14.8	14.8
Fitted final yield, Y_F (t·ha^{-1} at 15.5%)	7.4	7.5	7.8
Measured time of initial yield, x_0 (yr)[b]	2012	2012	2012
Fitted time of final yield, x_F (yr)[b]	2026.9	2021.8	2014.9
Time domain of yield change, x_D (yr)	14.9	9.8	2.9
Scale constant: k_W, k_M, or k_R (yr^{-1})	0.9748	0.9395	1.0582
Shape constant, c (—)	0.8299	1	2
Maximum yield change, R_M (t·ha^{-1}·yr^{-1})	5.3[c]	6.9[c]	6.3
Time of maximum yield change, x_M (yr)	2012[c]	2012[c]	2012.7
Integral aver. yield change, A_I (t·ha^{-1}·yr^{-1})	0.50	0.75	2.42
Fisher population skewness, S_F (—)	2.6569	2	0.6311
Fisher population kurtosis, K_F (—)	11.2258	6	0.2451
Adjusted coefficient of determination, R_A^2 (—)	0.9057	0.9266	0.8639
Normalized mean bias error, MBE_N (%)	0.03	0.11	0.09
Normalized std. err. of regression, SER_N (%)	9.16	8.08	11.00
Significance level of Weibull fit: $F_{W\text{-}M}$ or $F_{W\text{-}R}$	—	0.7576	0.1946
Corrected Akaike inform. criterion, AIC_C (—)	9.56	2.63	6.94
Prob. that best model: P_W, P_M, or P_R (%)	2.72	87.22	10.07

Note: Fitted parameters included Y_F, Weibull c, and k_W, k_M, or k_R.
[a]Measured corn grain yield (15.5% moisture content) in fall 2012 (x_0) after compost application in spring 2012.
[b]Time (x-axis) is measured in calendar years.
[c](x_M,R_M) occur at (x_0,Y_0) for $c \leq 1$.

The lowest AIC_C alone does not indicate how much more probable the "best estimator" model is relative to the other models, especially if differences among AIC_C values are small. To alleviate this deficiency, likelihood of being the "best estimator" model can be assessed using normalized probabilities, P_X, derived from differences among AIC_C weights (Banks and Joyner 2017; Vandekerckhove et al. 2015), i.e.,

$$(18.1)\quad P_W = \frac{100\omega_W}{(\omega_W + \omega_M + \omega_R)}$$

$$(18.2)\quad P_M = \frac{100\omega_M}{(\omega_W + \omega_M + \omega_R)}$$

$$(18.3)\quad P_R = \frac{100\omega_R}{(\omega_W + \omega_M + \omega_R)}$$

where P_W, P_M, and P_R are normalized probabilities (%), and ω_W, ω_M, and ω_R are the corresponding AIC_C weights for the Weibull, Mitscherlich, and Rayleigh models, respectively. The AIC_C weights are in turn given by (Banks and Joyner 2017)

$$(19.1)\quad \omega_W = \frac{e^{-(\Delta_W/2)}}{\left[e^{-(\Delta_W/2)} + e^{-(\Delta_M/2)} + e^{-(\Delta_R/2)}\right]}$$

$$(19.2)\quad \omega_M = \frac{e^{-(\Delta_M/2)}}{\left[e^{-(\Delta_W/2)} + e^{-(\Delta_M/2)} + e^{-(\Delta_R/2)}\right]}$$

$$(19.3)\quad \omega_R = \frac{e^{-(\Delta_R/2)}}{\left[e^{-(\Delta_W/2)} + e^{-(\Delta_M/2)} + e^{-(\Delta_R/2)}\right]}$$

where $\omega_W + \omega_M + \omega_R = 1$, and

$$(20.1)\quad \Delta_W = AIC_{CW} - AIC_{CMIN}$$

$$(20.2)\quad \Delta_M = AIC_{CM} - AIC_{CMIN}$$

$$(20.3)\quad \Delta_R = AIC_{CR} - AIC_{CMIN}$$

where AIC_{CW}, AIC_{CM}, and AIC_{CR} are the AIC_C values of the Weibull, Mitscherlich, and Rayleigh models, respectively, and AIC_{CMIN} is the smallest AIC_C value of the three models. The relative magnitudes among P_X values indicate if the AIC_C values are sufficiently different to select any one model as being better than the others for estimating or representing a particular Y vs. x dataset (Banks and Joyner 2017).

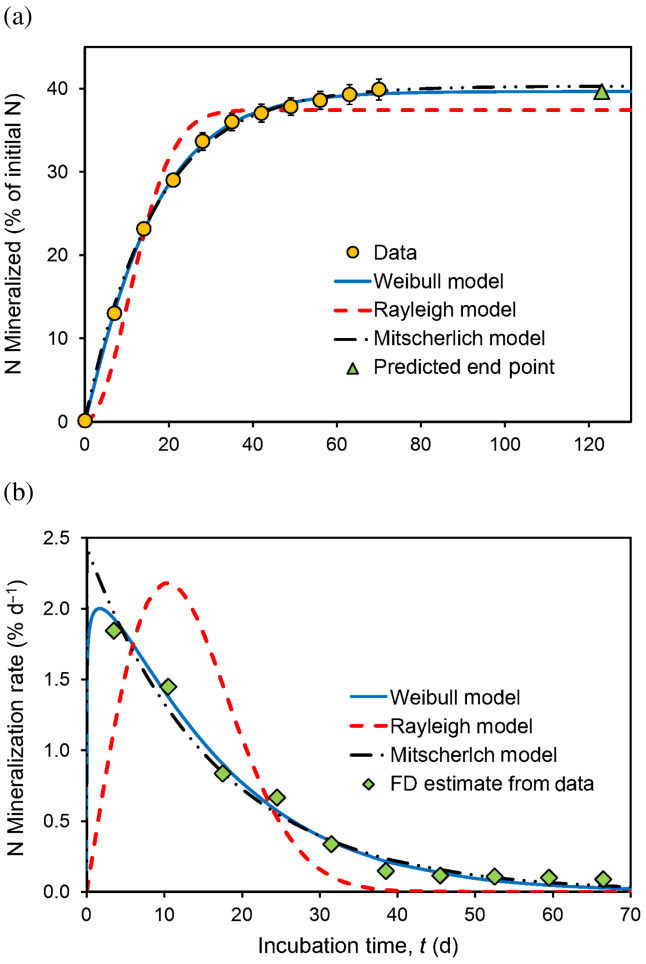
Model suitability can also be assessed by comparing a plot of the model-predicted derivative function, dY/dx vs. x, to a finite difference estimate of the actual derivative function, $\frac{\Delta Y}{\Delta x}\big|_i$ vs. $\bar{x}|_i$, where

$$(21)\quad \left.\frac{\Delta Y}{\Delta x}\right|_i = \frac{(Y_i - Y_{i+1})}{(x_i - x_{i+1})}; \qquad \bar{x}|_i = \frac{(x_i + x_{i+1})}{2}; \qquad i = 0, 1, 2, \ldots, (B-1);$$

and B is the number of measured (x_i,Y_i) data pairs. The comparison can be informal (e.g., visual) or quantified using goodness of fit criteria (e.g., eqs. 12–14, 18–20).

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Fig. 6. Weibull, Mitscherlich, and Rayleigh functions fitted to laboratory-incubated nitrogen mineralization data: (*a*) inorganic nitrogen mineralized (N) vs. time (*t*); (*b*) N mineralization rate, *dN*/*dt* vs. *t*. Circles are measured cumulative inorganic N; diamonds are finite difference (FD) estimates of *dN*/*dt* based on measured cumulative N (eq. 21); predicted end point (triangle) is estimated time when net mineralization ceased (x_F,Y_F). Vertical "T bars" are standard error ($n = 4$). Data from Yang et al. (2020). [Colour online.]



Although such comparisons are absent (or rare) in the literature, they can be critically important as good correlation between the model-predicted and estimated derivative function implies model compatibility with underlying biological–physical–chemical processes. To be useful, this approach obviously requires relatively low random variability in the (x_i,Y_i) measurements.

3. Example Applications

3.1. Seedling emergence

Seedling emergence typically occurs several days after planting of crop seeds or plow-down of weed seeds. Cumulative seedling emergence versus time graphs usually have a sigmoid or concave shape (e.g., Haj Seyed Hadi and Gonzales-Andujar 2009; Forcella et al. 2000), which is determined largely by seed viability and (or) dormancy and seedbed conditions. Both the time delay and emergence versus time patterns of seedling emergence are amenable to characterization using the Weibull–Mitscherlich–Rayleigh models and are illustrated here for emergence of grain corn (*Zea mays* L.) seedlings. Figure 3 shows corn emergence data and fitted models, where Y_0 and x_0 were specified constants, Y_F and k were fitted for the Mitscherlich and Rayleigh models, and Y_F, k, and c were fitted for the Weibull model. Table 3 gives 95% confidence limits for slope and intercept of Y_i^M vs. Y_i^P; Table 4 gives parameter values and fit metrics.

The Weibull and Rayleigh models produced better visual fits than Mitscherlich to both cumulative emergence (Fig. 3*a*) and estimated emergence rate (Fig. 3*b*). Note in particular that Weibull and Rayleigh tracked the estimated emergence rate data quite well, whereas Mitscherlich did not (Fig. 3*b*), suggesting that the emergence mechanism was not a Mitscherlich-type first-order exponential. The fit statistics for Weibull were excellent ($R_A^2 = 1.0000$, $MBE_N = -3.56 \times 10^{-5}$%, $SER_N = 0.05$%), whereas those for Rayleigh were somewhat poorer ($R_A^2 = 0.9988$, $MBE_N = 0.14$%, $SER_N = 2.17$%), and those for Mitscherlich were substantially poorer ($R_A^2 = 0.9617$, $MBE_N = 0.59$%, $SER_N = 12.4$%) (Table 4). In addition, the 95% confidence limits for slope and intercept of Y_i^M vs. Y_i^P were very narrow for Weibull, whereas those for Mitscherlich and Rayleigh were much wider (Table 3). Evidently, Weibull's excellent model-data fit was sufficient to over-ride its parsimony penalty (due to having one more fitting parameter), as the Weibull fit was highly significant ($F_{W\text{-}M} < 0.001$, $F_{W\text{-}R} < 0.001$), and Weibull was by far the most probable of the three models ($P_W \approx 100$%, $P_R \approx 0$%, $P_M \approx 0$%). The fitted Weibull model indicated a final seedling emergence (Y_F) of 104%, a final emergence time (x_F) of 8.3 d after planting, a maximum emergence rate (R_M) of 77.3% d^{-1} at day 6 after planting (x_M), and an integral average emergence rate (A_I) of 31.1% d^{-1} over the 3.3 d (x_D) duration of seedling emergence (Table 4). Note also that since $c = 2.3$ was obtained from the Weibull fit (Table 4), likelihood of corn seedling emergence increased with time at an increasing rate, suggesting viable non-dormant seed and favourable seedbed conditions. The Weibull-fitted emergence rate (derivative) function (Fig. 3*b*) was moderately right skewed ($S_F = 0.4536$), but only slightly less tailed ($K_F = -0.0353$) than a normal distribution (Table 2).

3.2. Nitrous oxide emissions

Nitrous oxide (N_2O) emissions from agricultural soil occur primarily as a result of microbial nitrification and denitrification of applied nitrogen sources, such as fertilizers and organic amendments (e.g., Signor and Cerri 2013). Although cumulative N_2O emission versus time curves can assume a wide variety of shapes, by far the most common are concave and sigmoid

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Table 7. Parameter values and associated metrics for Weibull, Mitscherlich, and Rayleigh models fitted to laboratory-incubated nitrogen mineralization data (Figs. 6*a*, 6*b*).

Parameter or metric	Weibull model	Mitscherlich model	Rayleigh model
Fitted initial mineralized N, Y_0 (% of N_0)[a]	0.06	0.00	3.41
Fitted final mineralized N, Y_F (% of N_0)[a]	39.65	40.29	37.72
Fitted initial time, x_0 (d)[b]	0.0	0.0	0.0
Fitted final time, x_F (d)[b]	123.2	151.6	49.0
Time domain of mineralization, x_D (d)	123.2	151.6	49.0
Scale constant: k_W, k_M, or k_R (d^{-1})	0.0622	0.0607	0.0619
Shape constant, c (—)	1.0902	1	2
Maximum mineralization rate, R_M (% d^{-1})	2.01	2.45[c]	1.82
Time of maximum mineralization rate, x_M (d)	1.6	0.0[c]	11.4
Integral aver. mineralization rate, A_I (% d^{-1})	0.3212	0.2657	0.6998
Fisher population skewness, S_F (—)	1.7574	2	0.6311
Fisher population kurtosis, K_F (—)	4.4937	6	0.2451
Adjusted coefficient of determination, R_A^2 (—)	0.9988	0.9979	0.9504
Normalized mean bias error, MBE_N (%)	5.92×10^{-8}	0.097	-2.55×10^{-8}
Normalized std. err. of regression, SER_N (%)	1.49	1.96	9.59
Significance level of Weibull fit: $F_{W\text{-}M}$ or $F_{W\text{-}R}$	—	0.0669	<0.001
Corrected Akaike inform. criterion, AIC_C (—)	−0.60	0.08	35.00
Prob. that best model: P_W, P_M, or P_R (%)	58.37	41.63	0.00

Note: Fitted parameters included Y_0, Y_F, x_0, Weibull c, and k_W, k_M, or k_R.
[a]N_0 is initial amount of mineralizable nitrogen in the incubated soil.
[b]Time is days after initiation of incubation.
[c](x_M,R_M) occur at (x_0,Y_0) for $c \leq 1$.

forms, which are readily characterized using the Weibull–Mitscherlich–Rayleigh models. Figure 4 shows cumulative N_2O emissions and fitted models for corn production on a clay loam soil that had received starter fertilizer (8–32–16) on Julian day 136 and side-dress nitrogen fertilizer (28% UAN) on Julian day 170 (see Drury et al. 2021 for details). Because there were more than twice as many data points (21) as fitting parameters (4–5), Y_0, Y_F, x_0, and k were fitted for the Mitscherlich and Rayleigh models, and Y_0, Y_F, x_0, k, and c were fitted for Weibull. Tables 3 and 5 give the parameter values and fit metrics.

The measured N_2O emissions formed a sigmoid cumulative curve (Fig. 4*a*) and a bell-shaped derivative, or emission rate, curve (Fig. 4*b*). The Weibull and Rayleigh models produced visually good model-data fits, but Mitscherlich was clearly not competitive. The 95% confidence limits for slope and intercept of Y_i^M vs. Y_i^P (Table 3), and the R_A^2, MBE_N, SER_N, $F_{W\text{-}M}$, and $F_{W\text{-}R}$ values (Table 5), further showed that the Weibull fit was excellent and significantly better than Rayleigh and Mitscherlich. As a result, Weibull was highly probable ($P_W = 100\%$) despite its parsimony penalty, while Rayleigh and Mitscherlich were highly improbable ($P_R \approx 0\%$, $P_M \approx 0\%$) (Table 5). The fitted Weibull model indicated maximum cumulative N_2O emission (Y_F) of 16 769 g N_2O-N·ha^{-1} at Julian day 231 (x_F), a maximum emission rate (R_M) of 424 g N_2O-N·ha^{-1}·d^{-1} at Julian day 178 (x_M), and an average emission rate (A_I) of 170 g N_2O-N·ha^{-1}·d^{-1} which occurred over a 98 d emission period (x_D) (Table 5).

Note as well that since $c = 3.2669$ was obtained (Table 5), likelihood of emission increased with time, and emissions were almost normally distributed ($S_F = 0.0872$, $K_F = -0.2884$, Table 2).

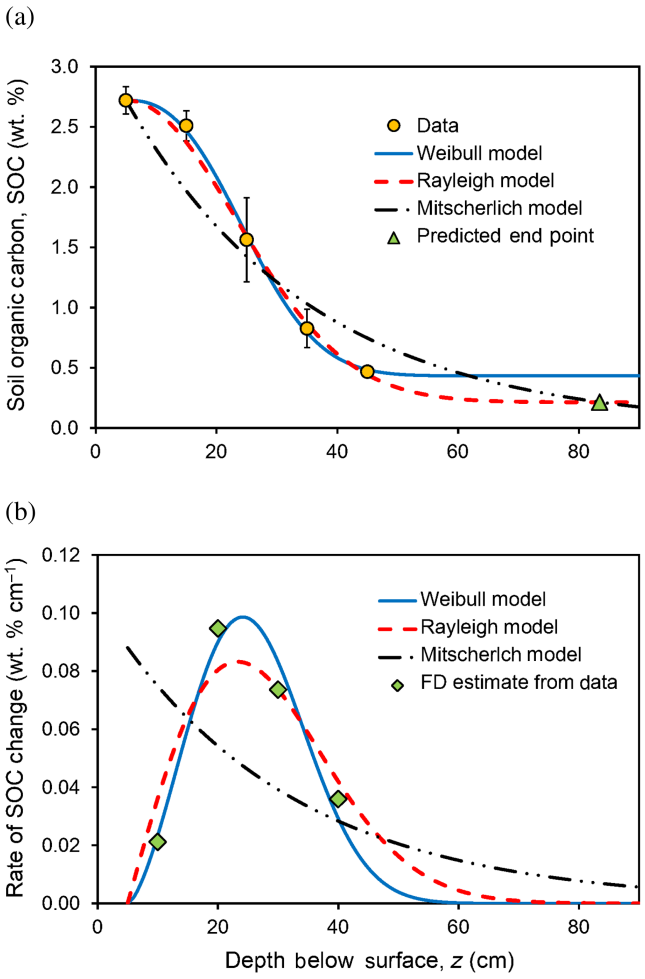
3.3. Yield change

Corn is highly sensitive to the soil environment, and annual yields decline when root zone fertility and biophysical health start to deteriorate (e.g., Bennett et al. 2012). Figure 5 shows annual grain yields and fitted models for unfertilized monoculture corn grown under no-tillage on a sandy loam soil after a one-time application (in 2012) of yard waste compost (experimental details given in Reynolds et al. 2020). Parameters Y_0 and x_0 were specified constants, whereas Y_F and k were fitted for Mitscherlich and Rayleigh, and Y_F, k, and c were fitted for Weibull. Parameter values and fit metrics appear in Tables 3 and 6.

The measured yields decline with time (Fig. 5*a*), which presumably reflects deteriorating soil fertility and health as the oxidizing compost loses its ability to supply nutrients and maintain a good biophysical environment. The data-estimated rate of yield decline (eq. 21, Fig. 5*b*) was too scattered to be useful, and it may reflect annual variation in weather. The fitted Weibull and Mitscherlich models predicted yield (Fig. 5*a*) and yield change (Fig. 5*b*) curves which were similar and convex in shape. The fitted Rayleigh model, in contrast, produced a sigmoid curve for yield (Fig. 5*a*) and a right-skewed bell curve for yield change (Fig. 5*b*), which was

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Fig. 7. Weibull, Mitscherlich, and Rayleigh functions fitted to soil organic carbon versus depth data: (*a*) soil organic carbon content (SOC) vs. depth (*z*); (*b*) rate of SOC change, $-d(\mathrm{SOC})/dz$ vs. *z*. Circles are measured SOC content; diamonds are finite difference (FD) estimates of $-d(\mathrm{SOC})/dz$ based on measured SOC (eq. 21); predicted end point (triangle) is estimated depth where minimum SOC occurs (x_F,Y_F). Vertical "T bars" are standard error ($n = 4$). Data from Reynolds et al. (2014). [Colour online.]



expected as this model's shape parameter is fixed at $c = 2$ (eqs. 10 and 11; Figs. 1*a*, 1*b*). The $R_A{}^2$ and $\mathrm{SER_N}$ fit metrics (Table 6) and confidence limits for $Y_i{}^M$ vs. $Y_i{}^P$ (Table 3) were similar between Weibull and Mitscherlich, and better than those for Rayleigh. However, the Weibull fit was not significantly better than the others ($F_{W\text{-}M} = 0.7576$, $F_{W\text{-}R} = 0.1946$); and this combined with the Weibull parsimony penalty, caused Mitscherlich to be most probable ($P_M = 87.2\%$), with Rayleigh second ($P_R = 10.1\%$), and Weibull third ($P_W = 2.7\%$). The fitted Mitscherlich model indicated that unfertilized grain yield declined over a period of about 9.8 yr (x_D) and stabilized at about 7.5 t·ha^{-1} (Y_F). The model also indicated a maximum rate of yield decline of 6.9 t·ha^{-1}·yr^{-1} (R_M) during the first year after compost addition ($x_M = 2012$), and an average rate of yield decline of 0.75 t·ha^{-1}·yr^{-1} (A_I) over the 9.8 yr yield stabilization period (x_D). Since $c = 1$ for Mitcherlich (Table 6), yield decline was predicted to be memoryless (i.e., likelihood of yield decline was constant and independent of time). As expected, the Mitscherlich rate of yield change (derivative) function (Fig. 5*b*) was strongly right skewed ($S_F = 2$) and substantially more tailed ($K_F = 6$) than a normal distribution (Table 2).

3.4. Nitrogen mineralization

Potential supply and release dynamics of crop-available nitrogen are often determined using laboratory incubations, wherein nitrogen release vs. time is measured for soil amended with fertilizers or organic materials (e.g., Yang et al. 2020). Nitrogen release curves can have a variety of shapes, but they are most often sigmoid or monotone, and thereby amenable to characterization using the Weibull–Mitscherlich–Rayleigh family of models. Figure 6 shows example mineralized nitrogen (N) release data (from Yang et al. 2020) and model fits for soil amended with crimson clover shoots. Because there were more than twice as many data points (11) as fitting parameters (4–5), Y_0, Y_F, x_0, and k were fitted for Mitscherlich and Rayleigh, and Y_0, Y_F, x_0, k, and c were fitted for Weibull. Tables 3 and 7 give parameter values and fit metrics.

All three models generated visually reasonable fits to cumulative N released vs. time data (Fig. 6*a*); however, only Weibull and Mitscherlich produced plausible estimates of mineralization rate (dN/dt) vs. time (Fig. 6*b*). Hence, Rayleigh was deemed non-competitive and not considered further. The Weibull fit was most probable of the three models ($P_W = 58.37\%$, $P_M = 41.63\%$, $P_R = 0.0\%$) (Table 7), and its 95% confidence limits for slope and intercept of $Y_i{}^M$ vs. $Y_i{}^P$ were narrowest (Table 3). The Weibull fit also had excellent metrics ($R_A{}^2 = 0.9988$, $\mathrm{MBE_N} = 5.92 \times 10^{-8}$ %, $\mathrm{SER_N} = 1.49\%$) which were better than Mitscherlich ($R_A{}^2 = 0.9979$, $\mathrm{MBE_N} = 0.097\%$, $\mathrm{SER_N} = 1.96\%$) (Table 7). The two models may actually be equivalent; however, as Weibull's probability, fit metrics, and confidence limits were only slightly better than Mitscherlich, and the partial *F*-test for Weibull vs. Mitscherlich was not significant ($F_{W\text{-}M} = 0.0669$). In addition, Weibull and Mitscherlich gave similar values for Y_0, Y_F, k, c, R_M, and A_I (Table 7), and both indicated that mineralization was effectively memoryless (i.e., $c \approx 1$, Table 7), strongly right skewed ($S_F = 1.8$–2, Tables 2 and 7), and moderate to strongly leptokurtic ($K_F = 4.5$–6, Tables 2 and 7). Note also that although both models track N mineralization rate (dN/dt vs. t) equally well (Fig. 6*b*), Weibull indicates that $dN/dt = 0$ at $t = 0$ (since $c = 1.09 > 1$), whereas Mitscherlich indicates that $dN/dt = \mathrm{R_M} = 2.45\%$ d^{-1} at $t = 0$ (since $c = 1$) (Table 7). Hence, the better model of the two may actually be the one that gives the most physically plausible representation of N mineralization rate at zero and near-zero time,

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Table 8. Parameter values and associated metrics for Weibull, Mitscherlich, and Rayleigh models fitted to soil organic carbon content (SOC) profile data (Figs. 7*a*, 7*b*).

Parameter or metric	Weibull model	Mitscherlich model	Rayleigh model
Measured maximum SOC, Y_0 (wt. %)[a]	2.72	2.72	2.72
Fitted minimum SOC, Y_F (wt. %)[b]	0.43	0.00	0.21
Measured minimum depth, z_0 (x_0) (cm)	5.0	5.0	5.0
Fitted maximum depth, z_F (x_F) (cm)	62.0	289.5	83.4
Depth domain, z_D (x_D) (cm)	57.0	284.5	78.4
Scale constant: k_W, k_M, or k_R (cm^{-1})	0.0426	0.0324	0.0387
Shape constant, c (—)	2.5027	1	2
Maximum rate of SOC change, R_M (% cm^{-1})	0.10	0.09[c]	0.08
Depth of max. rate of SOC change, x_M (cm)	24.1	5.0[c]	23.3
Integral aver. rate of SOC change, A_I (% cm^{-1})	0.04	0.01	0.03
Fisher population skewness, S_F (—)	0.3574	2	0.6311
Fisher population kurtosis, K_F (—)	−0.1444	6	0.2451
Adjusted coefficient of determination, R_A^2 (—)	0.9950	0.7827	0.9890
Normalized mean bias error, MBE_N (%)	−0.30	−2.44	−1.17
Normalized std. err. of regression, SER_N (%)	4.36	28.69	6.45
Significance level of Weibull fit: $F_{W\text{-}M}$ or $F_{W\text{-}R}$	—	0.0685	0.3171
Corrected Akaike inform. criterion, AIC_C (—)	−6.57	1.74	−13.18
Prob. that best model: P_W, P_M, or P_R (%)	3.53	0.06	96.41

Note: Fitted parameters included Y_F, Weibull c, and k_W, k_M, or k_R.
[a]Measured SOC at minimum soil profile depth, x_0.
[b]Predicted SOC at predicted maximum soil profile depth, x_F.
[c](x_M,R_M) occur at (x_0,Y_0) for $c \leq 1$.

given that most of the other parameters and metrics were similar.

3.5. Soil organic carbon depth profile

Soil organic carbon content (SOC) is perhaps the single most important soil attribute, as it affects crop productivity, most soil properties, and most soil quality and health indicators (e.g., Gregorich et al. 1997). Accurate and detailed characterization of SOC depth profiles can, therefore, be critically important for analytical and numerical models describing carbon sequestration, water transmission and storage, solute transport, crop growth, and environmental impact. Figure 7 gives an example SOC profile and corresponding Weibull–Mitscherlich–Rayleigh fits for a clay loam soil under a long-term corn–oat–alfalfa–alfalfa rotation. Parameters Y_0 and x_0 were specified constants, whereas Y_F and k were fitted for Mitscherlich and Rayleigh, and Y_F, k, and c were fitted for Weibull. The corresponding parameter values and fit metrics are in Tables 3 and 8.

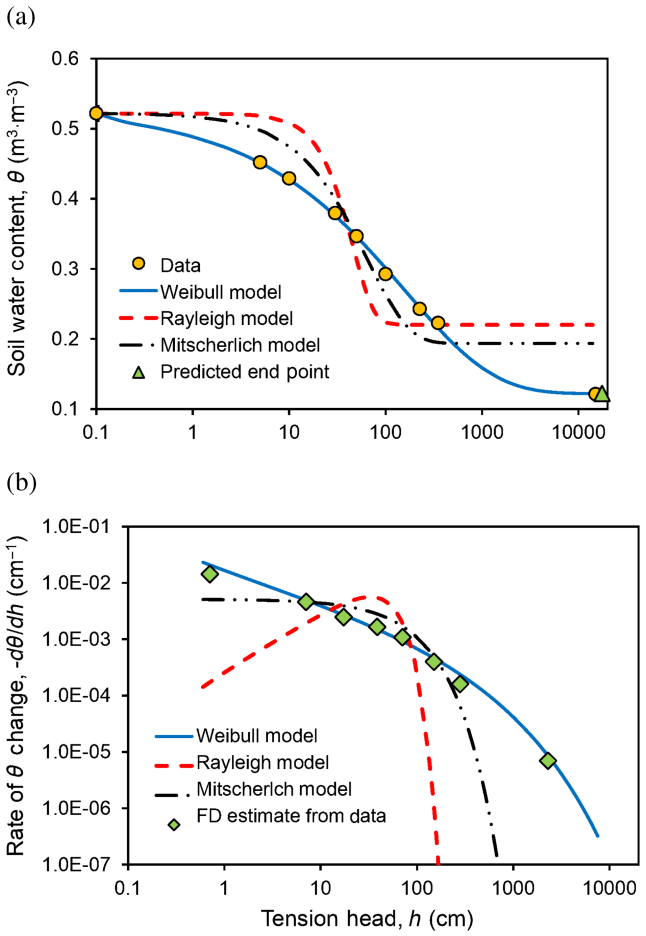
The SOC data profile is clearly sigmoidal (Fig. 7*a*) with a bell-shaped derivative function (Fig. 7*b*) — hence, well fitted by the Weibull and Rayleigh models but poorly fitted by Mitscherlich. The model-data fit metrics were excellent for Weibull ($R_A^2 = 0.9950$, $MBE_N = -0.30\%$, $SER_N = 4.36\%$), whereas Rayleigh was a close second ($R_A^2 = 0.9890$, $MBE_N = -1.17\%$, $SER_N = 6.45\%$), and Mitscherlich was a distant third ($R_A^2 = 0.7827$, $MBE_N = -2.44\%$, $SER_N = 28.69\%$) (Table 8). However, the Weibull fit was not significantly better ($F_{W\text{-}M} = 0.0685$, $F_{W\text{-}R} = 0.3171$); and when parsimony is taken into account, Rayleigh was most probable by far ($P_R = 96.41\%$), with Weibull a distant second ($P_W = 3.53\%$), and Mitscherlich a very distant third ($P_M = 0.06\%$) (Table 8). Although this clearly indicates that Rayleigh has the best balance of fit and parsimony, there may be other factors to consider. For example, Weibull not only produced better fit metrics than Rayleigh but also narrower confidence limits for slope and intercept of Y_i^M vs. Y_i^P (Table 3), and a better visual fit to the $-d(SOC)/dz$ vs. z data (Fig. 7*b*). Hence, choosing the most appropriate model may require subjective reasoning (e.g., visual model-data fits) as well as objective criteria (fit metrics). In any case, the Rayleigh (most probable) model indicated a minimum SOC (Y_F) of 0.21 wt. % at 83 cm depth (x_F), maximum rate of SOC change of 0.08 wt. %·cm^{-1} at 23 cm depth (x_M), and an average rate of SOC change (A_I) of 0.03 wt. %·cm^{-1} over a 78.4 cm depth range (x_D) (Table 8). Since $c = 2$ for Rayleigh, its derivative function for SOC (Fig. 7*b*) was moderately right skewed ($S_F = 0.6311$) and slightly more tailed ($K_F = 0.2451$) than a normal distribution (Table 2).

3.6. Soil water desorption curve

The soil water desorption curve, $\theta(h)$, describes the decrease in soil volumetric water content, θ [$L^3 \cdot L^{-3}$], with increasing soil water tension head, h [L]. It is a fundamental hydraulic property that underpins and regulates virtually all mechanistic descriptions of transmission and storage of water and gases in soils and other geologic porous media. Accurate representations of the $\theta(h)$ curve are consequently essential for valid and realistic

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Fig. 8. Weibull, Mitscherlich, and Rayleigh functions fitted to soil water desorption data: (*a*) water content (θ) vs. tension head (h); (*b*) rate of θ change, $-d\theta/dh$ vs. h. Circles are measured θ; diamonds are finite difference (FD) estimates of the differential water capacity relationship obtained using $-\Delta\theta/\Delta h$ vs. geometric mean h (eq. 21); predicted end point (triangle) is estimated tension head where minimum θ occurs (x_F,Y_F). Vertical "T bars" are standard error ($n = 10$). [Colour online.]



characterizations of a wide range of agriculturally relevant processes such as infiltration, drainage, irrigation scheduling, tillage, trafficking, and depth profiles of root zone temperature, aeration, and moisture. Due to the extreme complexity of soil–water dynamics, $\theta(h)$ is typically represented by one of several empirical or semi-empirical equations; but to the authors' knowledge, Weibull–Mitscherlich–Rayleigh expressions have never been applied. Figure 8 gives an example of Weibull, Mitscherlich, and Rayleigh model fits to a $\theta(h)$ curve from an agricultural sandy loam soil, with $(x_0,Y_0) = (h_S,\theta_S) = (0.1,0.52)$, where θ_S is measured water content at saturation (0.52 $m^3 \cdot m^{-3}$), and h_S is estimated tension head at saturation (0.1 cm). Fitting parameters included Y_F and k for Mitscherlich and Rayleigh, and Y_F, k, and c for Weibull. The resulting parameter values and fit metrics appear in Tables 3 and 9.

As is common for soil, the θ vs. h data produced a sigmoid, diminishing response curve (Fig. 8*a*, h on $\log_{10}$ scale), whereas $-d\theta/dh$ vs. h produced a concave curve (Fig. 8*b*, both θ and h on $\log_{10}$ scales). Although all three models could produce the same basic shapes as the data, it is clear that only Weibull was flexible enough (because of its variable c parameter) to achieve accurate and realistic model-data fits. Specifically, the Weibull line went through, or close to, every data point, whereas Mitscherlich and Rayleigh deviated systematically and substantially (Figs. 8*a*, 8*b*). Weibull consequently had much more favourable Y_i^M vs. Y_i^P confidence limits (Table 3) and fit metrics (Table 9) than Mitscherlich and Rayleigh; and as a result, P_W was 100% and both $F_{W\text{-}M}$ and $F_{W\text{-}R}$ were <0.001 (Table 9). Hence, Weibull was highly probable and produced accurate fits to the $\theta(h)$ and $-d\theta/dh$ data, whereas Mitscherlich and Rayleigh were both improbable and inaccurate. The Weibull fit indicated a minimum water content (Y_F) of 0.12 $m^3 \cdot m^{-3}$ at a tension head of 17 558 cm (x_F), a maximum desorption rate (R_M) of 2.23·cm^{-1} at 0.1 cm tension head (x_M), and an integral average desorption rate (A_I) of 2.28×10^{-5}·cm^{-1} over a 17 557.9 cm desorption range (x_D) (Table 9). Since $c = 0.4720$ was obtained (Table 9), the desorption curve was extremely right skewed ($S_F = 7.5$) and extremely leptokurtic ($K_F = 113.1$) (Tables 2 and 8).

Given its apparent success, Weibull was compared with two of the most popular and versatile functions for describing the $\theta(h)$ relationship — the "van Genuchten" equations (van Genuchten 1980) and the "Groenevelt-Grant" equations (Groenevelt and Grant 2004). The empirical van Genuchten $\theta(h)$ and $d\theta/dh$ equations are given by (van Genuchten 1980)

$$(22.1)\quad \theta(h) = \theta_R + \frac{(\theta_S - \theta_R)}{[1 + (\alpha h)^n]^m};\ h \geq 0$$

$$(22.2)\quad \frac{d\theta}{dh} = -\frac{nm\alpha^n h^{(n-1)}(\theta_S - \theta_R)}{[1 + (\alpha h)^n]^{(m+1)}};\ h \geq 0$$

where α [L^{-1}] and n [—] are empirical curve-fitting parameters, $m = 1 - (1/n)$ [—], θ_R [$L^3 \cdot L^{-3}$] is residual soil water content (treated as a fitting parameter), and θ_S [$L^3 \cdot L^{-3}$] is saturated soil water content (usually specified or measured). The empirical Groenevelt–Grant $\theta(h)$ and $d\theta/dh$ equations can be written as (Groenevelt and Grant 2004)

$$(23.1)\quad \theta(h) = \theta_A + k_1[e^{-(k_0/h_A)^p} - e^{-(k_0/h)^p}];\ h > 0$$

$$(23.2)\quad \frac{d\theta}{dh} = -\frac{pk_1k_0^p}{h^{(p+1)}} e^{-(k_0/h)^p};\ h > 0$$

where k_0 [L], k_1 [$L^3 \cdot L^{-3}$], and p [—] are empirical curve-fitting parameters, and (h_A,θ_A) is a specified water

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Table 9. Parameter values and associated metrics for Weibull, Mitscherlich, and Rayleigh models fitted to soil water desorption curve, $\theta(h)$, data (Figs. 8*a*, 8*b*).

Parameter or metric	Weibull model	Mitscherlich model	Rayleigh model
Meas. initial water content, θ_S (Y_0) ($m^3 \cdot m^{-3}$)[a]	0.52	0.52	0.52
Fitted final water content, θ_F (Y_F) ($m^3 \cdot m^{-3}$)	0.12	0.19	0.22
Assumed initial tension head, h_S (x_0) (cm)	0.1	0.1	0.1
Fitted final tension head, h_F (x_F) (cm)	17 558	591	140
Desorption domain, h_D (x_D) (cm)	17 557.9	590.9	139.8
Scale constant: k_W, k_M, or k_R (cm^{-1})	0.0063	0.0156	0.0217
Shape constant, c (—)	0.4720	1	2
Maximum desorption rate, R_M (cm^{-1})	2.233[b]	0.0051[b]	0.0056
Tension at max. desorption rate, h_M (x_M) (cm)	0.1[b]	0.1[b]	32.7
Integral aver. desorption rate, A_I (cm^{-1})	2.28×10^{-5}	5.55×10^{-4}	2.16×10^{-3}
Fisher population skewness, S_F (—)	7.5246	2	0.6311
Fisher population kurtosis, K_F (—)	113.0904	6	0.2451
Adjusted coefficient of determination, R_A^2 (—)	0.9981	0.8640	0.7039
Normalized mean bias error, MBE_N (%)	−0.04	2.86	5.19
Normalized std. err. of regression, SER_N (%)	1.63	13.97	20.61
Significance level of Weibull fit: $F_{W\text{-}M}$ or $F_{W\text{-}R}$	—	<0.001	<0.001
Corrected Akaike inform. criterion, AIC_C (—)	−84.41	−49.39	−42.38
Prob. that best model: P_W, P_M, or P_R (%)	100.00	0.00	0.00

Note: Fitted parameters included θ_F, Weibull c, and k_W, k_M, or k_R.
[a]θ_S is measured saturated soil volumetric water content at assumed tension head, h_S.
[b](x_M,R_M) occur at (x_0,Y_0) for $c \leq 1$.

Table 10. Parameter values and associated metrics for the Weibull, van Genuchten, and Groenevelt–Grant models fitted to soil water desorption curve, $\theta(h)$, data (Figs. 9*a*, 9*b*).

Parameter or metric	Weibull model	van Genuchten model	Groenevelt–Grant model
Measured initial water content, θ_S ($m^3 \cdot m^{-3}$)[a]	0.52	0.52	0.52
Fitted final water content, θ_F ($m^3 \cdot m^{-3}$)	0.12	—	—
Fitted residual water content, θ_R ($m^3 \cdot m^{-3}$)	—	0.01	—
Assumed initial tension head, h_S (cm)	0.1	0.1	0.1
Fitted final tension head, h_F (cm)	17,558	—	—
Scale constant, k_W (cm^{-1})	0.0063	—	—
Scale constant, α (cm^{-1})	—	0.1954	—
Scale constant, k_0 (cm^{-1})	—	—	36.0438
Scale constant, k_1 ($m^3 \cdot m^{-3}$)	—	—	0.4520
Shape constant: c, n, or p (—)	0.4720	1.1971	0.3655
Adjusted coefficient of determination, R_A^2 (—)	0.9981	0.9891	0.9929
Normalized mean bias error, MBE_N (%)	−0.04	0.19	0.12
Normalized std. err. of regression, SER_N (%)	1.63	3.95	3.19
Corrected Akaike inform. criterion, AIC_C (—)	−84.41	−68.54	−72.40
Prob. that best model: P_W, P_{vG}, or $P_{G\text{-}G}$ (%)[b]	100.00	0.04	0.24

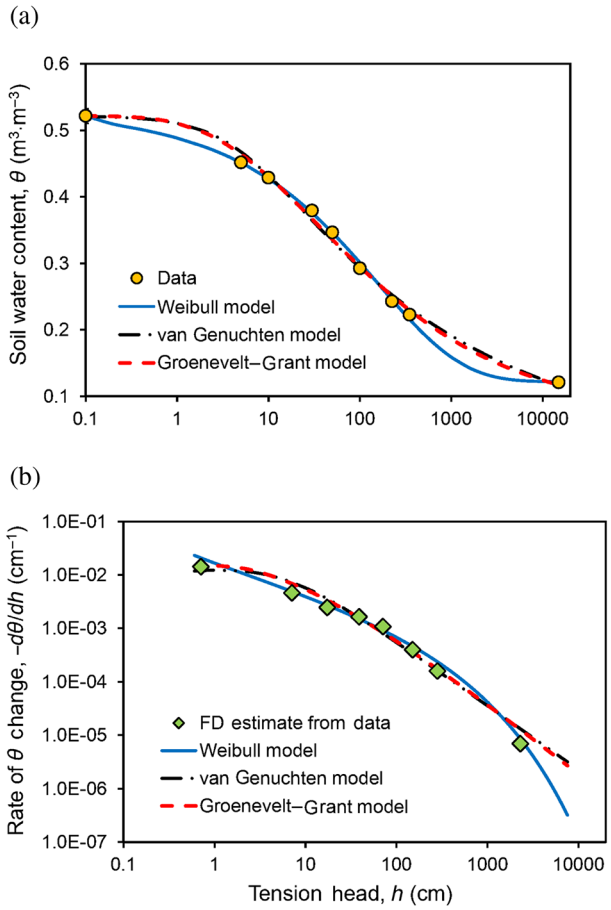
Note: Fitted parameters included θ_F, c, and k_W for Weibull; θ_R, α, and n for van Genuchten; and k_0, k_1, and p for Groenevelt–Grant.
[a]θ_S is measured saturated soil volumetric water content at assumed initial tension head, h_S.
[b]Calculated using eq. 18–20, where subscripts W, vG, and G-G represent the Weibull, van Genuchten, and Groenevelt–Grant models, respectively.

content-tension head coordinate. Equations 22.1 and 23.1 were fitted to the $\theta(h)$ data in Fig. 8*a* using the protocol in Section 2.5.1., with $\theta_S = \theta_A = 0.52\ m^3 \cdot m^{-3}$, and $h_A = 0.1$ cm. It should also be noted that eqs. 22.2 and 23.2 are traditionally referred to as the "differential water capacity" or "specific moisture capacity" relationship.

The van Genuchten and Groenevelt–Grant functions produced nearly identical visual fits to $\theta(h)$ and $-d\theta/dh$ (Figs. 9*a*, 9*b*), as well as similar fit metrics (Tables 3 and 10). The fits were not as good as Weibull; however, as their fit metrics and confidence limits for Y_i^M vs. Y_i^P were less favourable (Tables 3 and 10), and the

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Fig. 9. Comparison of Weibull, van Genuchten, and Groenevelt–Grant fits to soil water desorption data: (*a*) water content (θ) vs. tension head (h) (eqs. 4, 22.1, 23.1); (*b*) rate of θ change, $-d\theta/dh$ vs. h (eqs. 1, 22.2, 23.2). All three models were anchored to maximum water content and pressure head, i.e., $(\theta,h) = (0.52,0.1)$; Weibull fitting parameters were Y_F, k, and c; van Genuchten fitting parameters were θ_R, α, and n, with $m = 1 - (1/n)$; Groenevelt–Grant fitting parameters were k_0, k_1, and p. Circles are measured θ; diamonds are finite difference (FD) estimates of the differential water capacity relationship obtained using $-\Delta\theta/\Delta h$ vs. geometric mean h (eq. 21). Vertical "T bars" are standard error ($n = 10$). [Colour online.]



models did not "track" the wet and dry ends of the $-d\theta/dh$ vs. h relationship as well as Weibull (Fig. 9*b*). As a result, Weibull was by far the most probable of the three models with $P_W = 99.72\%$, whereas van Genuchten and Groenevelt–Grant were $P_{vG} = 0.04\%$ and $P_{G\text{-}G} = 0.24\%$, respectively (Table 10). Better tracking of $-d\theta/dh$ vs. h by Weibull (at least for this example) is particularly interesting, as it implies potentially better representations of water transmission and storage processes than the other two models, as well as more accurate estimates of the differential water capacity term ($d\theta/dh$) in the physically based Richards equation for soil moisture flow (Richards 1931).

4. Discussion

Although the above agricultural applications are limited in scope, they nonetheless provide a good demonstration of the potential and utility of the adapted Weibull–Mitscherlich–Rayleigh family of functions. In every case, at least one of the three models provided a highly plausible model-data fit (Figs. 3–9) with excellent fit metrics ($R_A^2 \geq 0.9266$, $|MBE_N| \leq 1.49\%$, $SER_N \leq 8.08\%$; Tables 4–10), and 95% confidence intervals that were narrow and near-symmetric about the Y_i^M vs. Y_i^P regression line (Table 3). In addition, the partial F test (F_P) showed when Weibull produced a significantly better fit ($P < 0.05$) than the simpler Mitscherlich and Rayleigh functions; and the normalized suitability metric (P_X) provided clear probability rankings among the fitted models (Tables 4–10). However, as has been noted by others (e.g., Archontoulis and Miguez 2015), there is currently no single metric (or suite of metrics) for definitive determination of the "best fitting", "most appropriate" or "most probable" model. It also needs to be remembered that P_X is a relative measure only, and therefore, cannot determine the best model in an absolute sense. In other words, some other model (e.g., Gompertz, Logistic, Gumbel, etc.) may provide better representations of the datasets than the Weibull–Mitscherlich–Rayleigh family. At present, estimating the best absolute representation of a dataset requires fitting a wide range of diverse (but plausible) models, and then applying the P_X metric as above (e.g., Banks and Joyner 2017).

Of the six example datasets, Mitscherlich was deemed most suitable for change in corn grain yield (Fig. 5, $P_M = 87.2\%$), Rayleigh was best for SOC distribution (Fig. 7, $P_R = 96.4\%$), and Weibull was best for corn seedling emergence (Fig. 3, $P_W = 100\%$), nitrous oxide emissions (Fig. 4, $P_W = 100\%$), nitrogen mineralization (Fig. 6, $P_W = 58.4\%$), and soil water desorption (Fig. 9, $P_W = 100\%$ and Fig. 10, $P_W = 99.7\%$). The greater success rate for Weibull is perhaps not surprising, as it has one more fitting parameter (c) and thereby much greater flexibility. However, goodness of fit and parsimony metrics may not necessarily be the only criteria for deciding model suitability. Perhaps consideration should also be given to how well the fitted model tracks the data derivative function (dY/dx), or to the model that yields more physically realistic end points for the derivative function. For example, the derivative function for SOC distribution (Fig. 7*b*) was not tracked as well by the best-fit Rayleigh model ($P_R = 96.4\%$) as it was by the second-best Weibull model ($P_W = 3.5\%$). Also, zero initial nitrogen mineralization rate indicated by the best-fit Weibull model ($P_W = 58.4\%$, Fig. 6*b*) may not be as physically realistic as maximum initial mineralization rate indicated by second-best Mitscherlich ($P_M = 41.6\%$, Fig. 6*b*). It may even be appropriate in some cases to select a model primarily on its ability to track the data

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

derivative function, or to provide physically plausible derivative function end points, with goodness of fit and parsimony metrics reduced to secondary consideration. This possibility warrants further investigation.

The results of this study suggest that the time/space/quantity behaviour of a wide range of agricultural processes is amenable to characterization and parameterization using the Weibull–Mitscherlich–Rayleigh family of functions. Some likely (or already partially demonstrated) candidates include seed germination and seedling emergence versus time or thermal time (e.g., Fig. 3); greenhouse gas production and emissions from soil versus time, soil moisture, or fertilizer source/rate/timing/placement (e.g., Fig. 4); crop yield and biomass production versus time, soil fertility, rainfall, temperature, or soil biophysical condition (e.g., Fig. 5); nitrogen mineralization dynamics versus time, temperature, or substrate composition (e.g., Fig. 6); SOC sequestration versus time, management practice, or depth (e.g., Fig. 7); transmission and storage of soil air and water (e.g., Fig. 8); and soil microbial population dynamics versus time or depth below surface. Further investigation of the usability of Weibull–Mitscherlich–Rayleigh functions for characterizing agricultural data and processes appears well justified.

5. Conclusions

The Weibull–Mitscherlich–Rayleigh family of functions was adapted to increase their utility for characterizing agricultural processes (Figs. 1 and 2; Tables 1 and 2). Using non-linear least-squares curve fitting, goodness of model-data fit metrics, and model selection metrics, it was demonstrated that the adapted functions were capable of producing accurate and unbiased fits to a wide range of agricultural data including corn seedling emergence, nitrous oxide emissions, change in corn grain yield, nitrogen mineralization, SOC distribution, and soil water desorption (Figs. 3–9). The fitted functions in turn provided several parameters (e.g., Y_0, Y_F, x_0, k, and c) and metrics (e.g., x_F, x_D, x_M, R_M, A_I, S_F, and K_F) that were useful or potentially useful for characterizing agricultural data (Appendix A, Tables 1 and 2). It was therefore concluded that the adapted Weibull–Mitscherlich–Rayleigh family of functions has substantial potential for informative application to a wide range of agricultural data and processes.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as an actual or potential conflict of interest.

Author Contributions

Author D.R. conceived the research and conducted the mathematical and computer analyses, while authors D.R., C.D., X.Y., and I.A. supervised data collection. All authors contributed to data interpretation and writing of the manuscript.

Funding

Funding for this work was provided by the Science and Technology Branch of Agriculture and Agri-Food Canada, A-Base Study 2380.

Acknowledgements

We gratefully acknowledge W. Calder, J. Gignac, J. Huffman, M. Reeb, and various summer students for data collection; Essex-Windsor Solid Waste Authority, Windsor, Ontario for provision of yard waste compost (Section 3.3.); and the AAFC-Harrow farm crew for ongoing operation and maintenance of the field sites.

References

Abernethy, R.B. 2010. The new Weibull handbook, 5th ed. Robert B. Abernethy, North Palm Beach, FL, USA.

Aboutalebian, M.A., Nazari, S., and Gonzales-Andujar, J.L. 2017. Evaluation of a model for predicting *Avena fatua* and *Descurainia sophia* seed emergence in winter rapeseed. Span. J. Agric. Res. **15**(2): 1–7. doi:10.5424/sjar/2017152-10572.

Agomoh, I.V., Drury, C.F., Reynolds, W.D., Woodley, A., Yang, X., Phillips, L.A., and Rehmann, L. 2021. Stover harvest and tillage effects on corn seedling emergence. Agronomy J. 1–9. doi:10.1002/agj2.20738.

Archontoulis, S.V., and Miguez, F.E. 2015. Nonlinear regression models and applications in agricultural research. Agron. J. **107**: 786–798. doi:10.2134/agronj2012.0506.

Banks, H.T., and Joyner, M.L. 2017. AIC under the framework of least squares estimation. Appl. Math. Lett. **74**: 33–45. doi:10.1016/j.aml.2017.05.005.

Beauchamp, E.G., Reynolds, W.D., Brasche-Villeneuve, D., and Kirby, K. 1986. Nitrogen mineralization kinetics with different soil pretreatments and cropping histories. Soil Sci. Soc. Am. J. **50**: 1478–1483.

Bennett, A.J., Bending, G.D., Chandler, D., Hilton, S., and Mills, P. 2012. Meeting the demand for crop production: the challenge of yield decline in crops grown in short rotations. Biol. Rev. **87**: 52–71. doi:10.1111/j.1469-185X.2011.00184.x.

Brown, R.F., and Mayer, D.G. 1988. Representing cumulative germination. 2. The use of the Weibull function and other empirically derived curves. Ann. Bot. **61**: 127–138.

Chu, P.C. 2013. Weibull statistics in ocean analysis and prediction. OCEANS-San Diego, San Diego, CA, USA. pp. 1–4. doi:10.23919/OCEANS.2013.6741376.

Chu, Y.-K., and Ke, J.-C. 2012. Approaches for parameter estimation of Weibull distribution. Math. Comput. Appl. **17**(1): 39–47.

Cousineau, D. 2009. Fitting the three-parameter Weibull distribution: review and evaluation of existing and new methods. IEEE Trans. Dielectr. Electr. Insul. **16**(1): 281–288. doi:10.1109/TDEI.2009.4784578.

Cousineau, D. 2011. The fallacy of large shape parameters when using the two-parameter Weibull distribution. IEEE Trans. Dielectr. Electr. Insul. **18**(6): 2095–2102.

Drury, C.F., Reynolds, W.D., Yang, X.M., McLaughlin, N.B., Calder, W.C., and Phillips, L.A. 2021. Diverse rotations impact microbial processes, seasonality and overall N_2O emissions from soils. Soil Sci. Soc. Am. J. (accepted). doi:10.1002/saj2.20298.

Evans, J.W., Kretschmann, D.E., and Green, D.W. 2019. Procedures for estimation of Weibull parameters. General Technical Report FPL-GTR-264. U.S. Department of

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Agriculture, Forest Service, Forest Products Laboratory, Madison, WI, USA. 17 p.
Folk, R.L. 1980. Petrology of sedimentary rocks. Hemphill Publishing Co., Austin, TX, USA.
Forcella, F., Benech Arnold, R.L., Sanchez, R., and Ghersa, C.M. 2000. Modeling seedling emergence. Field Crops Res. **67**: 123–139.
Gan, Y., Stobbe, E.H., and Njue, C. 1996. Evaluation of selected nonlinear regression models in quantifying seedling emergence rate of spring wheat. Crop Sci. **36**: 165–168.
Gardarin, A., Durr, C., and Colbach, N. 2011. Prediction of germination rates of weed species: Relationships between germination speed parameters and species traits. Ecol. Model. **222**: 626–636. doi:10.1016/j.ecolmodel.2010.10.005.
Gill, G.S., Cousens, R.D., and Allan, M.R. 1996. Germination, growth and development of herbicide resistant and susceptible populations of rigid ryegrass (*Lolium rigidum*). Weed Sci. **44**: 252–256.
Gillis, J.D., and Price, G.W. 2011. Comparison of a novel model to three conventional models describing carbon mineralization from soil amended with organic residues. Geoderma, **160**: 304–310.doi:10.1016/j.geoderma.2010.09.025.
Glen, S. 2020*a*. Memoryless property. From StatisticsHowTo.com https://www.statisticshowto.com/memoryless-property/ [accessed 5 Sept. 2020].
Glen, S. 2020*b*. Rayleigh distribution: definition, uses, mean, variance. From StatisticsHowTo.com https://www.statisticshowto.com/rayleigh-distribution/ [accessed 5 Sept. 2020].
Glen, S. 2020*c*. Expected value in statistics: definition and calculating it. From StatisticsHowTo.com https://www.statisticshowto.com/probability-and-statistics/expected-value/ [accessed 5 Sept. 2020].
Gregorich, E.G., Carter, M.R., Doran, J.W., Pankhurst, C.E., and Dwyer, L.M. 1997. Biological attributes of soil quality. Pages 81–114 in E.G. Gregorich and M.R. Carter, eds. Soil quality for crop production and ecosystem health. Developments in soil science. Vol. 25. Elsevier, New York, NY, USA.
Groenevelt, P.H., and Grant, C.D. 2004. A new model for the soil-water retention curve that solves the problem of residual water contents. Eur. J. Soil Sci. **55**: 479–485. doi:10.1111/j.1365-2389.2004.00617.x.
Haj Seyed Hadi, M.R., and Gonzales-Andujar, J.L. 2009. Comparison of fitting weed seedling emergence models with nonlinear regression and genetic algorithm. Comput. Electr. Agric. **65**: 19–25. doi:10.1016/j.compag.2008.07.005.
Harmsen, K. 2000. A modified Mitscherlich equation for rainfed crop production in semi-arid areas: 1. Theory. Neth. J. Agric. Sci. **48**: 237–250.
Hoben, J.P., Gehj, R.J., Millar, N., Graces, P.R., and Robertson, G.P. 2011. Nonlinear nitrous oxide (N_2O) response to nitrogen fertilizer in on-farm corn crops of the US Midwest. Global Change Biol. **17**: 1140–1152. doi:10.1111/j.1365-2486.2010.02349.x.
Horton, R.E. 1940. An approach towards a physical interpretation of infiltration capacity. Proc. Soil Sci. Soc. Am. **5**: 399–417.
Izquierdo, J., Bastida, F., Lezaun, J.M., Sa Nchez Del Arco, M.J., and Gonzales-Andujar, J.L. 2013. Development and evaluation of a model for predicting Lolium rigidum emergence in winter cereal crops in the Mediterranean area. Weed Res. **53**: 269–278. doi:10.1111/wre.12023.
Jamieson, P.D., Porter, J.R., and Wilson, D.R. 1991. A test of the computer simulation model ARCWHEAT1 on wheat crops grown in New Zealand. Field Crops Res. **27**: 337–350.
Johnson, J.B., and Omland, K.S. 2004. Model selection in ecology and evolution (review). Trends Ecol. Evol. **19**(2): 101–108. doi:10.1016/j.tree.2003.10.013.
Karadavut, U., and Tozluca, A. 2005. Growth analysis some characters in rye (*Secale cereal* L.): growth of root and upper ground parts. J. Crop Res. **2**: 1–10.
Karadavut, U., Kokten, K., and Kavurmaci, Z. 2010. Comparison of relative growth rates in silage corn cultivars. Asian J. Anim. Vet. Adv. **5**(3): 223–228.
Lai, C.-D., and Xie, M. 2006. Chapter 5 — Weibull related distributions. Pages 139–166 in Stochastic ageing and dependence for reliability. Springer, New York, NY, USA. doi:10.1007/0-387-34232-x5.
Mahanta, D.J., and Borah, M. 2014. Parameter estimation of Weibull growth models in forestry. Int. J. Math. Technol. **8**(3): 157–163.
Monahan, A.H. 2006. The probability distribution of sea surface wind speeds. Part 1: theory and sea winds observations. J. Clim. **19**: 497–520.
National Institute of Standards and Technology/SEMATECH. 2013. e-Handbook of statistical methods. http://www.itl.nist.gov/div898/handbook/. doi:10.18434/m32189 [accessed 5 Sept. 2020].
Navarro, M., Febles, G., Torres, V., Mesa, A.R., and Jay, Y.O. 2013. Use of the Weibull function to evaluate the emergence of *Albizia lebbeck* (L.) Benth seedlings. Pastos y Forrajes, **36**(2): 222–226.
NCSS, Weibull. 2020. NCSS statistical software, Weibull module. NCSS, LLC, Kaysville, UT, USA.
Piñeiro, G., Perelman, S., Guerschman, J.P., and Paruelo, J.M. 2008. How to evaluate models: observed vs. predicted or predicted vs. observed? Ecol. Model. **216**: 316–322. doi:10.1016/j.ecolmodel.2008.05.006.
Quinn, J.B., and Quinn, G.D. 2010. A practical and systematic review of Weibull statistics for reporting strengths of dental materials. Dental Mater. **26**: 135–147. doi:10.1016/j.dental.2009.09.006
Reliability Workbench, Weibull Analysis. 2020. Isograph Inc., Alpine, UT, USA.
ReliaSoft, Weibull 7.0. 2020. HBM Prenscia Inc., Southfield, MI, USA.
Relyence, Weibull. 2020. Relyence Corporation, Greensburg, PA, USA.
Reynolds, W.D., Drury, C.F., Yang, X.M., Tan, C.S., and Yang, J.Y. 2014. Impacts of 48 years of consistent cropping, fertilization and land management on the physical quality of a clay loam soil. Can. J. Soil Sci. **94**: 403–419. doi:10.4141/CJSS2013-097.
Reynolds, W.D., Nurse, R.E., Phillips, L.A., Drury, C.F., Yang, X.M., and Page, E.R. 2020. Characterizing mass–volume–density–porosity relationships in a sandy loam soil amended with compost. Can. J. Soil Sci. **100**: 289–301. doi:10.1139/cjss-2019-0149.
Richards, L.A. 1931. Capillary conduction of liquids through porous mediums. J. Appl. Phys. **1**: 318–333. doi:10.1063/1.1745010.
Schjønning, P., McBride, R.A., Keller, T., and Obour, P.B. 2017. Predicting soil particle density from clay and soil organic matter contents. Geoderma, **286**: 83–87. doi:10.1016/j.geoderma.2016.10.020.
Signor, D., and Cerri, C.E.P. 2013. Nitrous oxide emissions in agricultural soils: a review. Pesq. Agropec. Trop. Goiânia, **43**: 322–338.
Srivastava, S., Subba Rao, A., Alivelu, K., Singh, K.N., Raju, N.S., and Rathore, A. 2006. Evaluation of crop responses to applied fertilizer phosphorus and derivation of optimum recommendations using the Mitscherlich–Bray equation. Comm. Soil Sci. Plant Anal. **37**(05-06): 847–858. doi:10.1080/00103620600564182.
Vandekerckhove, J., Matzke, D., and Wagenmakers, E.-J. 2015. Model comparison and the principle of parsimony. Pages

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

300–318 in J.R. Busemeyer, Z. Wang, J.T. Townsend, and A. Eidels, eds. The Oxford handbook of computational and mathematical psychology. Oxford University Press, New York, NY, USA.
van Genuchten, M.Th. 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci. Soc. Am. J. **44**: 892–898.
Weibull, W. 1951. A statistical distribution function of wide applicability. ASME J. Appl. Mech. **18**: 293–297.
Weisstein, E.W. 2020. Rayleigh distribution. From MathWorld — a Wolfram web resource. https://mathworld.wolfram.com/RayleighDistribution.html [accessed 5 Sept. 2020].
Weisstein, E.W. 2004. Exponential distribution. From MathWorld — a Wolfram web resource. https://mathworld.wolfram.com/ExponentialDistribution.html [accessed 5 Sept. 2020].
Yang, X.M., Drury, C.F., Reynolds, W.D., and Phillips, L.A. 2020. Nitrogen release from shoots and roots of crimson clover, hairy vetch, and red clover. Can. J. Soil Sci. **100**: 179–188. doi:10.1139/cjss-2019-0164.
Zok, F.W. 2017. On weakest link theory and Weibull statistics. J. Am. Ceram. Soc. **100**: 1265–1268. doi:10.1111/jace.14665.

Appendix A

Potentially useful metrics of the adapted Weibull derivative and cumulative functions are derived below, illustrated in Fig. 2, and summarized in Table 1.

A1. Weibull derivative function: mode, median, skewness, kurtosis, and integral average

The derivative function mode gives the coordinates, (x_M,Y_M), and value of the derivative function peak, R_M (Fig. 2*a*). The mode is obtained by setting $x = x_M$ and $d(dY/dx)/dx = 0$:

$$\text{(A1)}\quad \frac{d}{dx}\left(\frac{dY}{dx}\right) = (c-1)(x_M - x_0)^{-1}\left(\frac{dY}{dx}\right) - kc(x_M - x_0)^{(c-1)}\left(\frac{dY}{dx}\right) = 0$$

then solving for x_M:

$$\text{(A2)}\quad x_M = x_0 + \left[\frac{1}{k}\left(\frac{c-1}{c}\right)\right]^{1/c};\ c \geq 1$$

and then back substituting x_M into eqs. 1 and 4 to produce

$$\text{(A3)}\quad R_M = \left.\frac{dY}{dx}\right|_M = ck(x_M - x_0)^{c-1}|Y_F - Y_M|;\ c \geq 1$$

where

$$\text{(A4)}\quad Y_M = Y_F - (Y_F - Y_0)e^{[(1-c)/c]};\ c \geq 1$$

Interestingly, Y_M (eq. A4) is independent of x and k. Note also that when $c = 1$, R_M occurs at $(x_M,Y_M) = (x_0,Y_0)$; and when $0 < c < 1$, $R_M \to \infty$as $x_M \to x_0$ (see e.g., Fig. 1*a*).

The median$(x_N,\ dY/dx|_N)$ bisects the area under the derivative function curve into two equal halves (Fig. 2*a*); and is obtained by setting $x = x_N$, then equating a normalized form of eq. 4 to the 50th percentile, i.e.,

$$\text{(A5)}\quad e^{-k(x_N - x_0)^c} = \frac{(Y_F - Y)}{(Y_F - Y_0)} = 0.5$$

then solving for x_N:

$$\text{(A6)}\quad x_N = x_0 + \left[-\frac{1}{k}\ln\left(\frac{Y_F - Y}{Y_F - Y_0}\right)\right]^{1/c} = x_0 + \left[-\frac{1}{k}\ln(0.5)\right]^{1/c}$$

and then back-substituting x_N into eq. 1 to obtain:

$$\text{(A7)}\quad \left.\frac{dY}{dx}\right|_N = ck(x_N - x_0)^{c-1}|Y_F - Y_N|$$

where

$$\text{(A8)}\quad Y_N = Y_F - (Y_F - Y_0)e^{-k(x_N - x_0)^c}$$

As might be expected, the median is less than the mode when the derivative function is left skewed, equal to the mode when the derivative function is symmetrical, and greater than the mode when the derivative function is right skewed (Fig. 2*a*).

The Fisher population skewness (S_F) and kurtosis (K_F) of the derivative function are given by (e.g., Chu 2013):

$$\text{(A9)}\quad S_F = \frac{\Gamma_3 - 3\Gamma_1\Gamma_2 + 2\Gamma_1^3}{(\Gamma_2 - \Gamma_1^2)^{3/2}}$$

and

$$\text{(A10)}\quad K_F = \frac{\Gamma_4 - 4\Gamma_1\Gamma_3 + 6\Gamma_1^2\Gamma_2 - 3[\Gamma_1^4 + (\Gamma_2 - \Gamma_1^2)^2]}{(\Gamma_2 - \Gamma_1^2)^2}$$

where $\Gamma_i = \Gamma(1 + (i/c))$ is the Gamma distribution (available as a built-in function in most computer spreadsheets — given as GAMMA(x) in Excel®). The S_F is a measure of function symmetry. Negative skewness indicates left skew (excess of small x-values relative to a normal distribution), positive skewness indicates right skew (excess of large x-values relative to a normal distribution), and zero skewness indicates a symmetrical (normal) distribution (National Institute of Standards and Technology 2013). The K_F, in contrast, is a measure of the degree of function tailing relative to a normal distribution. Positive kurtosis (leptokurtic derivative function) indicates more/heavier tailing (more small and large "outlier" x-values) than a normal distribution, and negative kurtosis (platykurtic derivative function) indicates less/lighter tailing (fewer small and large outlier x-values) than a normal distribution (National Institute of Standards and Technology 2013). The kurtosis of a normal distribution is zero when K_F is defined as in eq. A10. Table 2 gives proposed S_F and K_F categories

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

based on the inclusive graphic classifications of Folk (1980). As an example, the derivative function in Fig. 2*a* is moderately right skewed ($S_F = 0.9620$) and slightly leptokurtic ($K_F = 1.0440$) — indicating that it has both a moderate excess of large *x*-values and slightly more outlier *x*-values than the derivative function of a normal distribution.

The integral average of the Weibull derivative function (A_I) is obtained using the fundamental theorem of integral calculus, i.e.,

$$(A11)\quad A_I = \left.\frac{dY}{dx}\right|_{A_I} = \frac{1}{(x_F - x_0)}\int_{x_0}^{x_F}\frac{dY}{dx}dx$$

which readily simplifies to

$$(A12)\quad A_I = \frac{(Y_F - Y_0)}{(x_F - x_0)}\left[1 - e^{-k(x_F - x_0)^c}\right]$$

since:

$$(A13)\quad \int_{x_0}^{x_F}\frac{dY}{dx}dx = Y(x_F) - Y(x_0)$$

The integral average of the derivative function describes a rectangle of length $(x_F - x_0)$ and height A_I which has the same area as the area under the derivative function curve between x_0 and x_F (Fig. 2*a*). The A_I value can consequently be viewed as a mean rate of change of *Y* with respect to *x* (i.e., mean *dY*/*dx*) between x_0 and x_F.

A2. Weibull cumulative function: mean, inflection, quantiles, and domain

The mean, $(\overline{x},\overline{Y})$, is given by (Lai and Xie 2006; NCSS, Weibull 2020):

$$(A14)\quad \overline{x} = x_0 + \left[k^{(-1/c)}\right]\Gamma\left(1 + \frac{1}{c}\right)$$

$$(A15)\quad \overline{Y} = Y_F - (Y_F - Y_0)e^{-k(\overline{x} - x_0)^c}$$

where $\Gamma(1 + (1/c))$ is the Gamma function (Fig. 2*b*). The $\overline{x}$ value is an average *x* of the cumulative function weighted according to the corresponding derivative function (e.g., Glen 2020*c*), i.e.,

$$(A16)\quad \overline{x} = \int_{-\infty}^{\infty} x\frac{dY}{dx}dx$$

The $\overline{x}$ lies to the left of the modal *x* (x_N) for negative-skewed (left-tailed) derivative functions and to the right of x_N for positive-skewed (right-tailed) derivative functions (Fig. 2*a*). Note also that although $(\overline{x},\overline{Y})$ can be back-substituted into the derivative function (eq. 1), the resulting dY/dx value applies only at $\overline{x}$ and is, therefore, not the same as A_I (Fig. 2*a*).

The inflection, (x_I,Y_I), is derived in the same fashion as the derivative function mode, leading to

$$(A17)\quad x_I = x_0 + \left[\frac{1}{k}\left(\frac{c-1}{c}\right)\right]^{1/c};\ c \geq 1$$

$$(A18)\quad Y_I = Y_F - (Y_F - Y_0)e^{[(1-c)/c]};\ c \geq 1$$

and hence (x_I,Y_I) is identical to (x_M,Y_M). The inflection locates the point where the cumulative function changes from convex (upward bending) to concave (downward bending) (Fig. 2*b*), thus demarking the maximum slope of *Y* vs. *x* (i.e., maximum *dY*/*dx*), and thereby the peak (mode) of the corresponding derivative function (Fig. 2*a*).

The quantiles, (x_Q,Y_Q), are specified points on the cumulative function that represent fractions of the function maximum, e.g., $0.2|Y_F\text{-}Y_0|$, $0.6|Y_F\text{-}Y_0|$, etc. Quantiles are obtained by setting $x = x_Q$, then equating a normalized form of the cumulative function to the *Q*th quantile, i.e.,

$$(A19)\quad e^{-k(x_Q - x_0)^c} = \frac{(Y_F - Y)}{(Y_F - Y_0)} = 1 - Q;\ 0 \leq Q \leq 1$$

then solving for x_Q:

$$(A20)\quad x_Q = x_0 + \left[-\frac{1}{k}ln(1 - Q)\right]^{1/c}$$

and then back-substituting x_Q into eq. 4:

$$(A21)\quad Y_Q = Y_F - (Y_F - Y_0)e^{-k(x_Q - x_0)^c}$$

Therefore, $x_Q = x_{0.2}$ (for example) is given by

$$(A22)\quad x_{0.2} = x_0 + \left[-\frac{1}{k}ln(1 - 0.2)\right]^{1/c}$$

Interestingly, $x_Q = x_{0.5}$ is given by

$$(A23)\quad \begin{aligned} x_Q = x_{0.5} &= x_0 + \left[-\frac{1}{k}\ln(1-0.5)\right]^{1/c} \\ &= x_0 + \left[-\frac{1}{k}\ln(0.5)\right]^{1/c} = x_0 + \left[\frac{1}{k}\ln(2)\right]^{1/c}; \end{aligned}$$

and hence, the 0.5 quantile $(x_{0.5},Y_{0.5})$ (Fig. 2*b*) also locates the median $(x_N, dY/dx|_N)$ (Fig. 2*a*). Note as well that time, *t*, *Y*(*t*), and rate (*dY*/*dt*) at the 0.5 quantile are sometimes used as diagnostic metrics in seed germination, seedling emergence, and mineralization studies (e.g., Gardarin et al. 2011; Gan et al. 1996; Yang et al. 2020). That is, 0.5 quantile time is given by $x_{0.5}$ (eq. A23); germination, emergence, or mineralization rate at $x_{0.5}$ is given by

$$(A24)\quad R_{0.5} = \left.\frac{dY}{dx}\right|_{0.5} = ck(x_{0.5} - x_0)^{c-1}|Y_F - Y_{0.5}|$$

and cumulative germination, emergence, or mineralization at $x_{0.5}$ is given by $Y_{0.5}$ (eq. A21).

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

The domain, x_D, of our adapted Weibull function is defined here as

$$x_D = x_F - x_0 \quad \text{(A25)}$$

where x_0 is the initial (minimum) x-value, and x_F is the final (end-point) x-value (Fig. 2*b*). The corresponding Weibull function range, Y_R, is therefore $Y_R = Y_F - Y_0$, where $Y_F = Y(x_F)$ and $Y_0 = Y(x_0)$. It is clear from eq. 4 and Fig. 2*b* that the adapted Weibull cumulative function is finite and defined at x_0 but asymptotic to the x-axis at x_F (i.e., $x \to \infty$ as $Y \to Y_F$). Hence, x_0 is readily obtained, but x_F must be estimated. We estimate x_F here by equating it to the 0.9999 quantile in eq. A20, i.e.,

$$x_F \approx x_{0.9999} = x_0 + \left[-\frac{1}{k}\ln(1 - 0.9999)\right]^{1/c} \quad \text{(A26)}$$

which is the x-value that corresponds to

$$\frac{(Y_F - Y)}{(Y_F - Y_0)} = 0.0001 \quad \text{(A27)}$$

or

$$Y = 0.9999Y_F + 0.0001Y_0 \quad \text{(A28)}$$

Our estimated x_F consequently approximates $Y(x_F)$ to ≥99.99 % of Y_F (Fig. 2*b*). Note also that the derivative function departs from zero at $x = x_0$ and returns to zero at $x = x_F$ (Fig. 2*a*).

Downloaded From: https://complete.bioone.org/journals/Canadian-Journal-of-Soil-Science on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use