12 Figures, 2 Tables, 0 Appendices, 0 Highlighted Boxes, 0 Footnote

Supplement 1. Mathematical Concepts

S.1. Linear Versus Quadratic Functions

A linear function reflects constant proportionality between two variables and contains a constant offset of one variable relative to the other. Linear functions are written in the general form:

$$y = m\,x + b \qquad \text{(S1)}$$

where m represents the proportionality constant, or the ***slope***, and b represents the offset constant, or the ***y-intercept***. When plotted on a graph, Equation (S1) will reflect a straight line (Fig. S1). In a linear relationship, the slope of the line is the ratio of vertical distance to horizontal distance (often referred to as 'rise over run'). For a truly vertical line, the slope is ∞. For a truly horizontal line, the slope is 0. The slope is the same for all points on the line. Linear functions are ***first-degree polynomials*** because the independent variable (in this case x) has an implied exponent of 1. (A polynomial is defined as an algebraic expression of the general form $y = ax^k$, where k is a positive integer and a is a real number.)

Quadratic functions are ***second-degree polynomials***, meaning they describe a parabolic relationship between two variables (Fig. S2). Quadratic functions are written in the general form:

$$y = ax^2 + bx + c \qquad \text{(S2)}$$

where a, b, and c are constants. Different parabolic shapes, degrees of curvatures, and sizes can be defined by differences in the values of a, b and c. At $y = 0$, all values of x that satisfy the relationship:

$$ax^2 + bx + c = 0 \qquad \text{(S3)}$$

are referred to as the *roots of the quadratic*. At the roots, Equation (S3) can be rearranged to define x in terms of a, b and c:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \qquad \text{(S4)}$$

The ± sign in Equation (S4) reflects the existence of two possible roots.

S.2 Exponents and Logarithms

An exponent (m) is defined as a positive integer indicating the factorial multiplication of a real number (a). Thus, a^m is defined as a multiplied by itself m times. By convention, a^{-m} represents $1/a^m$. Thus, $3^4 = 3*3*3*3 = 81$ and $3^{-4} = 1/(3^4) = 1/81$. From the last example, it is clear that $3^2*3^2 = 3^4 = 81$. Exponential terms can be combined according to a series of rules. Thus, $a^m * a^n = a^{m+n}$, $a^m/a^n = a^{m-n}$, and $a^m a^n/a^q = a^{(m+n)-q}$.

When present as a fraction, exponents reflect the root of a number as follows:

$$y = \left(a^{1/m}\right) = \sqrt[m]{a} \qquad \text{(S5)}$$

Thus, y is equal to the m^{th} root of a. As an example, $3^4 = 81$ can be rewritten as $81^{1/4} = 3$. Using Equation (S5) to define a root:

$$a^{m/n} = \left(\sqrt[n]{a}\right)^m \qquad \text{(S6)}$$

Logarithms are defined with respect to the exponential series of base 10. Thus, log 100 = 2 means that base 10 must be raised by the power of 2 to equal 100; taking the example further, $\log 10^2 = 2 \log 10$. (Note that the log of 10 is 1, since 10 raised to the power of 1 equals 10.) In general terms, if $\log a = b$, then $10^b = a$, and $\log a^b = b \log a$. Logarithms can be manipulated algebraically according to a series of rules. Thus, $\log^{(rs)} = \log r + \log s$ and $\log^{(r/s)} = \log r - \log s$.

To this point, we have only considered logarithms within the context of base 10. Such functions are often called *common logarithms*. In the application of calculus, the logarithm base is often expressed in terms of the irrational number 2.71828, which is designated as *e*. When expressed in base *e*, logarithmic functions are referred to as *natural logarithms* and denoted as ln. This designation reflects the fact that *e* arises naturally from certain mathematical derivations, and can be applied effectively to the description of natural phenomena. Natural logarithms follow the same rules of algebra described above. Natural logarithms can be related to common logarithms by recognizing that $e^{\ln x} = x$, and taking the log of both sides of the equation:

$$\log\left(e^{\ln x}\right) = \ln x \log e = \log x \qquad \text{(S7)}$$

to derive the following relationship:

$$\ln x = \frac{\log x}{\log e} \qquad \text{(S8)}$$

S.3. Trigonometric Functions

Trigonometric functions relate the sides of a triangle to its angles. Three of the most useful trigonometric functions are the *sine*, *cosine*, and *tangent*. To illustrate, imagine angle θ which is formed as radius *r*, which is fixed to the vertex of a circle, and is rotated from the *x* axis in a counterclockwise direction (Fig. S3). Angle θ can be defined in units of *degrees* or *radians*, where a radian is defined as the length of arc *q* relative to radius *r* (1 radian = q/r). Trigonometric functions can be applied to the triangle that is formed by placing *r* within the context of *x, y* coordinates:

$$\sin\theta = y/r, \quad \cos\theta = x/r, \quad \tan\theta = y/x \qquad \text{(S9)}$$

Trigonometric functions are useful in quantifying the periodic phases of cyclic processes. Starting from the positive side of the *x*-axis and rotating the radius counterclockwise (through the quadrant defined by x, $y > 0$), the sine function will progress through positive values to a maximum of 1 at θ=90° (or $\theta = \pi/2$ radians). Continuing in the counterclockwise direction, the sine function will continue to reflect positive, but decreasing numbers to a value of 0 at θ= 180° (or π radians). (Keeping in mind that sin $\theta = y/r$, and noting that $y = 0$ at θ= 180°, it is clear why sin θ= 0 at θ= 180°.) As r is rotated still further in the counterclockwise direction, the sine function will reflect negative numbers (values of y are negative as r is rotated through the quadrant defined by x, y <0), reaching a minimum of –1 at θ= 270° (or 3π/2 radians). Finally, as r is rotated back to the x axis, the sine function will progress through diminishing negative numbers to a final minimum of 0 at θ= 360° (or 2π radians). The progression of the sine function through these cyclic phases can be represented by a periodic 'wave' pattern, with a periodicity of 360° (or 2π radians) (Fig. S4). Thus, cyclic processes can be described using trigonometric functions, and the trigonometric frequency (360° or 2π radians) can be calibrated to the frequency of cycles.

S.4. Differential Calculus

Differential calculus is the calculus of *derivative functions* – i.e., functions that define the rate of change in a dependent variable, as the span of the independent variable approaches an infinitesimally small value. Algebraically, the slope of a function is defined as $\Delta y/\Delta x$, given a function (f) that relates two variables as $y = f(x)$. For many functions, however, $f(x)$ cannot be defined at x. As an example consider the function:

$$y = f(x) = \frac{x^2 - 1}{x - 1} \qquad \text{(S10)}$$

In Equation (S10), $f(x)$ is undefined at $x = 1$. (At $x = 1$ the function yields a solution of 0/0 which cannot be mathematically defined.) One means of solving this function is to let values of x get as close as possible to 1 without actually equaling 1. In other words, let $(x - 1)$ approach 0 [denoted as $(x - 1 \rightarrow 0)$]. For values of $x \neq 1$, $f(x)$ in Equation (S10) can be defined as:

$$y = f(x) = \frac{x^2-1}{x-1} = \frac{(x+1)(x-1)}{(x-1)} = x+1 \qquad \text{(S11)}$$

As values of *x* approach 1, *f* (*x*) approaches 2. Thus, if we define *f* (*x*) with the caveat that (*x* – 1) approaches infinitesimally small values, then *f* (*x*) = 2. Under these conditions, *f* (*x*) is said to be evaluated ***within the limit*** as (*x* – 1)→0. The evaluation of a function as the independent variable becomes infinitesimally small is defined mathematically as:

$$\lim_{x \to a} f(x) = \mathrm{M} \qquad \text{(S12)}$$

which can be read, 'as the independent variable (*x*) approaches the value *a*, and the interval (*x*–*a*) becomes infinitesimally small, the dependent variable will equal M'.

The requirement to evaluate functions as they approach their limit is particularly relevant to the study of biotic and abiotic fluxes. Flux (F) is defined as the product between a proportionality coefficient (*K*) and the space-dependent concentration gradient (Δc/Δ*x*). Thus, F = *K* (Δc/Δ*x*). As long as c changes linearly as a function of *x*, Δc/Δ*x* can be evaluated algebraically. The evaluation becomes difficult, however, if c changes in a non-linear manner across distance Δ*x* (Fig. S6). With such circumstances, it is often most meaningful to evaluate Δc/Δ*x* (and thus flux) at a specific point as Δ*x* → 0. Mathematically, flux at a specific point can be represented as:

$$\mathrm{F} = K \lim_{x_2 \to x_1} \frac{\mathrm{c}_2 - \mathrm{c}_1}{x_2 - x_1} = K \lim_{\Delta x \to 0} \frac{\Delta \mathrm{c}}{\Delta x} = K \frac{d\mathrm{c}}{dx} \qquad \text{(S13)}$$

To appreciate the concept of a point-specific change in c, imagine a straight line drawn at a single point, A, whereby the ***tangent*** of that line represents the slope, or *d*c/*dx* at A (Fig. S6). In the language of differential calculus, point-specific slopes are denoted as *dy*/*dx*. This relationship reads that the ***derivative*** of *y* is evaluated with respect to *x*. Another way to refer to the derivative is to say that *y* is ***differentiated*** with respect to *x*.

It is important to keep in mind that in differential calculus, the dependent variable (y) is differentiated with respect to a *function* (f) defined in relation to the independent variable (x). The use of first principles to mathematically differentiate functions can be complex. Instead, most people use a series of rules that form a reliable foundation for differentiation problems. The simplest rule involves the function $f(x) = a$, where a is a constant. Imagine the graphical representation of this relationship – i.e., for all values of x, $y = a$. The slope of this relationship will be 0, leading to the formal statement:

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} 0 = 0 \qquad \text{(S14)}$$

Thus, the derivative of $f(x)$ = constant is 0. This result has an important implication – a constant can be added to any function without effect on the derivative of that function.

As another example, consider $f(x) = ax$. Graphically, this function would depict a straight line with slope a and y-intercept = 0. Conceptually, it is clear that if dy/dx reflects the slope of a function and if the slope of this function is a, then:

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{a\Delta x}{\Delta x} = a \qquad \text{(S15)}$$

Thus, the derivative of $f(x) = ax$ is a.

Several additional rules of differentiation are commonly used to solve differential equations. Some of these rules are listed in Table S1. A more complete discussion of the rules of differentiation can be found in a basic calculus book. One rule that is worth considering in more detail is the *chain rule*, one of the most commonly used tools in differential calculus. The chain rule is stated as:

$$\frac{df}{dx} = \frac{df}{du}\frac{du}{dx} \qquad \text{(S16)}$$

The chain rule is especially useful in the differentiation of functions that can be factored into separate entities, but linked with related variables. An example is presented in Appendix 9.1 (Chapter 9). In the latter case, $f(x) = e^{-GL/\cos\theta}$, and $f(x)$ is differentiated with respected to L. The resultant derivative is not straightforward because of the complexity of carrying multiple variables in the exponent, only one of which (L) is the independent variable driving the differentiation. If we let $c = -G/\cos\theta$, and $u = cL$, the chain rule can be applied as:

$$\frac{de^u}{dL} = \frac{de^u}{du}\frac{du}{dL} = \frac{de^u}{du}\frac{d(cL)}{dL} = ce^{cL} \qquad \text{(S17)}$$

It is worth considering two additional types of derivatives. *Higher-order derivatives* are derivatives of a derivative. Thus, the derivative of dx/dy is referred to as a *second derivative* and is denoted as:

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) \qquad \text{(S18)}$$

A second derivative can be thought of as the slope of a slope. Recall that flux reflects a derivative function (i.e., $K\, dc/dx$). The second derivative of a flux reflects the rate by which the flux changes with respect to the x coordinate (in this case distance in the x spatial coordinate). Higher-order derivatives can be solved using the same rules for derivatives. First, the primary derivative is solved, and then the primary solution is differentiated to obtain the second derivative.

Partial derivatives refer to the case when a dependent variable is defined by a function with two or more independent variables. When a derivative function is evaluated with regard to its partial derivatives, it is first evaluated with respect to one independent variable, then the other variables in their turn, in each case with all other independent variables held constant. Thus:

$$\frac{dy}{dx} = \left(\frac{\partial y}{\partial x}\right)_{m,n} dx + \left(\frac{\partial y}{\partial m}\right)_{x,n} dm + \left(\frac{\partial y}{\partial n}\right)_{x,m} dn \qquad \text{(S19)}$$

where the ***total derivative*** (dy/dx) is broken into its partial derivatives. (The symbol "∂" is used to denote partial derivatives.) The subscripts in each term refer to the independent variables that are held constant while the function represented in the numerator is evaluated with respect to the function represented in the denominator. Partial derivatives are solved using the same rules that apply to total derivatives.

S.5 Integral Calculus

Whereas differential calculus has as its focus the instantaneous slope of functions, integral calculus focuses on determination of the area beneath the curve defined by a function. The process of determining the area beneath curves is called ***integration***. In the case of fluxes, integration yields the total flux with respect to a definitive span of the independent variable. To illustrate the concept of integration, imagine a function $f(x)$ that produces the curve presented in Figure S7. The area (A) beneath the designated portion of the curve will be defined by the distance along the x-axis ($x_2 - x_1$) at one boundary and $f(x)$ at the other boundary. Now imagine a change in x ($x_2 + \Delta x$), with an accompanying change in the area beneath the curve (ΔA). The value for ΔA will reflect Δx at one boundary and $f(x)$ at the other boundary [i.e., $\Delta \mathrm{A} = f(x)(\Delta x)$]. As was the case for determining the slope of complex functions, determination of the area beneath the curve can be difficult. The task is made easier as $\Delta x \rightarrow 0$. In that case, $\Delta \mathrm{A} \rightarrow d\mathrm{A}/dx$, and $d\mathrm{A}/dx \rightarrow f(x)$. Stated formally:

$$\frac{d\mathrm{A}}{dx} = \lim_{\Delta x \to 0} \frac{f(x)\Delta x}{\Delta x} = f(x) \qquad \text{(S20)}$$

The aim of integration is to find the function that when differentiated equals $f(x)$. This function will equal A, the area beneath the curve represented by $f(x)$ when determined as $\Delta x \rightarrow 0$. The area under the curve at values outside the limit can be determined by multiplying the integration by dx. This is represented by:

$$\mathrm{A} = \int f(x)\,dx \qquad \text{(S21)}$$

Equation (S21) reads: the area beneath the curve (A) equals the integral of $f(x)$ determined across dx. The area defined in Equation (S21) is called the *indefinite integral*. As for the case of differentiation, integration is often done using a series of rules. Some of the more commonly used rules for indefinite integrals are listed in Table S2.

To gain a better conceptual grasp of integration, imagine finding the area beneath a curve by using a series of progressively smaller rectangles, all positioned next to each other, with the sum of their independent areas representing the total area beneath the curve (Fig. S8). The greater the number of rectangles that is squeezed under the curve, the more accurate the estimate of the area beneath the curve, since each rectangle has a flat upper boundary, and is thus only an approximation of its representative area beneath the curve. Using the concept of a limit, the area beneath the curve can be determined as:

$$\mathrm{A} = \lim_{\Delta x \to 0} \sum_{i=1}^{N} f(x_i)\ \Delta x \qquad \text{(S22)}$$

where N is the number of rectangles squeezed into the area beneath the curve and Δx is the thickness of each rectangle. Thus, within the limit as the thickness of each rectangle approaches 0 and the number of rectangles approaches ∞, the sum of the independent areas of all of the rectangles should equal the area beneath the curve. The limit defined in Equation (S22) is called the *definite integral*. The definite integral is related to the indefinite integral – i.e., the definite integral is the indefinite integral defined at its limits. The definite integral is simply another way of defining the area under a curve, in this case by quantification of the area of a simpler geometric relation defined within its limit. The definite integral is written as:

$$\int_{\mathrm{a}}^{\mathrm{b}} f(x)\, dx = \lim_{\Delta x \to 0} \sum_{i=1}^{N} f(x_i)\ \Delta x \qquad \text{(S23)}$$

where a and b on the left side of Equation (S23) define the points on the x-axis bounding the curve of interest.

Current computers make integration easy through a process known as *numerical integration*. Numerical integration involves an estimate of the integral between two points (b –

a) after arbitrarily choosing some number N of rectangular areas. Using a computer, $y_i = f(x_i)$ is evaluated for each interval ($b - a/N = \Delta x$) and multiplied by the total distance of the integral ($b - a$). The accuracy of the integral estimate will depend on the thickness of each rectangular area (i.e., how well the assumption $\Delta x \rightarrow 0$ is satisfied).

Occasionally, a variable is dependent on two or more separate independent variables, rather than one variable as has been assumed to this point. Integration of the relationship between these multiple variables requires use of a *multiple integral*. As an example, take the relationship $z = f(x,y)$, where z is a variable dependent on the values of x and y. Because there are two independent variables, integration is determined within the bounds of a curve defined in two-dimensions (x and y), rather than one (x alone). The formal definition of a multiple integral in two dimensions is:

$$\iint f(x,y)\ d\mathrm{A} = \lim_{\Delta A \to 0} \sum_{i=1}^{N} f(x_i, y_i)\ \Delta \mathrm{A}_i \qquad \text{(S24)}$$

Note that in Equation (S24) the relevant variable upon which the integration is based is change in area (dA), a two-dimensional variable, not change in Δx alone.

S.6 Analytical versus Numerical Solutions to Equations

In the biophysical relations used in this book, equations represent models of the behavior of a biological or physical system. In order to predict the behavior of the system under a specified set of conditions, the set of model equations must be solved. If we know *a priori* how the components of the modeled system work (e.g., Newton's Second Law that force equals mass times acceleration, f = ma), we can solve the model with an *exact solution*, which is referred to as an *analytical solution*. An analytical solution is also often referred to as a *closed-form solution* because it results in complete *mathematical closure* of all terms in the model. This is the preferred manner for resolving biophysical models. Often, however, the model is complex and some of the relations represent approximations of how the components of the system interact. In that case, an analytical solution may not be possible. Instead, we have to form an *approximate solution*; this is typically accomplished through numerical methods leading to a *numerical solution*. Numerical solutions often require the creation of higher-order terms that allow approximation of the lower-order terms for which numerical definition is sought; but which leave the higher order terms without exact definition. This process of creating higher-order terms creates what is often referred to as the ***model closure problem***. Numerical models are eventually closed through approximation of the higher-order terms. As long as these higher-order terms are far enough removed from the lower-order terms, the model can be resolved with confidence in the lower orders.

One of the most frequently-used methods for the numerical solution of ordinary differential equations is to define the model for an initial condition, then use the model equations to predict change in the dependent variable over very short span of change in the independent variable. For example, imagine the fundamental relation (dx/dt), where x represents a relatively complex model describing the time-dependent behavior of a system. We can initiate a numerical solution for the model by stating from $t = 0$. The behavior of x as a function of t can then be predicted through a series of iterated steps in which x is predicted as a function of t, producing a short interval from which x can be evaluated. The exact nature of the dependence of x on t will only be solved in the limit as $t \rightarrow 0$. However, we can approximate the solution to this dependence by characterizing the short interval. This approach is referred to as ***Euler's method***. It produces an accurate estimate of the relation between the dependent and independent variables near the initial condition, e.g., at $t = 0$, and as long as the intervals are short; the shorter the interval, the more accurate the approximated solution. Nonetheless, numerical solutions produce an unavoidable error associated with truncation of the interval at an arbitrary point. This ***truncation error*** can be estimated through comparison of the defined interval against a polynomial approximation through a ***Taylor expansion series***, as described in the next section. There are alternatives to Euler's method for numerical estimation (e.g., the Runge-Kutta method), but we will not discuss those methods in detail here.

S.7 Polynomial Approximations of Elementary Functions (The Taylor Polynomial Theorem)

In some cases, complex functions that are not resolvable by computational methods, are more easily evaluated by ***polynomial approximation***. As in the case for numerical methods as applied to ordinary differential equations, the polynomial model will not be an exact solution, but it does provide a means of evaluating the degree of error in model closure, and it thus provides some degree of control over the approximation. To illustrate the polynomial approximation, let $f(x)$ represent an exact function, and let $P_n(x)$ represent the polynomial approximation of $f(x)$. Accordingly:

$$f(x) = P_n(x) + R_n(x) \qquad \text{(S25)}$$

where $R_n(x)$ represents the error (or remainder) between the approximation and the exact function. This error is the truncation error, or closure error, associated with non-analytical methods. The polynomial expansion and resulting error can be quantified using the *Taylor Polynomial Theorem*:

$$f(x) = f(c) + \frac{f'(c)}{1!}(x-c) + \ldots + \frac{f^{(n)}(c)}{n!}(x-c)^n + R_n(x) \qquad \text{(S26)}$$

where:

$$R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-c)^{n+1} \qquad \text{(S27)}$$

and ξ represents a number in the interval between c, the value for the polynomial approximation, and x, the value for the exact function, and f represents a differentiable function through $n+1$ derivatives.

Polynomial approximation is especially useful in dealing with the many non-linearities that emerge during the derivation of flux and energy-balance relationships. As with the numerical methods discussed above, a non-linear function relating two variables can be approximated as a first-degree (linear) polynomial using the Taylor expansion theorem. The error in making the approximation (R_n) will be minimized if the relationship is applied across a small interval for the two variables. [The difference between a non-linear function $f(x)$ and linear approximation $P_n(x)$will decrease as Δx decreases, minimizing the degree of non-linearity in $f(x)$.] A good example of use of the Taylor Polynomial Theorem is found in Chapter 9 in which linear approximations greatly facilitates derivation of the Penman-Monteith relation, which describes surface evaporation.

S.8 Probability Density Functions

The statistical foundation for many of the models used in predicting ecosystem-atmosphere fluxes is based on the probability of an event occurring. Probability (P) refers to the fractional number of successes in making an observation or predicting an event. Thus, P = number of successes/total number of attempts, where $0 \le P \le 1$. When P is expressed as a function of an independent variable [e.g., number of attempts, time (t), distance (x), etc.], a ***probability density function*** (pdf) and its associated graph can be described (Fig. S9).

Derivation of the pdf is essentially a problem in integral calculus. Given the distribution of probabilities across a defined interval of values for the independent variable, the total probability can be described by:

$$\mathrm{P} = \int_a^b f(x)\,dx \qquad \text{(S28)}$$

where a and b define the interval bounds and $f(x)$ represents the pdf. Given that P is defined as a fractional variable and has an upper limit of 1, we can write:

$$\mathrm{P} = \int_0^{+\infty} f(x)\,dx = 1 \qquad \text{(S29)}$$

One of the most fundamental pdf's is the ***Gaussian (normal) distribution*** (Fig. S10). The Gaussian distribution requires only two parameters for definition; the mean (μ), which is the central tendency of the distribution of events, and the standard deviation (σ), which is the tendency for dispersion about the mean. In formal terms, the Gaussian probability distribution is defined as:

$$f(x) = \frac{1}{\sqrt{(2\pi\sigma)}} e^{-[(x-\mu)^2/2\sigma^2]} \qquad \text{(S30)}$$

where σ^2 is the variance.

Statistical moments of a pdf are expressions of dispersion about the mean; they define the symmetry of the pdf. Mathematically, statistical moments are stated as:

$$\mu_n = (x - \mu)^n \qquad \text{(S31)}$$

where μ_n denotes the order of the moment (e.g., μ_1 is the first moment and μ_2 is the second moment). The first moment is defined as the mean itself. The second moment is defined as the variance (σ^2). Thus, to define a Gaussian distribution, one only needs the first and second moments. The third moment, which is defined as the third power deviation [$(x - \mu)^3$] is often used to evaluate *skewness*, which reflects asymmetry in the distribution around the mean (Fig. S10). The fourth moment is the fourth power deviation [$(x - \mu)^4$], and is often used to evaluate *kurtosis*, the tendency for peakedness in the distribution.

Skewness in the distribution of x (Sk_x) is defined as:

$$Sk_x = \frac{\overline{x'^3}}{\sigma_x^3} \qquad \text{(S32)}$$

where x' is the deviation from the mean (i.e., $x - \mu$) and σ is the standard deviation of x. A Gaussian distribution would exhibit skewness equal to 0. When deviations from the mean are multiplied to the third power, the extremes become weighted most heavily. If the most extreme deviations tend to be negative more frequently than positive, a third-moment analysis will reveal negative skewness, and vice versa if the most frequent extreme deviations are positive. If the positive and negative deviations are evenly distributed (i.e., a Gaussian distribution), a third-power product will equal σ_x^3. Kurtosis in the distribution of x (Kr_x) is defined as:

$$Kr_x = \frac{\overline{x'^4}}{\sigma_x{}^4} \qquad \text{(S33)}$$

Expressing deviations as the fourth power causes the distribution to flatten. This flattened condition is ideal for analyzing aspects of peakedness. A Gaussian distribution will reveal a Kr_x of 3. *Leptokurtic* distributions have $Kr_x > 3$, and exhibit "fatter" tails in the distribution (i.e., there is a higher probability of an event occurring in the tails), compared to a Gaussian distribution. *Platykurtic* distributions have $Kr_x < 3$, and refer to the case where the flanks of the distribution are heavier, but the tails are thinner. The peak of a platykurtic distribution is flatter than the Gaussian case. In general, pdf's with significant skewness and/or kurtosis are referred to as *non-Gaussian*. Models of non-Gaussian functions are complex given the difficulty of predicting third and fourth moment terms.

In the statistical analysis of turbulence, *covariances* are frequent. A covariance is a 'double correlation'; i.e., two independent variables are multiplied together to reveal correlated patterns of variance. An example, is the eddy flux which consists of the covariance between the fluctuating components of vertical wind velocity (w') and scalar concentration (c'). (The term 'fluctuating components' is derived from Reynolds averaging, which is defined in Appendix 10.1, Chapter 10.) In formal terms, the time-averaged eddy-flux ($\overline{w'\mathrm{c}'}$) is a second-order statistical moment (a variance). When models are derived to predict $\overline{w'\mathrm{c}'}$, skewness emerges as a required term. In the case of eddy flux, skewness emerges as a *mixed third-order moment*. A mixed third-order moment refers to a 'triple correlation' among mixed variables. The mixed third-order moment that often emerges in modeling the eddy flux is $\overline{w'w'\mathrm{c}'}$. In order to define this third-order moment, kurtosis emerges as a required term. Thus, models of pdf's typically require that lower-order moments be defined in terms of higher-order moments. This exemplifies the same problem we described above with regard to mathematical closure of a model using numerical methods. In this case, using statistical modeling, the *closure problem* occurs because of our effort to define statistical moments in terms of higher-order moments, leading to more unknowns than equations, and thus preventing complete mathematical closure.

S.9 Scalars, Vectors and Tensors

Mathematical approaches in the field of physics often require that variable quantities be defined with regard to a coordinate system. Coordinate systems allow objects or points of reference to be explicitly defined in space. The most common coordinate system used in atmospheric physics is the *Cartesian coordinate system*; a system of coordinate axes that are mutually perpendicular and, by convention, referred to as the *x*, *y* and *z* coordinates. A fourth coordinate that is often referenced in the field of atmospheric physics is time, *t*. In fact, several models that are used in both diagnostic and prognostic roles in atmospheric physics are referred to as *4-dimensional models*, meaning that they include the Cartesian spatial coordinates and time as fundamental reference frames. Variables that are independent of a coordinate system are called *scalars*. Examples of scalars include mass, molecular density, temperature and humidity. These variables are represented by a single number that has magnitude, but is independent of direction in space or time. Variables that are defined within the context of a coordinate system, and thus have dimensions of *both* magnitude and direction, are called *vectors*. Examples of vectors include wind velocity, force and acceleration. Vectors represent quantitative entities that can be mathematically manipulated to provide the sum influence of a variable with contributions from different coordinate axes. Vectors are geometrically summed, when defined by the same coordinate axis, by placing them head to tail. When defined by different coordinate axes, the gross contributions of each vector quantity to a net vector quantity can be determined within a geometric framework (e.g., Pythagorean geometry).

The derivation and expression of equations to represent physical processes that involve vectors becomes complicated when forced to deal with multiple coordinates. In order to simplify such derivations, they are often written in *vector notation*, a type of 'shorthand' wherein the coordinates attached to a vector quantity are represented as subscript indices. Thus, velocity (v) can be represented in vector notation as v_i, where the subscript "i" represents the three coordinate components of the vector, i = 1, 2 and 3 (or *x*, *y* and *z* in the Cartesian coordinate system). Using summation notation, equations can be developed for the single variable v_i, with the understanding that behind the equation is an expansion of the relationship into all three component coordinates.

Most fluxes that are relevant to ecosystem-atmosphere interactions are defined with respect to a *vector field*, a field of vector quantities associated with each point in a defined spatial domain. The vector field will define the *net* coordinate components to any given flux. For example, the flux of heat across a surface will have components in the *x*, *y* and *z* directions that, when summed, provide a net heat flux from one side of the surface to the other. Some variables cannot be derived in only three coordinates. An example is momentum flux. Momentum is defined as mass times velocity, a scalar quantity times a vector quantity. Momentum flux is defined as momentum transferred across some surface area (m^2) per unit time (s), so that momentum flux has combined units of (kg m s^{-1}) (m^{-2} s^{-1}); these units condense to N m^{-2} (force per unit area), the same units as stress. Momentum, being the product of a scalar and a vector, can be defined in the three Cartesian coordinates, *x*, *y* and *z* which, for the case of wind, are identified as *u*, *v* and *w*, respectively. Momentum flux, has the additional quality of being able to transfer these three wind coordinates in one of three different Cartesian directions, *x*, *y* and *z*. Thus, momentum flux has nine components, rather than the three typically used for vector quantities (Fig. S.11). Momentum flux is an example of a ***tensor*** quantity. (Technically, one can equate the concepts of scalars, vectors and tensors by recognizing that a scalar is actually a zero-order tensor, a vector is a first-order tensor and momentum flux is a second-order tensor.)

Once again, in the spirit of trying to simplify the derivation of equations for tensor quantities such as momentum flux, we often rely on a type of shorthand, in this case referred to as ***Einstein's summation notation***. In this case, a tensor quantity (F) is represented as F_{ij}, where the two subscript indices represent i = 1, 2 and 3, j = 1, 2 and 3. In matrix algebra, these indices represent the cross multiplication of all nine coordinates.

Vector algebra is often written with special operators which define specific functions. One example that is particularly relevant to atmospheric physics is the gradient operator, ∇. The gradient operator defines the divergence or convergence in a vector field, for example:

$$\nabla \cdot \mathrm{v} = \frac{\partial \mathrm{v}_x}{\partial x} + \frac{\partial \mathrm{v}_y}{\partial y} + \frac{\partial \mathrm{v}_z}{\partial z} = \frac{\partial \mathrm{v}_i}{\partial x_i} \qquad \text{(S34)}$$

where the right-hand side of Equation S34 represents the divergence or convergence written in vector notation. In this case, the divergence or convergence is an expression of how much the velocity vectors (v_i) spread apart or come together, respectively, around a single point in Cartesian space.

S.10 Coordinate rotation

When considering the role of wind in transporting scalar quantities to produce a flux, we often run into a problem when undulating terrain forces the various wind vectors to take on differential importance. For example, when the wind is forced over an upward-oriented slope, even though the net direction of the wind remains predominantly horizontal (in the *x* coordinate), the vertical component of the wind (in the *z* coordinate) increases in importance. When the wind is forced down the other side of the slope, the vertical coordinate remains important, but the mathematical sign of the important component will be reversed. Thus, in situations where wind-driven processes are considered (as in many studies of turbulent flux) we ***rotate*** the 'local' coordinate system to provide a consistent, or standardized, frame of reference. In other words, we rotate the coordinate system to reflect ideal, perfectly flat terrain, irrespective of the nature of the real terrain.

To understand coordinate rotation, imagine a wind vector (v) with three orthogonally-oriented, Cartesian vectors (v_x, v_y and v_z) above flat terrain (Fig. S.12). Now, imagine that we need to rotate the coordinate system to define a new vector (v'), accounting for the up-and-down undulations (in the *z*-coordinate) above sloped terrain. In this case, we need to rotate the system within vertical space, through angle θ, which defines the angle of the undulation. The result will be two new coordinate axes, *x*', *z*', formed from the original axes, *x*, *z*. Vector v' can be defined in the new coordinate frame by drawing perpendiculars from the tail of the vector through the *x*-axis (Fig. S.12). In geometric terms, we can define v' with respect to axis *x*' as:

$$\begin{aligned} v_x' &= OA + AB \\ &= \frac{v_x}{\cos\theta} + \left(v_z - v_x \tan\theta\right)\sin\theta \\ &= \frac{v_x}{\cos\theta}(1-\sin^2\theta) + v_z\sin\theta \\ &= v_x\cos\theta + v_z\sin\theta \end{aligned} \qquad \text{(S35)}$$

A similar approach can be taken to define v_z' as:

$$v_z' = -\,v_x\sin\theta + v_z\cos\theta \qquad \text{(S36)}$$

In the rotation described above, the side-stream, lateral component (v_y) is unaffected. However, for situations where the terrain undulates laterally a similar rotation can be performed to define v_y'.

S.11 Inferential statistical modeling

Statistical analysis provides a means of organizing information about a process (or other type of phenomenon) into a quantitatively consistent model that can be used to predict the probability of a future outcome. Statistical models are a type of ***empirically-based model*** (see Section 1.D). Given the large amounts of data being generated by automated sensor networks statistical modeling has become an important activity in the earth system sciences. Here, we highlight some of the common approaches taken in statistical modeling, and we use the modeling of soil respiration as an example by which to illustrate the overall approach.

Let's begin by assuming that we have access to a set of observations of soil respiration rate, which were made at different spots on the landscape and at different times during the growing season. We know that soil temperature is a principal controlling variable and our set of observations covers a broad range of soil temperatures. Thus, one obvious application of statistical modeling might involve organizing the observations into a framework capable of predicting respiration rate from any given temperature. Using a classical inferential approach, we might develop a statistical regression with soil temperature represented as the independent variable and soil respiration rate as the dependent variable. We can then estimate the best-fit mathematical function to describe the relation; we refer to that function as our *process model*. Most likely, the model will be some type of exponential relation with a specified slope and y-intercept, which we refer to as *model parameters*. An assumption implicit in the analysis is that our method has *optimized* the model parameters given the condition of the observed data. In formal terms, we are predicting the *joint probability distribution* of a dependent variable (y) and a covariate (x):

$$y_i = \beta_0 + e^{\beta_1 x_i} + \varepsilon_i \qquad \text{(S.36)}$$

where y_i and x_i are the variables associated with independent observation i, β_0 and β_1 are the intercept and slope parameters in the model, respectively, and ε_i is the error referenced to observation i.

Now, having conducted this exercise, we assume that there is some level of uncertainty in the model parameters. The data we have collected are likely to provide an imperfect picture of the true relation between temperature and respiration rate. If we had assembled a different mix of observation days or measurement spots, we would have likely derived slightly different parameters. Thus, much of the uncertainty resides in the fact that we have arbitrarily chosen the set of data on which to base the model. We can use hierarchichal, mixed effects statistical approaches to partition the random effects in the observations, and thus quantify the uncertainty. In fact, if we want to control for these effects, we can create an even more complex model and include these effects as additional parameters. These types of activities would fall completely within the realm of *classical statistics*.

In recent years, it has become popular to deploy a second type of statistical inference known as *Bayesian analysis*. Bayesian analysis takes advantage of *prior knowledge* to shape an initial estimate of the model parameters. We then use the observations we have collected to 'hone' our initial estimate of the parameters. Prior knowledge about the distribution of parameters can narrow the range of estimates that are permitted and thus decrease uncertainty. Let's return to our example of soil respiration rate. Let's assume that we want to explore more fully the uncertainty in one parameter of the model, the slope of the regression; referred to here as β_1. If we treat temperature as a fixed covariate – in other words, we assume that it is actually measured and prescribed as an explicit value in the model – we can focus the analysis on the relationship between β_1 and the data that we've collected. In Bayesian terms, we would represent the parameter estimation process as:

$$p\,(\beta_1, \text{model} \mid \text{data}) \propto p\,(\text{data} \mid \text{model}, \beta_1)\; p\,(\text{model} \mid \beta_1)\; p\,(\beta_1) \qquad \text{(S.37)}$$

We can state Equation S.37 in words as: 'the joint probability distribution of β_1 and the model, given the observed data, is proportional to the probability of the data, given the model and β_1, multiplied by the probability of the model, given β_1, multiplied by the probability of β_1'.

Working from the right to the left, Equation S.37 states that: (1) if we have an estimate of the distribution of β_1, (2) if we know how the model responds to that distribution, and (3) if we can come up with a way to determine how the data compare to the model output using our estimated distribution of β_1, then (4) we can use the data to improve our initial estimate of the distribution of β_1. In order for Bayesian analysis to be most effective we should have access to both prior knowledge about the distribution of a parameter *and* a model that is dependent on the parameter. The aim of formulating Equation S.37 is to 'explore the parameter space' and determine the optimum estimate for the parameter(s); by optimum we mean the value of β_1 that provides the least error between the modeled data and observed data.

How do we find the parameter optimum? We begin with our previous knowledge about the distribution of β_1, which we will call our ***prior parameter distribution***, and designate it as p (β_1) (see Fig. S.13). We then sample values of β_1 from that prior distribution using a random sampling protocol; i.e., techniques such as Monte Carlo Markov Chain (MCMC) sampling or Gibbs sampling are often used. We enter these sampled values of β_1 into our model and predict a set of values that can be compared to the observed values. A ***merit function***, such as maximum likelihood, is then used to evaluate the total error between the sets of predicted and observed values. This error is often called the ***model-data mismatch***. The mismatch will be initially high, but using feedback between the merit and sampling functions, the sampled range of β_1 is progressively narrowed. Eventually, a set of parameter values is obtained which provide approximately equal amounts of model-data error. Through the numerous iterations (often numbering thousands) during which the sampling program is randomly choosing values of β_1, and given that the sampled range of β_1 narrows with each successive iteration, multiple selections of the same β_1 values will occur, and the frequency by which the program samples each value of β_1 can be represented as a probability distribution. The ultimate probability distribution that emerges from the process is called the ***posterior parameter distribution***. Provided that the data provides adequate constraint on the parameter estimation, the posterior distribution will be narrower than the prior distribution. An optimal value for β_1 can be selected from the peak of the posterior distribution and formal estimates of uncertainty can be generated from the distribution tails on either side of that peak. This is just one example of how Bayesian techniques can be applied to observed data. More detailed discussions of statistical aspects of the topic are presented in Clark and Gelfand (2006), Ogle and Barber (2008), and Cressie et al. (2009).

Table S1. Some of the most commonly used relationships for differentiation.

(1) $\frac{d\,(a)}{dx} = 0$, where a is a constant

(2) $\frac{d\,(ax)}{dx} = a$, where a is a constant

(3) $\frac{d\,(x^m)}{dx} = mx^{(m-1)}$

(4) $\frac{d\,(e^m)}{dx} = e^m$

(5) $\frac{d\ln x}{dx} = \frac{1}{x}$

(6) $\frac{d\,(au)}{dx} = a\frac{du}{dx}$, where $u = f(x)$

(7) $\frac{d\,(e^u)}{dx} = e^u\frac{du}{dx}$, where $u = f(x)$

Table S2. Some of the most commonly used relationships for integration.

(1) $\int a\ dx = a\,x$, where a is a constant

(2) $\int \frac{dx}{x} = \ln x$

(3) $\int (r+s)\ dx = \int r\ dx + \int s\ dx$

(4) $\int a\ f(x)\ dx = a \int f(x)\ dx$

(5) $\int e^x\ dx = e^x$

(6) $\int \sin x\ dx = -\cos x$

(7) $\int \cos x\ dx = \sin x$.

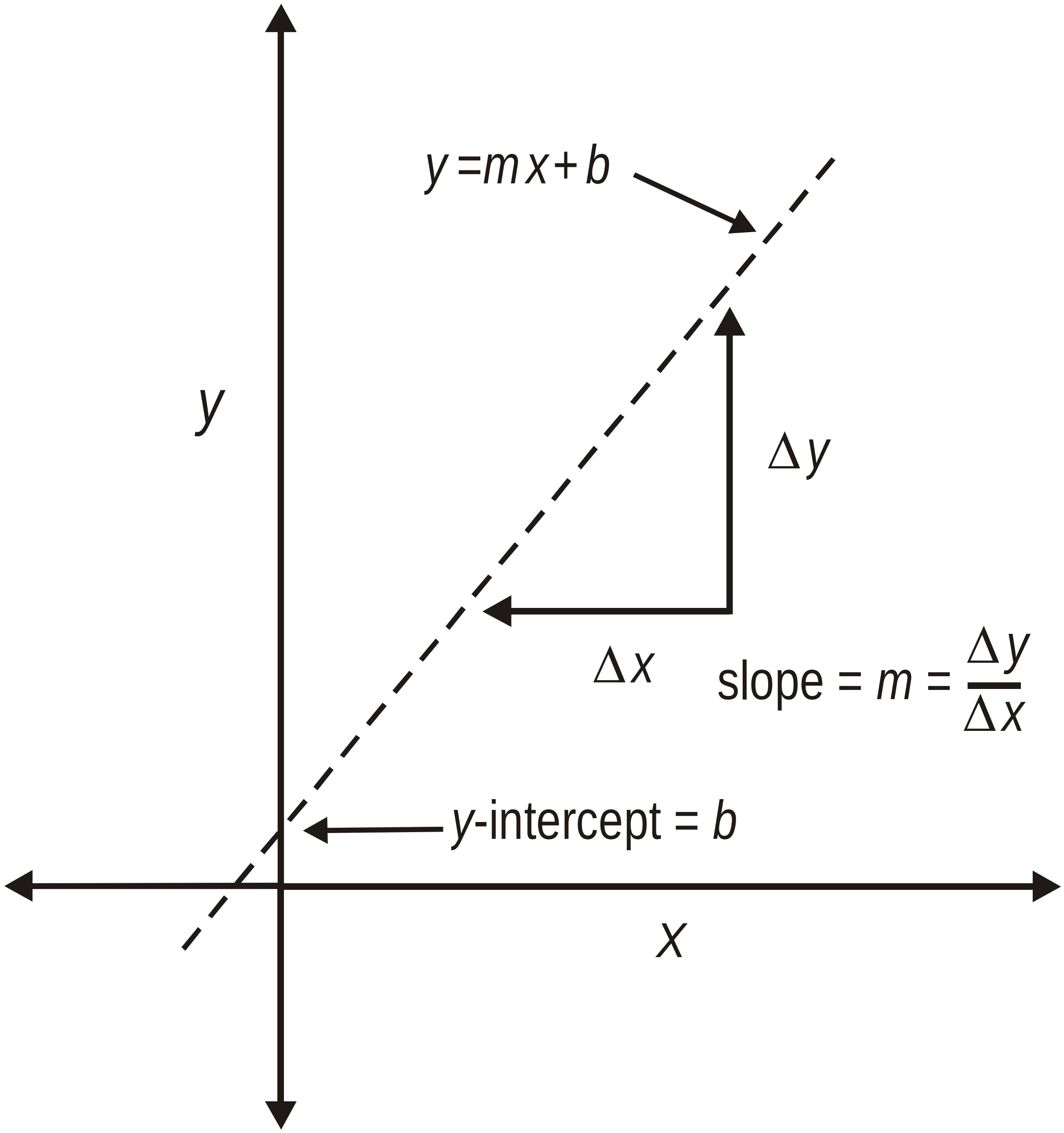

Figure S1. A linear relationship between independent (x) and dependent (y) variables.

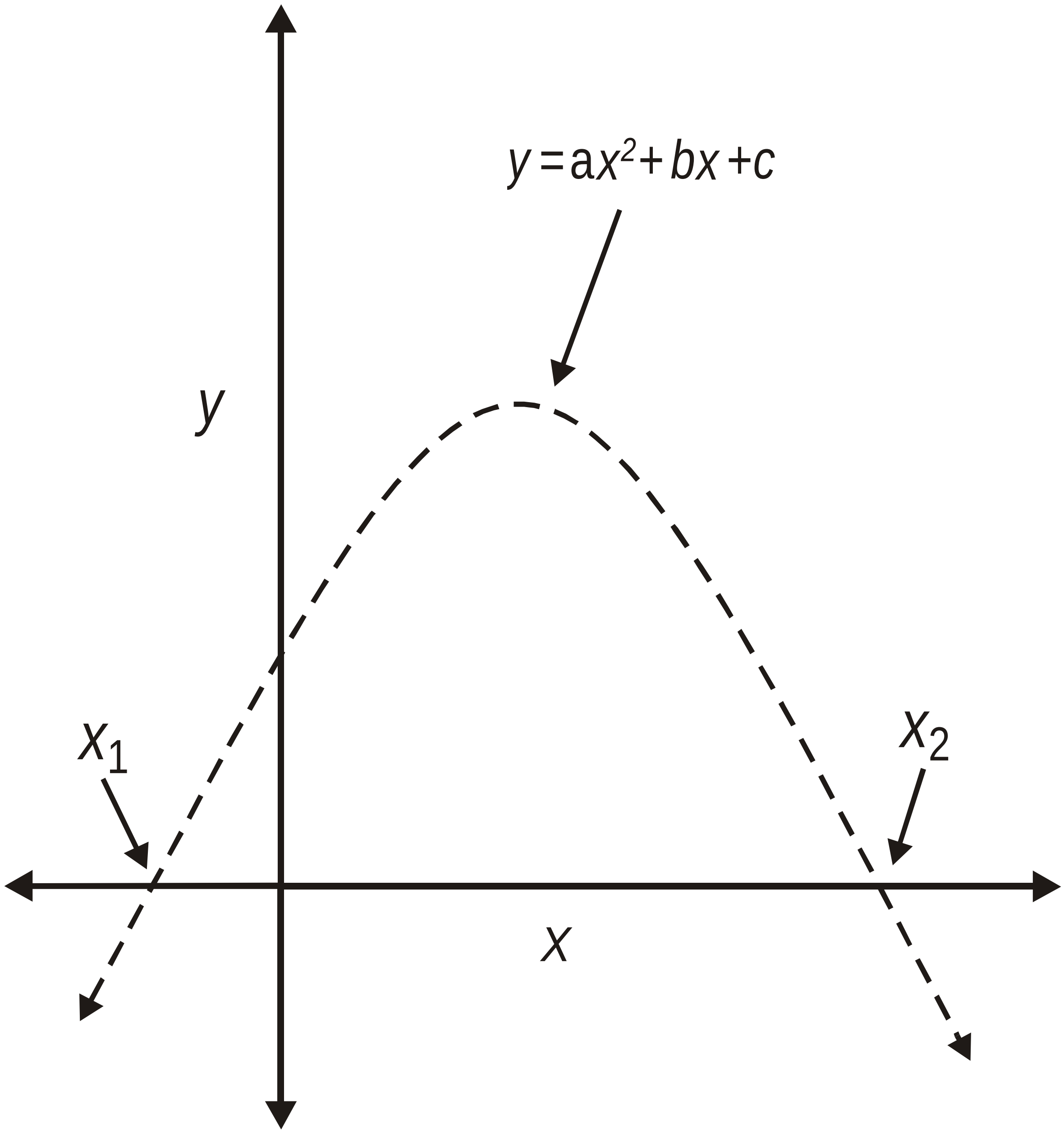

Figure S.2. A quadratic relationship between independent (x) and dependent (y) variables. The minimum and maximum roots are indicated as x_1 and x_2, respectively.

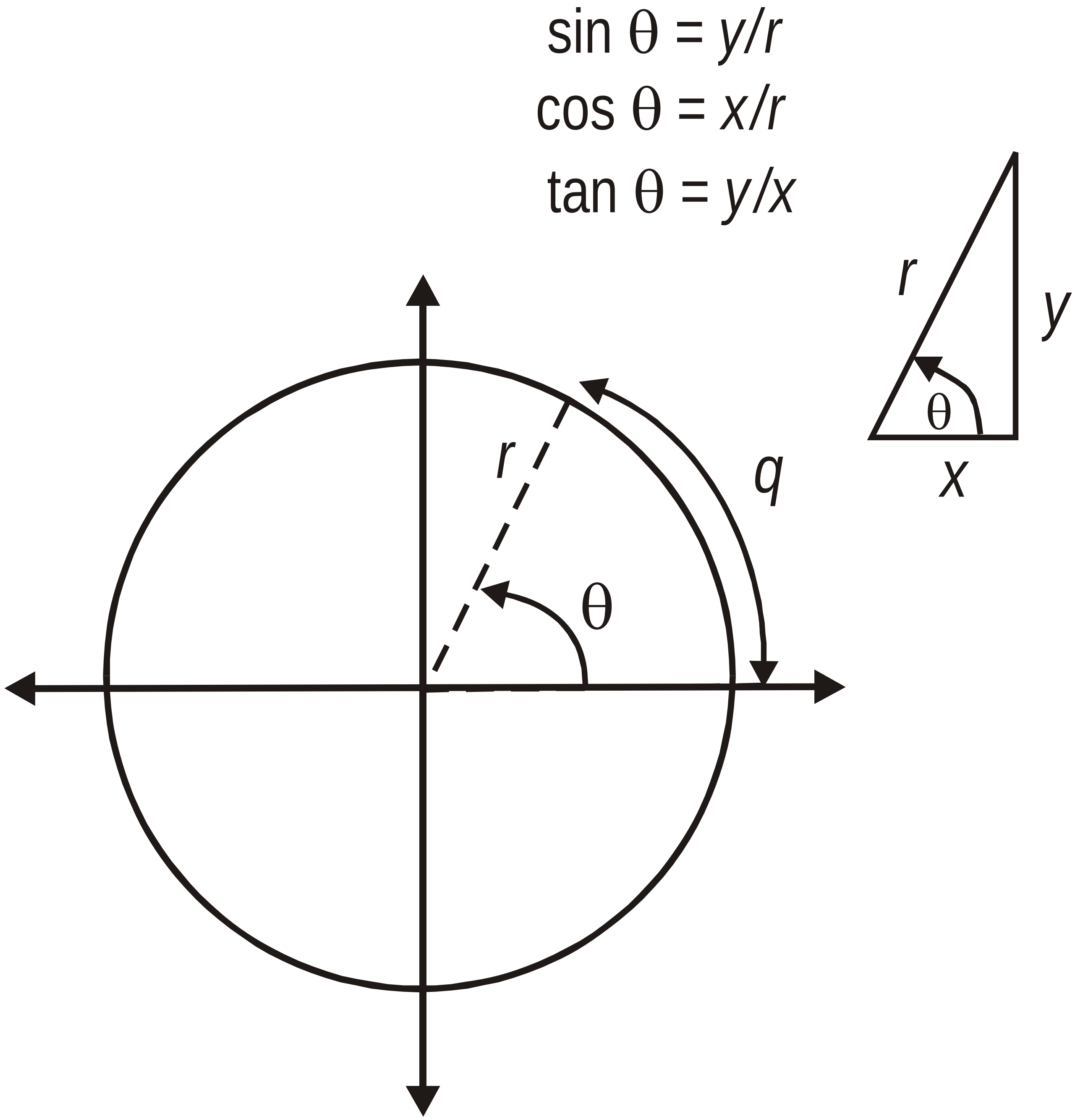

Figure S.3. The cyclic foundation for three of the primary trigonometric functions determined from triangular ratios. The triangle formed from the counterclockwise rotation of radius *r* and defined around angle θ is used to determine the *sin*, *cos* and *tan* according to the ratio of sides.

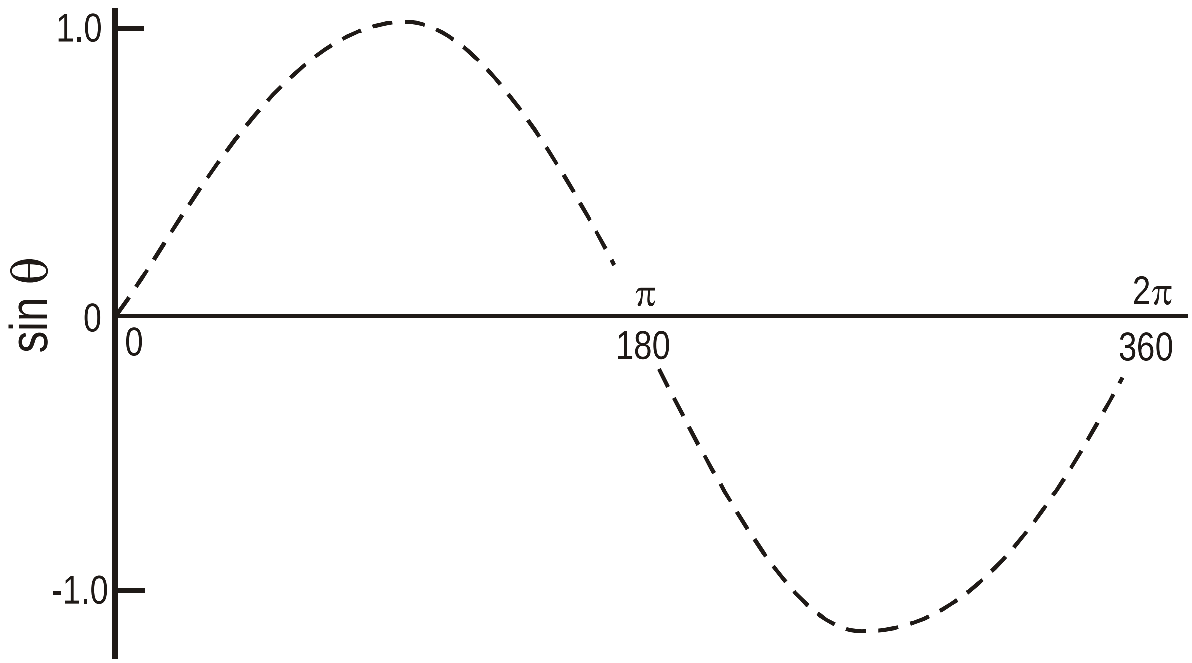

Figure S.4. The sine wave pattern formed from the progressive counterclockwise rotation of radius r plotted according to the angle θ in degrees and radians.

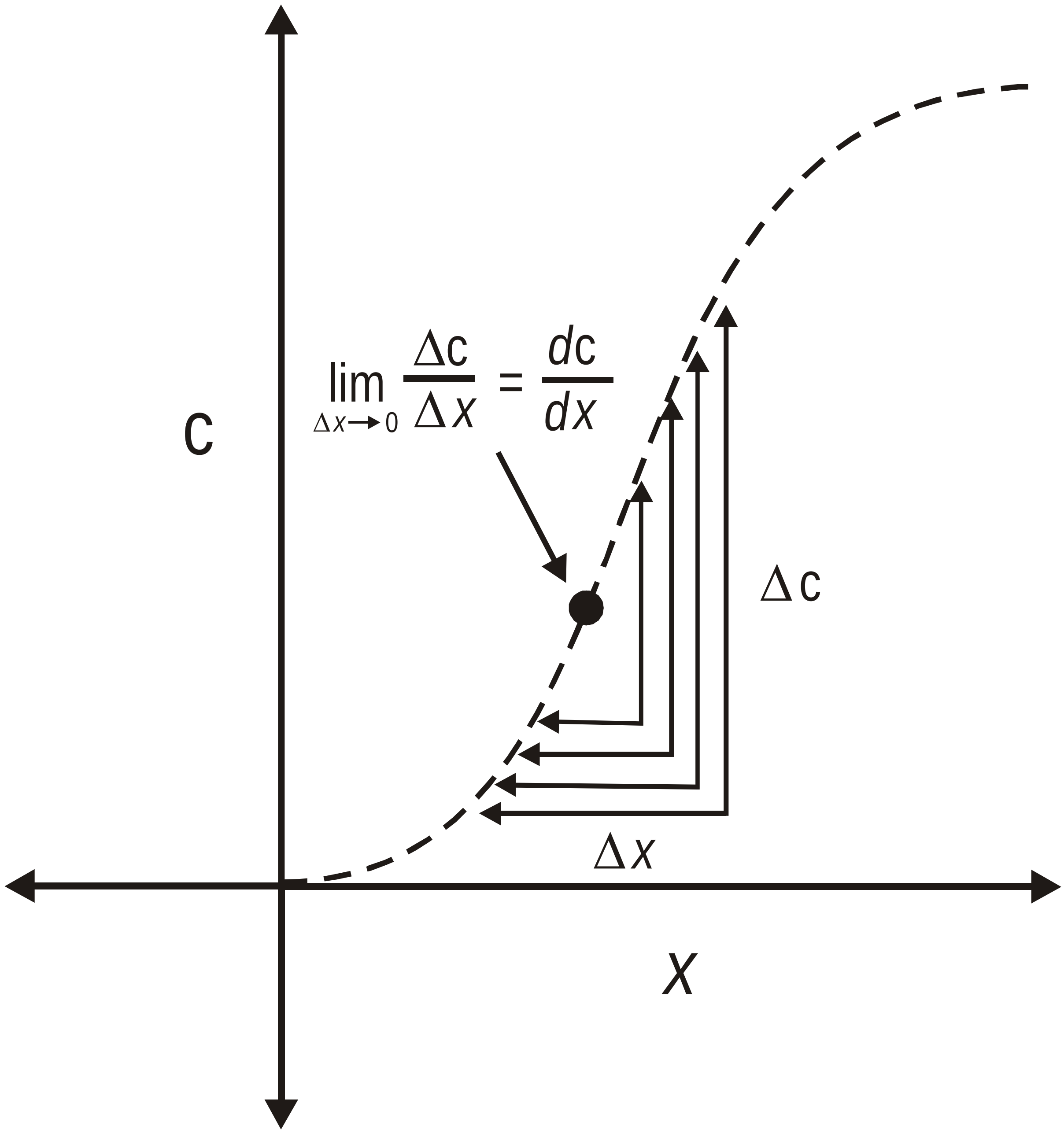

Figure S.5. The graphical representation of a limit in the relationship between scalar concentration (c) and distance (*x*). As Δx gets progressively smaller and eventually approaches 0 (i.e., $\Delta x \rightarrow 0$) the limit is reached.

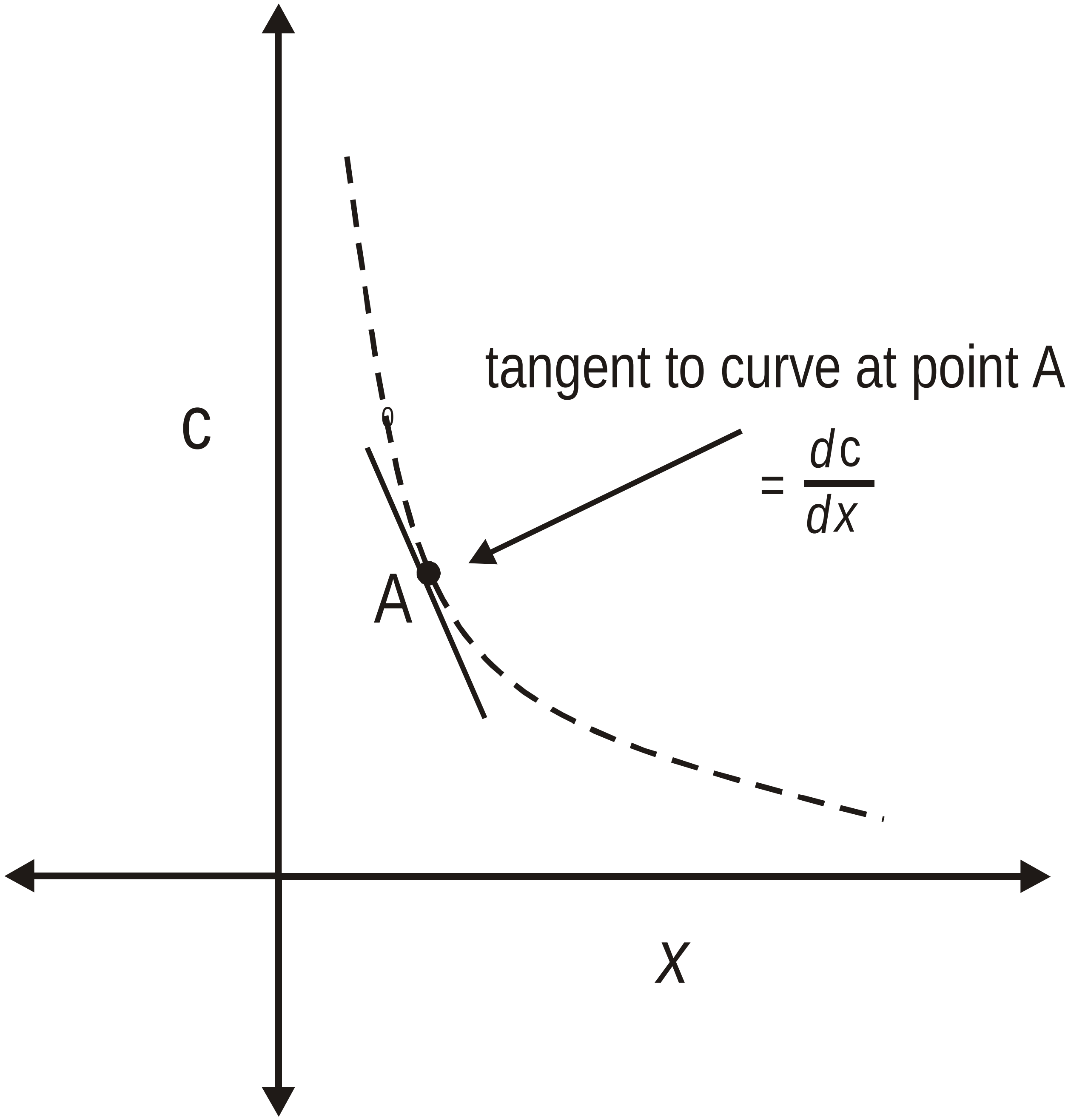

Figure S.6. The graphical representation of a point-specific slope illustrating the concept of the derivative of c with respect to *x* (*d*c/*dx*) as the tangent located at point A.

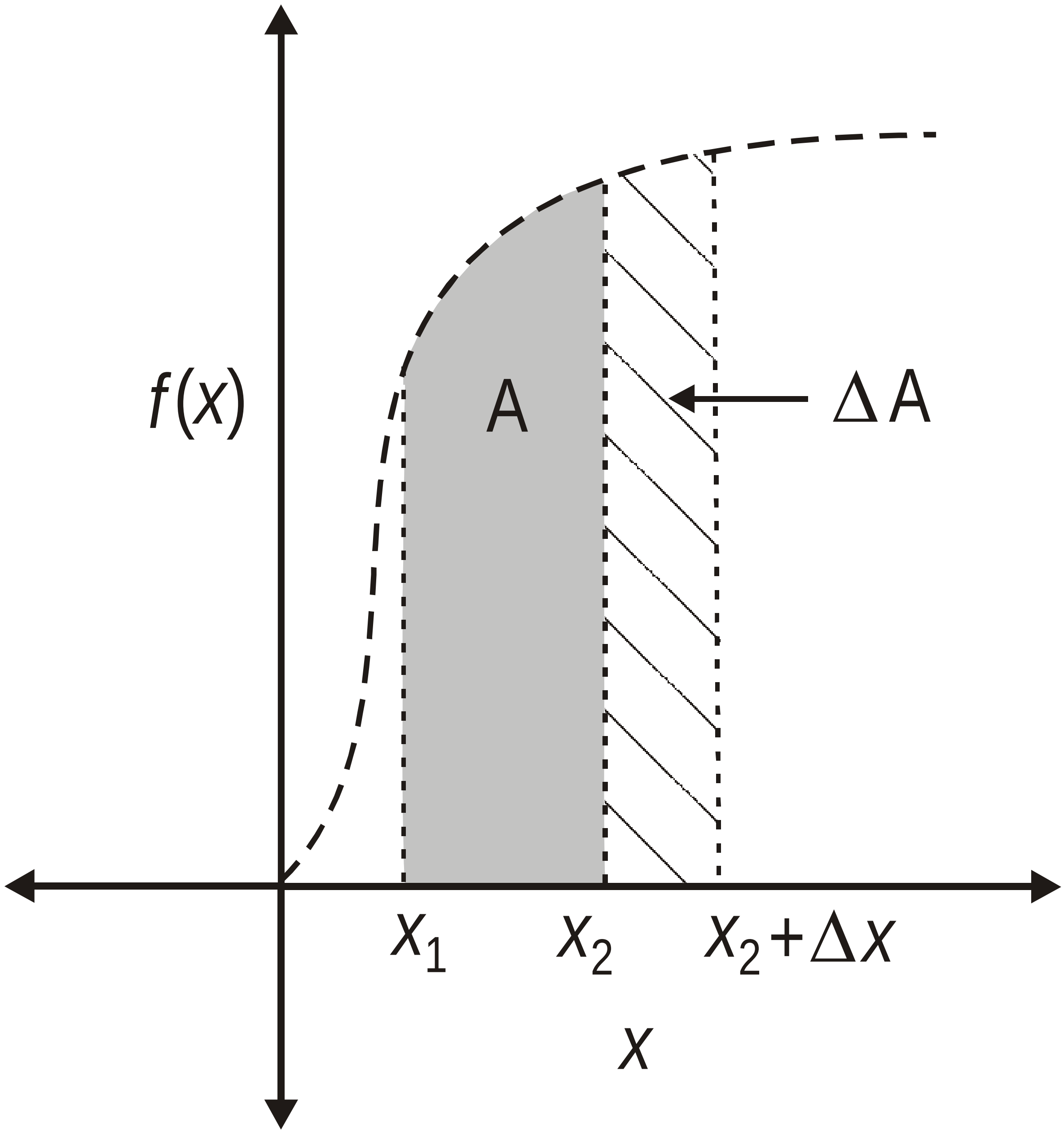

Figure S.7. The graphical representation of integration. The aim of integration is to find the area beneath the curve (A) defined by the interval between two values of the independent variable [i.e., bounded by $f(x_1)$ and $f(x_2)$]. This is accomplished by determining the function that when differentiated (i.e., as $\Delta x \rightarrow 0$) equals $f(x)$. The necessity to integrate within the limit is illustrated by designation of a change in area (ΔA), which approaches dA/dx as $\Delta x \rightarrow 0$.

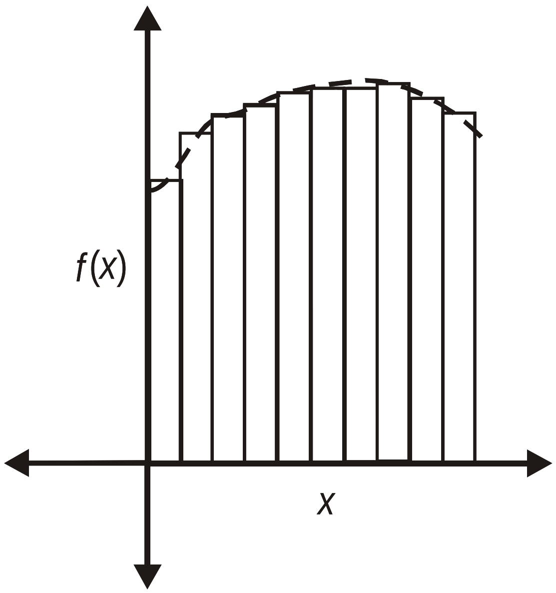

Figure S.8. An alternative perspective on integration. Integration can be thought of as the cumulative sum of areas for a series of rectangles arranged beneath the curve of interest. As the width (Δx) of the rectangles becomes smaller, the accuracy of the integration improves. The true area beneath the curve is determined within the limit as $\Delta x \rightarrow 0$.

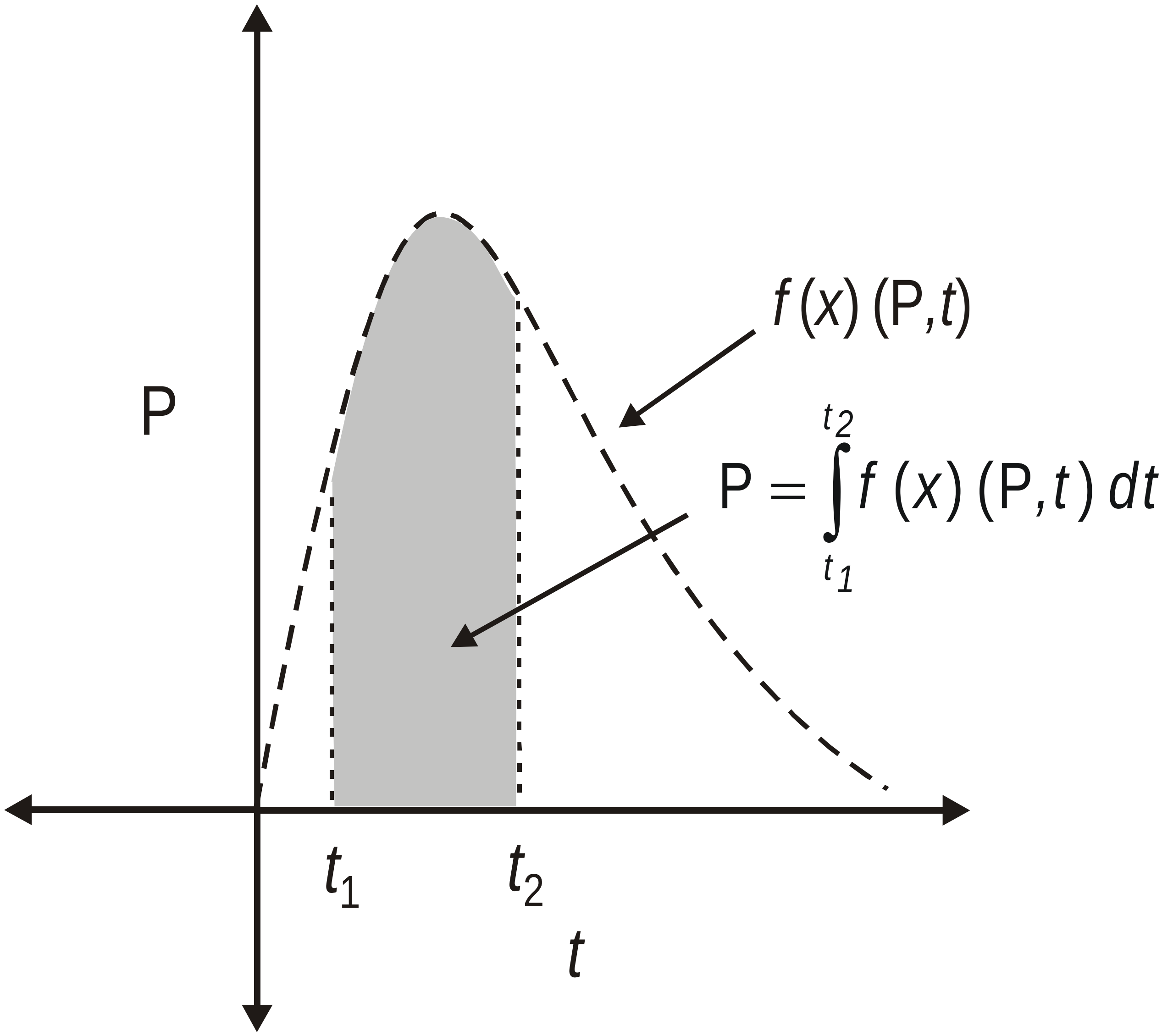

Figure S.9. Graphical representation of a probability density function [*f(x)(*P,*t)*] expressed in relation to time (*t*). The probability of a successful event at any point in time can be estimated by the function. The total probability of a successful event within the time interval $t_2 - t_1$ can be evaluated as the definite integral as shown.

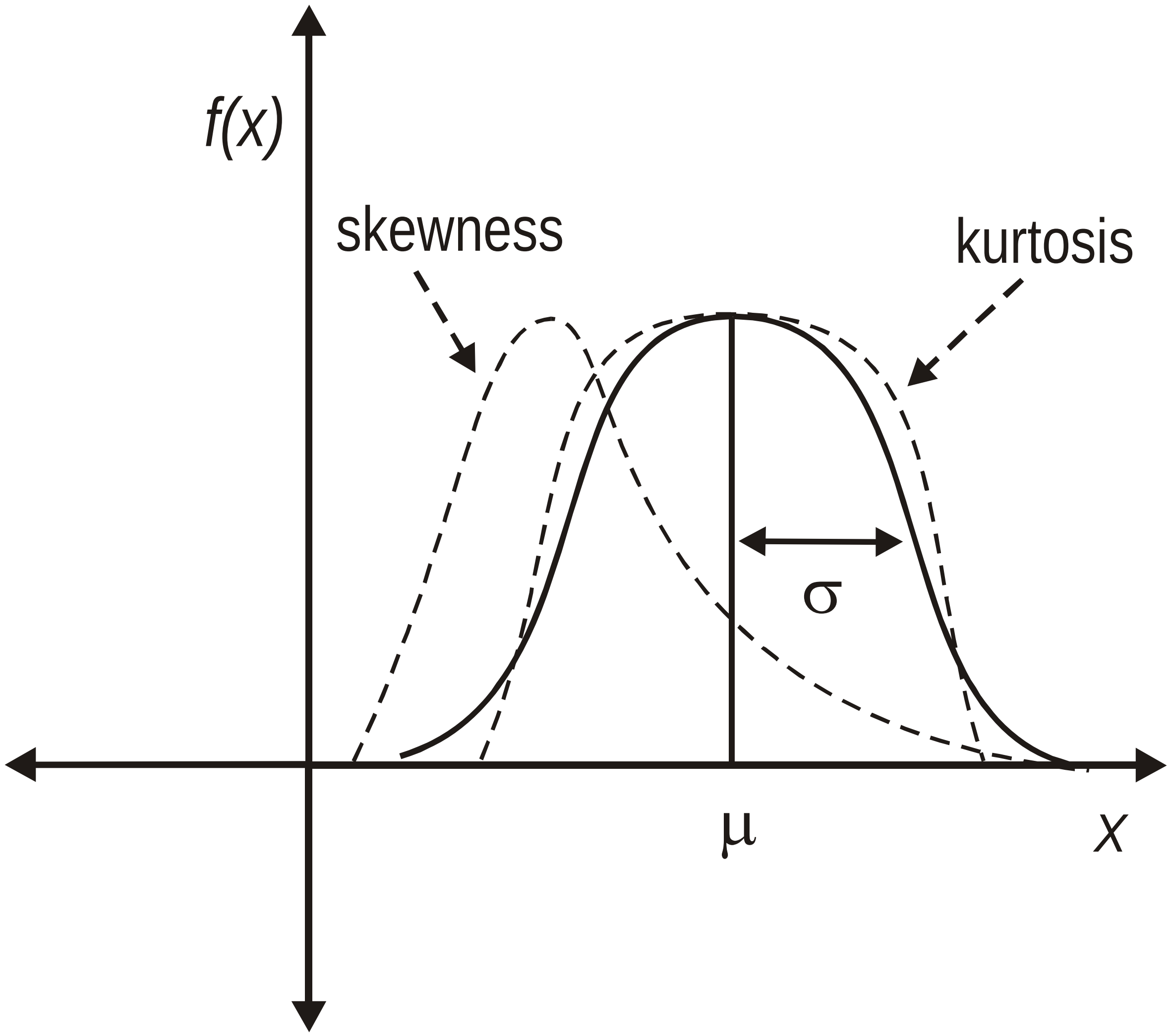

Figure S.10. A Gaussian probability distribution (shown as the solid curve) with the mean (μ) and standard deviation (σ) indicated. Distributions displaying significant kurtosis (platykurtosis) and negative skewness are also shown.

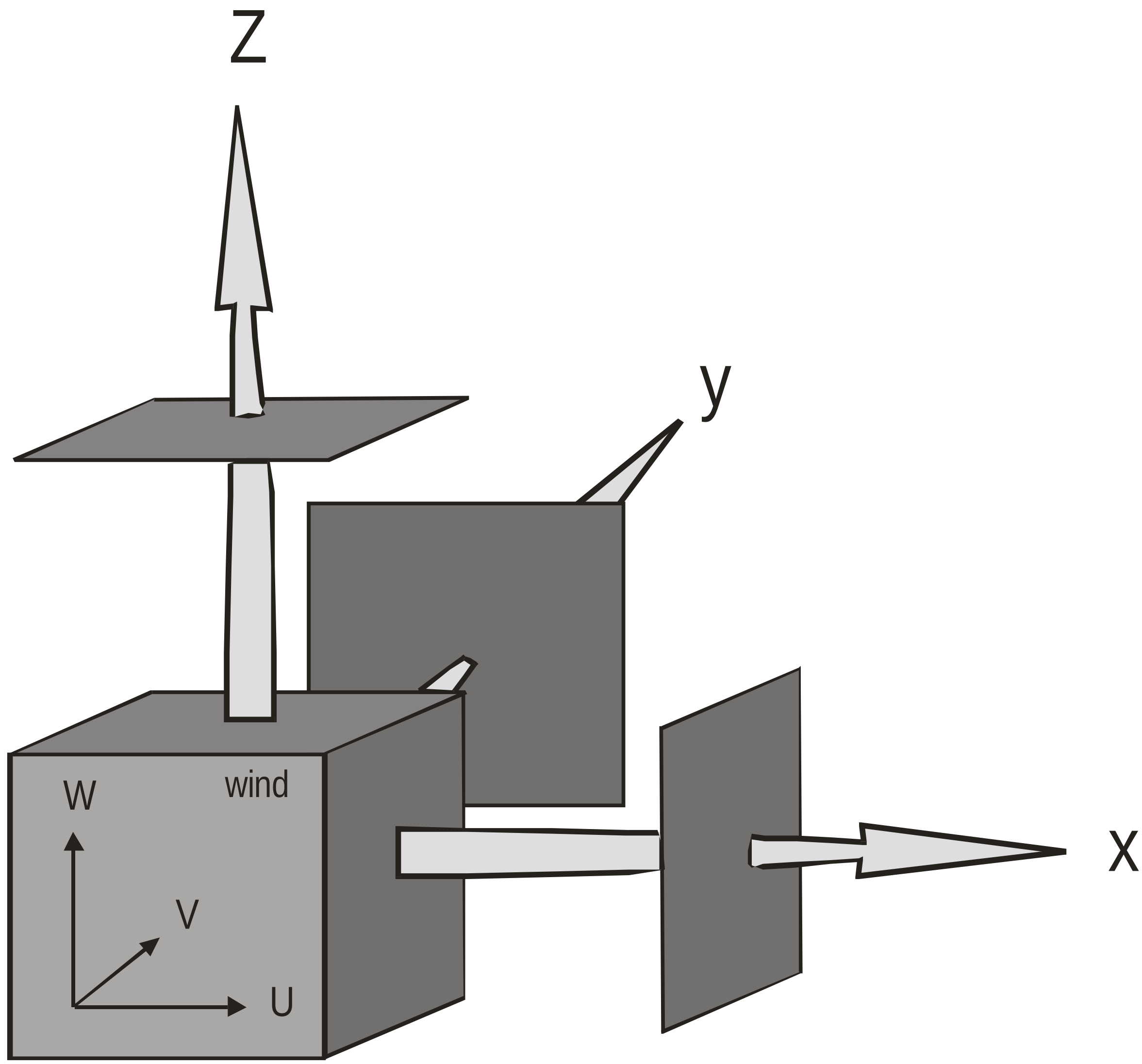

Figure S.11. Momentum flux occurs when wind, with its three Cartesian coordinates (*u*, *v* and *w*), is transferred in three possible Cartesian directions (*x*, *y* and *z*), yielding nine possible vector components. (Redrawn from Stull 1988).

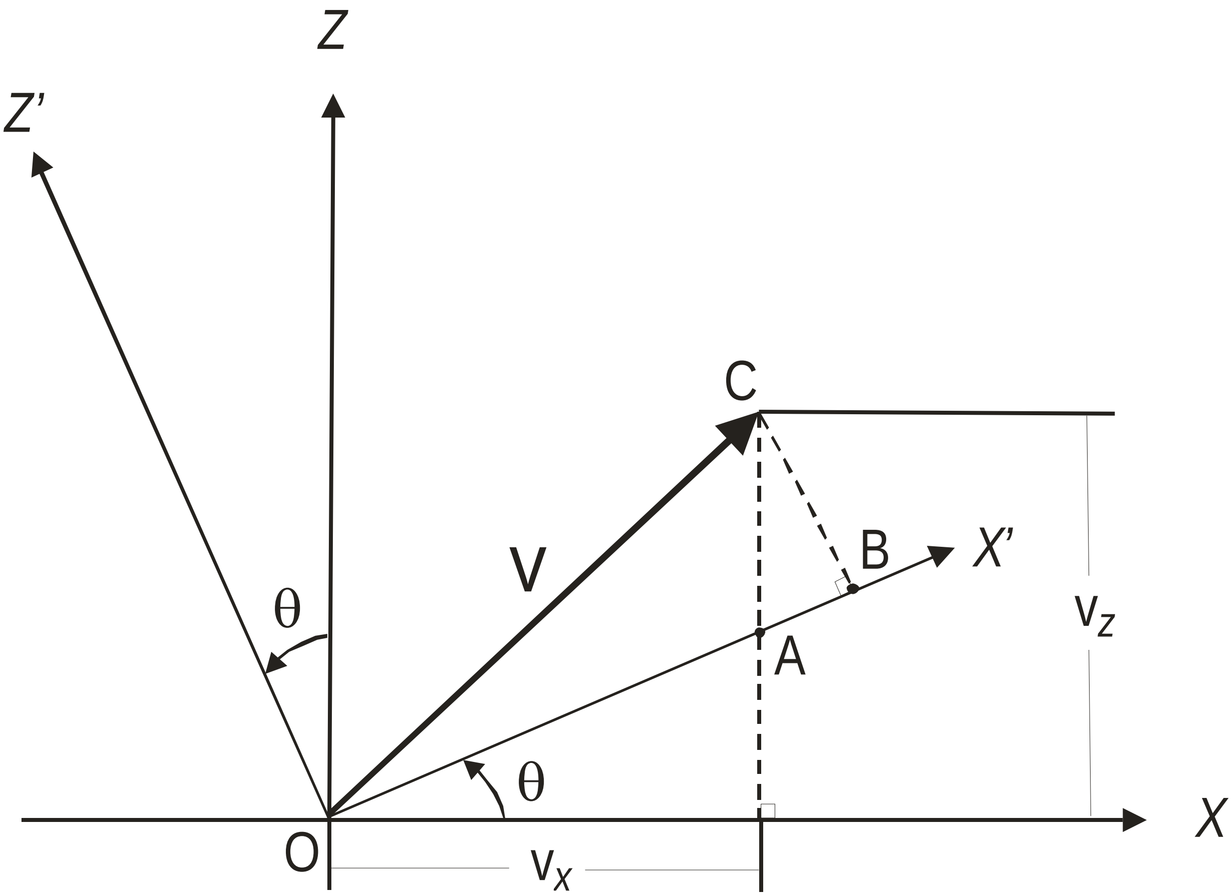

Figure S.12. Rotation of Cartesian axes around the y-axis. The original vector (v) is drawn relative to the original x,z axes. The rotation is intended to redefine v relative to the new, rotated x',z' axes. See text for more details. (Redrawn from Lea 2004).

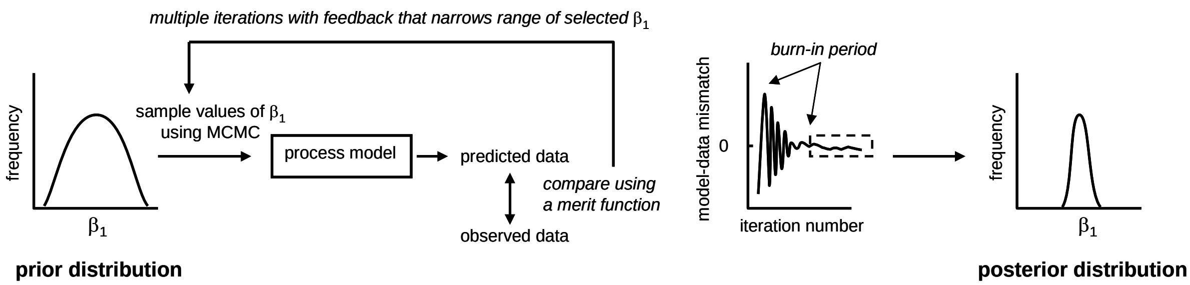

Figure S.13. Scheme showing steps taken in a Bayesian process whereby estimates of parameter values for a representative process model are optimized and uncertainties reduced in producing a posterior parameter distribution from a prior (estimated distribution.