esa

ECOSPHERE

Equitable transform applied to phenology and temperature in a changing climate: Scaling to maintain individuality

CASSANDRA ELPHINSTONE† AND GREG H. R. HENRY

Department of Geography, University of British Columbia, 1984 West Mall, Vancouver, British Columbia V6T 1Z2 Canada

Citation: Elphinstone, C., and G. H. R. Henry. 2020. Equitable transform applied to phenology and temperature in a changing climate: Scaling to maintain individuality. Ecosphere 11(2):e03036. 10.1002/ecs2.3036

Abstract. We describe an equitable transform that can be used to estimate missing data points, determine systematic patterns in data, observe baseline changes, and detect different amplitudes in replicated sequences. It is applicable to short discrete two- or three-dimensional data sets such as biological life cycles or water content in similar media. The technique is independent of the continuity or ordering of sequences but is distinct from methods such as principal component analysis. It is ideally used to preprocess noisy or incomplete data sets prior to analysis with other well-established techniques. This transformation maintains systematic differences between individual sequences when the underlying pattern is a separable function in two variables added to another function in one of these variables. The equitable transform partitions the original noisy data into the underlying signal determined from the data and its residuals. Points from one sequence can be transformed to any other sequence without losing any information. Information about one or more sequences can be used to infer others with missing data. A link to a github R package is provided so the transform can easily be run on any two-dimensional data set. Simulated two-dimensional data sets are used to demonstrate its utility in recovering missing data and scaling/offsetting in one of the dimensions. We used the transform to determine that winter temperatures at a High Arctic site have warmed by 1.8° ± 0.4°C/decade and summer temperatures by 1.1° ± 0.2°C/decade from 1986 to 2007. Applied to 18 yr of phenology data for the tundra plant *Dryas integrifolia* at the same site, we determined that the annual cycles of phenology events could be modeled accurately. Phenology, in some circumstances, can be described as offsetting and scaling the rate at which life cycle events occur. We introduce the idea of scaling and shifting the seasonal cycle of a reference plant via the equitable transform in order to approximate the behavior of multiple phenological cycles. Relative phenology rates of *Dryas integrifolia* were found to have increased over time indicating duration of phenological stages have become shorter in recent years, likely in response to the warming climate.

Key words: Arctic plant; averaging; Boltzmann transformation; climate change; equitable matrix; fractal; missing data; noise reduction; phenology; principal component analysis; scale.

Received 23 April 2019; accepted 29 April 2019; final version received 26 December 2019. Corresponding Editor: Debra P. C. Peters.

Copyright: © 2020 The Authors. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

† **E-mail:** cassandra.elphinstone@shaw.ca

INTRODUCTION

In ecology, knowing the spatial and temporal scales at which various processes occur is essential for understanding how ecosystems function (Wiens 1989). Patterns at different scales can be linked in complex ways, and finding methods to quantify the scaling of observations is an important aspect of ecology (Levin 1992). By making comparisons between multiple sequences that follow a similar pattern, each sequence can be defined by a scaling factor relative to the other sequences. For example, time sequences that cycle each year such as precipitation and temperature often have a repeating seasonal pattern that can be compared between

years through scaling. Alternatively, life cycle stages of an organism often follow a set pattern that repeats from one individual to another but the scale and timing of this pattern is often highly individualized (see Fig. 1). These set patterns may change character within different groups thereby altering the functional form of the life cycle trait development. Biologists would like to not only describe the average behavior of a generic organism but also the individual deviations from this behavior, ultimately linking the behavior of one group to another. In this paper, we describe a technique to characterize this individual behavior and suggest a means by which the linkage of different groups can be accomplished. This equitable transformation is equally applicable to data sets of abiotic phenomenon.

The comparison of sequences in a data set is often complicated by poor or irregular sampling and/or by noise superimposed on the large scale underlying pattern. With the assumption that the multiple sequences have an identical underlying signal, averaging is often employed to spot trends. The variance around the average is believed to reflect the extraneous noise in the system. However, any systematic variation between sequences can cause misinterpretation of results from averaging. Quantifying this systematic variation or individuality found in ecological data such as population dynamics or the life cycle of organisms is an important aspect of understanding the underlying patterns in these ecological systems (Chave 2013). The trait itself could be scaled and offset as in Fig. 1a, or, as in Fig. 1b, the times at which the events take place in the cycle of an individual could be scaled and offset. For example, events might be defined (as in Fig. 1b) by the times at which quantile levels are reached in a cycle. This might be the time the trait value attains its maximum value (quantile level of 100%) or the time at which the trait value has a quantile level of 25% after reaching a maximum (see Fig. 1b). These times may vary widely from individual to individual even though all the profiles are in an equitable sense equivalent to each other. An application of simple averaging of either the trait as a function of time (Fig. 1a) or the time at which each event (Fig. 1b) takes place for all individuals would both yield large systematic errors. We describe a method whereby these data sets can be identified and described even when concealed by large amounts of noise and missing information.

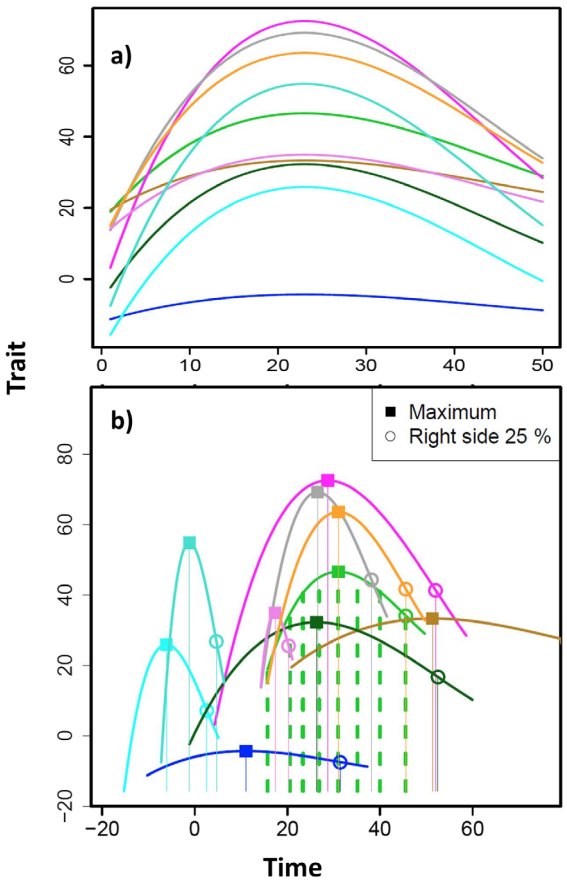


Fig. 1. Idealized individual patterns for life cycle attributes through a season: (a) change over a whole season where the attribute itself is scaled and offset (e.g., population changes, greenness index); (b) change within a season (e.g., flower phenology, germination, egg laying) where, instead, the timing of the life cycle events is scalable and offset relative to each other. Dashed vertical green lines represent "quantile level events (0, 25, 50, 75, 100%)" for the green curve. Squares show event maximum and circles, the event "25% quantile level on the right side of the curve." There is no preferential individual as all profiles are equivalent indicating that they are equitable with respect to each other. These patterns could also represent abiotic attributes such as temperature or soil moisture.

In standard data analysis, when two or more sequences are believed to be related by some underlying function, researchers often investigate

relationships between the sequences by fitting them to linear regressions, higher level polynomials or by finding common frequencies or wavelengths through autocorrelative functions and spectral analysis techniques such as wavelet analysis and the maximum entropy method or MEM (Ulrych and Bishop 1975). When both the correlated data sequences have associated errors, special techniques take this into account to avoid regression dilution biases (Berglund 2012) and the analyses can become specialized and complex.

A moving average (Kenney and Keeping 1954), unlike a typical average, does take into account systematic variations across the sequences in a data set. However, a moving average is only applicable to an ordered set of attributes (e.g., a chronological time sequence) and is generally used to smooth out variations over a narrow window of time and not throughout the entire data set. More sophisticated filtering techniques such as those involving autoregressive processes or spectral analysis (Brockwell and Davis 2002) can also be applied but these methods are often specialized for use in specific circumstances and are frequently limited by constraints of continuity and sample lengths.

Principal component analysis is also used extensively (Wold 1966, Andrecut 2009) and has the advantage of not generally being restricted by continuity and sequence length. In addition, it can separate the data into multiple orthogonal components based on a minimization of successive variances. Our technique is significantly different from this method although the two techniques resemble each other on a superficial level.

The method presented here, called an equitable transformation, takes into account systematic variations between the sequences that are not accounted for when averaging. Equitable in the context of this technique refers to the ability to convert (equitably or fairly) between one sequence and another and back again. Originally, the concept of this matrix was introduced by Parker (1965*a*, *b*) as a pure mathematical construct that allowed one to quickly determine whether a finite set of objects under a binary operation was a group or not. Parker (1965*b*) termed a matrix with these properties an equitable matrix and showed that such a matrix could be used in currency exchange interactions to equitably move currencies through foreign currency exchanges without loss. These matrices allow the conversion over many scales of measure while always being able return to the initial currency without loss or gain. The equitable nature of our transform allows this conversion process to take place and is what makes the technique so powerful. Elphinstone et al. (1986) later found a means to apply this equitable matrix concept to data analysis where they linked large-scale airglow structures in the upper atmosphere to planetary waves. Elphinstone et al. (1988) further developed this idea enabling a velocity spectrum to be constructed that proved useful for detecting the propagation characteristics of atmospheric gravity waves. These latter two papers dealt solely with equitable matrices and how to determine them from two-dimensional data sets. In the current paper, we show why their technique for determining the equitable matrix is successful, build on that success to construct a second equitable shift matrix, and finally determine an equitable transform that allows information to be equitably transferred between individuals that share the same underlying pattern. The equitable technique is applicable to a wide variety of data sets where the characteristics of individual data sequences play an essential role.

In the *Methods* section below, we first outline the equitable transform technique and what is meant by an equitable data set. We introduce relevant equations and provide a proof for why the convergence to an equitable matrix can occur. Non-technical readers can focus on the intuitive description provided and skip to the results that summarize the usefulness of the technique. An example of a small data set transformed is also presented in the Appendices to illustrate precisely the meaning of an equitable transformation. A link to an R package is given so that researchers can use the technique on their own two-dimensional data sets or explore prebuilt ones that we provide. In the *Results* section, we apply the equitable technique to simulated data sets in order to demonstrate its usefulness in coping with missing, discontinuous, and noisy data. We demonstrate how an equitable transformation can be applied to ecological data sets in order to better understand the patterns within them. Using this method, instances or individual

series in a data set can be scaled and offset relative to each other in a systematic way that preserves the integrity of the data while estimating missing data and reducing noise that usually interferes with interpretation.

We then use the technique to analyze temperature data and phenology events of *Dryas integrifolia* plants measured over an 18-yr period at a research site in the Canadian High Arctic. We show that individual phenology cycles can be thought of as scaled (stretched or compressed) and offset versions of each other or of some average cycle. Similarly, daily temperatures for a given year were stretched or compressed and offset versions of each other or some average or reference year. The *Discussion* section includes a comparison with principal component analysis and how this technique differs in philosophy. The implications of the phenology results are discussed for their relevance to climate change as well as for their implications for the study of organism life cycles. We also briefly introduce the concept of multiple-level equitable systems and their relation to fractal like systems.

Methods

Technique intuition

The equitable matrix technique can be applied to any data set containing multiple sequences measuring some attribute that are related via the stretching and shifting of an underlying function. The five sequences in the columns of Table 1 are the attribute values, I, associated with a set of events numbered in this case from $n = 0$ to 5. The function chosen as an example is $g(n) = n$. These so-called events represent different circumstances for the individuals involved, and the recorded values are the individual measurements. The columns, *A–E*, represent individuals whose trait (quantified by some particular attribute I) is recorded under six different circumstances (the rows $n = 0, 5$). These circumstances could be treatments, event times, events themselves, spatial locations, points in time, or other attributes. What is key here is that all the individuals are responding (via the attribute I) in a consistent way to the different events, and this underlying manner is expressed by some function of the events, $g(n)$. They do not, however, all respond identically. Individual A has attributes $I_A(n)$ which are equal to the generalized function $g(n)$. Individual D responds like individual A but its response is shifted by a constant value ($I_D(n) = g(n) + \text{shift}$). Individual *B* is twice as sensitive to the stimuli than is expressed by $g(n)$ but it is not shifted relative to individual *A* ($I_B(n) = 2\ g(n)$). Individual C is three times as responsive as is implied by $g(n)$ and is shifted to a larger value from $g(n)$ by 5 ($I_D(n) = 3\ g(n) + 5$). Individual E is a special case where the behavior of its attribute (I_E) is unaffected by the different events resulting in a flat line.

In this example, because the attribute of individual *A* was identical to $g(n)$, $I_A(n)$ and $g(n)$ are interchangeable and one could think of all of the other individuals responding to the events relative to the attribute of individual A instead of relative to the function $g(n)$. In fact, it seems logical that any individual might be used as a reference for comparison to others. We wondered whether a system of individuals which were not perfectly related but all followed the same general underlying function could be described in a rigorous

Table 1. An equitable data set (arbitrary units): columns *A–E* represent the individuals or replicates and rows 0–5 represent the events, different circumstances under which the same attribute, I, is measured.

Event number (n)	Individual *A*	Individual *B*	Individual *C*	Individual *D*	Individual *E*
$g(n) = n$	$I = 1 \cdot g(n) + 0$	$I = 2 \cdot g(n) + 0$	$I = 3 \cdot g(n) + 5$	$I = 1 \cdot g(n) + 2$	$I = 0 \cdot g(n) + 3$
$n = 0$	0	0	5	2	3
$n = 1$	1	2	8	3	3
$n = 2$	2	4	11	4	3
$n = 3$	3	6	14	5	3
$n = 4$	4	8	17	6	3
$n = 5$	5	10	20	7	3

Note: The general response for each individual with respect to the function g(n) is given in the second row.

manner without defining a single individual as an absolute reference. The result was what we refer to as an equitable transform where one individual does not have preference over another.

If all of the individuals acted nearly identically to individual A, a simple average of the attribute, I, at each event, $n = 0, 1, 2\ldots$, would adequately describe all of the individuals' behavior under each circumstance. It is clear for the example in Table 1, however, that simple averaging is not always appropriate and could lead to a poor description of a system that is really quite well-defined. Note that when an individual does not respond to the varying conditions (rows or events), this represents a special case that shall be referred to as nodes and will be excluded from our equitable analysis. The nodes have a simple description on their own since measurements of the attribute are invariant with respect to the changing conditions.

A typical data set of this nature could be many sine waves stretched and shifted relative to each other. The sine wave describes a quantitative measurement (such as temperature) discretely sampled over a set of different events (such as time). The sine wave structure (Fig. 2a) is scaled and shifted for the different individual sequences. For example, the quantitative measure along the y-axis (shown in Fig. 2a) could represent temperature. The three different sequences could represent three different locations where temperatures were measured at the times A–J on the x-axis. The attributes along the x-axis are typically either time or space coordinates although they are by no means restricted to them. These three sequences (Fig. 2a) have a linear relationship with the arbitrarily chosen reference sequence (Fig. 2b).

If nothing is known concerning the sequences, data like these might be averaged over the sequences to determine whether there is a trend in the measurements over the attributes. In the space/time/temperature example, this would involve averaging over spatial location to infer temporal changes in temperature. Due to the systematic changes (shifting and stretching) occurring in the sequences, averaging across the spatial locations would mask variations of the sinusoidal signal in each sequence. The equitable transform technique is a systematic method that accounts for the individual character of each

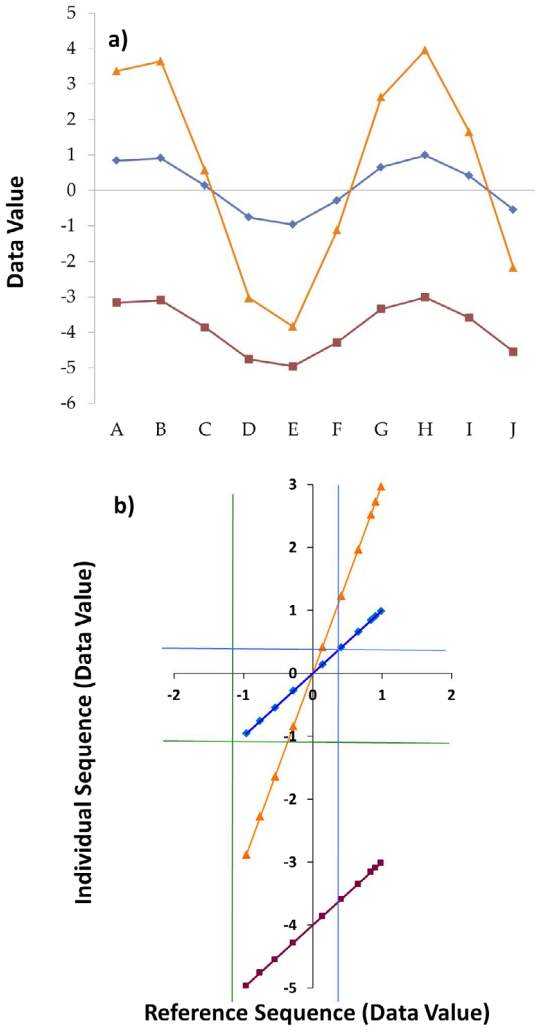


Fig. 2. (a) Three sequences (yellow, blue, red) of some quantitative measure (such as temperature at three spatial locations) as a function of arbitrary attributes A-J (such as times the temperature was measured at the three locations). (b) The three sequences from (a) as a function of a reference sequence, chosen in this case to be the blue sequence in (a). By shifting the origin (0, 0) along the blue reference line ($y = x$), the new y-intercepts of the lines of the other locations yield the quantitative variation of the attribute when the reference sequence has the value of the new zero (new reference frames with different origins shown in green and blue).

sequence. The data are equitable because they all have the same underlying function.

Data sets of this equitable nature (Figs. 1, 2 as well as Table 1) correspond to a discrete

sampling of a function of two dimensions, for example, space (x) and time (t), that can be separated into the product of two functions (a spatial function $f(x)$ and a time function $g(t)$) combined with a time-independent function $u(x)$. Knowing one reference sequence in a perfectly equitable data set is sufficient to deduce all other sequences, if it is known how the other sequences are shifted and scaled relative to the reference. This is similar to how the Boltzmann transformation is used in soil science to transform various soil water content distributions to a reference soil distribution (Nielsen et al. 1997). We cannot emphasize enough that although we use the variables x and t frequently in this paper, they represent arbitrary variables and are in no way limited to the conventional associations of spatial and temporal coordinates.

In most natural systems, the observed sequences are not perfectly equitable (as they are in Figs. 2, 3) and variations from this perfectly equitable system are inevitable. Temperature, for example, is equitable on large spatial and temporal scales (e.g., at high latitudes it is generally warmer in summer than in winter) while on smaller scales, temperature becomes much less equitable (e.g., temperatures at a specific microsite location cannot be determined accurately by knowing the specific day of the year). The technique presented here yields a parameter that describes the degree to which a data set has this equitable form and includes errors associated with the transformed data. The technique can also be used to determine the scale at which a system becomes equitable.

The equitable matrix technique yields several concrete results. First, it establishes whether or not sequences within a data set are equitable; that is, do the sequences have the same underlying function and differ only by a shifting or stretching of this function? Second, after determining that the raw data collected are reasonably equitable, an equitable transform can be constructed that better represents the underlying function. Raw data are rarely perfectly equitable as a result of the noise and error that are found in any measured system. Non-equitable data points in the sequences are assumed to be a result of noise or error and will be averaged out leaving a much clearer equitable function (a function separable in two variables). Extreme outliers may or may not be removed depending on whether or not the outliers are equitable with respect to the equivalent points in the other data sequences. Thirdly, this transformation process allows missing data to be estimated. Finally, the equitable slope and intercept matrices define the general function underlying all of the data sequences that allow any sequence to be transformed into any other. These matrices determine the amount of shifting (intercept) or stretching (slope) required of a reference sequence to reproduce any other replicate sequence. This concept of relating equitable matrices to data analysis was first described by Elphinstone et al. (1986) while investigating high latitude airglow observations. The following outline of the equitable slope matrix and its relation to a least squares linear regression matrix closely follows their formulation. We show why their method for determining an equitable slope matrix works and then we use this matrix as a starting point for developing a novel equitable transform method that involves a second matrix termed a shift matrix. This matrix also has special properties.

Fundamental equations

In what follows, a data set $L(t,x)$ is assumed to consist of M observations in variable t at N values of another variable x resulting in a data set of $M \times N$ observations. The two variables t and x can be thought of as time and space coordinates, respectively, if the reader finds it more intuitive to understand the formulation this way. However, these variables could equally be any other parameters such as a set of different species, events in a life history, experimental treatments, or types of soil. No assumptions are necessary regarding the ordering or distribution of the N coordinate values of x. When interpreting them as spatial coordinates, any points distributed in 3D space could be used. Each of the M observations at a given x value should be evaluated at the same t coordinate as for all the observations at other x coordinates (coincidental in time if time is the variable in question or coincidental in event if event is the coordinate label for t). These also do not need to be ordered in any way (i.e., if interpreted as a temporal variable they do not have to be in chronological order). There may also be points on the $M \times N$ grid where no data exist.

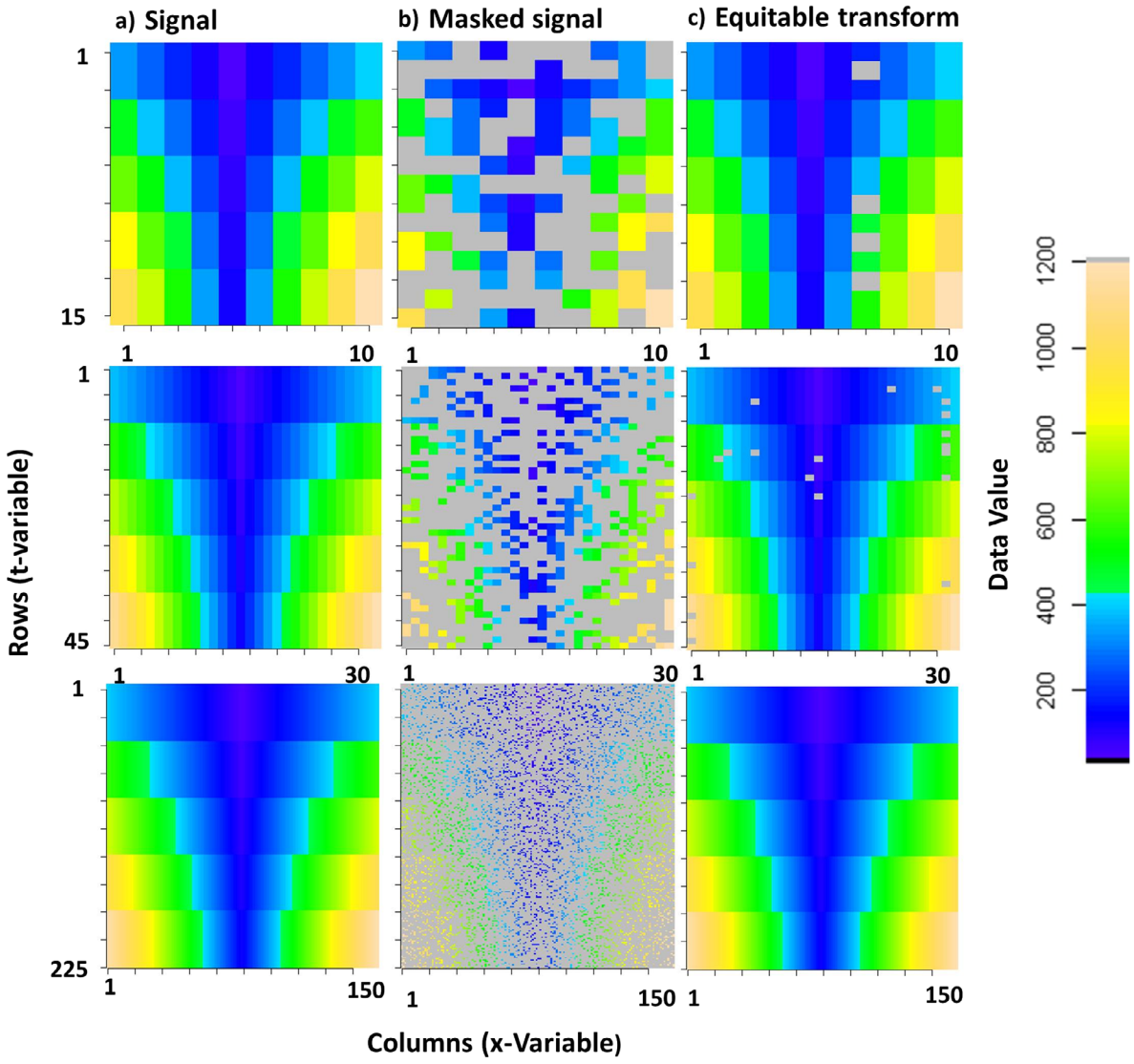


Fig. 3. A false color data set consisting of a step function in t and an absolute value of a linear function in x. The columns show full sampling (a), missing samples (b), and the recovered data from an equitable transform (c). Rows show the data sets with increasing resolutions: The top row has 15 samples in the variable t taken at equal intervals and 10 samples in the x dimension (for each t); the resolution in the middle row is increased by a factor of 3 (45 time samples $\times$ 30 spatial samples) and the bottom row by a factor of 15 (225 time samples $\times$ 150 spatial samples). The fraction of randomly missing data increased from 47% in the top row to 70% in the middle row and 85% in the bottom row.

When the data set of the attribute L has a perfectly equitable form, the N sequences of M points will correlate perfectly with the reference sequence and any sequence can be chosen as this reference sequence. We define a perfectly equitable data set, measuring some quantity L, as a product of a function $f(x)$ and a function $g(t)$ plus a second function of x, $u(x)$. That is,

$$L(t,x) = g(t)f(x) + u(x). \quad (1)$$

There are assumed to be N samples in variable x. M samples in variable t exist for each value of

x. Eq. 1 can be written in vector/matrix form by letting g be a row vector (length M) and f, u being column vectors (length N). Then, $L = g^*f + (u'^*\underline{1})'$ where $\underline{1}$ is a row vector (length M) consisting of 1s. The symbol * represents the Cayley product (Eves 1980), and the symbol $'$ represents transpose.

To put Eq. 1 in the context of our example data set (Table 1), the variable t is associated with the six events 0 to 5 sampled for all of the five individuals labeled from $x = A$–E. That is, there are five individual sequences, each with observed values for six events. The connection to equitable matrices is made when two specific individual sequences, say individual values at z and y, are correlated together. The two specific sequences (at locations/individuals z and y) can then be written as:

$$L_z(t) = f_z g(t) + u_z \quad \text{and} \quad L_y(t) = f_y g(t) + u_y, \tag{2}$$

where the factors f_z and u_z represent a stretching and shifting of a generic function, $g(t)$, at the specific location/individual z. The generic function, $g(t)$, represents the underlying function observed in all of the N sequences in the data set. By stretching (via f_y) and shifting (via u_y) $g(t)$, all of the sequences in the data set can be determined. The differences between sequences in Fig. 2a can then be understood as stretching due to the function f and shifting due to the function u at/for various locations/individuals.

We can make the linear connection between sequences z and y explicit (as in Fig. 2b), by solving the first part of Eq. 2 for $g(t)$ and substituting it into the latter part of (2). This gives

$$L_y(t) = a_{yz} L_z(t) + b_{yz}. \tag{3}$$

Eq. 3 exhibits the linear relationship between the sequence at location/individual y and the sequence at reference location/individual z (any location/individual can be used as the reference). The terms a_{yz} and b_{yz} are the independent slope and intercept, respectively, and are independent of the variable t. They relate the sequences z and y to each other. The slope and intercept are related to the original functions ($f(x)$ and $u(x)$ at specific locations/individuals $x = z$ and y) by

$$a_{yz} = f_y/f_z, \quad \text{and} \quad b_{yz} = u_y - a_{yz} u_z . f_z \neq 0. \tag{4}$$

Correlating all possible pairs of sequences in a perfectly equitable data set (described by Eq. 2) will result in perfect correlations, with the slopes and intercepts described by Eq. 4. If there are N sequences, an $N \times N$ matrix of the slopes (a_{yz}) and another $N \times N$ matrix of the intercepts (b_{yz}) (b_{yz}) will result. These slope (stretching) and intercept (shifting) matrices allow any reference sequence in the data set to be related to any other sequence. The values in the diagonal of the slope matrix are all one and the values in the diagonal of the intercept matrix are all zeros because the diagonals are the slopes and intercepts of the reference sequences compared with themselves (resulting in a slope = 1 and intercept = 0).

The matrix of slopes has particular properties that can be derived from Eq. 4. That is,

$$a_{zz} = 1, a_{zy} = 1/a_{yz}, \text{ and } a_{xy} = a_{xz} a_{zy} \text{ for all indices } x, y \text{ and } z. \tag{5}$$

Matrices with these properties have a rank of 1 and were first used as a method to quickly determine whether or not a set of objects were a mathematical group under some binary operation (Parker 1965*b*, Eves 1980). As well, the matrix of intercepts or the equitable shift matrix also has special properties. These include the following:

$$b_{yx} = b_{yz} - a_{yx} b_{xz}, a_{yx} b_{xz} = -a_{yz} b_{zx} \quad \text{and} \quad a_{yx} b_{xy} + b_{yx} = 0, \tag{6}$$

for all indices (x, y and z) in the matrix. The properties of this intercept matrix use addition and multiplication to relate the equitable slope matrix (a_{yx}) elements with the intercept elements (b_{yx}). These shift matrices normally have rank 2 and depend both on the equitable slope matrix and the values of $u(x)$.

Eqs. 3–6 describe relationships that any discrete data set with the equitable form given in Eq. 1 must have. The perfectly equitable sequences at the N locations/individuals, when correlated together, will produce a matrix of slopes and intercepts/shifts given by Eq. 4 with the slope properties given by Eq. 5 and the intercept properties given by Eq. 6. In reality, however, sequences are rarely perfectly equitable and usually do not result in perfect linear fits having these slopes and intercepts. A data set consisting of N sequences can be investigated to see

whether it follows the equitable form of Eq. 2 by calculating linear regression fits between all possible pairs of unique sequences. If two sequences z and y are perfectly equitable, exactly of the form in Eq. 1, the correlation between the sequences will result in a correlation of determination $r_{zy}^2 = 1$. As the two sequences become less equitable, meaning they do not exactly follow the form described by Eq. 1, their r_{zy}^2 decreases. Any r_{zy}^2 value less than one gives an estimate of the portion of noise that could be associated with relating sequence z with sequence y. An average (R^2) of all N^2 coefficients of determination (r_{zy}^2) provides an estimate of the equitability of the data set; that is, how closely the system fits the form of Eq. 1 and if there is the same underlying function $g(t)$ in all of the sequences in the data set (Elphinstone et al. 1986).

If two time series, x and y, are perfectly linearly related, the product of the associated slopes, a_{xy} (x vs. y) and a_{yx}, (y vs. x), is equal to one. If the x and y series are less than perfectly correlated, the product of the least squares regression slopes is equal to the square of the correlation coefficient, r^2. The method of conversion from least squares matrices of slopes and intercepts to their equitable equivalent is described in detail in Appendix S1. The slope matrix (A) is determined in a manner similar to that of Elphinstone et al. (1986) whereby the ratios of least squares slopes, a_{yz}'/a_{xz}', relative to the reference location z will approximate the equitable slopes associated with a true signal slope of a_{yx}. In Appendix S1, we demonstrate for the first time that this ratio approximates the signal slope due to a cancellation of the variances associated with the reference location z. A new method for determining the shift or intercept equitable matrix (B) from the least squares matrices is given, and the detailed weighted method for creating the equitable transform is outlined. Without this shift matrix having the properties in Eq. 6, any resulting transformation process would not be practical as the intercepts involved with the different sequences would depend inappropriately on the slope values. With both of these special matrices determined, the equitable transform, T, can be defined that converts values of the data, $I(z, t)$, for locations/individuals z to values for x. That is,

$$I(t, x) \sim T[I(t, z)] = 1/N(A * I' + B * [1]'), \quad (7)$$

where [1] is a matrix consisting of 1′s with the dimensions of I ($M \times N$). In practice, weighting factors (see Appendix S1 for details) help ensure sequences more related to each other are given enhanced effect in the calculation. This equitable transform is somewhat similar to averaging the data. If the equitable slope matrix, A, was all 1s and the shift matrix, B, was all 0s, then the equitable transform of the data would be a simple average over the N locations. However, if systematic changes occur, then the unique aspects of each sequence (at location/individual x) consistent with the function $f_x * g(t) + u_x$ in Eq. 1 are retained. The linear fits in an equitable transform allow an average to be given at every point in the data set rather than yielding a single point from averaging over some selected range. This allows the individual character of each sequence to be maintained while incorporating its relationship to all the other sequences. A by-product of this analysis is that if there were no direct observations at a particular point in variables t and x, the equitable transform gives an estimate (with error) of the value based on the sequences at the other locations/individuals. The unusual character of the shift matrix allows one to describe the spatial/individual variations of the data set for some t_c relative to the reference spatial/individual location (z) simply by shifting the origin of the data set to that reference x and t coordinates (Fig. 2; Appendix S1). A least squares transform, T_{ls}, can also be defined in the same manner as an equitable one (via Eq. 7) simply by replacing the equitable matrices with the least squares equivalents.

In order to illustrate how the equitable slope and intercept matrices are calculated, we give an example of this transform applied to a small simply defined (5×4) data set in Appendix S1.

The equitable transform allows the individuality of sequences to be retained and lets sequences be equitably interconverted between each other. The transform, $T[L(\underline{t}, z)]$, takes sequence values at one or more fixed points, $\underline{z}$, and converts them equitably to all other x variable points. It is also possible to let $\underline{z}$ denote a fictitious location/individual associated with a temporal profile formed by averaging over all z,

and the transform could be evaluated including this extra location/individual. Knowing one value (perhaps an average temperature at other locations/individuals) allows the rest of the sequences to be determined based on the sequence values associated with $\underline{z}$. If an average sequence profile is used, then such a point must be included in the calculation of the equitable matrices. If the average trait (or trait at some location/individual) has a zero or small variation over the sequence interval of M values, then these locations/individuals become nodes and should generally not be used.

Code availability

The equitable transform package is given on GitHub at: https://github.com/celphin/Equitable

Also, a pdf manual of available functions and data files containing temperatures and Dryas phenology are to be found in https://github.com/celphin/Equitable-Project.

In R, type:

```
install.packages("devtools")
library(devtools)
install_github("celphin/Equitable") #to
install the package
library(Equitable) # to load the package
??Equitable #for more information on the
functions
eg4(1,1) # to see one example dataset
```

Multiple examples are given in the manual and help to show how the software can be used and the effectiveness of the transform under diverse conditions. Functions for creating the synthetic data sets are provided, including those used in the examples below. After creating their own real (or synthetic) data sets containing both noise and missing information at various resolutions, researchers can explore the data sets using the plotting functions in this package.

Results

Analysis techniques for two-dimensional data sets often cannot deal effectively with short sequences, data with missing values, or data that do not implicitly involve continuous functions. To illustrate the effectiveness of the equitable transform, we first simulated three different two-dimensional data sets that discretely sample simple separable functions (of the form in Eq. 1) at different spatial/temporal resolutions with varying degrees of missing, noisy, and discontinuous data. We then use the equitable transform to analyze temperature and plant phenology data from a long-term Arctic research site.

Example 1: dealing with missing data

In this first example, we chose a temporal function, $g(t)$, to be a step function with five levels over the full temporal interval beginning at 2 and incrementing by 1 for each level. A spatial function, $f(x)$, is chosen to be an absolute value function minimizing at 20 near the center and increasing linearly with x on each side. Three discrete sampling resolutions were chosen resulting in three-two-dimensional data sets with 15 × 10, 45 × 30, and 225 × 150 points. False color representations of these full data sets are shown in the left panels of Fig. 3.

Sometimes, problems arise when sampling and some data go missing or are not collected. This problem is simulated here by randomly removing some fraction of the discretely sampled equitable data. Missing data can be problematic for many standard data analysis techniques. However, since the equitable transform does not rely on continuity or equal intervals between samples, it can be used to recover missing information from the attenuated data. The R^2 value for a case with missing data will still be 1 since the data set is separable and all possible pairs of time series are perfectly correlated. The equitable transform can completely recover the original pattern until the fraction of missing data reaches a critical threshold. If too many data are missing, some values in the transformed data set will not be able to be estimated (e.g., blank spaces the right panels of Fig. 3). For data sets with higher resolution, a larger fraction of missing data can be recovered. In the low-resolution data set (top row of panels in Fig. 3), 47% of the data are missing and the equitable transform accurately recovers all but four points of the original data (top right panel of Fig. 3). At three times higher resolution (45 × 30 points), 70% of the data can be removed and almost the entire original data set is recovered (middle row of Fig. 3). At 15 times the sampling resolution (225 × 150 points), 85% of the original data can be missing and the entire data set is recoverable (bottom row of Fig. 3). The equitable transform can recover the missing

data even though the original data sets were discrete, discontinuous, and sampled poorly. This recovery of the data using the equitable transform would now allow a more detailed analysis using standard techniques to proceed.

Example 2: dealing with noisy data

An underlying signal can be unrecognizable in the presence of too much noise. Often, however, an equitable transform can (1) demonstrate that there is an underlying signal and (2) recover the signal with relatively little bias. Appendix S2: Fig. S1 gives one example of such a signal embedded in noise, and Appendix S2: Fig. S2 illustrates the bias inherent in a least squares transform vs. the equitable one. The least squares transform generally overestimates low values of the signal and underestimates high values of the signal. This bias is related to the product of the least squares slopes being the square of the correlation coefficient ($a_{xy}' \cdot a_{yx}' = r_{yx}^2$) instead of 1, thereby compressing the transformed data toward its mean value. It is a result of error in the independent variable of the linear least squares regression fit.

Example 3: Dealing with both noise and missing values

A third example incorporates both missing data points and adding normally distributed noise onto a separable signal. A separable signal (left panel Fig. 4) was created with a resolution of 50 × 75. Then, noise was added to this signal and 50% of the data removed creating a noisy incomplete data set, $I(x, t)$ (middle panel of Fig. 4). The equitable transform of this noisy incomplete data set was determined, and all missing data were replaced with transform values (right panel of Fig. 4). It took 3 iterations to change the least squares regression matrix of slopes (with an $R^2 = 0.40$ and $1 - P = 0.77$) into an equitable matrix (agreeing with the properties in Eq. 5 corresponding to an $R^2 = 1.00001$). Note that no information concerning the signal was assumed.

For each point in the equitable transform, $T[I(t, x)]$, the standard deviation of the transformed point can be determined. If zero, then the value at that point is accurately determined by its relationship to the other time sequences. In this case, the pattern of the equitable transform accurately

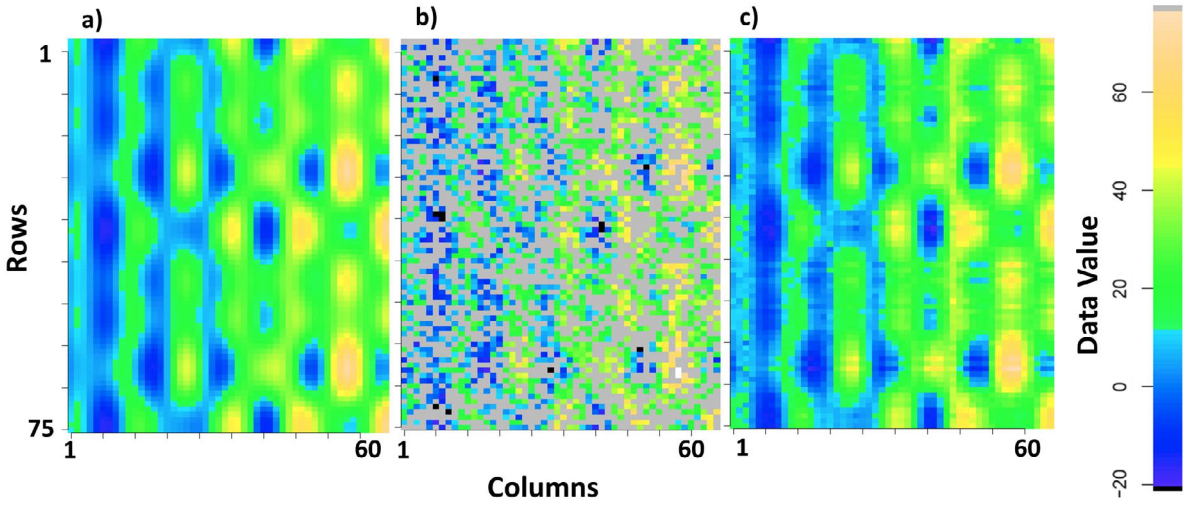


Fig. 4. False color images of (a) the signal, (b) the signal data + noise ($\sigma_G = 0.5\sigma_S$) without 50% of the data to be analyzed (middle), and (c) the equitable transform of the noisy incomplete data (right). No information about the signal was used to infer the equitable transform in (c). The signal function (L, the left panel) was defined using $f(x) = \sqrt{x} \cdot \sin(2\pi(x + 20)/90)$, $g(t) = \sin(2\pi(t + 30)/180) + \sin(2\pi(t + 30)/90)$, and $u(x) = 0.1x{+}10 \sin(2\pi x/60)$. Both x and t had 360 units that were sampled at regular intervals with 50 spatial locations and 75 time points. Noise with a standard deviation, σ_G, of one-half of the original signal standard deviation (σ_L) was added to the system. In the middle panel, 50% of the data were randomly was replaced with n/a values.

reflects the original separable signal and the standard deviations of the transform data reflect the noise resulting from $I(t,x)$ deviating significantly from this signal (Fig. 5). Without approximating the missing information, it would be very difficult to interpret some of the noisy incomplete original data (the black squares in Fig. 5a,b) as corresponding to the separable signal (the black lines in Fig. 5c,d). The equitable transform (triangles in Fig. 5c,d) reproduces the signal very well. The complexity of the signal would make relationships between time sequences or spatial patterns difficult to determine by other methods. For example, there is no clear connection between the spatial (column) variation at row 60 (black squares in Fig. 5a) and the spatial pattern at row 75 (black squares in Fig. 5b) but the two are related (in the absence of noise) via Eq. 1. Note that these two sequences are not linearly related since they are the spatial variations and

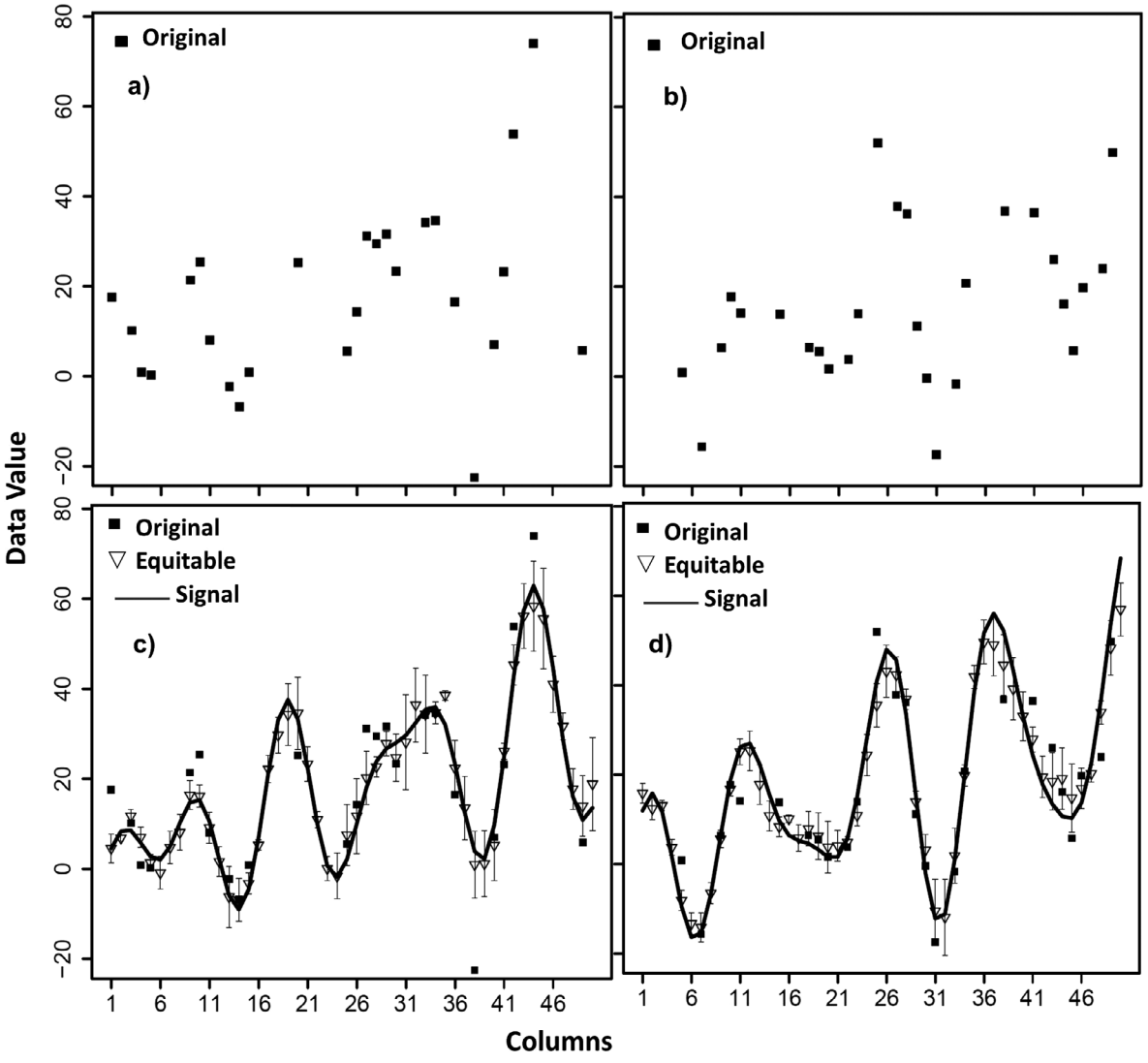


Fig. 5. Horizontal cross sections of Fig. 4 showing column (x) variation at row = 60 (a, c) and 75 (b, d). Only the noisy/incomplete data (solid squares) are shown in panels (a) and (b). Panels (c) and (d) include the signal (solid line), the noisy/incomplete data (solid squares), and the equitable transform values (triangles with standard deviation error bars).

described by $f(x)x007E;\cdot g_t + u(x)$ for each specific time (t). Similar results were found for the row (temporal) sequences at all column locations; these however, are linearly related.

Comparison of the equitable transform of a noisy signal to the separable signal

In the simulation example shown in Appendix S2, we found that the standard deviation of the difference between the original (signal (L) + noise (G)) and the equitable transform, T ($L + G$), was close to the standard deviation of the noise by itself $\sigma(T_{L+G} - L) \sim \sigma_G$. This indicated that the transformation process might be partitioning the original data ($I = L + G$) into the separable signal (L) and the original noise (G). In this section, we explore this possibility and show a more precise description of this partitioning. Those readers not interested in these details can skip to the long-term temperature variations section where we analyze High Arctic temperatures using this equitable transform.

If a data set is finite with N time sequences each containing M points, the superimposed noise, $G(t,x)$, has mean values over both the t and x variables, $G(x)_{\underline{t}}$ and $G(t)_{\underline{x}}$ (where the variable underlined is the variable over which the averaging takes place) whose σ values are approximately $\sigma_G/\sqrt{M}$ and $\sigma_G/\sqrt{N}$, respectively. The set of $M \times N$ points containing only the averaged noise, $H(x, t) = G(t)_{\underline{x}} + G(x)_{\underline{t}}$, takes the form of Eq. 1 and so matches the definition of equitable (with $f(x) = 1$). It has a standard deviation of $\sigma_H = \sigma_G \cdot \sqrt{(1/M + 1/N)}$. When the noise is superimposed on the signal, $L(t, x)$, the variable t-independent, x-dependent noise term $G_{\underline{t}}(x)$, cannot be distinguished from the $u(x)$ term in Eq. 1 and so contributes as part of the separable signal. The connection of $G_{\underline{t}}(x)$ to the separable form is less clear but the equitable transformation of $I(t, x) = L(t,x) + G(t,x)$ will combine slopes with the time-dependent x-average of the noise $G_{\underline{x}}(t)$. Thus, the residuals (RE) between the equitable transformation $T[L + G]$ and the original signal, L, should include a term that resembles H (the t-variable and x-variable noise averages). For the second simulation shown in Appendix S2: Fig. S1, using $N = 150$ and $M = 225$, σ_H becomes 0.13 units and $\sigma_{RE} = 0.15$. The residual variance from comparing the signal to the equitable transform appears to be dominated by the standard deviations of the x and t averages of the noise (which decreases by $\sqrt{(1/M + 1/N)}$. This indicates (neglecting variations associated with the spatial and temporal noise averages) that the equitable transformation of a signal (of the form in Eq. 1) superimposed with noise does approximately partition the data set into the original separable signal (the transform) and the noise (the residuals). We explore this possibility further by examining these standard deviations under a variety of conditions but we first consider what the results from a least squares transform should look like.

Least squares transform: regression dilution bias

In the second simulation (Appendix S2: Fig. S2), we found that the regression or least squares transform had significant bias compared to the equitable transform. Linear regression fits yield slopes such that $a_{ij} \cdot a_{ji} = r^2$ where r^2 is the coefficient of determination. As r^2 approaches 1 for all pairs of sequences, the system becomes more equitable, whereas when it is close to zero, the product of the slopes also becomes zero. This is known as regression dilution bias (Berglund 2012) and is associated with error existing in the independent variable of a linear regression fit. Thus, correlating two series both of which contain error will result in biased estimates. When noise is superimposed on two sequences, the r^2 from the linear regression fit comparing the two sequences tends toward zero. This occurs even if the two sequences both contain the same underlying function $g(t)$ and fit the form of Eq. 2. One consequence of this for a regression or least squares transform is that the variance associated with the residuals between this type of transform and an underlying signal is made up primarily by the bias itself. It therefore is not a useful means by which to transform the data.

As seen in the middle panel of Appendix S2: Fig. S2, the least squares transform (T_{ls}) is approximately linearly related to the separable signal (L). A line can be fitted comparing the two. In general, higher signal values imply underestimating by the least squares transform results and the smallest signal values are generally overestimated by the transform. This approximately linear relationship ($mI + p$) relating the transform to the data has a slope (m) of <1 (Appendix S2: Fig. S2). The slope (m) and

intercept (p) from this fit can be used to approximate the variance of the difference between the least squares transform and the signal (σ_{Tls-L}). We also assume the means, μ, of both the T_{ls} and the data are about the same. The variance of the residuals between the transform and the signal is $\sigma_{Tls-L}{}^2 \sim \sigma_{mI+p-L}{}^2 \sim \sigma^2 \cdot [(m-1) \cdot L + mG + p] \sim (m-1)^2 \cdot (\sigma_L{}^2 + \mu^2) + m^2 \ \sigma_H{}^2 + p^2 + 2(m-1)p\mu$. When there is no bias ($m = 1$, $p = 0$), the residual variance should be only $\sigma_H{}^2$ or approximately the variance associated with the averages of the noise term. We shall see that, for the equitable transform, this is close to the case indicating that the equitable transform has little bias and in general approximates the signal quite well. With this information, we can better investigate the effects of noise and sampling on the equitable and least squares transforms.

Effect on the transforms of changing the resolution, amount of noise, and missing data

To understand how well an equitable transform can distinguish between a separable signal and a random noise component, it is informative to vary the resolution of the sampling, the amount of noise, and the fraction of missing data (Fig. 6). Increasing the length of the sequence (increasing M) decreases both σ_H (solid lines in Fig. 6b) and $\sigma(T - L)$ (black squares in Fig. 6) such that it appears that for large N and M, $\sigma_H \sim \sigma(T - L)$. That is, the σ associated with the difference between the transform and the original signal approaches σ_H or the standard deviation of the combined noise averages (or is just slightly larger). These average standard deviations in turn decrease with $1/\sqrt{M}$ and $1/\sqrt{N}$. In contrast, for the least squares transform, $\sigma(T_{ls} - L)$ (triangles in Fig. 6) remains relatively constant at somewhat more than ½ of the noise standard deviation. Increasing the variance of the noise relative to the signal while fixing N and M at 100 and 150 points, respectively (Fig. 6a), increases the residual σ_H (solid line in Fig. 6) which is connected closely to changes in $\sigma(T–L)$ (black squares in Fig. 6). Note that this close relation holds with cases that include the noise standard deviation being 1.3 times larger than the signal standard deviation. The residuals from comparing the least squares transform also increase remaining at a level of about $0.5\sigma_G$. The standard deviations for lower noise levels are somewhat larger than predicted by the formula above (dashed line in Fig. 6a) but are similar at very high noise levels where the noise is considerably larger than the signal. Increasing the missing data fraction, F, of the total number of grid points has the expected effect of increasing $\sigma(T–L)$ since this effectively decreases the number of data points with information associated with them.

In summary, Fig. 6 shows that the standard deviation between the transform and the signal is slightly larger than the standard deviation of a data set of the noise averages and decreases with it as the data sampling increases. There is little apparent bias. In contrast, the equivalent least squares transform performs rather poorly in comparison and does not show this trend. Increasing the noise (fixing other parameters) increases these averages, and this is reflected in the equitable transformed values (Fig. 6a). Increasing the amount of missing data plays a more important role as the fraction of missing data exceeds about 50% (Fig. 6c).

Principal component analysis and its relation to an equitable transform

There is a superficial resemblance between the equitable transform in Eq. (7) and that of principal component analysis (PCA). Usually, PCA and the equitable transform are applied under very different circumstances. Principal component analysis is generally applied when it is desirable to cluster groups of individuals together based on various quantitative measurements. Principal component analysis results in multiple eigenvectors representing independent properties of the data set and relates the various measurements and individuals to these principal components. If a data set has a single dominant principal component, then the results of PCA and the equitable transform can be compared. The equitable transform makes a model using a single property or component of the data and relates individuals by scaling or offsetting this property. Principal component analysis is best used when there are multiple different factors of a system measured such as temperature, snowmelt, and soil moisture over a landscape. The equitable system is applicable when investigating only one of these factors. If only one of these factors explains the variation in the system, PCA

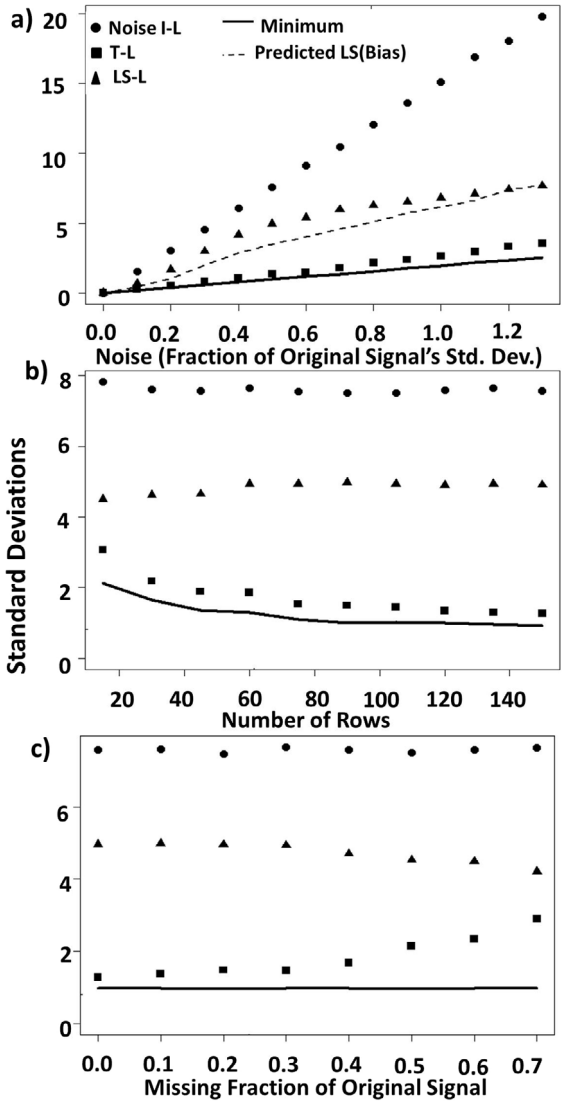


Fig. 6. Standard deviations of the difference between the separable signal described in Fig. 4 and the original data, $\sigma_{I\text{-}L} = \sigma_G$, the equitable transform, $\sigma_{T\text{-}L}$, the regression least squares transform, $\sigma_{Tls\text{-}L}$, the standard deviation of the row and column noise averages, σ_H, (solid lines), and the least squares standard deviation due to the bias, σ_{ls}, (dashed line). (a) Standard deviations listed above vs. the fraction of noise (f) relative to the signal ($\sigma_G = f{\cdot}\sigma_L$) with 150 rows and 100 columns. (b) Standard deviations listed above vs. the length of the sequences in the data set (M = number of rows) for a fixed amount of noise and columns ($\sigma_G = 0.5\sigma_L$, 100 columns). (c) Standard deviations listed above vs. the fraction of missing data (F) from the 150 × 100 data set with $\sigma_G = 0.5\sigma_L$.

can be used to make a model via its predictive capabilities and this can be compared to the results an equivalent equitable model.

We can picture the data sets described prior as cases in which a single eigenvector dominates the data and investigate predictions that result from a PCA. It is therefore informative to compare the two techniques for some of the cases we have shown. The non-linear iterative partial least squares (NIPALS) implementation of PCA (Andrecut 2009) can deal with missing information, and so we use it to compare results between a PCA transform and the equitable transform. We wanted to test both the effectiveness of the algorithms to deal with missing information as well as noise.

For the missing information case, we used the signal (step function combined with an absolute value of a linear function) from Example 1 (Fig. 3) with the lower two resolutions and missing data percentages (47% and 70% missing data). As expected, after centering the data, when no missing information existed, the PCA reproduced the signal perfectly. However, when missing information was introduced, the effect of centering the data led to inaccuracies in the centering itself and that in turn led to inaccuracies in the PCA compared with the signal. In contrast, since no centering was necessary for the equitable transform (or the least squares transform), they corresponded exactly to the signal. The top panels of Fig. 7 show the inaccuracies associated with the centering process (due to missing information affecting the averages) while the bottom panels illustrate the overall mismatch between the PCA results and the signal.

The situation when noise is introduced is even more dramatic. For the signal, we used the one from the noise example 2 (Appendix S2) with 30 sequences (N) each of length 45 (M). An arbitrary high level of noise (70% of the standard deviation of the signal, $\sigma_G = 0.7\sigma_S$) was introduced. The three transforms (PCA, least squared, and equitable) were calculated and compared with both the original data or signal + noise ($L + G$) and just the signal (Fig. 8). Comparisons with the original data (not shown) indicated that only the least squares transform showed much bias with both the other transforms being qualitatively similar. When compared with the signal, the variation in all the transforms was greatly reduced compared

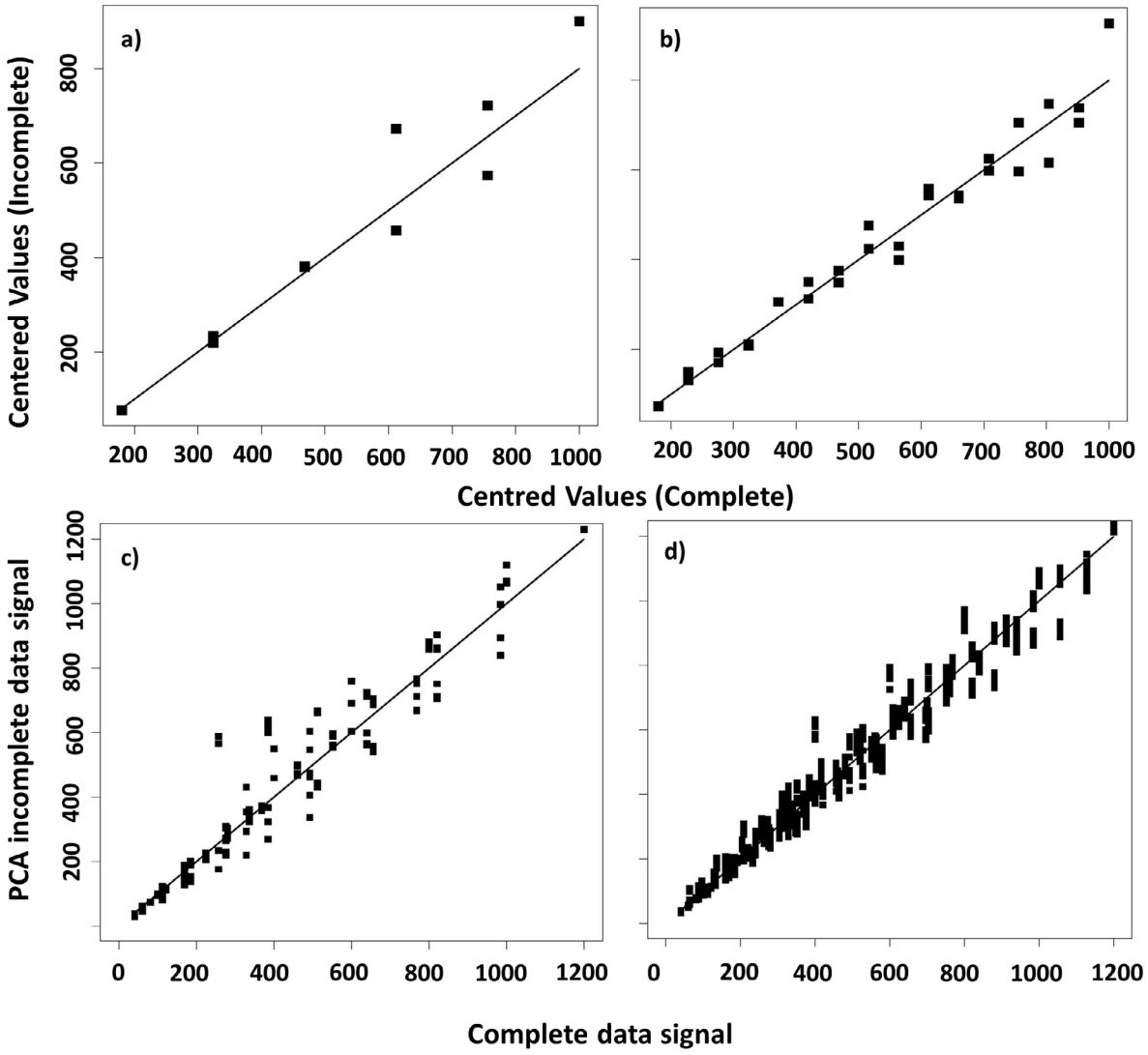


Fig. 7. Using data sets having missing data (47% (panels a and c) and 70% (panels b and d)) equivalent to those found in Fig. 3, an equitable transformation gave a perfect correspondence with the signal. Comparisons are shown with principal component analysis. Top panels (a and b): Centered values calculated from the incomplete data sets vs. the data with the complete signal. Bottom panels (c and d): Data constructs from principal component analysis on the incomplete data sets compared to the signal data.

to the signal + noise (Fig. 8a). Interestingly, however, the equitable transform showed little bias (Fig. 8c) whereas both of the other transforms showed significant bias but in opposite directions. The PCA transform (Fig. 8d) shows a bias in the opposite sense to that found in the least squares transform (Fig. 8b).

Long-term temperature variations.—A daily air temperature (at 1.5 m) record from an automatic climate station at Alexandra Fiord on Ellesmere Island from 1986 to 2007 was used to determine temperature changes at the site. The daily pattern of temperatures should be replicated to some extent each year. This repetition is ideally suited to the equitable transform analysis as we would expect daily temperature patterns to be shifted each year and the daily variation to be amplified or stretched for different years. The long-term temperatures, $W(n)$, (where n is the total day number beginning in 1986) were

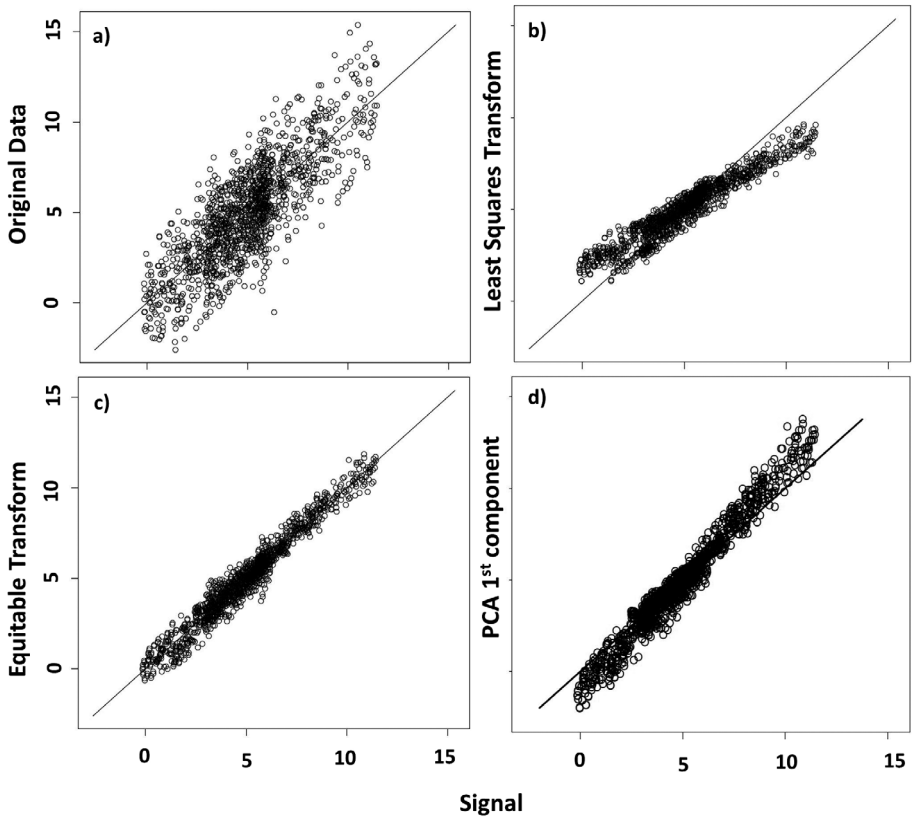


Fig. 8. Data transforms compared to the signal data for (a) original data or signal+noise, (b) least squares transform, (c) equitable transform, and (d) a principle component analysis (first component) transform. Signal used is the same as in Appendix S2: Fig. S1. Bias is evident in both the least squared transform and the principal component analysis.

reorganized into a function of two variables: year (y) and day number, d (from 1 to 365). A function $g(d)$ is replicated each year and n can be written as $n = 365 \cdot (y\text{-}1986) + d$. The long-term temperature record can then be written as $W(y{:}d)$ where the colon symbol (:) is used to denote a separation of one parameter (time), into two scales. The smaller scale (d) is replicated for each value of the larger scale (y). A first step in the analysis of $W(y{:}d)$ is to determine whether this 2D data set (in year and day number) takes the form $W(y{:}d) = f(y)\, g(d) + u(y)$. For example, if the summers were colder than usual and the winters were warmer than usual in a certain year, then $f(y)$ and the slope for that year (relative to some reference) might be less than in a normal year. Alternatively, if summer and winter temperatures increase in a particular year, then the baseline temperature ($u(y)$) and the intercept will be greater than those values in a normal year. A false color image of the original long-term temperature data, $W(y{:}d)$, shows the daily temperature measures repeat reasonably well each year (Fig. 9a). A composite plot of all the years also demonstrates the repetitive nature of the yearly cycle (Fig. 9b) although there are evident differences between years. An equitable transform applied to the data retains the portions of the data that, with some scaling and shifting, fit the annually repeated function $g(d)$. With 365 time points in each sequence and 22^2 correlated pairs of sequences, the value for R^2 was 0.87 and 1-P = 1. As in the simulation cases, the transformation allows the gross trends to emerge more clearly (Fig. 9c, d). These trends show an overall warming during the 22-yr time interval as well as a longer summer thaw period with temperatures above freezing.

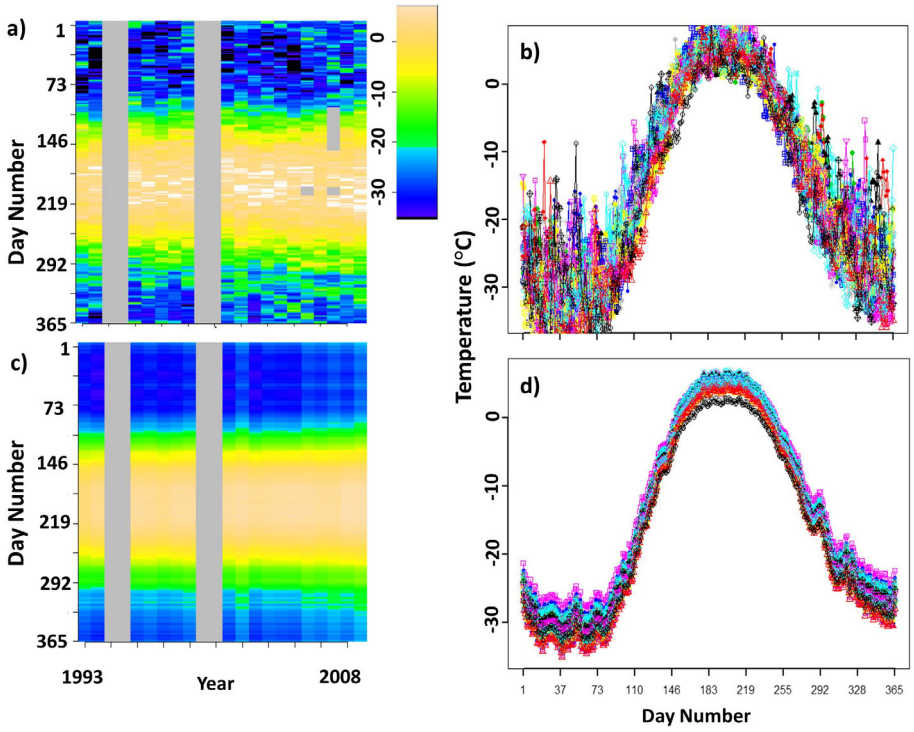


Fig. 9. (a) False color images of climate station temperature $W(y{:}d)$, (b) daily temperature variations for $W(y{:}d)$, (c) the equitable transform $T[W(y{:}d)]$ as a function of year (y-axis) and day number (x-axis), and (d) the equitable transform $T[W(y{:}d)]$ with each year color coded separately. Gray indicates data are missing.

Winter and summer warming rates

It was shown in Appendix S1 that the matrix of equitable intercepts changes with the zero value of the data set. This change of the zero value is equivalent to shifting the x-axis and y-axis (Fig. 2b). If a zero value (c) is chosen to correspond to a temperature on a particular day number, $c = W(z{:}d_c)$, for some reference year, z, then the intercept matrix (from Appendix S1: Eq. S5) associated with that zero is $b(y)_z^c = W(y{:}d_c) - c$. That is, the yearly variation of the intercept values describes the yearly temperature variation (relative to the reference year) on the day number when the temperature at the reference location equals c. In this case, these intercepts represent changes relative to a particular temperature (c) that is tied to a particular day number based on the reference year z. Choosing different zeroes for the temperature record at Alexandra Fiord will show how temperatures have changed over the last 22 yr in the different seasons (Fig. 10). The different zero values will also give different errors (Appendix S1) depending on the equitability of the system at that day number (Fig. 10).

Fitting straight lines to the weighted intercept values for temperatures of −27°, −20°, −15°, and 0°C gave significant correlations with probabilities for no correlation of 0.0007, 0.0002, 7×10^{-5}, and 4×10^{-5}, respectively. When the zero value was set to −27°C, there was the greatest warming (1.8° ± 0.4°C/decade). This warming trend decreases to 1.6° ± 0.3°C/decade at −20°C, 1.5° ± 0.28°C/decade at −14.7°C, and 1.1° ± 0.2°C/decade for 0°C. A decrease in the equitable slopes from 1986 to 2007 was also observed but was not significant. This decrease in slopes, if real, would be associated with greater warming in winter than in summer.

Changing length of summer season

Because stochastic fluctuations are for the most part removed from the equitable transform and the underlying seasonal pattern is emphasized, it

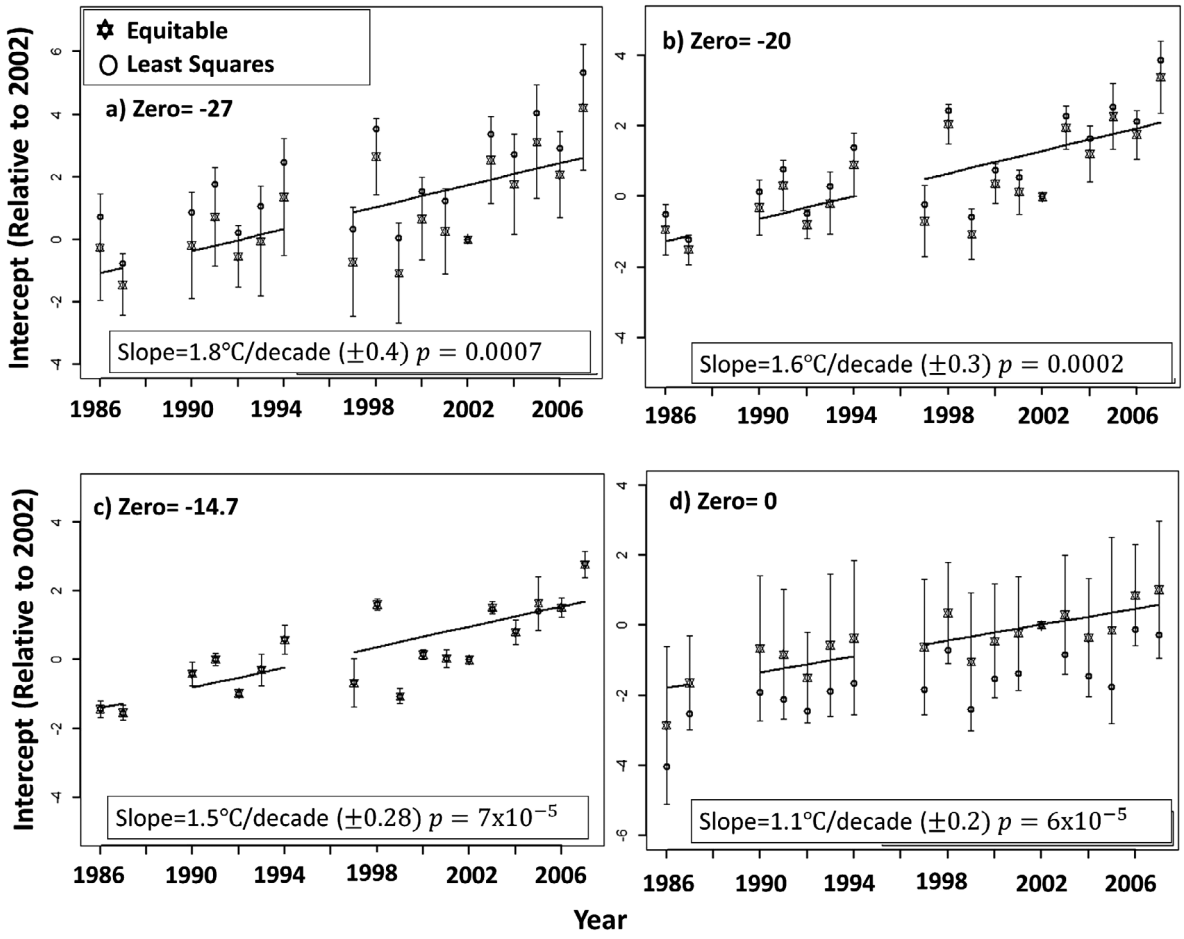


Fig. 10. Relative to the reference year = 2002, several different long-term trends for day numbers associated with the following temperatures: (a) −27°C (slope = 1.8° ± 0.4°C/decade, n = 18, P-value = 0.0007); (b) −20°C (slope = 1.6° ± 0.3°C/decade, n = 18, P-value = 0.0002); (c) the average temperature = −14.7°C (slope = 1.5° ± 0.28°C/decade, n = 18, P-value = 7×10^{-5}); and (d) 0°C (slope = 1.1° ± 0.2°C/decade, n = 18, P-value = 4×10^{-5}).

becomes possible to determine how long temperatures stay above 0°C each year (Fig. 11). The equitable transform deals with the entire data set at once, uniquely scaling and shifting the annual daily temperature sequences but eliminating fluctuations that do not follow the underlying seasonal pattern temperatures generally followed each year (i.e., the transform removes fluctuations that do not fit Eq. 1). We determined the first day and the last day of >0°C daily mean air temperatures from the equitable transform for each year and found a strong trend toward longer summers in later years (Fig. 11). Springtime thawing temperatures arrived earlier by about 9 d and autumn freezing arrived later by 10 d, indicating the growing season has lengthened by 19 d over 22 yr. The linear fits had P-values of 3×10^{-4} and 4×10^{-5}, respectively, for spring and fall.

Spatial temperature patterns

The spatial structure of temperature (from June to August) across the lowland at Alexandra Fiord was recorded in 2000 and again in 2010 using 30 temperature sensors located in a systematic grid pattern across the 7-km^2 lowland. Combining the spatial structure with the long-term temperature record measured at one

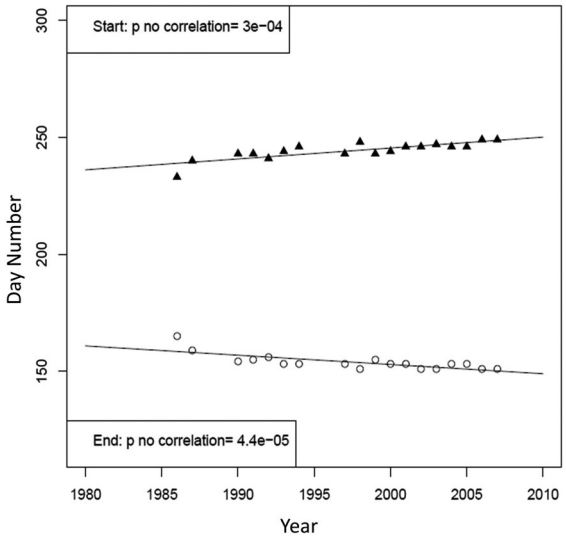


Fig. 11. Day numbers of the first day (open circles) and last day (solid triangles) with air temperatures (1.5 m) above 0°C as a function of year. Linear fits show a longer growing season by 19 d over a 22-yr period. ($n_{\text{firstday}} = 18$, $n_{\text{lastday}} = 18$, $P\text{-value}_{\text{firstday}} = 3 \times 10^{-4}$, $P\text{-value}_{\text{lastday}} = 4 \times 10^{-5}$).

location for the last 22 yr, we have a data set $W(x, y{:}d)$ where x represents the spatial locations where temperatures were recorded. If c represents the location of the long-term climate station, then we know $W(c, y{:}d)$, $W(x, 2000{:}d)$, and $W(x, 2010{:}d)$. Using these data, we determined that the daily variation in temperature in year 2000, $W(x, 2000{:}d)$, followed the same overall trend at all 30 temperature sensors and thus could be reasonably represented by Eq. 1. This result indicated a time-independent spatial structure of temperature in the year 2000 represented by the transform $T_{2000}[W(x, 2000{:}d)]$. Northerly areas near the coast were consistently cooler and more moderated than the inland areas near glaciers. Based on the year 2000 spatial structure, the 2010 spatial temperatures, $W(x, 2010{:}d)$, were predicted from the long-term climate station data for the year 2010, by assuming the spatial equitable matrices remained unchanged from 2000 to 2010 ($W(x, 2010{:}d) \sim T_{2000}[W(c, 2010{:}d)$. Comparing 2010 data with predictions, we found that, indeed, the spatial structure of temperatures remained mostly unchanged through time. The exception to this correspondence was an area in the central part of the lowland where 2010 temperatures during the summer were consistently about 3–4°C warmer than expected based on predictions inferred from the 2000 equitable system.

Tundra plant phenology as an equitable system

A long-term phenology data set of White Mountain Avens (*Dryas integrifolia*) was recorded at Alexandra Fiord, Ellesmere Island between 1993 and 2009. We explored whether or not it was an equitable system and found that plant phenological cycles in the Arctic appear to be reasonably approximated using Eq. 1. We can rewrite Eq. 1 as

$$t_i(n) = f_i g(n) + u_i, \tag{8}$$

where $t_i(n)$ represents the day number of a given event (n) for an individual plant (i). The function $g(n)$ is some general phenological curve involving events from $n = 1$ to $n = N$ (e.g., new leaf frequently corresponds to the first event $n = 1$). Eq. 8 can be reformulated as an equitable system as

$$t_i(n) = a_{ir} \cdot t_r(n) + b_{ir}. \tag{9}$$

In such an equitable phenology system, a linear relation exists between individual plants where the cycle of their events is scaled and offset in much the same manner as given in Fig. 1b. A variety of phenology rates could occur with larger equitable slopes indicating slower progression through phenology stages. This actually is the case for many of these Arctic plants at Alexandra Fiord. Using 227 plant phenology profiles (including an average profile) from a drymesic site, the R^2 value was 0.82 with a $1 - P$ value of 0.98. Convergence of the matrices occurred in two iterations.

The equitability of the data set is illustrated in Fig. 12 for two *D. integrifolia* plants relative to a 1996 reference plant. The slopes in the graphs of Fig. 12 describe the rate a plant progresses through the phenology stages. Plants with larger slopes (e.g., the 1993 plant Fig. 12a) indicate slower progression through phenology stages compared to the reference plant selected. The 1997 plant in Fig. 12b shows an earlier start time and a faster progression than either of the other two plants. It is difficult to convey the complexity of all of the relationships as there are ~25,000 separate relationships between the various plants

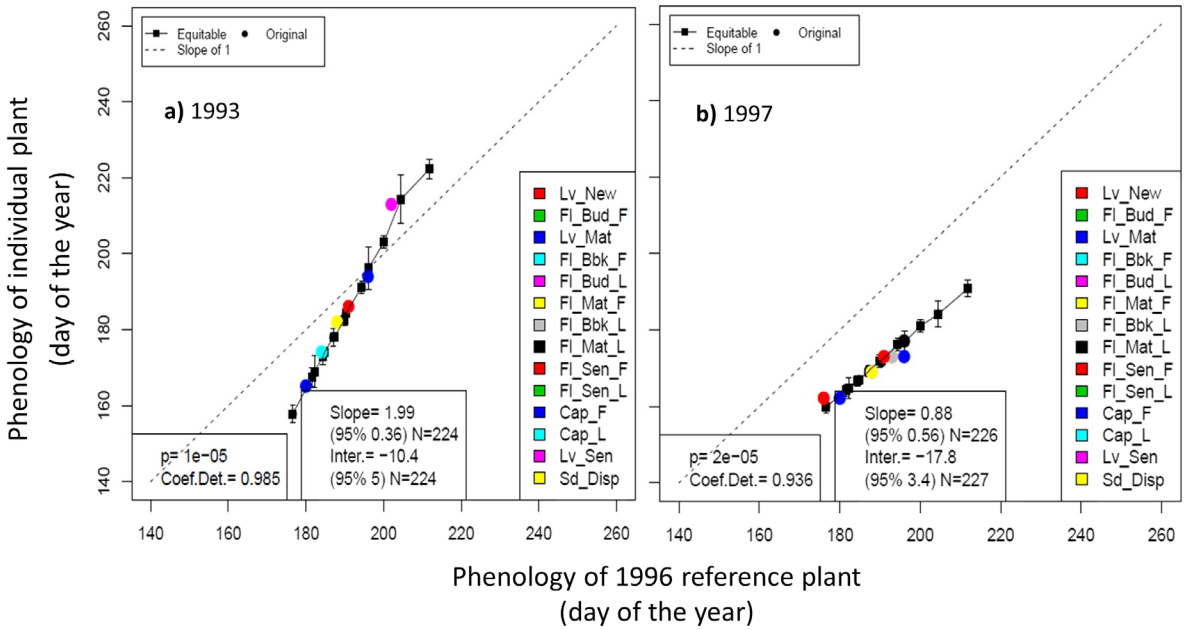


Fig. 12. Phenology of two *D. integrifolia* plants (a) year 1993 and (b) year1997 from a dry-mesic site at Alexandra Fiord plotted against a 1996 reference plant. The colored points are the raw data, and the solid squares with error bars correspond with the equitable model. The error bars represent the standard deviation of the model value based on all other plants generating that value. The dotted line shows the reference plant plotted against itself. Phenological stages are colored as in the legend. (Lv_New = First new leaf, Fl_Bud_F = First flower bud, Lv_Mat = First mature leaf, Fl_Bbk_F = First flower bud break, Fl_Bud_L = Last flower bud, Fl_Mat_F = First mature flower, Fl_Bbk_L = Last flower bud break, Fl_Mat_L = Last flower mature, Fl_Sen_F = First flower senescence, Fl_Sen_L = Last flower senescence, Cap_F = First twisting of the awns, Cap_L = Last twisting of the awns, Lv_Sen = First leaf senescence, Sd_Disp = First seed dispersal).

but it is significant to note that the phenology of this Arctic plant and perhaps plants in general can be viewed in this new perspective. A simple illustration of the usefulness of this view is shown in Fig. 13. It illustrates the long-term variation of the slope (relative to the average profile) for the years from 1993 to 2009. Plants proceeded more rapidly through their phenological stages in the later years in conjunction with the warmer temperatures described in the previous section.

Discussion

We have shown that a relatively simple concept of linearly fitting all possible pairs of sequences can be modified from linear regression fits to equitable slope and shift matrices that show little bias relative to both the original data and an underlying signal. Using an equitable transform, variations between sequences can be retained and attributed to an underlying function pervading the entire data set. One sequence in a system can be transformed equitably to any other sequence. This method keeps the individuality of each sequence intact while partitioning the data into noise and an underlying function that takes the form of Eq. 1. Averaging ignores systematic changes and individuality that may or may not be occurring within a two-dimensional data set. Rather than assuming that the multiple sequences are identical (except for noisy variations), a more general assumption is that they can be scaled and shifted relative to each other. By determining the details of this scaling and shifting, the relationships that exist between the multiple sequences that make up the data set can be determined. Thus, the individual character of each sequence can be retained while describing it in terms of the other sequences and an underlying base function. This is the essence of equitable

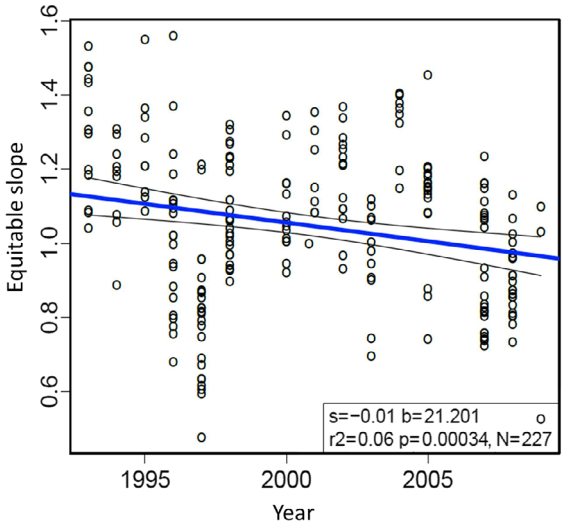


Fig. 13. Equitable slopes (relative to the average) of the *D. integrifolia* phenology system plotted against the year (1993–2009). The blue line is the regression fit, and the thin black lines represent 95% confidence intervals. The equitable system was determined from 227 individual *D. integrifolia* plants from a dry-mesic site at Alexandra Fiord between 1993 and 2009. Note that an increase in the equitable slope represents a decrease in the phenology rate.

transformation. When systematic changes are not occurring, the equitable slopes and intercepts cannot be distinguished from 1 and 0, respectively. The transform would then devolve into a simple arithmetic weighted mean (i.e., Eq. 7 with all $a_{yx} = 1$ and $b_{yx} = 0$).

An equitable two-dimensional data set of M rows (e.g., time or event number) and N columns (e.g., space or individual) results in $M \times N$ points whose large-scale pattern is described by Eq. 1. The equitable matrices A and B each require at most N parameters to construct them (from $f(x)$ and $u(x)$). The other function, $g(t)$, needs at most M parameters, and no more than $2N + M$ parameters are required to describe this equitable system. These parameters in turn could be approximated via other standard techniques further compressing the required number of parameters.

Ordering and discontinuity of data

Data analysis techniques (e.g., spectral analysis) often have assumptions regarding the continuity/ordering of a sequence and/or the data sampling frequency. The prevalence of such techniques means that data sets based on continuous functions or ones in which data points are assumed to be contiguous are usually used to test a given technique. On the other hand, the equitable transform (similar to principal component analysis) is not limited by discontinuity considerations, can deal with high frequency patterns, and the component parts do not need to be contiguous. A data set could be formed by randomly scrambling the indices (x and t). The new data set would look discontinuous at most points, and few of them would be contiguous. Imagine, for example, randomly scrambling the x and t indices in Fig. 4. The analysis of either this discontinuous data set or the one in Fig. 4 would give identical results (subject to the reordering of course). For the equitable transform, low frequency patterns are not simpler or easier to analyze than other configurations. This arbitrary ordering of rows and columns also means that N locations, in the case of a spatial variable, $x_1, x_2, \ldots, x_N$, can be randomly distributed on a two-dimensional surface. These points define the function $f(x)$ in Eq. 1. A regular grid is not required and neither is equal spacing because the ordering of the spatial locations is irrelevant for the transformation. This property makes it ideal as a technique when the x variable in question is not contiguous or linked in a systematic way. In this, it resembles PCA.

Principal component analysis, least squares, and equitable transformations

The results from comparing the equitable transform, the PCA transform, and the least squares transform are summarized in Table 2 (For the equitable and least squares transforms, centered values are not required and are simply correct after the transforms have been constructed whereas, for the PCA transform, the centered values are found pre-transform and affect the transform itself). Neither the PCA transform or the equitable transform show bias with respect to the original signal + noise while the least squares transform does. Both the PCA transform and the least squares transform do show bias with respect to the original signal whereas the equitable transform does not. As well, missing information can adversely affect

Table 2. A comparison of the variance and bias from the equitable, least squares, and principal component analysis (PCA; non-linear iterative partial least squares [NIPALS]) transforms using the data shown in Figs. 12, 13.

Variance ($T - I$)	Equitable T	Least squares T	PCA (NIPALS) T
$I = S$, $T(S)$	0	0	0
$I = S$, $T(S + \text{NA})$	0	0	Significant with little bias
$I = S + N$, $T(S + G)$	Small with little bias	Significant bias Dominated by Slope < 0	Small with little bias
$I = S$, $T(S + G)$	Small with little bias	Significant bias Dominated by Slope < 0	Significant bias Dominated by Slope > 0
$I = CV$, $CV(S + \text{NA})$	0	0	Significant with little bias

Notes: S, Signal; G, Noise; NA, Missing data; CV, Centering values; T, Variance of residuals is the difference between transform; and I, Data.

PCA results even in the absence of noise whereas no such problem was seen for either the least squares transform or the equitable transform. The PCA and equitable transforms are applicable to somewhat different circumstances.

Principal component analysis generates eigenvectors representing the data whereby the first component minimizes the variance between it and the original data (Wold 1966). The data need to be centered before the analysis will be effective. It is ideally used when more than one set of attributes (x) are recorded for many replicates (t) and when multiple competing processes are in play. In contrast, the equitable system consists of a single trait measured in multiple states for individuals each having their own sensitivity to the conditions. The equitable system is associated with a single eigenvector dominating a data set and no centering being performed beforehand. The equitable matrix having a rank of 1 (Eves 1980) is related to a single eigenvector associated with $f(x)$. The two techniques operate under very different assumptions and should not be confused with each other.

When to use the Equitable Transform?

Ideally, the transform should be applied when there is a theoretical reason to believe the data set has the form of Eq. 1. Partial differential equations often have solutions that involve separable functions of the form in Eq. 1. For example, diffusion processes can be used to model population dispersal (Levin 1992) and can take on a separable form. These processes are also used extensively to model the transport of water through different soil types (Nielsen et al. 1997, Timm et al. 2004, Reichardt et al. 2010). In latter case, the relation between distance travelled (x) and time (t) for the water content profiles follows just such an equitable form, $x = f(i)\cdot t^{1/2}$, and was used by Reichardt et al. (1972) to determine dimensionless scaling between some arbitrary reference soil (R) and other soils (i). In the context of the equitable system, their dimensionless ratio, λ, was the square of the ratio $f(i)/f(R)$. This ratio would the equitable slope matrix element between soil i and soil R found by analyzing the data set of spatial profiles, $x(i,\ t) = f(i)\cdot t^{1/2}$. The ratio, $f = x/t^{1/2,}$ is the Boltzmann transformation that is used to make soil profiles essentially equivalent to each other (Nielsen et al. 1997). More generally, this transformation is used to simplify a differential equation in two variables, x and t, to one involving only a single variable f. The equitable transform can be applied in cases where Boltzmann-like transformations are useful to convert between similar media, since it allows the conversion of one commodity to another via an equitable exchange.

If no theoretical reason is known for applying the transform, the process of changing least squares slope and intercept matrices into their equitable equivalents will help determine if the system is equitable. In practice, 2–4 iterations are required for the convergence of the slope matrix into its equitable equivalent and 1–3 iterations are required to transform the intercept matrix into its equitable form. More than this number of iterations is an indication convergence will either not occur or is inaccurate. Combining this information with the average value, R, the error bars of the transform and individual coefficients of determination allow the rapid identification of poor data and whether or not the system has a large-scale pattern consistent with Eq. 1. The calculation of the transform also allows errors to be

estimated at each point that reflect how well the system acts as an equitable one. This is important in evaluating which, if any, of the sequences are equitable with respect to each other and allows the researcher to potentially identify these outlier sequences. When a PCA analysis indicates the dominance of one principal component (or is linear with respect to two components), then the equitable transform can be used to more accurately extract that component since it does not involve the centering of data and explicitly requires the sequences to be linked in an equitable way. The errors associated with the resulting transform enable a researcher to accurately evaluate the extent to which the sequences and points in the transform are truly equitable. These errors will therefore help identify sequences different from the other more equitable sequences that are either contaminated by noise or simply inappropriately linked with the others.

The equitable transform can be used to preprocess data in order to determine underlying patterns prior to the application of other techniques such as spectral analysis so that the underlying spectra of the pattern can be more easily determined. Elphinstone et al. (1988) have shown examples of the effectiveness of equitable matrices in finding underlying signal frequencies within noisy data, and similar results will occur using the equitable transform itself.

Different forms of equitable transform: no diagonal/diagonal, zero, and average

The equitable transform can be calculated in somewhat different forms depending on the end goal of a researcher. Table 3 lists some of these options. Including the diagonals of these matrices in the transformation (Eq. 7) will result in the inclusion of the data at a particular location to be used in the calculation of the transformed value. There may be circumstances where it is preferable to determine the results purely due to other individuals/locations. As well, there are circumstances when one might want to know the relationships of all sequences relative to some average sequence (or not if that average sequence is all zeros). This is useful as it allows the average profile to substitute for the entire data set in some other multilevel system (see below). Alternatively, already having the transform matrices, it may be desirable to calculate a data set based on limited observations scattered over one or more values of x and new values of t (similar to what was done in determining temperatures in 2010 from those in 2000). The value of the equitable transform for prediction in this way still needs investigation.

Table 3. Software implementation of the equitable transform has the options listed below.

Options	Details/usefulness of options
Include/Exclude Diagonal	Use of diagonal elements implies the use of data at a location to contribute to transformed values at that same location
Include/Exclude average column	Use of the average profile implies an additional "column" profile in the equitable calculation. This is useful only if it is nonzero. Including it gives a profile with less error that can be used to link this system to other larger scale equitable systems
Limited elements, Single reference, all elements	Once the matrices are constructed, limited elements either distributed over many columns or limited to a single reference can rebuild the entire system
Zero value	Changing the zero of the data enables the shift matrix to be used to show variations relative to that zero value

Notes: The options allow the transform to be run under different conditions. The conditions for each option are described.

Multilevel systems of equitable transforms

We have seen that the spatial and long-term structure of the temperatures could be described with equitable transforms in two different ways. One transform partitioned time into a two-dimensional structure in variables year and day number. A second transform related the spatial structure to day number. Predictions were made by combining these two transforms, and it was found that modifications to the spatial structure of the temperatures occurred between the year 2000 and 2010 at the Arctic research site. As well, because each system can be exactly described by the matrices and a single average reference sequence, systems at different scales can in principle be linked together through their average profiles. The averages at each scale provide a new set of sequences upon which a new equitable system can be constructed. This implies a profoundly different view of how patterns at different scales can be combined in a manner somewhat reminiscent of fractal systems.

Equitable transforms can be applied to life cycles such as phenology patterns which can be imagined as stretched or offset cycles of generic events to characterize individual cycles. In this case, the phenology patterns between many plants of the same species are complex but the underlying principle of a cycle simply being stretched and offset is a simple one precisely described by the equitable system. We found through this method that plants are progressing through their entire phenological cycle more rapidly in recent years than in the early 1990s. Other techniques that focus on one or two events in the cycle cannot make firm conclusions about the behavior of the entire cycle as is possible here through the determination of slope in the equitable system. This rapid progression corresponds with the warming trend found at this location also through this method and may relate to observations by Høye et al. (2007) of *Dryas* plants having shorter pre-floration intervals in warmer average temperatures. The application of this principle to phenology may be very useful in other ways as well. It will allow us to investigate the synchronicity of stages in plant development.

In principle, the equitable transform may be applicable in many other situations involving two variables in which the variables are not contiguous. For example, it may be applicable when comparing the frequencies of codon and amino acid sequences between multiple organisms since this system can be thought of potentially as a set of stretched and offset frequencies relative to some reference sequence.

Conclusions

A technique has been outlined that retains individuality of sequences by describing each sequence through the scaling and shifting of a general underlying function. The equitable transform can fill in missing data and partition a two-dimensional data set consisting of noise and an underlying signal of the form in Eq. 1 into a function describing the signal and the residuals which are similar to the original noise portion of the data set. Importantly, in contrast to techniques such as principal component analysis, it describes the system in such a way that the sequence or pattern of any individual can be used to determine all the other individuals.

Many data sets are missing some data or contain some values that were recorded incorrectly. The equitable technique can approximate these missing data and identify outliers based on the general underlying pattern defined by all the sequences in the data set. Slope dilution bias is avoided using the equitable technique. This bias normally occurs when running linear regression fits on a data set that has error in the independent variable. It results in inaccurate slopes (and thus predictions) due to the noise in the system. Using the equitable technique, terms used in the calculation of regression slopes approximately cancel out, and as a result, one can obtain slopes implied from the large-scale pattern of Eq. 1 and that have the properties of an equitable matrix. The equitable shift (intercept) matrix allows the conversion between different origins of the system.

Application of this equitable transform to long-term temperature data on Ellesmere Island showed winters were warming faster than summers. We determined that winter temperatures have warmed by 1.8° ± 0.4°C/decade and summer temperatures by 1.1° ± 0.2°C/decade from 1986 to 2007. The thaw season above 0°C has extended by 19 d over the 22 yr from 1986 to 2007. A two-level equitable system was found to reasonably describe the daily temporal variations extending over a twenty-year time period covering several square kilometers for the years 2000 and 2010. Patterns of phenology in *D. integrifolia* indicate that phenology events follow this principle of equitability and indicate that the rate of progression through phenological events has increased as the climate warms. Applying this principle to plant phenology as either single or multilevel equitable systems may help us better understand phenomena such as the synchronicity of flowering. Researchers are invited to use the equitable transform package given at the end of the *Methods* section to investigate their own real or simulated data sets.

Acknowledgments

Cassandra Elphinstone would like to acknowledge Robert Elphinstone for some R programming as well

as introducing her to equitable matrices and the beauty of mathematics in general. Without this introduction to equitable matrices, she would never have thought of applying them to ecological systems. We would also like to thank Esther Frei and Robert Elphinstone for reading through the (long) first drafts and making many helpful suggestions. The research at Alexandra Fiord was supported by grants to Greg Henry from the Natural Sciences and Engineering Research Council of Canada, ArcticNet, the Canadian International Polar Year Program, and to Cassandra Elphinstone from the Northern Scientific Training Program, Polar Knowledge Canada. Logistical support for the field research was provided by the Polar Continental Shelf Program and the Royal Canadian Mounted Police. We also acknowledge the many students and assistants that have participated in collecting phenology data at the site over the years.

Literature Cited

Andrecut, M. 2009. Parallel GPU implementation of Iterative PCA algorithms. Journal of Computational Biology 16:1593–1599.

Berglund, L. 2012. Regression dilution bias: tools for correction methods and sample size calculation. Upsala Journal of Medical Sciences 117:279–283.

Brockwell, P. J., and R. A. Davis. 2002. Introduction to time series and forecasting. Second edition. Springer-Verlag, New York, New York, USA.

Chave, J. 2013. The problem of pattern and scale in ecology: What have we learned in 20 years? Ecology Letters 16:4–16.

Elphinstone, R. D., L. L. Cogger, J. S. Murphree, G. S. N. Murty, and J. W. Meriwether. 1988. The analysis of hydroxyl rotational temperatures to characterize moving thermal structures near the mesopause. Journal of Atmospheric and Terrestrial Physics 50:657–669.

Elphinstone, R., J. S. Murphree, and L. L. Cogger. 1986. A technique for the analysis of two-dimensional data: application to OI 5577Å airglow. Journal of Atmospheric and Terrestrial Physics 48:213–220.

Eves, H. 1980. Similarity and congruence. Pages 274–275 *in* Elementary matrix theory. Dover Publications, New York, New York, USA.

Høye, T. T., S. Mølgaard Ellebjerg, and M. Philipp. 2007. The impact of climate on flowering in the High Arctic—the case of Dryas in a hybrid zone. Arctic, Antarctic, and Alpine Research 39:412–421.

Kenney, J. F., and E. S. Keeping. 1954. Moving averages. Pages 221–223 *in* Mathematics of statistics, Part One. Third edition. D. Van Nostrand Company, Princeton, New Jersey, USA.

Levin, S. A. 1992. The problem of pattern and scale in ecology: the Robert H. MacArthur Award Lecture. Ecology 73:1943–1967.

Nielsen, D. R., J. W. Hopmans, M. Kutílek, and O. Wendroth. 1997. A brief review of soil water, solute transport and regionalized variable analysis. Scientia Agricola 54:89–115.

Parker, F. D. 1965*a*. Matrices in the market place. Mathematics Magazine 38:125–128.

Parker, F. D. 1965*b*. When is a loop a group. American Mathematical Monthly 72:765–766.

Reichardt, K., O. O. S. Bacchi, and L. C. Timm. 2010. Dimensional analysis and scaling. Pages 1–23 *in* College on soil physics: soil physical properties and processes under climate change. The Abdus Salam International Centre for Theoretical Physics, Trieste, Italy.

Reichardt, K., D. R. Nielsen, and J. W. Biggar. 1972. Scaling of horizontal infiltration into homogeneous soils. Soil Science Society of America Journal 36:241–245.

Timm, L. C., K. Reichardt, and O. Oliveira Santos Bacchi. 2004. Dimensional analysis, scaling and fractals. International Atomic Energy Agency (IAEA), Trieste, Italy.

Ulrych, T. J., and T. N. Bishop. 1975. Maximum entropy spectral analysis and autoregressive decomposition. Reviews of Geophysics 13:183–200.

Wiens, J. A. 1989. Spatial scaling in ecology. Functional Ecology 3:385–397.

Wold, H. 1966. Estimation of principal components and related models by iterative least squares. Pages 391–420 *in* Multivariate analysis. Academic Press, New York, New York, USA.

Supporting Information

Additional Supporting Information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecs2.3036/full