B

B-spline An abbreviation of *basis spline*, introduced in Schoenberg (1967): A chain of **polynomial**s of fixed degree (usually cubic **function**s are used) ordered in such a way that they are continuous at the points at which they join (*knots*). The knots are usually placed at the *x*-**coordinates** of the data points. The function is fitted in such a way that it has continuous first- and second-**derivative**s at the knots; the **second derivative** can be set to zero at the first and last data points. **Spline**s were first described by the Romanian-American mathematician, Isaac Jacob Schoenberg (1903–1990) (Schoenberg 1946, 1971). Other types include: **quadratic**, **cubic** and **bicubic spline**s (Ahlberg et al. 1967). Jupp (1976) described an early application of B-splines in geophysics. See also: **piecewise function**, **spline**, **smoothing spline regression**.

Back Projection Tomography (BPT) An early method used in **seismic tomography**. It has its origins in the work of the Australian-born American physicist, radio astronomer and electrical engineer, Robert Newbold Bracewell (1921–2007) who showed theoretically (Bracewell 1956) how an image of a celestial body (e.g. the brightness distribution over the Sun) could be obtained by "line integration" of the observations obtained by a narrow beam sweeping across it. In exploration seismology, the aim is to determine the velocity structure in a region which has been sampled with a **set** of rays. In the basic *back projection tomography* approach (Aki et al. 1977), a reference velocity structure (e.g. a laterally-**average**d **plane**-layer **model** for the region studied) is assumed, and deviations from the travel times are inverted to obtain the **slowness** (i.e. reciprocal velocity) **perturbation**s of the blocks. Only the assumed velocity structure is used to guide the ray's path. The **least squares** solution to the problem is found by solving the **normal equations** $\mathbf{L}^{\mathrm{T}}\mathbf{L}\boldsymbol{s} = \mathbf{L}^{\mathrm{T}}\boldsymbol{t}$, where $\boldsymbol{t} = \mathbf{L}\boldsymbol{s}$. $\boldsymbol{t}$ are the time **delay**s, $\boldsymbol{s}$ are the slowness perturbations associated with the blocks, $\mathbf{L}$ is an N by M **matrix** of lengths ($\boldsymbol{l}$) of the ray segments associated with each block, N are the number of travel-time data, and M are the number of blocks in the model. Because

© Springer International Publishing AG 2017

R.J. Howarth, *Dictionary of Mathematical Geosciences*,
DOI 10.1007/978-3-319-57315-1_2

most of the blocks are not hit by any given ray, the majority of the elements of **L** are zero. Considering only the diagonal of $\mathbf{L}^{\mathrm{T}}\mathbf{L}$, $\boldsymbol{s} = \mathbf{D}^{-1}\mathbf{L}^{\mathrm{T}}\boldsymbol{t}$, where $\mathbf{D} = \mathrm{diag}(\mathbf{L}^{\mathrm{T}}\mathbf{L})$, each ray is projected back from its receiver, one at a time. For each block encountered, the contributions to the sums of $\boldsymbol{tl}$ and $\boldsymbol{l}^2$ are accumulated separately. Once all the rays have been back-projected, each block's slowness is **estimate**d using $\boldsymbol{s} = \Sigma tl/\Sigma l^2$. The method is fast, but provides rather blurred results (Humphreys and Clayton 1988). This problem can be overcome using techniques which iterate on the basis of the travel-time **residual**s, such as the **Algebraic Reconstruction Technique** and the **Simultaneous Iterative Reconstruction Technique**.

Background

1. In geophysics: The **average** systematic or **random noise** level of a time-varying **waveform** upon which a desired **signal** is superimposed (Dyk and Eisler 1951; Sheriff 1984).
2. In exploration geochemistry: a **range** of values above which the magnitude of the concentration of a geochemical element is considered to be "**anomalous**." The term was adopted following the work of the pioneering American geochemist, Herbert Edwin Hawkes (1912–1996) (Hawkes 1957). See recent discussion by Reimann et al. (2005).
3. In computing, a *background process* is one which does not require operator intervention but can be run by a **computer** while the workstation is used to do other work (International Business Machines [undated]). See also: **anomaly**.

Backus-Gilbert method, Backus-Gilbert inversion A numerical method for the solution of **inverse problem**s in geophysics (less frequently known as *Backus-Gilbert inversion*; Google Research 2012) first introduced into geophysics by the American geophysicists, George Edward Backus (1930–) and James Freeman Gilbert (1931–2014) in a series of papers (Backus and Gilbert 1967, 1968, 1970) to infer the internal density structure, bulk **modulus**, **shear** modulus, etc. of the Earth from seismically-derived observations of its vibration **frequencies**. Their method aims to optimise the **resolution** of undetermined **model parameter**s. See also: **trade-off curve**; Parker (1977), Menke (1989, 2012), Eberhart-Phillips (1986), Snieder (1991), Buttkus (1991, 2000), Press et al. (1992), Koppelt and Rojas (1994) and Gubbins (2004).

Backward elimination A method of **subset selection** used in both **multiple regression** and **classification** (**discriminant analysis**) in which there may be a very large number (N) of potential predictors, some of which may be better than others. *Backward elimination* begins with all N predictors; each one is temporarily eliminated at a time, then the best-performing subset of the remaining ($N - 1$) predictors is retained. Selection stops when no further improvement in the **regression** fit or classification success rate is obtained. See Berk (1978), and in an earth science context, Howarth (1973a).

BALGOL An acronym for "Burroughs ALGOL." ALGOL is itself an acronym for **Algorithm**ic Oriented Language, a **computer programming language** originally developed by a group of European and American computer scientists at a meeting in Zurich in 1958 (Perlis and Samelson 1958). It was subsequently refined and popularised as ALGOL-60 (Naur 1960), assisted by the work of the computer scientists, Edsger Wybe Dijkstra (1930–2002) and Jaap A. Zonneveld (1924–) in the Netherlands; and (Sir) Charles Antony Richard Hoare (1934–), then working with the computer manufacturers, Elliott Brothers, in England. Later variants used in geological studies included BALGOL, developed by the Burroughs Corporation in the USA. Early examples of its use in the earth sciences include Harbaugh (1963, 1964) and Sackin et al. (1965), but it was soon replaced by **programming** in **FORTRAN**.

Band A **range** of **frequencies** such as those passed (**band-pass**) or rejected (**band-reject)** by a **filter**. Electrical **low-pass**, **high-pass** and **band-pass** "wave filters" were initially conceived by the American mathematician and telecommunications engineer, George Ashley Campbell (1870–1954) between 1903 and 1910, working with colleagues, physicist, Otto Julius Zobel (1887–1970) and mathematician Hendrick Wade Bode (1905–1982), but the work was not published until some years later (Campbell 1922; Zobel 1923a, 1923b, 1923c; Bode 1934). The term *band pass* was subsequently used in Stewart (1923) and Peacock (1924); see also: Wiggins (1966) and Steber (1967). See: **frequency selective-filter**.

Band-limited function A **function** whose **Fourier transform** vanishes, or is very small, outside some finite interval, i.e. **band** of **frequencies**. The term was introduced into **digital signal processing** by the American statistician, John Wilder Tukey (1915–2000) and communications engineer, Ralph Beebe Blackman (1904–1990) (Blackman and Tukey 1958). For discussion in a geophysical context, see: Grillot (1975) and Boatwright (1978).

Band-pass filter *Filters* are **algorithm**s for selectively removing **noise** from a **time series** (or spatial **set** of data), **smoothing**, or for enhancing particular components of the **signal** by removing components that are not wanted. A *band-pass filter* attenuates all **frequencies** except those in a given **range** between two given cut-off frequencies and may also be applied to smoothing of a **periodogram**. A **low-pass filter** and a **high-pass filter** connected in **series** is one form of a band-pass filter. Information in the **passband** frequencies are treated as signal, and those in the **stopband** are treated as unwanted and rejected by the filter. There will always be a narrow frequency interval, known as the **transition band**, between the **passband** and **stopband** in which the relative **gain** of the passed signal decreases to its near-zero values in the stopband. Electrical low-pass, high-pass and band-pass "wave filters" were initially conceived by the American mathematician and telecommunications engineer, George Ashley Campbell (1870–1954) between 1903 and 1910, working with colleagues, physicist, Otto Julius Zobel (1887–1970) and mathematician Hendrick Wade Bode (1905–1982), but the work was not published until some years later (Campbell 1922; Zobel 1923a, 1923b, 1923c; Bode 1934). Equivalent filters were introduced into **digital signal processing** by the American statistician, John Wilder Tukey (1915–2000) and mathematician Richard Wesley Hamming (1915–1998) (Tukey and Hamming 1949). Parallel theoretical background was provided by the work of the American physicist, George W. Steward (1876–1956), who worked on acoustics between 1903 and 1926 and solved the fundamental **wave equation**s involved in acoustic filter design (Crandall 1926; Stewart 1923). See Buttkus (1991, 2000), Camina and Janacek (1984), Gubbins (2004) and Vistelius (1961) for discussion in an earth sciences context.

Band-reject filter, band-stop filter A **filter** which is designed to remove (reject) a narrow **band** of **frequencies** in a **signal** while passing all others. It is also known as a *notch* or **rejection filter** (Sherriff 1984; Wood 1968; Buttkus 2000; Gubbins 2004). The opposite of a **band-pass filter**. See: Steber (1967) and Ulrych et al. (1973).

Banded equation solution This refers to the solution of a system of **linear equation**s involving a square **symmetric matrix** in which the *band* referred to is a symmetrical area on either side of, and parallel to, the matrix diagonal which itself contains nonzero values. Outside this band, all entries are zero. See: Segui (1973), Carr (1990) and Carr and Myers (1990).

Bandwidth

1. The width of the **passband** of a **frequency selective-filter**; the term was introduced into **digital signal processing** by the American statistician, John Wilder Tukey (1915–2000) and mathematician Richard Wesley Hamming (1915–1998) (Tukey and Hamming 1949).
2. A term introduced by the British statistician, Maurice Stevenson Bartlett (1910–2002), in the context of the **smoothing parameter** used in smoothing a **periodogram** (Bartlett 1950). *Bandwidth* always exceeds the **Rayleigh frequency**.
3. It has more recently been applied to the **smoothing parameter** used in **kernel density estimation** (Chin 1991). Mentioned in an earth science context by: Sheriff (1984), Buttkus (1991, 2000), Weedon (2003) and Gubbins (2004).

Bandwidth retention factor A criterion used in **normalizing** the **taper** coefficients in designing a multitaper **filter**, it is the **ratio**: (energy within the chosen spectral **frequency band**)/(energy in the entire **band**). It was called the *bandwidth retention factor* by Park et al. (1987). See: **multi-tapering method**.

Bar chart A **graph** in which either the absolute **frequency** or relative frequency of occurrence of a category is shown by the proportional-length of a vertical bar for each category in a **data set**. Since they are categorical **variables**, ideally, the side-by-side bars should be drawn with a gap between them. Not to be confused with a **histogram**, which shows the binned **frequency distribution** for a continuous- or discrete-valued variable. The earliest *bar chart*, based on absolute amount, was published by the English econometrician, William Playfair (1759–1823) (Playfair and Corry 1786). An early earth science use was by Federov (1902) to show relative mineral birefringences. In a *divided bar chart*, each bar is divided vertically into a number of proportional-width zones to illustrate the relative proportions of various components in a given **sample**; total bar-length may be constant (e.g. 100% **composition**) or vary, depending on the type of **graph**. These were first used by the German scientist, Alexander von Humboldt (1769–1859) (Humboldt 1811). In geology, divided bars were first used by the Norwegian geologist, metallurgist and experimental petrologist, Johan Herman Lie Vogt (1858–1932) (Vogt 1903–1904). The Collins (1923) bar chart uses double divided bars to show the cationic and anionic compositions of a water sample separately; each **set** is recalculated to sum to 100% and plotted in the left- and right-hand bars respectively. Usage in geology increased following publication of Krumbein and Pettijohn's *Manual of sedimentary petrography* (1938).

Bartlett method, Bartlett spectrum, Bartlett taper, Bartlett window, Bartlett weighting function Named for the British statistician, Maurice Stevenson Bartlett (1910–2002) who first **estimate**d the **power spectrum** density of a **time series**, by dividing the data into a number of contiguous non-overlapping segments, calculating a **periodogram** for each (after **detrending** and **tapering**), and calculating the **average** of them (Bartlett 1948, 1950). The term *Bartlett window* (occasionally misspelt in recent literature as the "Bartlet" window), was introduced into **digital signal processing** by the American statistician, John Wilder Tukey (1915–2000) and communications engineer, Ralph Beebe Blackman (1904–1990) (Blackman and Tukey 1958) and has remained the most frequently used term since the mid-1970s (Google Research 2012). It is used in the operation of **smoothing** a periodogram with a **lag window** of **weight**s applied to a discrete time **waveform**. N, the width of the Bartlett window, is typically even and an **integer** power of 2, e.g. 2, 4, 8, 16, 32, etc.; for each point, $n = 0, \cdots, N$, the weight $w(n)$ is given by

$$w(n) = \frac{2n}{N-1}; 0 \leq n \leq \left(\frac{N}{2}\right) - 1$$

and

$$w(n) = 2 - \frac{2n}{N-1}; \frac{N}{2} \leq n < N;$$

otherwise zero. It is also known (Harris 1978) as the *triangle*, **triangular**, or **Fejér window**, named for the Hungarian mathematician, Lipót Fejér (1880–1959) (Fejér 1904). See also Blackman and Tukey (1958) and, for a comprehensive survey, Harris (1978). Mentioned in an earth science context by: Buttkus (1991, 2000) and Weedon (2003). See also: **spectral window**.

Barycentric coordinates The **percentage**-based coordinate system used today in an equilateral **ternary diagram** is equivalent to the *barycentric coordinate system* introduced by the German mathematician, August Ferdinand Möbius (1790–1886) (Möbius 1827). Imagine three masses w_A, w_B and w_C, placed at the apices A, B, C of a triangle and all joined by threads to an interior point P at equilibrium, then the areas of the subtriangles $BPC = a$, $APC = b$ and $APB = c$ are proportional to w_C, w_B and w_A respectively. The barycentric coordinates $\{a, b, c\}$ may be normalised so that $a + b + c = 1$. However, Möbius never seems to have used the idea as the basis for a **graphical** tool.

BASIC Acronym for *Beginner's All-purpose Symbolic Instruction Code*, a general-purpose interactive **computer programming language** (i.e. interpreted on the fly, rather than compiled and run) originally developed in 1963–1964 by American mathematicians and computer scientists, John George Kemeny (1926–1992), and Thomas Eugene Kurtz (1928–) at Dartmouth College, New Hampshire, USA, as a teaching tool for non-scientist undergraduates (Kemeny and Kurtz 1964). It was partly based on **FORTRAN** II and **ALGOL**, with additions to make it suitable for timesharing use. Because of its ease of use, it was subsequently adopted for use on mini**computer**s, such as the DEC PDP series, Data General and Hewlett Packard in the late 1960s and early 1970s, but it was the development of BASIC **interpreter**s by Paul Allen (1953–) and William Henry Gates III (1955–), co-founders of Microsoft, and Monte Davidoff (1956–) for the Altair and Apple computers, and its subsequent take-up in many other dialects by other manufacturers which popularised its use in the personal-computing environment of the 1980s. Early applications in the earth sciences include: Till et al. (1971), McCann and Till (1973) and Jeremiasson (1976).

Basin analysis The quantitative modelling of the behaviour of sedimentary basins through time has become an important tool in studying the probable hydrocarbon potential of a basin as an aid to exploration. Modelling generally embraces factors such as basement subsidence, compaction and fluid flow, burial history, thermal history, thermal maturation, and hydrocarbon generation, migration and accumulation. The aim is to determine the relative timing of hydrocarbon evolution in relation to the development of traps and their seals, and the continuing integrity of the sealed traps following petroleum entrapment. The methods used have been largely developed by the British-American physicist, theoretical astronomer, and geophysicist, Ian Lerche (1941–); see: Lerche (1990, 1992), Dore et al. (1993), Harff and Merriam (1993) and Lerche et al. (1998).

Basin of attraction A region in **phase space** in which solutions for the behaviour of a **dynamical system** approach a particular **fixed point**; the **set** of initial conditions gives rise to trajectories which approach the **attractor** as time approaches infinity. The term was introduced by the French topologist, René Thom (1923–2002) in the late 1960s and published in Thom (1972, 1975). For discussion in an earth science context, see Turcotte (1997). See also: **phase map**.

Basis function, basis vector

1. An element of a particular *basis* (a **set** of **vector**s that, in a **linear** combination, can represent every vector in a given vector **space**, such that no element of the set can be represented as a linear combination of the others) for a **function** space (a set of functions of a given kind). *Basis function* has been the most frequently used spelling since the 1980s (Google Research 2012). Examples include the **sine** and **cosine** functions which make up a **Fourier series**, **Legendre polynomial**s, and **spline**s.
2. **Algorithm**s which form the basis for numerical modelling and for methods of approximation (Sheriff 1984; Gubbins 2004)

Batch processing The execution of a series of "jobs" (**program**s) on a **computer**, established so that they can all be run to completion without manual intervention. Used on mainframe computers since the 1950s, it ensures the maximum level of usage of the computer facilities by many users. Early examples of geological programs for such an environment are those of Krumbein and Sloss (1958), Whitten (1963) and Kaesler et al. (1963). By the 1970s, "time-shared" operations enabled input/output via remote Teletype terminals which offered both keyboard and punched paper-tape readers as means of input and, in the latter case, output also. An early example of a suite of statistical **computer program**s for geological usage written for a time-sharing environment is that of Koch et al. (1972).

Batch sampling

1. An alternative name for *channel **sampling***, a means of physical **sampling** in a mine environment in which a slot, or channel, of given length is cut into the rock face in a given alignment (generally from top to bottom of the bed, orthogonal to the bedding plane); all the rock fragments broken out of the slot constitute the sample.
2. In statistical **sampling**, it is a method used to reduce the volume of a long data **series**: the **arithmetic mean** of all the values in a fixed non-overlapping **sampling interval** is determined and that value constitutes the channel sample. See: Krumbein and Pettijohn (1938) and Krumbein and Graybill (1965); **composite sample**.

Baud In asynchronous transmission, the unit of modulation rate corresponding to one unit interval per second; e.g. if the duration of the interval is 20 ms, the modulation rate is 50 **baud** (International Business Machines [undated])

Bayes rule, Bayesian methods Given a *prior* **frequency distribution** of known (or sometimes assumed) **function**al form for the occurrence of the **event**, the *posterior* frequency distribution is given by Bayes' rule, named after the English philosopher and mathematician, Thomas Bayes (1702–1761). Expressed in modern notation as:

$$p(S|X) = [p(X|S)p(S)]/\{[p(x_1|S)p(S)] + [p(x_2|S)p(S)] + \cdots + [p(x_n|S)p(S)]\},$$

where p(S|X) is the posterior distribution of a given state (or **model parameter**s) S occurring, given a **vector** of observations, X; p(S) is the prior distribution; and p(x|S) is the **likelihood**. However, this "rule" does *not* appear in Bayes (1763); John Aldrich in Miller (2015a) gives the first use of the term "la règle de Bayes" to Cournot (1843) but attributes its origin to Laplace (1814). The term *Bayesian* was first used by the British statistician, (Sir) Ronald Alymer Fisher (1890–1962) in Fisher (1950). See: Wrinch and Jeffreys (1919) and, in an earth science context: Appendix B in Jeffreys (1924), also: Rendu (1976), Vistelius (1980, 1992), Christakos (1990), Curl (1998), Solow (2001) and Rostirolla et al. (2003); **Bayesian inversion**, **Bayesian/maximum-entropy method**.

Bayesian inversion The application of **Bayesian methods** to solution of **inverse problem**s (e.g. the reconstruction of a two-dimensional cross-sectional image of the interior of an object from a **set** of measurements made round its periphery). For discussion in an earth science context, see: Scales and Snieder (1997), Oh and Kwon (2001), Spichak and Sizov (2006), Hannisdal (2007), Gunning and Glinsky (2007) and Cardiff and Kitandis (2009).

Bayesian/Maximum-Entropy (BME) method A methodological approach to the incorporation of **prior information** in an optimal manner in the context of spatial and spatio-temporal **random field**s: given measurements of a physical **variable** at a limited number of positions in space, the aim is to obtain **estimate**s of the variable which are most likely to occur at unknown positions in space, subject to the *a priori* information about the spatial variability characteristics. Introduced by the Greek-born American environmental scientist and statistician, George Christakos (1956–), Christakos (1990, 2000). See also: **Bayes rule**, **maximum entropy principle**.

Beach ball plot Given a **population** of compatible measurements of the characteristics of geological faults, determining the proportion of points in compression or **extension** in each direction enables the three orthogonal principal **stress axes** (σ_1, σ_2 and σ_3) to be located. The method involves placing a **plane** perpendicular to the plane of movement in a fault; dividing the fault into a **set** of 4 **dihedra** or quadrants. Two will be in compression (+) and two will be in extension (−). σ_1 and σ_3 will lie somewhere between the dihedra; if the directions of σ_1 and σ_3 can be determined, then the remaining stress **axis**, σ_2, can be calculated from them, as it must be perpendicular to them both, or **normal** to the plane they define. σ_1 will lie somewhere in the area of compression and σ_3 will lie somewhere in the area of extension. As faults are rarely isolated, other faults in the fault system can also be plotted on a **Lambert equal area projection**. As increasing numbers are plotted, a direction for σ_1 and σ_2 representing the entire fault system may be determined. The two compressional right **dihedra** and two extensional right dihedra shown on the **graph** may be coloured white and black respectively, leading to its being called a *beach ball plot*. Introduced by the French structural geologist, Jacques Angelier (1947–2010) and geophysicist, Pierre Mechler (1937–) (Angelier and Mechler 1977) when it was known as an **Angelier-Mechler diagram**. Their method was improved on by the British structural geologist, Richard J. Lisle (1987, 1988, 1992).

Beat If two **sinusoid**s of similar **wavelength**s are added together, the resultant **waveform** will have constant **wavelength** (equal to the **average** of the wavelengths of the two sinusoids), but the **amplitude** of the resulting waveform, the *beat*, will vary in a fixed manner which will be repeated over the *beat wavelength*. The term originally derives from the acoustics of music, and was used (battement) by the French mathematician and physicist, Joseph Sauveur (1653–1716) (Sauveur [1701] 1743); by 1909 it was in use in wireless telegraphy, first patented by Italian physicist, Guglielmo Marconi (1874–1937) in 1896. Mentioned in an earth science context by Panza (1976) and Weedon (2003). See also: **amplitude modulation**.

Belyaev dichotomy Named for the Russian statistician, Yuri Konstantinovich Belyaev (1932–), who proved (Belyaev 1961, 1972) that with a **probability** of one, a **stationary Gaussian process** in one **dimension** either has continuous **sample path**s, or else almost all its paths are unbounded in all intervals. The implication for a **Gaussian random field** is that if it is smooth it is very smooth, but if it is irregular, it is highly irregular and there is *no* in-between state. This concept was applied to the topography of a soil-covered landscape by the British theoretical geomorphologist and mathematician, William Edward Herbert Culling (1928–1988) in Culling and Datko (1987) and Culling (1989) who used it to justify the view that the **fractal** nature of a landscape renders "the customary geomorphic stance of phenomenological measurement, naïve averaging and mapping by continuous **contour** lines" both "inappropriate" and "inadmissible" (Culling 1989).

Bell-curve, bell-shaped curve, bell-shaped distribution An informal descriptive name for the **shape** described by a continuous **Gaussian** ("**normal**") **frequency distribution**. John Aldrich in Miller (2015a) says that although the term "bell-shaped curve" appears in Francis Galton's description of his *Apparatus affording Physical Illustration of the action of the Law of Error or of Dispersion*: "Shot are caused to run through a narrow opening among pins fixed in the face of an inclined **plane**, like teeth in a harrow, so that each time a shot passes between any two pins it is compelled to roll against another pin in the row immediately below, to one side or other of which it must pass, and, as the arrangement is strictly symmetrical, there is an equal chance of either event. The effect of subjecting each shot to this succession of alternative courses is, to disperse the stream of shot during its downward course under conditions identical with those supposed by the hypothesis on which the law of **error** is commonly founded. Consequently, when the shot have reached the bottom of the tray, where long narrow compartments are arranged to receive them, the general outline of the mass of shot there collected is always found to assimilate to the well-known bell-shaped curve, by which the law of error or of dispersion is mathematically expressed," Galton demonstrated his apparatus at a meeting of the Royal Institution in February 1874 (Committee of Council on Education 1876), but did *not* actually use the term in his many statistical publications. Nevertheless, the term began to be used in the early 1900s and by Thompson (1920), but it gained in popularity following its appearance in textbooks, such as Uspensky (1937) and Feller (1950).

Bending power law spectrum An energy **spectrum** which is a modification of the linear ($1/f$) **power law spectrum** (f is **frequency**) which includes an element enabling to bend downwards, steepen, at high frequencies: It has the form:

$$E(f) = \frac{Nf^{-c}}{1 + \left(\frac{f}{f_B}\right)^{d-c}}$$

where N is a factor which sets the amplitude, f_B is the frequency at which the bend occurs, and c (usually in the **range** 0 to 1) and d (usually in the range 1 to 4) are constants which govern the **slope** of the spectrum above and below the bend. Vaughan et al. (2011) discuss the problems inherent in choice of a first-order autoregressive, AR(1), process as a **model** for the spectrum in cyclostratigraphy and recommend use of the **power law**, *bending power law* or **Lorentzian power law** models as alternatives. See also **power spectrum**.

Bernoulli model, Bernoulli variable A Bernoulli **random variable** is a **binary** variable for which the **probability** that e.g. a species is present at a site, $\Pr(X = 1) = p$ and the probability that it is not present, $\Pr(X = 0) = 1 - p$. Named for the Swiss mathematician, Jacques or Jacob Bernoulli (1654–1705), whose book, *Ars Conjectandi* (1713), was an important contribution to the early development of probability theory. A statistical **model** using a variable of this type has been referred to since the 1960s as a *Bernoulli model* (Soal 1965; Merrill and Guber 1982).

Bernstein distribution A family of **probability distribution**s of the form

$$F(x; m) = \Phi\left\{\frac{(x - m)}{\sqrt{f(x)}}\right\},$$

where $\Phi\{\bullet\}$ is the **normal distribution**; m is the **median**; and $f(x)$ is a **polynomial function** in x, (e.g. $ax^2 - 2bx + c$; where a, b, and c are constants), whose value is greater than zero for all x. Introduced by the Russian mathematician, Sergei Natanovich Bernštein (1880–1968) (Bernštein 1926a, b; Gertsbakh and Kordonsky 1969); for discussion in an earth science context, see Vistelius (1980, 1992).

Bessel function A **set** of **function**s that are solutions to **Laplace's equation** in cylindrical **polar coordinates**. Named (Lommel 1868) for the German astronomer and mathematician, Friedrich Wilhelm Bessel (1784–1846). The first spherical *Bessel function* is the same as the unnormalised **sinc** function, i.e. sin(x)/x. Mentioned in an earth science context by Buttkus (1991, 2000).

Best Linear Unbiased Estimator (BLUE) A **linear** estimator of a **parameter** which has a smaller **variance** associated with it than any other estimator, and which is also **unbiased**, e.g. the **ordinary least squares** estimator of the coefficients in the case of fitting a linear **regression equation**, as shown by the French mathematician, Pierre Simon, Marquis de Laplace (1749–1827), (Laplace 1812, p. 326), or the use of "ordinary" **kriging** to **estimate** concentration values for spatially distributed data in applied **geostatistics** (Journel and Huijbregts 1978; Isaaks and Srivastava 1989).

Beta diagram Introduced by the Austrian structural geologist, Bruno Sander (1884–1979) (Sander 1948; Sander 1970), the **β-axis** is the line of intersection between two or more **plane**s distinguished by a parallel fabric (e.g. bedding planes, foliation planes). If the attitudes of these planes in a folded structure are plotted in **cyclographic** form on a **stereographic projection**, the **unimodal** ensemble of intersections statistically defines the **location** of the **mean** **β-axis**, which may correspond to a cylindrical fold **axis** (in certain types of complex folding they may not represent a true direction of folding). Also called a **pole diagram**. See: Turner and Weiss (1963), Robinson (1963) and Ramsay (1964, 1967).

Beta distribution, Beta function A family of continuous **probability distribution**s of the form

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)},$$

where $0 < x < 1$, $0 < \alpha, \beta < \infty$ and $B(\alpha, \beta)$ is the *Beta function:* $B(\alpha,\beta) = \int_0^1 u^{\alpha-1} (1-u)^{\beta-1}\,\mathrm{d}u$, first studied by the Swiss mathematician, Leonhard Euler (1707–1783) (Euler 1768–1794), and by the French mathematician, Pierre Simon, Marquis de Laplace (1749–1827), (Laplace 1774). Subsequently given its name by the French mathematician, Jacques Phillippe Marie Binet (1776–1856) (Binet 1839). The distribution is J-shaped if α or β lie between 0 and 1, and U-shaped if both are within this **range**. Otherwise, if α and β are both greater than 1, then it is **unimodal** with the peak of the distribution (**mode**) falling at $(\alpha - 1)/(\alpha + \beta - 2)$. It is frequently used to fit data on a finite interval and has been applied to the modelling of the proportions of microlithotype data in coal (Cameron and Hunt 1985). A **Beta** distribution **scaled** to the observed maxima and minima is known as a

stretched Beta distribution, which is now being used for distribution-fitting in petroleum resource estimation studies (Senger et al. 2010). See also **incomplete Beta function**.

Beta test To test a pre-release version of a piece of **software** by making it available to selected users (International Business Machines [undated]).

Bias, biased

1. In statistical terms, *bias* is the difference between the **estimate**d value of a **parameter**, or **set** of parameters, and the true (but generally unknown) value. The terms *biased* and **unbiased error**s were introduced by the British econometrician, (Sir) Arthur Lyon Bowley (1869–1957) (Bowley 1897). Typically, the estimated value might be inflated by erroneous observations or the presence of an **outlier**, or outliers, in the data. In **time series analysis**, it may be applied to the incorrect estimation of the **periodogram** as a result of the **leakage effect**. For discussion in an earth science context, see: Miller and Kahn (1962), Buttkus (1991, 2000) and Weedon (2003).
2. In geochemical analysis, or similar measurement processes, it is the difference between a test result (or the **mean** of a **set** of test results) and the accepted reference value (Analytical Methods Committee 2003). In practice, it is equivalent to **systematic error**. In analytical (chemical) work, the magnitude of the *bias* is established using a standard reference material, and it is generally attributable to instrumental **interference** and/or incomplete recovery of the analyte. See also: **accuracy**, **precision**, **inaccuracy**, **blank**.

Bicoherence This is a **measure** of the proportion of the **signal** energy at any **bifrequency** that is quadratically **phase**-coupled. **Nonlinear frequency modulation** of a signal will be indicated by the presence of phase- and frequency-coupling at the **frequencies** corresponding to the **sideband**s, e.g. where a signal is composed of three **cosinusoid**s with frequencies f_1, f_2, and $f_1 + f_2$ and phases φ_1, φ_2 and $\varphi_1 + \varphi_2$. This will be revealed by peaks in the *bicoherence*, a squared normalised version of the **bispectrum** of the **time series**, $B(f_1,f_2)$:

$$b(f_1,f_2) = \frac{|B(f_1,f_2)|^2}{E\left[|P(f_1)P(f_2)|^2\right]E\left[|P(f_1+f_2)|^2\right]},$$

plotted as a **function** of f_1 and f_2; where $P(f)$ is the complex **Fourier transform** of the **time series** at frequency f; and $E(\bullet)$ is the **expectation operator**. The term was introduced into **digital signal processing** by the American statistician, John Wilder Tukey (1915–2000) about 1953 (Brillinger 1991; Tukey 1953). See also: Brillinger (1965), Brillinger and Rosenblatt (1967a, b) and Brillinger and Tukey (1985); discussed in an earth science context by: Elgar and Sebert (1989), Mendel (1991), Nikias and Petropulu (1993), Persson (2003) and Weedon (2003).

Bicubic spline A chain of **polynomial**s of fixed degree (usually cubic **function**s are used) in such a way that they are continuous at the points at which they join (knots). The knots are usually placed at the *x*-**coordinates** of the data points. The function is fitted in such a way that it has continuous first and **second derivative**s at the knots; the second derivative can be set to zero at the first and last data points. **Spline**s were discovered by the Romanian-American mathematician, Isaac Jacob Schoenberg (1903–1990) (Schoenberg 1946). See also: Schoenberg (1971), Ahlberg et al. (1967) and Davis and David (1980); **smoothing spline regression**, **spline**, **piecewise function**.

Bifrequency A reference to two **frequencies** of a single **signal**. The term was introduced into **digital signal processing** by the American statistician, John Wilder Tukey (1915–2000) (Brillinger 1991; Tukey 1953). See: **bicoherence**, **bispectrum**.

Bifurcation A sudden change in the behaviour of a **dynamical system** when a control **parameter** (p) is varied, resulting in a **period-doubling**, quadrupling, etc. with the onset of **chaos**. A system of behaviour that previously exhibited only one **mode**, which subsequently exhibits 2, 4, etc. It shows on a **logistic map** as a splitting of the trace made by the **variable** representing the behaviour of the system when plotted as a **function** of p; the splitting becomes more and more frequent, at progressively shorter intervals, as p increases in magnitude. The term was coined by the French mathematical physicist and mathematician, Jules Henri Poincaré (1854–1912) (Poincaré 1885, 1902) but was first used in this context by the Austrian-born German mathematician, Eberhard Frederich Ferdinand Hopf (1902–1983) (Hopf 1942; Howard and Kopell 1976), and the Russian mathematician, Lev Davidovich Landau (1908–1968) (Landau 1944). For earth science discussion see: Turcotte (1997) and Quin et al. (2006). See also: **Andronov-Hopf bifurcation**, **period-doubling bifurcation**, **pitchfork bifurcation**.

Bi-Gaussian approach A method of **geostatistical** estimation (Marcotte and David 1985) in which the conditioning is based on the simple **kriging estimate** of the **mean** value of the **Gaussian variable** representing the grades of a point, or block, rather than the actual data values. See also: **multi-Gaussian approach**.

Bilinear interpolation A two-**dimension**al **interpolation** method in which values are first interpolated in one direction and then in the orthogonal direction. It was originally used for interpolation in tables, e.g. Wilk et al. (1962). Konikow and Bredehoeft (1978) used the method in computing solute transport in groundwater, and Sheriff (1984) gives the example of first interpolating in time between picks at velocity analysis points and then spatially between velocity analysis positions.

Bilinear mapping, bilinear transform A **stability**-preserving **transform** used in **digital signal processing** to transform continuous-time system representations (**analogue** signal) to discrete-time (**digital** signal) and *vice versa*. It is often used in the design of **digital**

filters from an analogue prototype. Sometimes known as the **Tustin transform** or **Tustin's method**, after the British electrical engineer, Arnold Tustin (1899–1994), who first introduced it (Tustin 1947). See Buttkus (1991, 2000) for discussion in an earth science context.

Billings net This **graphical** net (a **Lambert equal-area** (polar) **projection** of the sphere) is used as an aid to plotting structural data (e.g. **pole**s to joint **plane**s). Named for the American structural geologist, Marland Pratt Billings (1902–1996), whose textbook (Billings 1942) greatly helped to promote its use in analysis of geological structures. This seems a little surprising, as the **stereographic net**, which appeared in the second edition (Billings 1954) is acknowledged by him as being reproduced from a paper by the American structural geologist, Walter Herman Bucher (1888–1965) (Bucher 1944). However, it was the Austrian mineralogist, Walter Schmidt (1885–1945) who was the first to adopt the use of the Lambert projection in **petrofabric** work in structural geology (Schmidt 1925), and it was first used in macroscopic structural work by Fischer (1930), but it was undoubtedly Billing's work which popularised its use in macro-scale structural geology (Howarth 1996b).

Bimodal distribution A **variable** with two local maxima in its **probability density**. Use of the term goes back to about 1900. The first attempt to decompose a *bimodal distribution* into two normally distributed components in the geological literature appears to be that of the British petrologist, William Alfred Richardson (1887–1965) who, in 1923, applied it to the **frequency distribution** of silica in igneous rocks (Richardson 1923), using the **method of moments** originally described by the British statistician, Karl Pearson (1857–1936) (Pearson 1894). Jones and James (1969) discuss the case of bimodal **orientation data**. See also: **frequency distribution decomposition**.

Bin

1. One of a **set** of fixed-interval divisions into which the **range** of a **variable** is divided so as to count its **frequency distribution**. The term is believed to have been first used the British statistician, Karl Pearson (1857–1936) in his lectures at Gresham College, London, probably in 1892/1893 when he introduced the **histogram** (Bibby 1986).
2. Sherriff (1984) uses the term for one of a **set** of discrete areas into which a survey region is divided (it is also used in this sense in astronomical surveys).

Binary coefficient Statistical models for the analysis of **binary**-coded (presence/absence) data were reviewed by Cox (1970). Cheetham and Hazel (1969) review 22 **similarity coefficients** for such data in the literature, some of which are discussed in more detail by Sokal and Sneath (1963) and Hohn (1976); see also Hazel (1970) and Choi et al. (2010). Of these, the **Dice coefficient**, **Jaccard coefficient**, **Otsuka coefficient**, **Simpson**

coefficient and **simple matching coefficient** were embodied in a **FORTRAN program** for biostratigraphical use by Millendorf et al. (1978).

Binary digit (bit) Usually known by its acronym *bit*, the term was coined by the American statistician, John Wilder Tukey (1915–2000) about 1946, because the two states of an element in a **compute**r's core can represent one digit in the binary representation of a number. It first appeared in print in an article by the American mathematician, Claude Elwood Shannon (1916–2001) (Shannon 1948), see also Koons and Lubkin (1949) and Shaw (1950). A **series** of 8 bits linked together are referred to as a **byte** (Buchholz 1981). It is mentioned in Davis and Sampson (1973).

Binary notation The representation of **integer** numbers in terms of powers of two, using only the digits 0 and 1. The position of the digits corresponds to the successive powers, e.g. in binary arithmetic: $0 + 0 = 0$, $0 + 1 = 1$, $1 + 0 = 1$, $1 + 1 = 10$; decimal $2 = 0010$, decimal $3 = 0011$, decimal $4 = 0100$, etc. and, e.g., decimal $23 =$ decimal $16 + 4 + 2 + 1$, i.e. $10000 + 00100 + 00010 + 00001 = 10111$ in *binary notation*. Although it has been asserted (Leibniz 1703, 1768) that binary arithmetic may have been used in the Chinese *I-king* [*Book of permutations*] which is believed to have been written by the Chinese mystic, Wön-wang (1182–1135 BC), it "has no historical foundation in the *I-king* as originally written" (Smith 1923–1925, I, 25). Binary arithmetic was discussed by the German mathematician and philosopher, Gottfried Wilhelm von Leibniz (1646–1716), (Leibniz 1703). In computing, the binary numbering system was used in a report (von Neumann 1945) on the EDVAC (*Electronic Discrete Variable Automatic* ***Computer***), developed under J. Presper Eckert (1919–1995) and John Mauchly (1907–1980) at the Eckert-Mauchly Computer Corporation, USA, in 1946 and its successor, Binac (Binary Automatic Computer) (Eckert-Mauchly Computer Corp. 1949). **Statistical model**s for the analysis of presence/absence data, often coded as $\{1,0\}$ values, were reviewed by Cox (1970). Binary notation in an earth science context is discussed by Ramsayer and Bonham-Carter (1974), who consider the **classification** of petrographical and palaeontological data when represented by strings of **binary variables**. See also: Sheriff (1984) and Camina and Janacek (1984); **binary coefficient**.

Binary variable A **variable** which may take one of only two discrete values, e.g. the presence of a particular lithology in a **map cell** might be coded: absent = 0, present = 1. Statistical methods for the analysis of such data were reviewed by Cox (1970). See also: Ramsayer and Bonham-Carter (1974); **binary coefficient**.

Bingham distribution A spherical **frequency distribution** first studied by the American statistician, Christopher Bingham (1937–) in 1964, but only published some years later (Bingham 1964, 1974). It is the distribution of a **trivariate vector** of **normal distribution**s, all with zero **mean** and an arbitrary **covariance matrix**, **C**, given that the length of the vector is unity. If the **random** vector is $\boldsymbol{x} = (x_1, x_2, x_3)$, the **probability distribution**, is given by:

$$f(\boldsymbol{x}; \boldsymbol{m}, \boldsymbol{k}) = \frac{1}{4\pi d(k)} e^{\left\{k_1(\boldsymbol{x}_1 m_1)^2 + k_2(\boldsymbol{x}_2 m_2)^2 + k_3(\boldsymbol{x}_3 m_3)^2\right\}}$$

where $\boldsymbol{k} = (k_1, k_2, k_3)$ is a **matrix** of constants, known as the concentrations; m_1, m_2 and m_3 are three **orthogonal** normalised vectors, the **principal axes**; $\boldsymbol{m} = (m_1, m_2, m_3)$; and $d(k)$ is a constant which depends only on k_1, k_2 and k_3 and e is **Euler's number**, the constant 2.71828. . . See Mardia (1972), Fisher et al. (1993), and Mardia and Jupp (2000) for further discussion. This distribution was popularised for use with paleomagnetic data by Onstott (1980). For other earth science applications, see: Kelker and Langenberg (1976) and Cheeney (1983). See also: *e*, **spherical statistics**, **Fisher distribution**, **Kent distribution**.

Binomial distribution, binomial model, binomial probability If p is the **probability** of an **event** occurring one way (e.g. a "success") and q is the probability of it occurring in an alternative way (e.g. a "failure") then $p + q = 1$, and p and q remain constant in n **independent** trials, then the **probability distribution** for x individuals occurring in a **sampling** unit is:

$$P(x; n, p) = \left[\frac{n!}{x!(n-x)!}\right] q^{n-x} p^n$$

where x is the number of individuals per sampling unit; and $k!$ means k **factorial**. The **arithmetic mean** is kp and the **standard deviation** is $\sqrt{kpq}$. Knowledge of this **distribution** goes back to the eighteenth Century, but the term *binomial* was introduced by the British statistician, George Udney Yule (1871–1951) (Yule 1911). For discussion in an earth science context, see: Miller and Kahn (1962), Koch and Link (1970–1971), Vistelius (1980, 1992), Agterberg (1984a) and Camina and Janacek (1984). See also **trinomial distribution**.

Biochronologic correlation A method of **correlation** between two or more spatial positions based on the dates of first and last appearances of taxa, reaching a particular evolutionary state, etc. For general reviews, see: Hay and Southam (1978) and Agterberg (1984c, 1990). See also: **biostratigraphic zonation**, **correlation and scaling**, **ranking and scaling**, **unitary associations**.

Biofacies map A **map** showing the areal distribution in the biological **composition** of a given stratigraphic unit based on quantitative measurements, expressed as **percentage**s of the types of group present (e.g. brachiopods, pelecypods, corals, etc.). The American mathematical geologist, William Christian Krumbein (1902–1979) and Laurence

Louis Sloss (1913–1996) used **isoline**s to portray the **ratio** of cephalopods/(gastropods + pelecypods) in the Mancos Shale of New Mexico (Krumbein and Sloss 1951). See also: **lithofacies map**.

Biometrical methods, biometrics Statistical and mathematical methods developed for application to problems in the biological sciences have long been applied to the solution of palaeontological problems. The term has been in use in the biological sciences since at least the 1920s, e.g. Hartzell (1924). The journal *Biometrics* began under the title *Biometrics Bulletin* in 1945 but changed to the shorter title in 1947 when the Biometrics Society became established in the USA under the Presidency of the English statistician, (Sir) Ronald Alymer Fisher (1890–1962), and a British "region" followed in 1948. Important early studies include those by the American palaeontologists, Benjamin H. Burma (1917–1982), followed by those of Robert Lee Miller (1920–1976) and Everett Claire Olsen (1910–1993) and by the English vertebrate palaeontologist, Kenneth A. Kermack (1919–2000) (Burma 1948, 1949, 1953; Miller 1949; Olsen and Miller 1951, 1958; Kermack 1954). The American geologist, John Imbrie (1925–) commented (Imbrie 1956) on the slowness with which palaeontologists were taking up such methods and he promoted the use of **reduced major axis regression** (Jones 1937), introduced into palaeontology by Kermack's (1954) study, while regretting (in the pre-**computer** era) that practicalities limited such studies to the use of one- or two-**dimension**al methods. In later years, they embraced **multivariate** techniques such as **principal components analysis**, **nonlinear mapping** and **correspondence analysis** (Temple 1982, 1992). See also: Sepkoski (2012); **biochronologic correlation**.

Biostratigraphic zonation A biostratigraphic *zone* is a general term for any kind of biostratigraphic unit regardless of its thickness or geographic extent. Use of microfossils as an aid to stratigraphic zonation in the petroleum industry dates from about 1925, and **graphical** depiction of microfossil assemblage **abundance**s as a **function** of stratigraphic unit position in a succession has been in use since at least the 1940s (Ten Dam 1947; LeRoy 1950a). Methods for achieving quantitative stratigraphic zonation are discussed in Hay and Southam (1978), Cubitt and Reyment (1982), Gradstein et al. (1985), Hattori (1985) and Agterberg (1984c, 1990).

Biphase The **phase** relationship of two **nonlinear**ly related **frequency** components. The term was introduced into **digital signal processing** by the American statistician, John Wilder Tukey (1915–2000) (Tukey 1953). See: **bispectrum**. See also: Brillinger (1965), Brillinger and Rosenblatt (1967a, 1967b) and Brillinger and Tukey (1985); and in an earth science context: Elgar and Sebert (1989), King (1996) and Weedon (2003).

Biplot, Gabriel biplot **Graphical** display of the rows and columns of a rectangular $n \times p$ data **matrix X**, where the rows generally correspond to the **sample compositions**, and the columns to the **variables**. In almost all applications, *biplot* analysis starts with performing some **transformation** on **X**, depending on the nature of the data, to obtain a transformed matrix ***Z***, which is the one that is actually displayed. The graphical representation is based on a **singular value decomposition** of matrix **Z**. There are essentially two different biplot representations: the **form biplot**, which favours the display of individuals (it does not represent the **covariance** of each variable, so as to better represent the natural form of the **data set**), and the **covariance biplot**, which favours the display of the variables (it preserves the covariance structure of the variables but represents the samples as a spherical cloud). Also known as the *Gabriel biplot*, named for the German-born statistician, Kuno Ruben Gabriel (1929–2003) who introduced the method (Gabriel 1971). See also: Greenacre and Underhill (1982), Aitchison and Greenacre (2002); and, in an earth science context, Buccianti et al. (2006).

Bispectral analysis, bispectrum The *bispectrum*, $B(f_1, f_2)$ of a **time series** measures the statistical dependence between three **frequency band**s centred at f_1, f_2, and $f_1 + f_2$: $B(f_1, f_2) = E[P(f_1)P(f_2)P^*(f_1 + f_2)]$, where $P(f)$ is the complex **Fourier transform** of the time series at **frequency** f; E(•) is the **expectation operator**; and $P^*(f)$ is the **complex conjugate**. Each **band** will be characterised by an **amplitude** and **phase**. If the sum or difference of the phases of these bands are statistically **independent**, then on taking the **average**, the bispectrum will tend to zero as a result of **random** phase mixing; but if the three frequency bands are related, the total **phase** will not be random (although the phase of each band may be randomly changing) and averaging will yield a peak at $\{f_1, f_2\}$ on a **graph** of $B(f_1, f_2)$ as a **function** of f_1 and f_2. The term was introduced by the American statistician, John Wilder Tukey (1915–2000) in an unpublished paper (Tukey 1953). See also: Tukey (1959b), Mendel (1991) and Nikias and Petropulu (1993); and, in an earth science context: Haubrich (1965), Hagelberg et al. (1991), Rial and Anaclerio (2000), Persson (2003) and Weedon (2003). See also: **bicoherence**.

bit An acronym for *bi*nary digi*t*. Coined by the American statistician, John Wilder Tukey (1915–2000) about 1946, because the two states of an element in a **computer** core can represent one digit in the **binary** representation of a number. In the binary system, representation of **integer** numbers is in terms of powers of two, using only the digits 0 and 1. The position of the digits corresponds to the successive powers. e.g. in binary arithmetic $0 + 0 = 0$, $0 + 1 = 1$, $1 + 0 = 1$, $1 + 1 = 10$; decimal $2 = 0010$, decimal $3 = 0011$, decimal $4 = 0100$, etc. and, e.g., decimal $23 =$ decimal $16 + 4 + 2 + 1$, i.e. $10000 + 00100 + 00010 + 00001 = 10111$ in *binary notation.* It first appeared in print in an article by the American mathematician, Claude Elwood Shannon (1916–2001) (Shannon 1948). A **series** of 8 bits linked together are referred to as a **byte**. Mentioned in Davis and Sampson (1973).

Bit-map A **set** of **bit**s that represent an image. Armstrong and Bennett (1990) describe a classifier for the detection of trends in hydrogeochemical **parameter**s as a **function** of time, based on the conversion of concentration-time curves into bit-strings.

Bivariate, bivariate frequency distribution

1. The term *bivariate* is used in the context of the analysis of data in which each observation consists of values from two **variables**. It came into usage following its use by the British statistician, Karl Pearson (1857–1936) (Pearson 1920).
2. A *bivariate frequency distribution* is the **probability distribution** corresponding to the simultaneous occurrence of any pair of values from each of two variables (x and y). It shows not only the **univariate** frequency distributions for x and y, but also the way in which each value of y is distributed among the values of x and *vici-versa*. It is also known as a two-way or **joint** frequency distribution. The distribution of the "joint chance" involving two variables was discussed by the British mathematician, mathematical astronomer and geophysicist, (Sir) Harold Jeffreys (1891–1989) (Jeffreys 1939). However, bivariate frequency distributions were actually used earlier in geology, in an empirical fashion, by the French mathematician and cataloguer of earthquakes, Alexis Perrey (1807–1882) (Perrey 1847) and subsequently by Alkins (1920) and Schmid (1934); see also Miller and Kahn (1962), Smart (1979), Camina and Janacek (1984) and Swan and Sandilands (1995); **joint distribution**, **multivariate**.

Black box A **conceptual model** which has input **variables**, output variables and behavioural characteristics, but without specification of internal structure or mechanisms explicitly linking the input to output behaviours. The term is used to describe an element in a statistical **model** which contains features common to most techniques of statistical inference and in which only the input and output characteristics are of interest, without regard to its internal mechanism or structure. Although attributed to the Canadian statistician, Donald Alexander Stuart Fraser (1925–) (Fraser 1968), the term was previously used by the American statistician, John Wilder Tukey (1915–2000) in a geophysical context (Tukey 1959a). For discussion in geoscience applications see: Griffiths (1978a, 1978b), Kanasewich (1981), Tarantola (1984), Spero and Williams (1989), Gholipour et al. (2004), Jiracek et al. (2007) and Cabalar and Cevik (2009).

Black noise Coloured (American English sp. colored) **noise** can be obtained from **white noise** by passing the **signal** through a **filter** which introduces a degree of **autocorrelation**, e.g. $x(t) = ax(t-1) + kw(t)$ where $w(t)$ is a white noise signal; a is a constant, $0 < a < 1$; k is the **gain**, and $x(t)$ is the output signal at time t. The **power spectrum** density for *black noise* is either characterised by predominantly zero power over most frequency ranges, with the exception of a few narrow **spike**s or **band**s; or increases **linear**ly as f_p, $p > 2$. The concept of white light as having a uniform power density over its spectrum was first discussed by the American mathematician, Norbert Wiener (1894–1964) (Wiener 1926), and taken up in **digital signal processing** by the American mathematician Richard Wesley Hamming (1915–1998) and statistician John Wilder Tukey (1915–2000) (Tukey and Hamming

1949); see also Blackman and Tukey (1958). For discussion in an earth science context, see Weedon (2003).

Blackman-Harris window, Blackman-Harris taper Used in the operation of **smoothing** a **periodogram** with a **lag window** of **weight**s applied to a discrete time **waveform**. N, the length of the window is typically even and an **integer** power of 2; for each point, $n = 0, \ldots, N$. The weight for a four-term window is given by

$$w(n) = 0.35875 - 0.48829 \cos\left(\frac{2\pi n}{N}\right) + 0.14128 \cos\left(\frac{4\pi n}{N}\right) - 0.01168 \cos\left(\frac{6\pi n}{N}\right); \text{where}$$

$n = 0, 1, 2, \cdot \ldots, (N - 1)$. Named for the American communications engineer, Ralph Beebe Blackman (1904–1990) (Blackman and Tukey 1958) and signal processing and communications specialist, Frederic J. Harris (1940–). Use of this window (Gubbins 2004) was introduced by Harris (1976) and subsequently became more widely known through industrial taught-courses (Harris 1977) and publication (Harris 1978; Rabiner et al. 1970). *Window* seems to be the preferred usage over *taper* (Google Research 2012). See also: **Bartlett window**, **boxcar taper**, **cosine taper**, **Daniell window**, **data window**, **Gaussian taper**, **Hamming window**, **Hann window**, **multi-tapering method**, **optimal taper**, **Parzen window**, **Thomson tapering**.

Blackman-Tukey method, Blackman-Tukey spectrum estimation Named for the American communications engineer, Ralph Beebe Blackman (1904–1990) and statistician, John Wilder Tukey (1915–2000) who introduced it (Blackman and Tukey 1958), this method of **power spectral density analysis** is based on the **Fourier transform** of the smoothed **autocovariance function**, which has been computed for **lag**s up to a certain value (the **truncation point**), so as to eliminate the most **noisy** values (which are based on only a small number of data) prior to the Fourier transform. The results were shown in one study (Edmonds and Webb 1970) to be similar in practice to those obtained using the **Fast Fourier transform** (FFT) method, although the latter was found to be superior from the point of view of flexibility of use and computation time. For discussion in an earth science context, see Buttkus (1991, 2000) and Weedon (2003); see also: **mean lagged product**.

Blake's method A method for determining the **ellipticity** (**strain ratio**) from measurements of the pressure-deformed spiral logarithmic growth curve in ammonites, goniatites and cephalopods. Named for the British geologist, John Frederick Blake (1839–1906) (Blake 1878). Mentioned in Ramsay and Huber (1983).

Blank

1. In analytical geochemistry, a dummy sample which has a chemical **composition** designed to contain a "zero" quantity of an analyte of interest. The term was in use in this sense in geochemistry by the early 1900s (Strutt 1908; Holmes 1911).
2. In geophysics, to replace a value by zero (Sheriff 1984).

Blind source separation, blind signal separation More usually known as **Independent Component Analysis**, this is a technique based on information theory, originally developed in the context of **signal** processing (Hérault and Ans 1984; Jutten and Hérault 1991; Comon 1994; Hyvärinen and Oja 2000; Hyvärinen et al. 2001; Comon and Jutten 2010) intended to separate **independent** sources in a **multivariate time series** which have been mixed in **signal**s detected by several sensors. After **whitening** the data to ensure the different channels are **uncorrelated**, they are rotated so as to make the **frequency distribution**s of the points projected onto each **axis** as near uniform as possible. The source signals are assumed to have non-**Gaussian probability distribution functions** and to be statistically independent of each other. Unlike **principal components analysis** (PCA), the **axes** do not have to be **orthogonal**, and **linearity** of the **mixture model** is not required. ICA extracts statistically independent components. Ciaramella et al. (2004) and van der Baan (2006) describe its successful application to seismic data. *Blind source separation* appears to be the most frequent usage (Google Research 2012).

Block averaging A technique for **smoothing** spatial distribution patterns in the presence of highly erratic **background** values, using the **mean** values of non-overlapping blocks of fixed size so as to enhance the presence of, for example, mineralized zones (Chork and Govett 1979).

Block diagram This is typically an oblique pseudo three-dimensional view of a **gridded** (**contour**ed) surface with cross-sectional views of two of its sides. It has its origins in diagrams to illustrate geological structure. Early examples were produced as a by-product in **computer mapping** packages such as SURF (Van Horik and Goodchild 1975) and SURFACEII (Sampson 1975).

Block matrix This is a **matrix** which is subdivided into sections called *blocks*. Each block is separated from the others by imaginary horizontal and vertical lines, which cut the matrix completely in the given direction. Thus, the matrix is composed of a **series** of smaller matrices. A block **Toeplitz matrix**, in which each block is itself a Toeplitz matrix, is used in Davis (1987b). It is also known as a **partitioned matrix**, but the term *block matrix* has become the more widely used since the 1990s (Google Research 2012).

Block model A method of modelling, say, a mineral deposit, by its representation as a **grid** of three-**dimension**al blocks. One approach is to use equal sized ("fixed") blocks.

Dunstan and Mill (1989) discuss the use of the **octree** encoding technique to enable blocks of different sizes to be used so as to better **model** the topography of the spatial boundary of the deposit by enabling the use of progressively finer **resolution** blocks as it is approached.

Blue noise Coloured [U.S. spelling, colored] **noise** can be obtained from **white noise** by passing the **signal** through a **filter** which introduces a degree of **autocorrelation**, e.g., $x(t) = ax(t-1) + kw(t)$, where $w(t)$ is a white noise signal; a is a constant, $0 < a < 1$; k is the **gain**, and $x(t)$ is the output signal at time t. The **power spectrum** density for *blue* (or **azure**) noise increases **linear**ly as f. The concept of white light as having a uniform power density over its spectrum was first discussed by the American mathematician, Norbert Wiener (1894–1964) (Wiener 1926), and taken up in **digital signal processing** by the American mathematician Richard Wesley Hamming (1915–1998) and statistician John Wilder Tukey (1915–2000) (Tukey and Hamming 1949); see also Blackman and Tukey (1958). For discussion in an earth science context, see Weedon (2003).

Bochner's theorem This theorem, used in Armstrong and Diamond (1984), is named for the American mathematician of Austro-Hungarian origin, Salomon Bochner (1899–1982). It characterizes the **Fourier transform** of a positive finite **Borel measure** on the real line: every positive definite **function** Q is the Fourier transform of a positive finite Borel measure.

Bochner window This is another name for a window named after the Austro-Hungarian-American mathematician, Salomon Bochner (1899–1982) used in the operation of **smoothing** a **periodogram** with a **lag window** of **weight**s applied to a discrete time **signal** (Parzen 1957, 1961). N, the length of the window is typically even and an **integer** power of 2; for each point $0 \le n \le N-1$, the weight is given by:

$$w(n) = \begin{cases} 1 - 6\left(\frac{n-N/2}{N/2}\right)^2 + 6\left(\frac{|n-N/2|}{N/2}\right)^3; 0 \le \left|n - \dfrac{N}{2}\right| \le \dfrac{N}{4} \\ 2\left(1 - \frac{|n-N/2|}{N/2}\right)^3; \dfrac{N}{4} < \left|n - \dfrac{N}{2}\right| \le \dfrac{N}{2} \end{cases}.$$

It is also named for the American statistician, Emanuel Parzen (1929–2016). Parzen (1962) applied a similar technique to estimation of a **density trace**. It is also known (Harris 1978) as the **Riesz window**. See also: Preston and Davis (1976), Buttkus (1991, 2000); **spectral window**.

Body rotation, body translation *Body rotation*: When a body moves as a rigid mass by **rotation** about some **fixed point**. *Body translation*: When a body moves without **rotation** or internal distortion. Both terms were used by Thomson and Tait (1878) and popularised in geology through the work of the English geologist, John Graham Ramsay (1931–) (1967, 1976). See also: Hobbs et al. (1976) and Ramsay and Huber (1983).

Boltzmann-Hopkinson theorem **Convolution** is the **integral** from $i = 0$ to t of the product of two **function**s, $\int_0^t f_{1i} f_{2t-i} \mathrm{d}x$. For two equal-interval discrete **time series** $a = \{a_0, a_1, a_2, ..., a_n\}$ and $b = \{b_0, b_1, b_2, ..., b_n\}$, the convolution, usually written as a^*b or $a \otimes b$, is $c = \{c_0, c_1, c_2, ..., c_n\}$, where

$$c_t = \sum_{i=0}^{t} a_i b_{t-i}.$$

The operation can be imagined as sliding *a* past *b* one step at a time and multiplying and summing adjacent entries. This type of **integral** was originally used by the French mathematician, Pierre Simon, Marquis de Laplace (1749–1827), (Laplace 1781). The Hungarian-born American mathematician, Aurel Friedrich Wintner (1903–1958) may have been the first to use the English term **convolution** (Wintner 1934), although its German equivalent *Faltung* (*folding*, referring to the way in which the coefficients may be derived from cross-multiplication of the *a* and *b* terms and summation of their products along diagonals if they are written along the margins of a square table) appeared in Wiener (1933). The operation has also been referred to as the **Boltzmann-Hopkinson theorem**, **Borel's theorem**, **Duhamel's theorem**, **Green's theorem**, *Faltungsintegral*, and the **superposition theorem** and a similar result may also be achieved in terms of ***z*-transform**s or **Fourier transform**s. It can also be applied in more than two **dimension**s (see: **helix transform**). See also: Tukey and Hamming (1949), Blackman and Tukey (1958), and in an earth science context: Robinson (1967b), Jones (1977), Vistelius (1980, 1992), Camina and Janacek (1984), Buttkus (1991, 2000) and Gubbins (2004); **deconvolution**.

Boolean algebra A version of standard algebra introduced by the British mathematician George Boole (1815–1864) (Boole 1854), based solely on use of the **integer** values zero (false) and unity (true). The usual algebraic operations of addition ($x + y$), multiplication (xy), and negation ($-x$) are replaced by the **operator**s: OR (disjunction, equivalent to the arithmetic result xy), AND (conjunction, equivalent to $x + y - xy$), and NOT (negation or compliment, equivalent to $1 - x$). Mentioned in an earth science context by Vistelius (1972).

Boolean similarity matrix This similarity criterion is named for the George Boole (1815–1864), a British mathematician who pioneered the use of **binary** logic in problem solving (Boole 1854). Each **attribute** (e.g. the occurrence of *n* indicator mineral species, at *m* mineralised districts to be compared), is coded as either zero for "absent" or unity for "present." The resultant *m* (row) × *n* (column) data **matrix** (**M**) is multiplied by its $n \times m$ **transpose** ($\mathbf{M}^{\mathrm{T}}$) to form a product matrix (**P**). The **square root**s of the sums of squares of the elements of the rows of **P** were called the *mineral typicalities* by the American geologist, Joseph Moses Botbol (1937–) (Botbol 1970). See also **characteristic analysis**.

Booton integral equation The American mathematician, Norbert Wiener (1894–1964) and Austrian-born American mathematician, Eberhard Frederich Ferdinand Hopf (1902–1983), who worked with Wiener at the Massachusetts Institute of Technology (1931–1936), devised a method for the solution of a class of **integral** equations of the form:

$$f(x) = \int_0^\infty k(x-y)f(y)dy,$$

where $x \geq 0$ (Wiener and Hopf 1931; Wiener 1949; Widom 1997). The solution for the **non-stationary** case was developed by American electrical engineer Richard Crittenden Booton Jr. (1926–2009) in the context of prediction of **random signal**s and their separation from **random noise** (Booton 1952). The objective is to obtain the specification of a **linear dynamical system** (**Wiener filter**) which accomplishes the prediction, separation, or detection of a random signal. For discussion in a geophysical context, see Buttkus (1991, 2000).

Bootstrap A technique which involves **computer**-intensive re**sampling** of a **data set**, in order to obtain **nonparametric estimate**s of the **standard error** and **confidence interval** for **median**s, **variance**s, **percentile**s, **correlation** and **regression coefficient**s etc. It is based on repeatedly drawing at **random**, with replacement, a set of *n* **sample**s from a pre-existing set of data values and determining the required statistics from a large number of trials. It was introduced by the American statistician, Bradley Efron (1938–) (Efron 1979; Efron and Tibshirani 1993). Examples of earth science applications include: Solow (1985), Campbell (1988), Constable and Tauxe (1990), Tauxe et al. (1991), Joy and Chatterjee (1998), Birks et al. (1990), Birks (1995) and Caers et al. (1999a,b); see also: **cross-validation**, **jackknife**.

Borehole log, well log A **graphical** or **digital** record of one or more physical measurements (or quantities derived from them) as a **function** of depth in a borehole; also known as a *well log* or *wireline log*, as they are often derived from measurements made by a instruments contained in a sonde which is lowered down the borehole (Nettleton 1940; LeRoy 1950b). The first geophysical log ("electrical coring") was made by Henri Doll (1902–1991), Roger Jost and Charles Scheibli over a 5 h period on September 5, 1927, in the Diefenbach Well 2905, in Pechelbronn, France, over an interval of 140 m, beginning at a depth of 279 m, using equipment designed by Doll following an idea for *Recherches Électriques dans les Sondages* [Electrical research in boreholes] outlined by Conrad Schlumberger (1878–1936) in a note dated April 28, 1927 (Allaud and Martin 1977, 103–108). The unhyphenated *well log* appears to be by far the most frequent usage (Google Research 2012).

Borel algebra, Borel measure The *Borel algebra* over any topological space is the **sigma algebra** generated by either the **open set**s or the **closed set**s. A **measure** is defined on the

sigma algebra of a topological space onto the **set** of **real number**s ($\mathbb{R}$). If the **mapping** is onto the interval [0, 1], it is a *Borel measure*. Both are named for the French mathematician, Félix Edouard Justin Émile Borel (1871–1956) and are mentioned in an earth science context by Vistelius (1980, 1992).

Borel's theorem **Convolution** is the **integral** from $i = 0$ to t of the product of two **function**s,

$$\int_0^t f_{1_i} f_{2_{t-i}} dx.$$

For two equal-interval discrete **time series** $a = \{a_0, a_1, a_2, \ldots, a_n\}$ and $b = \{b_0, b_1, b_2, \ldots, b_n\}$, the convolution, usually written as a^*b or $a \otimes b$, is $c = \{c_0, c_1, c_2, \ldots, c_n\}$, where

$$c_t = \sum_{i=0}^{t} a_i b_{t-i}.$$

The operation can be imagined as sliding a past b one step at a time and multiplying and summing adjacent entries. This type of **integral** was originally used by the French mathematician, Pierre Simon, Marquis de Laplace (1749–1827), (Laplace 1781). The Hungarian-born American mathematician, Aurel Friedrich Wintner (1903–1958) may have been the first to use the English term **convolution** (Wintner 1934), although its German equivalent *Faltung* (*folding*, referring to the way in which the coefficients may be derived from cross-multiplication of the a and b terms and summation of their products along diagonals if they are written along the margins of a square table) appeared in Wiener (1933). The operation has also been referred to as the **Boltzmann-Hopkinson theorem**, **Duhamel's theorem**, **Green's theorem**, *Faltungsintegral*, and the **superposition theorem** and a similar result may also be achieved in terms of ***z*-transform**s or **Fourier transform**s. It can also be applied in more than two **dimension**s (see: **helix transform**). See also: Tukey and Hamming (1949) and Blackman and Tukey (1958), and in an earth science context: Robinson (1967b), Jones (1977), Vistelius (1980, 1992), Camina and Janacek (1984), Buttkus (1991, 2000) and Gubbins (2004); **deconvolution**.

Boundary condition A **constraint** that a **function** must satisfy along a boundary. Knopoff (1956) and Cheng and Hodge (1976) are early examples of usage in geophysics and geology respectively.

Boundary value problem Solution of a **differential equation** with **boundary condition**s. The term was used in mathematics in Birkhoff (1908). Wuenschel (1960) and Cheng and Hodge (1976) are early examples of usage in geophysics and geology respectively.

B

Box-Cox transform A general method of **transformation** of a skewed (asymmetrical) **frequency distribution** into one which is more symmetrical, for the purposes of statistical analysis:

$$x^* = \begin{cases} \dfrac{x^{\lambda-1}}{\lambda}; \lambda \neq 0 \\ log_e(\lambda); \lambda = 0 \end{cases}$$

where e is **Euler's number**, the constant 2.71828... In practice, the value of λ is determined empirically so that it minimises one or more measures of the **asymmetry** of the distribution (e.g. **skewness**). Introduced by the British-born American chemist and mathematician, George Edward Pelham Box (1919–2013) and statistician, (Sir) David Roxbee Cox (1924–) (Box and Cox 1964); it is also known as the **power transformation**. Introduced into geochemical usage by Howarth and Earle (1979), its usage has been further developed by Joseph and Bhaumik (1997) and Stanley (2006a,b).

Box-count dimension This is a popular term for an estimator of **fractal dimension** (D; > 0) for a two-**dimension**al spatial point pattern. The area occupied by the **set** of points is covered with a square **mesh** of **cell**s, beginning with one of diameter d, sufficient to cover the whole of the area occupied by the point set. The mesh size is then progressively decreased, and the number of occupied cells, $N(d)$, at each size step is counted. Then, $N(d) = cd^{-\mathrm{D}}$, where c is a constant; a **graph** of log[$N(d)$] (y-**axis**) as a **function** of log(d) (x-**axis**) will be **linear** with a **slope** of –D. This is more properly known as the **Minkowski** or **Minkowski-Bouligand dimension**, named after the Russian-born German mathematician, Hermann Minkowski (1864–1909) and the French mathematician, Georges Louis Bouligand (1889–1979). See: Minkowski (1901), Bouligand (1928, 1929), Mandelbrot (1975a, 1977, 1982), Turcotte (1997) and Kenkel (2013) for a cautionary note on **sample**-size requirements for such dimensionality estimation methods. Taud and Parrot (2005) discuss methods applied to topographic surfaces. See also **Richardson plot**.

Box-Jenkins process A **stationary process** in which the value of a **time series** at time t is **correlated** in some way with the value(s) in the previous time steps. An **autoregressive moving average process**, ARMA(p, q) is:

$$x_t - m = \varphi_1(x_{t-1} - m) + \varphi_2(x_{t-2} - m) + \ldots + \varphi_p\left(x_{t-p} - m\right) + \varepsilon_t - \theta_1\varepsilon_{t-1}$$
$$-\theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q}$$

where m is the **mean** level; ε is a **white noise** process with zero mean and a finite and constant **variance**; φ_i, $i = 1, p$ and θ_j, $j = 1, q$ are the **parameter**s; and p, q are the **order**s. To obey the assumption of **stationarity**, the **absolute value**s of φ_1 and θ_1 should be less than unity. The basic idea was introduced by the Swedish statistician, Herman Ole Andreas Wold (1908–1992) (Wold 1938), and later developed by the British-born

American chemist and mathematician, George Edward Pelham Box (1919–2013) and statistician, Gwilym Meirion Jenkins (1933–1982) (Box and Jenkins 1970). For discussion in an earth science context, see: Camina and Janacek (1984), Sarma (1990), Buttkus (1991, 2000) and Weedon (2003); see also: **autoregressive process**.

Boxcar distribution A **probability density** in which the **probability** of occurrence of the value of a **variable** $f(x)$ is the same for all values of x lying between $x_{\min}$ and $x_{\max}$ inclusive and zero outside that **range** (Vistelius 1980, 1992; Feagin 1981; Camina and Janacek 1984). The distribution is named after the shape of a "boxcar" railway freight waggon, a term which has been used in U.S. English since at least the 1890s. It is also known as the **rectangular** or **uniform distribution**.

Boxcar taper, boxcar weighting function, boxcar window The *boxcar taper* or *window* (Blackman and Tukey 1958; Alsop 1968), is named after the shape of a "boxcar" railway freight waggon, a term which has been used in American English since at least the 1890s, and is used in the operation of **smoothing** a **periodogram** with a **lag window** of **weight**s applied to a discrete time **waveform**. N, the **half-width** of the window is typically even and an **integer** power of 2; for each point within $0 \le n \le N - 1$, the weight $w(n) = 1$, otherwise it is zero. Its shape contrasts with that of the smoothly changing weights in windows which are **tapered**. It is also known as a **Daniell window** (Blackman and Tukey 1958); **rectangular window** (Harris 1978); and **Dirichlet window** (Rice 1964; Harris 1978); see also: Camina and Janacek (1984) and Gubbins (2004).

Boxplot A **graphical** display, originally devised by the American statistician, John Wilder Tukey (1915–2000) (Tukey 1977; McGill et al. 1978), which is extremely useful for the simultaneous comparison of a number of **frequency distribution**s (e.g. concentrations of a trace element in a number of different **sample**d rock types). For each **set** of data, the top and bottom of a central "box" are given by the first and third **quartiles** (Q_1, Q_3), so the rectangle formed by the box (which is conventionally drawn parallel to the vertical **axis**, corresponding to increasing magnitude of the **variable** studied) encloses the central 50% of the frequency distribution. The position of the second quartile (the **median**) is shown by a horizontal line dividing the box. In the most useful **graph**, so-called *whiskers* are drawn outwards from the top and bottom of the box to the smallest data value lying within Q_1 and $Q_1 - 1.5R$, where $R = Q_3 - Q_1$; or to the largest data value lying within Q_3 and $Q_3 + 1.5R$; and any "further out" data values are deemed to be **outlier**s and are plotted individually. Less informative plots are produced by simply extending the whiskers out to the maximum and minimum of the data values. In a multi-group comparison, box-width can be made proportional to the **sample size** of each group. See Helsel (2005) for discussion of treatment of data containing **nondetect**s. Although the spelling *box-and-whisker-plot* was originally used, the contractions *boxplot* or *box plot* now appear to be equally frequent (Google Research 2012). See also: **notched boxplot** and Chambers et al. (1983), Kurzl (1988), Frigge et al. (1989), Helsel and Hirsch (1992) and Reimann et al. (2008) for examples of usage.

Branching process A **Markov process** that models a **population** in which each individual in generation n produces some **random** number of offspring in generation (n + 1), according to a fixed **probability distribution** which does not vary from individual to individual. The lines of descent "branching out" as new members are born. It has been applied to the study of the evolution of **population**s of individuals who reproduce independently. The mathematical problem was originally solved by the French statistician, Irénée-Jules Bienaymé (1796–1878), who published, without proof (Bienaymé 1845), the statement that eventual extinction of a family name would occur with a probability of one if, and only if, the **mean** number of male children is less than or equal to one (Heyde and Seneta 1977). The topic was revisited with the work of the English statistician, (Sir) Francis Galton (1822–1911) and mathematician, Rev. Henry William Watson (1827–1903) (Galton and Watson 1874; Watson and Galton 1875). Modern work began in the 1940s (Kolmogorov and Dmitriev 1947, 1992; Harris 1963; Jagers 1975). Discussed in the context of earthquake-induced crack-propagation in rocks by Vere-Jones (1976, 1977). See also Turcotte (1997).

Bray-Curtis coefficient A **measure** of the similarity of one sample to another in terms of their p-**dimension**al **composition**s. Given two samples j and k and **percentage**s of the i-th **variable** (e.g. in ecological or paleoecological studies, species abundance) in each sample the *Bray-Curtis* metric, named for American botanists and ecologists, J. Roger Bray (1929–) and John T. Curtis (1913–1961), is:

$$d_{jk}^{BC} = \left\{ \frac{2\sum_{i=1}^{p} \min\left(x_{ij}, x_{ik}\right)}{\sum_{i=1}^{p} \left(x_{ij} + x_{ik}\right)} \right\}$$

where min() implies the minimum of the two counts where a species is present in both samples (Bray and Curtis 1957). In their usage, the data were first **normalized** by dividing the percentages for each species by the maximum attained by that species over all samples. However, Bray and Curtis attribute this formulation to Motyka et al. (1950) and Osting (1956). Use of the minimum abundance alone was proposed as an "index of affinity" by Rogers (1976). An alternative measure:

$$d_{jk}^{S} = 100\left\{ 1 - \frac{\sum_{i=1}^{p} \left|x_{ij} - x_{ik}\right|}{\sum_{i=1}^{p} \left(x_{ij} + x_{ik}\right)} \right\},$$

where the difference without regard to sign (the absolute difference) replaces the minimum, has been used in Stephenson and Williams (1971) and later studies, but use of this measure has been criticised by Michie (1982). See also the comments by Somerfield (2008).

Breakage model, breakage process Theoretical **statistical model**s for the size **frequency distribution** which results from progressive breakage of a single homogeneous piece of material. First discussed by the Russian mathematician, Andrey Nikolaevich Kolmogorov (1903–1987), (Kolmogorov 1941a, 1992) the result of a *breakage process* (Halmos 1944; Epstein 1947) yielded size distributions which followed the **lognormal distribution**, but it was subsequently found that this **model** may not always fit adequately. Applied to consideration of the comminution of rocks, minerals and coal, see Filippov (1961) and more recently discussed in connection with the formation of the lunar regolith (Marcus 1970; Martin and Mills 1977). See the discussion in the context of particle-size distribution by Dacey and Krumbein (1979); see also: **Rosin's law**, **Pareto distribution**.

Breakpoint The point at which a statistically significant change in **amplitude** in the **mean** and/or **variance** of a **time series** occurs, indicating a change in the nature of the underlying process controlling the formation of the time series. Generally detected by means of a graph of the cumulative sum of mean and/or variance as a **function** of time (Montgomery 1991a) in which **changepoint**s are indicated by a statistically significant change in **slope**, e.g. Green (1981, 1982) but see discussion in Clark and Royall (1996). See also: Leonte et al. (2003); **segmentation**.

Breddin curves In structural geology, a **set** of curves of **angular shear strain** (ψ; ***y-axis***) as a **function** of orientation of the greatest principal **extension** direction (φ; *x*-**axis**) for differing values of the **strain ratio**, or **ellipticity**, (R). The **strain** ratio in a given case may be **estimate**d by matching a curve of observed ψ versus φ as found from field measurements of deformed fossils with original bilateral **symmetry**. Introduced by the German geologist, Hans Breddin (1900–1973) (Breddin 1956); see Ramsay and Huber (1983).

Briggsian or common logarithm (log) An abbreviation for the *common* (i.e. base-10) logarithm. If $x = z^y$, then y is the logarithm to the base z of x, e.g. $\log_{10}(100) = 2$ and $\log(xy) = \log(x) + \log(y)$; $\log(x/y) = \log(x) - \log(y)$, etc. The principle was originally developed by the Scottish landowner, mathematician, physicist and astronomer, John Napier, 8th Laird of Murchiston (1550–1617), who produced the first table of **natural logarithm**s of **sine**s, **cosine**s and **tangent**s, intended as an aid to astronomical, surveying and navigational calculations (Napier 1614; Napier and Briggs 1618; Napier and Macdonald 1889). "The same were transformed, and the foundation and use of them illustrated with his approbation" by the British mathematician, Henry Briggs (1561–1630), who following discussions with Napier whom he visited in 1615 and 1616, developed the idea of *common logarithms* (sometimes called *Briggsian logarithms*), defining $\log(1) = 0$ and $\log(10) = 1$, and obtaining the intermediate values by taking successive **root**s, e.g. $\sqrt{10}$ is 3.16227, so $\log(3.16227) = 0.50000$, etc. His first publication (Briggs 1617) consisted of the first 1000 values computed, by hand, to 14 decimal places (they are almost entirely accurate to within $\pm 10^{-14}$; see Monta (2015) for an interesting analysis). A full table was initially published in Latin (Briggs 1624). After Briggs death an English edition was published "for the benefit of such as understand not the Latin tongue" (Briggs 1631). Briggs logarithms were soon being applied in works on geophysics, e.g. by the English mathematician, Henry Gellibrand (1597–1637) who was studying terrestrial magnetism (Gellibrand 1635). The first extensive table of (Briggsian) anti-logarithms was made by the British mathematician, James Dodson (?1705–1757) (Dodson 1742). All the tables mentioned here were calculated by hand as mechanical calculations did not come into use until the beginning of the twentieth Century. Although 10 is the common or Briggsian base, others may be used, see: **Napierian logarithm** and **phi scale**.

Broken-line distribution This refers to the **shape** of the **cumulative distribution** of two complimentarily **truncated normal** or **lognormal distribution**s, which form two straight lines which join at an angle at the **truncation point**. **Parameter estimation** uses a numerical estimation of **maximum likelihood**. Applied by the British physicist, Cecil Reginald Burch (1901–1983) to analysis of major and trace element geochemical distributions (Burch and Murgatroyd 1971).

Brown noise Coloured (colored, American English sp.) **noise** can be obtained from **white noise** by passing the **signal** through a **filter** which introduces a degree of **autocorrelation**, e.g. $x(t) = ax(t-1) + kw(t)$ where $w(t)$ is a white noise signal; a is a constant, $0 < a < 1$; k is the **gain**, and $x(t)$ is the output signal at time t. The **power spectrum** density for *brown noise* decreases **linear**ly as $1/f^2$. The concept of white light as having a uniform power density over its spectrum was first discussed by the American mathematician, Norbert Wiener (1894–1964) (Wiener 1926), and taken up in **digital signal processing** by the American mathematician Richard Wesley Hamming (1915–1998) and statistician John Wilder Tukey (1915–2000) (Tukey and Hamming 1949); see also Blackman and Tukey (1958). For discussion in an earth science context, see Weedon (2003).

Brownian motion, Brownian walk Now generally considered in the context of a one-**dimension**al **time series** in which over a fixed interval (T) the **variance** is proportional to T and the **standard deviation** is proportional to $\sqrt{T}$. In fractional *Brownian motion* (**fractal**), the variance is proportional to $2H$ and standard deviation to H, where H is the **Hurst exponent**. It is named for the British botanist, Robert Brown (1773–1858), who first described the phenomenon (Brown 1828), which he observed in 1827 in microscopic examination of the **random** movement of pollen grains suspended in water. In 1905, the German-American physicist, Albert Einstein (1879–1955), unaware of Brown's observations, showed theoretically (Einstein 1905, 1926) that the random difference between the pressure of molecules bombarding a microscopic particle from different sides would cause such movement, and that the **probability distribution** of a particle moving a distance d in a given time period in a given direction would be governed by the **normal distribution**. His theory of Brownian motion was verified in emulsions by the

French physicist, Jean-Baptiste Perrin (1870–1942), following invention of the ultramicroscope (Perrin 1908; Newburgh et al. 2006). See also: Wiener (1923, 1949) and Weedon (2003); **random walk**.

Buffon's needle problem This was first posed in an editorial comment by the French natural historian and mathematician, Georges-Louis Leclerc, Comte de Buffon (1707–1788) in 1733. It seeks the **probability** P(x) with which a needle of given length l, dropped at **random** onto a floor composed of parallel strips of wood of constant width d, will lie across the boundary between two of the strips. He showed (Buffon 1777, 46–123) that if $d \geq 1$ then

$$P(x) = \frac{2l}{\pi d},$$

and if $d < l$, then

$$P(x) = \frac{2l}{\pi d}(1 - \cos\theta) + \frac{\pi - 2\theta}{\pi},$$

where $\theta = \arcsin(d/l)$. In modern times, it has been used as a **model** for an airborne survey seeking a **linear** target and flying along parallel, equi-spaced, flight lines (Agos 1955; McCammon 1977). Chung (1981) solved the problem for the case of search using unequally-spaced parallel strips and a needle with a **preferred orientation**.

Bug An **error** in a **computer program**, or hardware (International Business Machines [undated]) which causes it to produce erroneous, or unexpected, results. Although use of the term in this context was popularised following work in engineering, radar and early **computer**s in the late 1940s (Shapiro 1987), its origins go back to nineteenth Century telegraphy and its use by Thomas Edison to indicate the occurrence of some kind of problem in electrical circuits (Edison 1878; Mangoun and Israel 2013).

Burg algorithm A method of **spectrum analysis**, also known as the **maximum entropy method**, introduced by the American geophysicist, John Parker Burg (1931–) in 1967–1968 (Burg 1967, 1968, 1975). It minimizes the forward and backward prediction **error**s in the **least squares** sense, with the autoregressive coefficients constrained to satisfy the **Levinson-Durbin recursion**. For earth science applications see: Ulrych (1972), Ulrych et al. (1973), Camina and Janacek (1984), Yang and Kouwe (1995), Buttkus (1991, 2000) and Weedon (2003).

Burnaby's similarity coefficient This is a ***weighted*** **similarity coefficient**. The English palaeontologist, Thomas Patrick Burnaby (1924–1968) discussed the use of **character weighting** in the computation of a similarity **coefficient** in a paper, originally drafted in

1965, which was only published posthumously (Burnaby 1970). See Gower (1970) for a critique of Burnaby's approach.

Burr distribution Named for the American statistician, Irving Wingate Burr (1908–1989), this right-**skew distribution** was introduced by Burr (1942), is

$$f(x) = ck\left[\frac{x(c-1)}{(1+xc)(k+1)}\right]$$

and the cumulative distribution

$$F(x) = 1 - \frac{1}{(1+xc)k},$$

where $x \geq 0$ and with **shape parameter**s $c \geq 1$ and $k \geq 1$. The **Weibull**, **exponential** and **log-logistic distribution**s can be regarded as are special cases of the *Burr distribution*. It has been widely applied in reliability studies and failure-time modelling. Discussed in an earth science context by Caers et al. (1999a,b).

Burr-Pareto logistic distribution A bivariate **distribution** which fits **bivariate** data which exhibit **heteroscedacity** quite well. Introduced by Cook and Johnson (1981) as a unifying form for the **multivariate** versions of the **Burr**, **Pareto** and **logistic distributions** and used by them (1986) in analysis of hydrogeochemical data from a large **regional** survey. Named for the American statistician, Irving Wingate Burr (1908–1989) and the French economist, Vilfredo Pareto (1848–1923).

Butterfly effect The property of **sensitivity** of a **dynamical system** to initial conditions. The idea was first popularised by the French mathematical physicist and mathematician, Jules Henri Poincaré (1854–1912) (Poincaré 1908), but the term itself apparently arose from the title of a talk given by the American meteorologist, Edward Norton Lorenz (1917–2008) to the American Association for the Advancement of Science in 1972: "Predictability: Does the flap of a butterfly's wings in Brazil set off a tornado in Texas?" See: **Lorenz attractor**.

Butterworth filter An electronic **filter** designed to have as flat as possible a **frequency response** (i.e. free of ripples) in the **passband**; the **gain** drops off in a **linear** fashion towards negative infinity away from the edges of the **passband**. Introduced by the British physicist, Steven Butterworth (1885–1958) (Butterworth 1930). Mentioned in an earth science context by Buttkus (1991, 2000) and Gubbins (2004).

Byte A **sequence** of eight **bit**s. The term was introduced by the German-American computer scientist, Werner Buchholz (1922–2003) in 1956, when he was working on the design of the International Business Machines (IBM) 7030 "Stretch" **computer**, their first transistorized supercomputer, to describe the number of bits used to encode a single character of text in a computer (Buchholz 1962, 1981). Mentioned in Davis and Sampson (1973).

Springer
http://www.springer.com/978-3-319-57314-4

Dictionary of Mathematical Geosciences
With Historical Notes
Howarth, R.
2017, XVI, 893 p., Hardcover
ISBN: 978-3-319-57314-4