Hydrology & Earth System Sciences

Balancing the books – a statistical theory of prospective budgets in Earth System science

J. Philip O'Kane

Environmental Institute and Department of Civil and Environmental Engineering, National University of Ireland, Cork, Republic of Ireland

Email: p.okane@ucc.ie

Abstract

An honest declaration of the error in a mass, momentum or energy balance, ε, simply raises the question of its acceptability: "At what value of ε is the attempted balance to be rejected?" Answering this question requires a reference quantity against which to compare ε. This quantity must be a mathematical function of all the data used in making the balance. To deliver this function, a theory grounded in a workable definition of acceptability is essential. A distinction must be drawn between a retrospective balance and a prospective budget in relation to any natural space-filling body. Balances look to the past; budgets look to the future. The theory is built on the application of classical sampling theory to the measurement and closure of a prospective budget. It satisfies R.A. Fisher's "vital requirement that the actual and physical conduct of experiments should govern the statistical procedure of their interpretation". It provides a test, which rejects, or fails to reject, the hypothesis that the closing error on the budget, when realised, was due to sampling error only. By increasing the number of measurements, the discrimination of the test can be improved, controlling both the precision and accuracy of the budget and its components. The cost-effective design of such measurement campaigns is discussed briefly. This analysis may also show when campaigns to close a budget on a particular space-filling body are not worth the effort for either scientific or economic reasons. Other approaches, such as those based on stochastic processes, lack this finality, because they fail to distinguish between different types of error in the mismatch between a set of realisations of the process and the measured data.

Keywords: balance, budget, sampling, hypothesis test, closing error, Earth System

Introduction

This paper is a contribution to theory, in the belief that theory justifies practice.

BALANCES

Making mass, energy, or momentum balances is a 'research method of choice' in many areas of Earth System science. Goldberg (1976a and b) has suggested that mass balances "can pinpoint areas where surveillance measurements must be made and assist in the determination of the frequency of such measurements". For some, balances have the additional attraction of requiring a minimum of theory. They are also eclectic and liberal in the following sense. The members of each scientific community contributing to a balance carry out their measurements and reduce their data in accordance with their own traditions. On the day of reckoning, there is a brief coming together, when the balance is debated and closed. The compulsion to process the data to deliver a perfect balance is justified by an appeal to theory, namely, a law of balance or conservation. Nevertheless, there is always a closing error!

THE PROBLEM – THE ERROR IN THE BALANCE

An honest declaration of the error in the balance, ε, simply raises the question of its acceptability: "How small should ε be for an acceptable balance?" or, alternatively, "At what value of ε is the attempted balance to be rejected?" Answering these questions requires a reference quantity against which to compare ε. This quantity must be a mathematical function of all the data used in making the balance. To deliver this function, a theory grounded in a workable definition of acceptability is essential. The theory proposed here answers the second form of the question, but not the first. In other words, acceptability will mean failure to reject the balance. The theory presented in this paper will incorporate the relevant law of balance in a manner which leaves

open the possible rejection of the attempted balance. The goal of balancing the books is not to test the conservation law itself. If that were the case, the physics laboratory would be a far better facility for doing so. The objective of balancing the books is to gain insight into relative magnitudes of fluxes in the study of the Earth System in particular circumstances of place and time. Sampling variability, and errors of omission or commission, can cloud such insight. This paper is a theoretical contribution to the better design of measurement campaigns where the primary objective is to balance the books.

BALANCE AND CONSERVATION LAWS

Scientific balances in Earth System science are suggested by the seven independent principles of rational mechanics that govern the dynamics of space-filling bodies:

1. Conservation of mass
2. Balance of linear momentum
3. Balance of rotational momentum
4. Conservation of energy
5. Conservation of electric charge
6. Balance of electric field
7. Irreversibility

Discussion of these principles from very different points of view may be found in Feynman (1965, 1989, 1992), Truesdell and Toupin (1960), and Truesdell (1980, 1984). Peixoto and Oort (1992) have applied five of them to make exemplary retrospective balances for understanding the earth's climate.

Modern physics insists that the fundamental conservation laws hold at all scales from the sub-atomic to the extra-galactic, and at all relative speeds less than the speed of light. Since mass depends on velocity in relativistic mechanics, conservation of mass is no longer a fundamental law of physics. However, on planet Earth, the speed of particles of the solid earth, its watery surface, the oceans, and the atmosphere, are so far below the speed of light that the relativistic correction to classical mechanics is negligible. In the case of the bio-geochemical cycles — among which the water cycle is pre-eminent — the measurement of mass can, in principle, be replaced by a physico-chemical counting of discrete molecules, atoms, ions, or isotopes. Dividing by Avogadro's number converts a counting balance to a mass balance measured in moles. The concepts developed in the remainder of this paper are presented in terms of a balance of mass but they are applicable to all other balances and conservation laws in Earth System science.

Mass balances

The five steps in making a mass balance are:

(1) The choice of a substance S to be studied and the elaboration of techniques for its measurement;
(2) The identification of the reservoir B, as a space-filling body, where S is accumulated;
(3) The identification of (a) all mass flows of S into and out of B and (b) all sources and sinks of S within B, during some interval of time $(0,T)$;
(4) The measurement of each mass flow rate, each source rate, each sink rate, and the quantities of S in the reservoir B at the beginning and end of the interval $(0,T)$;
(5) The processing of the measurements to produce the balance.

THE PERFECT MASS BALANCE

The reservoir to be balanced is idealised as a space-filling body B located in a three-dimensional Euclidean space with spatial coordinates (x,y,z). The boundary of B is a complete, continuous, smooth, oriented surface ∂B in the same space. In this paper, B does not move or change shape. Hence the space-filling body or reservoir is synonymous with the fixed space it occupies. The statement of *conservation of mass* is simplest in this case:

the instantaneous rate of increase of the mass of S contained in B is equal to the net mass flux of S into B across the boundary ∂B of B.

This statement may be written in the precise language of the vector calculus (Kaplan, 1952; Lin and Segel, 1974) as

$$\frac{d}{dt}\iiint_B s(x,y,z,t)\,dx\,dy\,dz =$$
$$-\oiint_{\partial B} \vec{n}(x,y,z)\cdot\vec{v}(x,y,z,t)\,s(x,y,z,t)\,ds \qquad (1a)$$

On the left-hand side, $s(x,y,z,t)$ is the instantaneous mass density of S at time t throughout B measured in $kg\ m^{-3}$. It is a dynamic scalar quantity. This is integrated over B to yield the total mass $S(t)$ in B at time t. Its derivative with respect to time is *the instantaneous rate of increase of the mass of S contained in B.*

On the right-hand side, n is the 'outward normal' vector perpendicular to the surface of B. $n(x,y,z)$ is a vector field defined on ∂B and is constant in time. The vector $v(x,y,z,t)$ is the dynamic velocity field transporting S into and out of B at all points in its surface. The vector dot-product of n and v is a dynamic scalar quantity equal in magnitude to the component of the velocity $[m\ s^{-1}]$, which is perpendicular to the surface. It is positive or negative depending on the

Table 1. The hydrological cycle [Peixoto and Kettani, 1973]. Units are 10^{12}m^3 per year.

		From		
		Atmosphere	Oceans	Land	Totals
	Atmosphere		361	62	423
	Oceans	324		37	361
To	Land	99	0		99
	Totals	423	361	99	883

Table 2. The hydrological cycle [Baumgartner & Reichel, 1975]. Units are 10^{12}m^3 per year.

		From		
		Atmosphere	Oceans	Land	Totals
	Atmosphere		424	71	495
	Oceans	385		40	425
To	Land	111	0		111
	Totals	496	424	111	1031

transport direction; outwards is the positive direction. Tangential flow in the boundary of B does not transport material through ∂B. Hence, the vector dot-product $\boldsymbol{n.v}$ is also the instantaneous volume flux per unit area *[m³ m⁻² s⁻¹]* through any infinitesimal surface element ds in ∂B. Multiplying by s, gives the instantaneous mass flux per unit area *[kg m⁻² s⁻¹]*, the so-called *mass flux density*, a dynamic scalar quantity, at all points in ∂B. Integrating this over ∂B, gives $M(t)$, the *instantaneous net rate of mass flux of S out of B*. Multiplying by minus one, converts it to a flux into B. The surface and volume integrals over ∂B and B respectively are limiting sums in the sense of Riemann (Kaplan, 1952; Bressoud, 1994). If chemical transformations of S are possible, each chemical pathway gives rise to additional internal 'source' or 'sink' rates at all points in B. These are excluded from expression (2a) in the interests of clarity. If the body B is in motion with respect to the coordinate system *(x,y,z)*, or if it changes shape with time, the statement of conservation of mass in (1a) becomes considerably more complicated.

Rewriting (1a) in terms of S and M yields

$$\frac{d}{dt} S(t) = -M(t) \tag{1b}$$

Integrating over the duration T of the mass balance gives

$$S(T) - S(0) = -\int_0^T M(t)dt \tag{2a}$$
$$= -T\overline{M}$$

where the average whole-body mass flux during the interval *(0,T)* is defined as

$$\overline{M} = (1/T)\int_0^T M(t)dt \tag{2b}$$

For convenience, the body B should be defined so that its boundary ∂B can be partitioned into planar surfaces, each having a fixed outward normal vector. This partitions M into a set of fluxes M_f that sum to M. These fluxes should be significant in themselves for some aspect of Earth System science. When T is large, the quantity *(S(T) – S(0))/T* is often assumed to be negligible, and the perfect mass balance (2a) simplifies to an algebraic sum of average fluxes

$$\overline{M} = \overline{M}_1 + \overline{M}_2 + ... + \overline{M}_f + ... = 0 \tag{2c}$$

where the subscript f runs over the set of planar surfaces into which ∂B has been partitioned.

If there is more than one substance S, or more than one reservoir B, multiple copies are made of Eqns. (2a–2c), writing balances for each substance in each reservoir and including the obvious additional terms for all mass flows between them.

Tables 1 and 2 provide an illustration. They contain two water balances for the Earth taken from Figure 12.2 of Peixoto and Oort (1992) and are referenced as Peixoto and Kettani (1973) and Baumgartner and Reichel (1975) respectively. There are three reservoirs in the balance: the atmosphere, the oceans and the land. The balance is made on the fluxes connecting the reservoirs, assuming no change in water storage. The tabulated data correspond to three copies of Eqn.(2c) and are in almost perfect balance with each other.

Prospective budgets

A *retrospective balance* is a balance calculated from data which is already to hand and which has not been collected following an appropriate statistical design. Given the eclectic provenance of such data, no compelling general theory is possible for the closing error on retrospective balances. In contrast, *prospective budgets* are planned in great detail before the measurements are made, in order to achieve the goal of judging the closure on the budget. Balances look to the past; budgets look to the future. This distinction between the *ex post* retrospective balance and the *ex ante* prospective budget is suggested by similar distinctions in comparative

Time interval	0									T
Mass fluxes	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
M_1			■		■	■				
M_2	■		■						■	
M_3		■			■					■
…										
M_f		■				■			■	
…										

Fig. 1. *The sampling frame for simple random sampling of several mass fluxes M_f. The duration of the mass balance is the interval (0,T). This is divided into N=10 sub-intervals each of duration T/10. One row in the array corresponds to one mass flux. Each unit in the frame is denoted by a rectangle. Each rectangle corresponds to one member of the "real" population of sampling sub-intervals. The black rectangles in each row show samples of three units drawn at random from the sampling frame without replacement, using, for example, a table of random numbers. From the table, it is clear that a maximum of two samples must be taken simultaneously in the case of this particular random sample.*

statistics (Anderson, 1980) and the theory of experiments (Fisher, 1966). It is not to be found in textbooks on applied statistics in hydrology.

A *prospective mass budget* for a space-filling body lays down procedures for the collection of data at unique places and at unique times in the future, together with their subsequent processing. The procedures are defined so that logically consistent statements may be made about the closing error on the budget. In the remainder of this paper, the word *budget* is used everywhere instead of *balance* to indicate concern with the design of a plan for a future campaign of measurement. While the goal is to close the budget, the plan must leave open the possibility of failure by allowing future data to determine the outcome. Also, the design process must recognise the possibility of abandoning the proposed campaign because (a) it cannot achieve its scientific objectives with respect to accuracy and precision, or (b) it costs too much. Consequently, the concepts presented below may find some degree of acceptance by both the scientific community and its funding agencies.

THE STATISTICAL MASS BUDGET

The theory of prospective mass budgets must be a statistical theory because it is impossible to measure (a) the simultaneous change in mass at *the infinitely large number of points* within a given space-filling body B, and (b) *every instantaneous* flux across its surface, during a unique finite interval of time. Hence, from now on, a finite, but often very large number of small sampling intervals in both space

$(\Delta x, \Delta y, \Delta z)$ and time Δt must be considered, instead of points in space and instants in time. These small sampling intervals are chosen to agree with the time it takes to make a measurement in a small volume of space using a particular instrument. For example, Δt could be the time it takes to fill a sampling bottle, the shape of which is the sampling interval $(\Delta x, \Delta y, \Delta z)$ in space. All measuring devices record measurements averaged, or accumulated, over some small, but finite, interval of both space and time. The size and response times of these devices allow appropriate sampling intervals to be chosen.

The great counter-intuitive discovery of classical statistics is that a properly conducted, prospective, sampling experiment can deliver a precise result with a surprisingly small number of samples (Cochran, 1977) selected from the very large number of possible samples.

The statistical mass budgets that correspond to the perfect balances Eqns. (2a and c) may be defined by the expressions

$$\hat{S}(T) - \hat{S}(0) + T \sum_f \hat{\overline{M}}_f = \varepsilon \qquad (3a)$$

$$\sum_f \hat{\overline{M}}_f = \varepsilon \qquad (3b)$$

They differ in three fundamental respects from the perfect balances:

(a) the error on the budget ε is equal to the numerical value of the sum on the left hand side,

(b) the caret symbol indicates an estimate of a population characteristic, made from a statistical sample taken from the appropriate population,

(c) the population characteristics are total mass at the beginning and end of the budget, $S(T)$ and $S(0)$, and the average mass fluxes during the interval (0,T) $\overline{M}_f$.

If the distribution of mass in B were perfectly uniform in space and steady in time, *i.e.* $S(T) = S(0)$, and if all the fluxes were constant, one measurement of each $\overline{M}_f$ at any time between *0* and T would yield the estimates $\hat{\overline{M}}_f$ with no sampling error. If the value of ε were different from 0, it would be concluded that either *(S(T) -S (0))/T* was not zero, or there was a systematic error in the measurement technique, or the mass flows were not specified correctly, or a combination of all three reasons. Failure to include a mass flow in the budget is called an error of specification.

However, bodies are rarely in dynamic equilibrium. Measurements at many points in time are required to estimate average mass flow rates. Similarly, when the distribution of S in B is not uniform, $S(T)$ and $S(0)$ must be estimated by measuring the concentration of S at many points in B. Hence, the detection of errors of measurement and of specification, is obscured by errors due to statistical sampling.

Consequently, it may appear that there are no objective grounds for rejecting any attempted budget, and that all are acceptable, as is the case with retrospective balances. It shall be shown that this is not so for prospective budgets.

THE POPULATION AND ITS SAMPLING FRAME

The key step in the development of the theory is the definition of the *population* to be sampled. Two different populations are defined, one for *B*, and one for the time interval *(0,T)*. In the case of the time interval *(0,T)*, its duration is divided into N_T sub-intervals each equal to the sampling interval Δt. Hence,

$$T = N_T \Delta t \tag{4a}$$

The population to be sampled is the set of N_T sub-intervals. A *sampling frame* is constructed by placing the sub-intervals in one-to-one correspondence with the integers, with one line of the frame for each mass flux. This is shown in Fig. 1 for four different mass fluxes with $N = N_T = 10$ and $\Delta t = T/10$ in each case. In practice, there would be a very much larger number.

In a similar manner, *B* is divided into N_B sub-bodies, ΔB_j, each of volume ΔV_j. Since *B* is a space-filling body, the following conditions on the union and intersection of the point-sets corresponding to the sub-bodies are imposed

$$B = \bigcup_{j=1}^{N_B} \Delta B_j, \quad \bigcap_{j=1}^{N_B} \Delta B_j = \varnothing \tag{4b}$$

The volume ΔV_j is determined by the measuring system. One may think of ΔV_j as a small cube in direct contact with the instrument while it makes a single measurement. The sub-bodies are mapped onto cubes in a Euclidean space with a suitable co-ordinate system and distance metric. The cubes stack without gaps to cover *B*. They can then be placed in one-to-one correspondence with the integers defining their sampling frame. Any sub-body selected from the frame for measurement can be located easily in the Euclidean space. This frame will have two lines, one for the mass in *B* at the beginning of the budget, and a second for the mass in *B* at the end. One line of the frame for a body varying in one space dimension is illustrated in Fig. 2 where $N = N_B = 10$. In this figure, the summation of ΔB_j represents the union along a line of the uniform sub-bodies, or elements.

Applying a sequence of random numbers to the sampling frames, whether taken from a published table of random numbers, or generated by a pseudo-random procedure, specifies when and where measurements are to be made. Figure 2 shows *n=3* random samples in the case of each of the four mass fluxes; the samples selected in each column must be measured simultaneously. Figure 1 shows *n=3* simultaneous random samples in the case of the mass *S* in *B* at one point in time. Population statistics are estimated from statistical samples such as these. Since two measurements cannot be made simultaneously at the same location, the sampling is without replacement in the population.

Depending on the detailed objectives of the budget, more complicated populations may be envisaged involving stratification in time and space, or Cartesian products of the point sets corresponding to the time interval of the budget and parts of the surface of *B*.

Such populations may be described as 'real' since observers can choose to make measurements at any place and time in the sampling frame. Every member of the population is accessible in this sense. The notional, or imaginary, populations that underpin the application of stochastic processes to Earth System data, lack this concrete reality. Such real and notional populations are compared again later in the paper.

The following section illustrates these ideas by presenting, in outline, the application of classical sampling theory to the estimation of precipitation on a forest floor.

THE APPROACH OF SAMPLING THEORY

To estimate the total volume of rainfall falling on the floor of a forest during some interval of duration *T*, as part of a prospective water budget, the forest floor *F* is divided into a grid of *N* non-overlapping squares ΔF, each equal in size to the footprint of the standard rain gauge, and which completely covers the floor. If *F* has an area of 1 km^2 and the rain gauge an area of 1 m^2, *N* is 1 000 000. The squares ΔF are placed in one-to-one correspondence with the points of an integer-valued coordinate system. The corresponding set of pairs of integers is the sampling frame for *F*. Selecting a pair of integers using a table of random numbers, or other similar device, locates the first rain gauge in the frame and this is positioned in the forest using differential GPS. Repeating this 'simple random sampling' *n* times without replacement, allocates *n* rain gauges to positions on the forest floor. Now consider the complication of the trees themselves and their stem-flow. The trees are separated by introducing a stratification in space, which partitions *F* into squares that can be covered with 'square (2 m × 2 m) stem-flow collars on trees', or 'square (1 m × 1 m) rain gauges on the floor'. A second sampling frame allocates *p* collars to trees.

Now, the time interval *T* is divided into *m* subintervals each of duration ΔT. Since rain gauges are integrating devices, ΔT is arbitrary in this case and may be taken as one day, one week, one month, or one season. During each subinterval, the random allocation of rain gauges to squares

B divided into 10 sub-units										
Index i	1	2	3	4	5	6	7	8	9	10
Mass in each subunit			21		3	55				

Fig. 2. *The sampling frame for simple random sampling of the mass S in B at time 0. The body B is divided into 10 sub-bodies which do not overlap and which together just cover B. They are enumerated with the index i to form the sampling frame. One unit in the frame corresponds to one member of the "real" population of sub-bodies. A sample of three units is drawn simultaneously without replacement from the population of 10 units in the frame using a table of random numbers. The measurements on the corresponding sub-bodies yield masses, for example (21, 3, 55), from which the total mass and its sampling error can be estimated. The masses of the seven other sub-bodies remain hidden and unknown. If an estimate of the mass of S in B is required at any other time t, a new sampling frame is constructed and sampled in a manner similar to that at time 0.*

may be repeated, so that the rain gauges perform a random walk. In other words, the sampling frame is now the Cartesian product of the set of squares times the set of subintervals. There is always some discrepancy in locating a sample in space and time; if there is a bias, it will appear in the closing error of the budget and will be included in the test on the budget.

Minimising the cost of estimating the total volume of rain during the interval T subject to a predefined sampling error, determines the values of p, n and m. The sampling error arises from the fact that not every unit of the sampling frame is chosen for measurement. A pilot study is necessary to provide the data (sampling variances) to solve this optimisation problem. Also, sampling schemes, other than simple random sampling, are possible, such as stratified or cluster sampling. The central assumption in the underlying probability model is that the measurement campaign is carried out with known probabilities assigned to each unit of the sampling frame. Consequently, the sampling procedure *guarantees* the assumptions used in estimating the sampling variance of the total rainfall over a unique area F during a unique interval of T since they are completely 'covered' by the sampling frame.

The word 'sample' has different meanings for the statistician and the environmental chemist. A statistical sample, in the context of this paper, is a unit selected from a sampling frame using a random device, such as a roulette wheel, table of random numbers, or pseudo-random algorithm, which specifies when and where a measurement is to be made in a space-filling body. A sample for an environmental chemist is a volume of water, or other medium, taken from the environment and analysed in accordance with a specific protocol.

STATISTICAL THEORY

The theory of the prospective budget rests on two classical statistical techniques applied to the sampling frames. They are
(1) probability sampling for estimating the terms of the budget, the closing error, $S, \hat{\bar{M}}_p, \varepsilon$ and their sampling precision;
(2) Hypothesis testing that any particular value of ε is due to chance alone.

Both techniques must satisfy R.A. Fisher's (1966; page 34) "vital requirement that the actual and physical conduct of experiments should govern the statistical procedure of their interpretation". This requirement is impossible to achieve in a retrospective balance.

Probability sampling is any sampling procedure governed by sets of random numbers. Applying the calculus of probability to the chosen procedure, one can calculate the theoretical probability distribution of the terms of the budget and the closing error, $S, \hat{\bar{M}}_f, \varepsilon$, *on the assumption that there are no systematic errors of measurement or missing fluxes*. In the simplest case, simple random sampling may be used. Random samples are statistically independent, by construction! Appropriate 'construction' is a key part of Fisher's "vital requirement". Consequently, all possible correlations between the estimators in relations (3a and b) are made zero, no matter what degree of correlation may be present in the underlying time series being sampled. When the number of independent samples (measurements) is sufficiently large, the central limit theorem implies that the closing error ε in 3a and 3b is normally distributed. The following argument delivers the mean (and variance) of this Normal Distribution. The expected value $E(\varepsilon)$ is *zero* because *(1) there are no systematic errors of measurement and no missing fluxes*, **by assumption**, *(2) the probability sampling used e.g. simple random sampling, delivers unbiased estimates of average fluxes and stored masses of S*, **by construction**, *and (3) the law of conservation of mass holds*, **by agreement**.

Definite statements which can now be made about any observed value of ε, are logically consistent with the law of conservation of mass, the measurement/sampling procedure, and the data reduction. If the future observed value of ε lies outside its theoretical distribution, the hypothesis that this could have occurred only by chance is rejected. It is much more likely that systematic errors are present so that the attempted budget should be rejected. If the observed value of ε lies within its distribution, there are no grounds for rejecting the hypothesis or the budget. However, by increasing the number of measurements, *Var(ε)* can be reduced, so refining the discrimination of the test.

All other types of measurement procedure lack this logical

and objective conclusion. Measurement campaigns designed without using a random device, or procedures that forbid measurements at weekends, at night or during lunchtime, allow only subjective judgements to be made about the closing error.

PROBABILITY SAMPLING

The simplest form of probability sampling is simple random sampling.

By using a table of random numbers, n equally likely sampling units of duration ΔT can be selected without replacement and this can be repeated independently for each mass flow. The result defines the sequence of measurements in the field. The mass flow rates which are subsequently measured are denoted by $m_f(i)$ where i is an index which spans the sampling units chosen from 1 to n.

Following Cochran (1977), lower case letters refer to samples and capitals to the population.

SIMPLE RANDOM SAMPLING

In this section S means sampling variance, not the mass of substance S in the reservoir. This change should not cause confusion since the next two sections are concerned exclusively with relation (3b).

Using simple random sampling, an unbiased estimate of each mean mass flow rate $\hat{\overline{M}}_f$ is given by the sample mean $\overline{m}_f$

$$\hat{\overline{M}}_f \underset{est}{=} \overline{m}_f \underset{def}{=} \sum_{i=1}^{n} m_f(i) / n \tag{5}$$

where "def" means "by definition" and "est" means "is estimated by". An unbiased estimate of the sampling variance of the sample means $\overline{m}_f$ is

$$v(\overline{m}_f) \underset{def}{=} s_{\overline{m}_f}^2 \underset{est}{=} \frac{s_f^2}{n} \frac{(N-n)}{N} \tag{6}$$

where s_f^2 is an unbiased estimate of the variance of the mass flow rate $\overline{M}_f$

$$\hat{S}_f^2 \underset{est}{=} s_f^2 \underset{def}{=} \sum_{i=1}^{n} (m_f(i) - \overline{m}_f)^2 / (n-1) \tag{7}$$

The quantity $S_{\overline{m}_f}$ (lower-case) is also called the standard error of the estimate of $\overline{M}_f$ Proofs can be found in Cochran (1977), page 26.

If $n=N$, the mass flow rates are measured completely and Eqns. (5) and (6) yield correctly $v(\overline{m}_f) = 0$, $\overline{M}_f \equiv \overline{m}_f$ and $S_f^2 \equiv s_f^2$. In most cases, however, N>>n>>1

and

$$v(\overline{m}_f) \underset{est}{=} s_f^2 / n \tag{8}$$

Hence, the precision with which $\hat{\overline{M}}_f \equiv \overline{m}_f$ is estimated depends directly on the variance of the mass flows themselves and, inversely, on the number of measurements. It does not depend on N, which for many is counter-intuitive. It appears, therefore, that n should not be the same for all mass flows but should vary in proportion to s_f^2. This is discussed further below.

TESTING THE CLOSING ERROR ON THE BUDGET

The application of independent random numbers to the sampling frame makes the sampling distributions of the estimates $\hat{\overline{M}}_f$ of $\overline{M}_f$ *mutually independent* by construction. Randomisation destroys any correlation that may be present during the measurement of the components of the budget and which might bias the estimation. Independence allows the means and variances of the estimates of the individual terms of the budget to be added *i.e.*

$$E(\varepsilon) = \sum_f \overline{m}_f \tag{9}$$

$$Var(\varepsilon) = \sum_f s_{\overline{m}_f}^2 = \sum_f s_f^2 / n \tag{10}$$

When n is large, say greater than 100, the central limit theorem suggests that the distribution of ε is approximately normal. See Cochran (1977), section 2.15, for a discussion of the effect of skewness on the speed with which the normal distribution is approached as n increases.

Since the $\overline{m}_f$ are *unbiased estimators* of $\overline{M}_f$, the law of conservation of mass requires that

$$E(\varepsilon) = 0 \tag{11}$$

when there are no errors of measurement or of specification.

Let ε_0 be a particular value of the closing error. Now compare it with a normal distribution of mean 0 and variance $\sum_f s_f^2 / n$. There are two possible results:

(1) ε_0 is an outlier and the hypothesis that ε_0 could have arisen solely from the mechanisms of chance which were applied to the sampling frame is rejected. It is much more likely that errors of measurement and of specification are present.

(2) ε_0 is not an outlier and there are no grounds for rejecting this hypothesis. Consequently, the mass budget as a whole cannot be rejected.

An outlier is defined to be any ε for which

$$|\varepsilon_0| > 2\tfrac{1}{2} \cdot \mathrm{Var}^{\frac{1}{2}}(\varepsilon) \qquad (12)$$

This gives a probability of $1\tfrac{1}{4}\%$ of a type I error, that is of rejecting the hypothesis when it is true. If the total number of measurements is less than 50, Student's t distribution should be used instead of the Normal distribution.

The discrimination of this test can be increased by increasing n. This decreases the variance of the normal distribution and improves its validity as an approximation. In addition, the precision of the estimated $\overline{M}_f$, is also improved. Hence, not only the precision but also the accuracy of the measurements can be controlled by closing the budget to within a chosen variance $Var(\varepsilon)$.

Campaign design – the simplest case

If the variable costs of measurement are

$$C = \sum_f c \; n_f \qquad (13)$$

where c is the unit cost of measurement and n_f measurements are to be carried out on each mass flow. Hence, the total number of measurements is

$$n = \sum_f n_f \; = \; C \,/\, c \qquad (14)$$

The design problem is

What allocation of measurements will minimise the sampling variance of the closure at a given cost?

It can be shown that $Var(\varepsilon)$ is minimised for fixed n and C by taking n proportional to s_f, not s_f^2 (Cochran, 1977, page 98). Hence, the optimal n_f^* are

$$n_f^* \; = \; n \, s_f \; / \; \left(\sum_f \; s_f \right) \qquad (15)$$

Substituting in $Var(\varepsilon) = \sum_f s_f^2 \; / \; n_f$ the corresponding minimum-closing-variance V^* is

$$V^* \; = \; \left(\sum_f \; s_f \right)^2 / \, n \qquad (16)$$

Eliminating n from Eqn. (16) using Eqn. (14), gives the hyperbolic **cost-effectiveness function for the measurement campaign**

$$V^* = \left(\sum_f s_f \right)^2 c/C \qquad (17)$$

To design the campaign, preliminary estimates of s_f are needed A pilot campaign can be used to find them. Given a money budget C — set by the funding agency — Eqn. (17) shows what minimum-closing-variance, V^*, will be delivered; or, alternatively, a closing variance V^* — set by the scientists — can be chosen and Eqn. (17) gives the lowest price at which it can be achieved. Then, n can be determined from Eqn.(14) and the n_f^* from Eqn. (15).

Analogous results are available when c varies with f as c_f (see Cochran, 1977, page 97). Further discussion on the problem of advance estimates of population variances may be found in section 4.7, page 78 of Cochran (1977).

The fixed costs have been ignored in this illustrative example of campaign economics. A part of the fixed cost sets the maximum possible number of *simultaneous* samples that can be made at each point in time during the campaign. This is an important constraint, which should be checked, or better, included in the optimisation.

Clearly, one outcome of such an analysis is the demonstration in advance that a mass, energy, or momentum budget on a particular body B may not be worth the effort for either scientific, or economic reasons. This possibility is a requirement for a useful theory of budgets.

STRATIFIED RANDOM SAMPLING

Probability sampling plans, which are more complex than simple random sampling, can be designed to give a smaller variance of the estimate using the same number of samples. For example, in *stratified sampling*, the population is divided into non-overlapping sub-populations or strata. It can be shown that the sampling variance of an estimated mean is a minimum for a specified cost, when the number of samples in each stratum is proportional to the true standard deviation of each sub-population. (Cochran, 1977; page 97).

This idea arose earlier, whereby the individual mass flows M_f can be regarded as a particular stratification in space of the mass of S, which flows across the open boundary of B. Stratification may also be applied in time, especially to flows that change direction across the boundary of B. Oscillating flows are especially difficult because they give rise to differences between large uncertain numbers — their sampling variances add when the estimates are independent. It is usually better to redefine the body B to avoid oscillating flows altogether. Details of the stratification technique can be found in Chapters 5 and 5A of Cochran (1977).

UNIFORM SAMPLING

Many modern measuring systems sample on a uniform grid in space or time e.g. remote sensors mounted on aeroplanes

or satellites. If the origin of the grid is chosen at random, uniform grids can sometimes be brought within the ambit of classical sampling theory. See Cochran (1977) for further details. Uniform sampling may not be used when components in the measured signal have periods less than the interval of the grid. These unknown components are said to alias the measurements in the frequency domain. Spectral analysis of data collected at very high frequencies in a pilot survey may show what uniform sampling interval should be used to avoid aliasing. This is very rarely done. Consequently, absence of aliasing is usually an unverified assumption.

MEASUREMENTS OF $M_F(I)$

Mass flows are often measured as a flow rate of water at a point, multiplied by a concentration of S. Rating curves are used to convert water level into discharge in rivers. The statistical theory of sampling can be applied to these too. See for example, ratio and regression estimators in Chapters 6 and 7 of Cochran (1977).

Other stories

The approach advocated here, based on sampling theory, is compared and contrasted with an alternative approach based on stochastic processes.

THE ALTERNATIVE PROBABILISTIC APPROACH

In what may be called the ***traditional approach***, samples are allocated according to preconceived ideas of *representativity*, e.g. placing rain gauges at the circumcentres of triangles formed by adjacent trees by analogy with Thiessen polygons in hydrological practice, or on a uniform quincunx or square grids, and so on. Since these are *representative* sites, they remain fixed in space. Summing the rain depths at all the gauges and collars during the interval T and multiplying by the appropriate areas delivers the total rainfall. However, there is no simple way in which the error in this total can be quantified. Consequently, the simple traditional approach cannot contribute to the judgement on the water balance without a probability model.

In the ***alternative probabilistic approach,*** a probability model, or stochastic process with certain stationarity and covariance properties, is assumed. The model may account for (a) the spatial variation in total canopy drip, or stem flow, during the interval T, or (b) the spatio-temporal variation in these quantities during T. The data are regarded as a single realisation of the chosen spatial or spatio-temporal process. See for example Chapter 6 *Multidimensional*

Hydrologic Processes in Bras, R.L. and Rodriguez-Iturbe, I. (1985), or Chapter 9 *The Modelling of Spatial Processes* in Clarke, R.T. (1994). Such models, when fitted to the data of the previous paragraph, can deliver sampling variances induced by the probability model, for both the canopy drip and stem flow at all points in the forest and, consequently, for the whole forest.

However, these models fail to meet R.A. Fisher's (1966; page 34) "vital requirement that the actual and physical conduct of experiments should govern the statistical procedure of their interpretation". For example, if uniform sampling is used, there is no way in which it can be planned *ex ante* to guarantee the absence of aliasing *ex post*. Absence of aliasing is always a more or less plausible assumption dependent on models, and very rarely on data, at frequencies greater than the Nyquist frequency.

One could make a numerical example to show that the two approaches deliver similar, or different, numerical results. But this would prove nothing! Ultimately, the choice lies between the probability models that underpin these two approaches and their usefulness in judging the closing error.

ALTERNATIVE PROBABILITY MODELS

In random sampling, the underlying probability model is a sampling device, or scheme, which produces independent samples, drawn from the uniform probability distribution. This device or scheme can be tested *before it is applied*, using *as many realisations and tests as necessary*, to ensure that it delivers random samples from the sampling frame with the required properties. In other words, there need be no errors in the probabilistic model that underpins random sampling; the population being sampled is real.

In the alternative probabilistic approach, a complex spatial or spatio-temporal probability model is based on *a single realisation*, which cannot be verified at those places in F where the possibility of measurement has been excluded. The underlying probabilistic model has unknown errors at all locations where there are no rain gauges or collars. In this case, the population being sampled is imaginary, or notional.

If the choice is so clear-cut, why are stochastic models so dominant in the hydrological and geological literature today? Because, in *retrospective analyses* of existing data, there is nothing better that can be done! This is not the case with *prospective* data intended for a well-defined goal. This can be done better. The distinction between *retrospective balance* and *prospective budget* was introduced to highlight this point. The closure of a future budget must be regarded as an experiment and Fisher's vital requirement of logical coherence between experiment and interpretation must be

satisfied. Random sampling, in all its many forms, and any subsequent hypothesis-testing, possess this coherence, whereas stochastic process models, at least in their present form, do not.

If the data are already available, the best that can be done is a retrospective balance, *ex post*. The steps are:

(a) calculate the closure on the *retrospective balance;*
(b) make plausible assumptions about stationarity and covariance and fit a probability model to the retrospective data;
(c) extract the variance of the sum of all the terms in the budget induced by the probability model;
(d) compare the closing error with the variance in (c).

The introduction to this paper called for a quantity such as (c) which is a function of all the data and which is delivered by a theory. However, no compelling conclusion is possible in (d) since one cannot separate the different contributions to the variance of the closing error, namely, errors in the assumed probability model, errors of specification in the balance, errors arising from the failure to measure at certain places and times and not at others, and errors of measurement; all are confounded together.

Stochastic differential equations may be regarded as the most advanced form of the approach based on stochastic processes. See for example Dagan (1989). When a conservation law is a part of the set of stochastic differential equations, a perfect balance is imposed on every realisation of the process and all errors are confounded in the mismatch between the state variables and measurements.

On the other hand, in planning a *prospective budget,* classical sampling theory can be applied once the sampling frame has been defined. Furthermore, the errors can be separated into those due to sampling and those due to everything else. Consequently, the null hypothesis that the closing error in the prospective budget is due to sampling error alone can be tested. A definite conclusion is the outcome: the null hypothesis *on a particular body during a unique interval of time* is either rejected, or not. Rejection leads to a search for systematic error: missing fluxes, instrument errors, etc. The test can be strengthened by taking a larger number of samples. The trade-off between cost and the precision and accuracy of the budget can be examined using optimisation. It can also show that the budget is not worth the effort. These are the fruits of Fisher's *vital requirement*.

In contrast to all that has been said, a perfect balance *(ex post)*, or a perfect budget *(ex ante),* can also be imposed on a particular space-filling body *B* during a unique interval of time, by building the conservation law into a model of *B* based on *deterministic* differential or integral equations. If the time-varying state variables in such equations are spatial or temporal averages, an approach similar to that presented in this paper is possible, which meets Fisher's *vital requirement*. If random sampling is used to estimate the spatial or temporal averages, one can, in principle, test the null hypothesis that the errors between the estimated averages in space and/or time and the corresponding values of the state variables are due solely to sampling. But that is another story for another day.

Conclusions

Classical sampling theory has been shown to be applicable to the problem of estimating and closing prospective budgets on a particular body during a unique interval of time. Retrospective balances cannot be brought within this theoretical framework. Only prospective budgets can satisfy R.A. Fisher's vital requirement that the actual and physical conduct of measurement campaigns be logically consistent with the statistical procedures used in their design and interpretation. The key step is the definition of the sampling frame to which randomisation is applied.

The theory provides a test, which rejects, or fails to reject, the hypothesis that the closing error on a budget, when realised, is due to sampling error only. Increasing the number of measurements improves the discrimination of the test, controlling both the precision and accuracy of the budget and its components.

The theory delivers a cost-effective design for a sampling campaign, which aims to close a budget on a particular space-filling body. Both scientific communities and their funding agencies would welcome knowing in advance when it is not worth the effort to estimate the closure on a budget for either scientific or economic reasons. Only prospective random sampling offers such finality.

The proof of this theory, its concepts and procedures, will be their first successful application in Earth System science. For the moment, it remains a counsel of perfection.

References

Anderson, S. et al. 1980. *Statistical Methods for Comparative Studies, Techniques for Bias Reduction*. New York. John Wiley, New York.

Baumgartner, A. and Reichel, E., 1975. *The World Water Balance*. Elsevier, Amsterdam, The Netherlands.

Bras, R.L. and Rodriguez-Iturbe, I., 1985. *Random Functions in Hydrology.* Dover, New York, USA.

Bressoud, D., 1994. *A Radical Approach to Real Analysis*. The Mathematical Association of America. Washington, DC, USA.

Clarke, R.T., 1994. *Statistical Modelling in Hydrology.* Wiley, Chichester, UK.

Cochran, W.G., 1977. *Sampling Techniques*. 3rd edition. Wiley, New York, USA.

Dagan, G., 1989. *Flow and Transport in Porous Formations*. Springer, Berlin, Germany.

Feynman, R.P., 1965, 1992. *The Character of Physical Law*. British Broadcasting Corporation and Penguin Books. London, UK.

Feynman, R.P., Leighton, R.B. and Sands, M., 1965, 1989. *The Feynman Lectures on Physics*. Commemorative Issue. Volumes 1-3. Addison-Wesley. Reading, Mass., USA

Fisher, R.A., 1966. *The design of experiments*. 8th edition. Oliver and Boyd. Edinburgh, UK.

Goldberg, E.D., 1976a. *The Health of the Oceans*. The UNESCO Press, Paris, France.

Goldberg, E.D., 1976b. *Strategies for Marine Pollution Monitoring*. Wiley, New York, USA.

Kaplan, W. 1952. Advanced Calculus. Reading, Mass. Addison-Wesley. p. 273 and 295.

Lin, C.C. and Segel, L.A., 1974. *Mathematics Applied to Deterministic Problems in the Natural Sciences.* Macmillan, New York. p.444.

Peixoto, J.P. and Oort, A.H., 1992. *Physics of Climate*. American Institute of Physics, New York, USA.

Peixoto, J.P. and Kettani, M.A., 1973. The control of the water cycle. *Sci. Amer.*, **228**, 46–61.

Truesdell, C. and Toupin, R.A., 1960. The Classical Field Theories. In: *Encyclopedia of Physics*, Vol.III/1, S. Flügge, (Ed.). Springer, Berlin, Germany.

Truesdell, C.A., 1980. Sketch for a history of Constitutive Relations, *Proc. 8th Int. Congress on Rheology* , G. Astarita and G. Marrucci (Eds.) **1**, 1–27. Plenum Press. New.York, USA.

Truesdell, C. *et al*., 1984. *Rational thermodynamics*. 2nd Edition. Springer, Berlin, Germany