

Does a Global Temperature Exist?

Christopher Essex¹, Ross McKittrick², Bjarne Andresen^{3,*}

¹ Department of Applied Mathematics, University of Western Ontario, London, Ontario, N6A 5B7, Canada

² Department of Economics, University of Guelph, Guelph, Ontario, N1G 2W1, Canada

³ Niels Bohr Institute, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark

*Corresponding author (andresen@fys.ku.dk)

Abstract

Physical, mathematical, and observational grounds are employed to show that there is no physically meaningful global temperature for the Earth in the context of the issue of global warming. While it is always possible to construct statistics for any given set of local temperature data, an infinite range of such statistics is mathematically permissible if physical principles provide no explicit basis for choosing among them. Distinct and equally valid statistical rules can and do show opposite trends when applied to the results of computations from physical models and real data in the atmosphere. A given temperature field can be interpreted as both “warming” and “cooling” simultaneously, making the concept of warming in the context of the issue of global warming physically ill-posed.

1. Introduction

It is widely held that the atmosphere and oceans have been warming over the past half century. The basis of this view is that there is an upward trend in the graph of a statistical measure or statistic called the “global temperature” [1, 2]. It arises from projecting a sampling of the fluctuating temperature field of the Earth onto a single number (e.g., [3, 4]) at discrete monthly or annual intervals. Proponents claim that this statistical measure or statistic represents a measurement of the *annual global temperature* to an accuracy of $\pm 0.05^\circ\text{C}$

(see [5]). Moreover, they presume that small changes in it, up or down, have direct and unequivocal physical meaning.

While that statistics is nothing more than an average over temperatures, it is regarded as *the* temperature, as if an average over temperatures is actually a temperature itself, and as if the out-of-equilibrium climate system has only one temperature. But an average of temperature data sampled from a non-equilibrium field is not a temperature. Moreover, it hardly needs stating that the Earth does not have just one temperature. It is not in global thermodynamic equilibrium – neither within itself nor with its surroundings.

It is not even approximately so for the climatological questions asked of the temperature field. Even when viewed from space at such a distance that the Earth appears as a point source, the radiation from it deviates from a black body distribution and so has no one temperature [6]. There is also no unique “temperature at the top of the atmosphere”. The temperature field of the Earth as a whole is not thermodynamically representable by a single temperature.

The global temperature statistic is also described as *the* average, as if there is only one kind of average. Of course there is an infinity of mathematically legitimate options. Indeed, over one hundred different averages over temperatures have been used in meteorology and climate studies [7] with more appearing regularly. For the case of temperature, or any other thermodynamic intensity, there is no physical basis for choosing any one of these from the infinite domain of distinct mathematical options.

The international standards organization ISO tried to choose one but failed [8]. The problem is not a mere absence of a convention for selecting one from among many mathematically different but physically equivalent measures for a single underlying property. The problem is that there is no single underlying property, because there is no global temperature. But this does not stop averages from being made.

There is no experimental or theoretical way to falsify any particular choice of averaging rule, if averages are (falsely) proclaimed to be temperatures. Proclaiming them to be temperatures leads to a paradox, as any two *ad hoc* choices applied to a particular out-of-equilibrium field can have mutually contradictory behaviors: the system can seem to be both warming and cooling simultaneously. Paradoxically, whether the system is “warming” or “cooling” becomes a property of the choice of average – a choice which is independent of the system.

The resolution of this paradox is not through adoption of a convention. It is resolved by recognizing that it is an abuse of terminology to use the terms “warming” and “cooling” to denote upward or downward trends in averages

of temperature data in such circumstances. Statistics might go up or down, but the system itself cannot be said to be warming or cooling based on what they do, outside of special circumstances.

By the same token, different statistics produced by different averaging rules should not be expected to behave the same, since they are not measuring a single, well-defined physical quantity. Unfortunately, it has long been considered a problem in climatological debates that the well-known global temperature averages behave differently from each other. Debates as to which is the correct one are fundamentally false, with no correct resolution. They have an interminable aspect, as has been illustrated in comparisons of satellite and ground-measured “global temperatures” [9], or of different versions of the satellite-measured series [10, 11], or in debates about whether the medieval era was “warmer” than the present, e.g. [12]. We argue herein that these disputes have their root in attempting to estimate a physical quantity that does not actually exist, and hence there is no prospect of resolution on scientific grounds.

But if the statistics are not actually temperatures, what then are they? Can they be used as indices to warn us of subtle dangers, like the proverbial canary in the coal mine? Or are they just examples of an infinity of *ad hoc* numbers extractable from measurements that might be followed to no end?

So far the proponents of global temperatures have been able to avoid this question. They have not substantiated exactly how changes upward or downward in their statistical measure or statistic might affect dynamics and local states within the atmosphere and oceans. However, that has not stopped many remarkable attributions of cause: small increases in their global statistical measure or statistic have been cited to explain hurricane formation [13], viral infections in frogs [14], encephalitis in horses, and even pulmonary disease, delirium, and suicide in humans [15].

The awkwardness of such claims is obscured because the concept of averaging is so routine. In fact, it is so pervasive that it may even seem implausible to mount a critique. But it should only be routine where it makes sense. Personal income and height are meaningful at the individual level, and no conceptual problems emerge when adding or averaging over a population. But there are certainly examples where averaging or adding destroys the meaning of a variable. In economics, for example, an exchange rate is meaningful when comparing two currencies, but the ideas of a “global exchange rate” or a sum over exchange rates are both nonsensical. Regardless of the fact that enough data exist to compute something analogous to a “global temperature” for the money markets, neither the level nor the trend in such a statistics would provide any meaningful information about the global economy. Another ex-

ample: Individual telephone numbers are both meaningful and useful, while the sum or average over telephone numbers in a directory have no meaning.

The notion of being globally “hotter” or “colder” for out-of-equilibrium systems is not altogether without merit. Miami in January, with temperatures ranging from 20 to 30°C, say, is certainly warmer than Toronto at, say, –15 to –5°C. However, this ranking of relative warmth is not based on averages, but on the ranges in respective temperature fields. Since the ranges do not overlap, all averages will agree which field seems to be the warmer. It is *independent* of the choice of average.

Not so for the case of comparisons of Earth’s temperature field at times a few years apart. The range over the globe (about –80°C to +40°C) is essentially the same for both when compared to statistical trends in averages (i.e., about $\pm 0.1^\circ\text{C yr}^{-1}$), which are three orders of magnitude smaller. In cases where ranges overlap, not all averages over a given set of actual observations agree on trends (see Section 3), throwing into doubt for this case what “warmer” or “cooler” means.

It is clear that there are many misconceptions about non-equilibrium temperatures fields. This paper serves to expose and identify them with specific reference to the measurement of climate change. They may be summarized by the following points, which are treated in detail later in the paper:

1. Sums or averages over the individual temperatures in the field are not temperatures. Neither are they proxies for internal energy.
2. Temperatures from a field (individually or averaged) neither drive dynamics nor thermodynamics. Instead, dynamics is driven by gradients and differences, in temperatures and other variables.
3. A global spatial average cannot be an index for local conditions, otherwise nonlocal dependence (i.e., “thermodynamics at a distance”) for local conditions would be required.
4. The utility of any global spatial average of the temperature field as an index for global conditions has been presumed but not demonstrated.
5. It is easily demonstrated that different spatial averaging rules over temperatures can have contrary trends in time (i.e., some increase while others decrease in time) when the two fields being compared have range-overlap, as they do in this context. This is demonstrated here in a basic example and subsequently with actual atmospheric temperature-field observations.
6. No ground has been provided for choosing any one such statistics over the rest as the one proper index for global climate.
7. If there are no physical or pragmatic grounds for choosing one over another, and one increases while the other decreases, there is no basis for concluding that the atmosphere as a whole is either warming or cooling.

2. Averaging and physical temperature

2.1. Temperature versus average of temperature data

Thermodynamic variables come in two varieties: extensive and intensive. Extensive variables are proportional to the size of the system. They are additive. In this category we find volume, mass, energy, entropy, particle number, etc. We can combine two systems and the values of extensive variables for the whole system will simply be the sum of the values from the two components. Correspondingly, a mean subsystem (loosely called the average) will have this sum divided by the number of components. Such an average over a quantity like mass is meaningful because the sum is meaningful. For example, average mass is of importance to airlines because it is helpful to estimate the total load of an aircraft without having to weigh every passenger.

Intensive variables, by contrast, are independent of system size and represent a quality of the system: temperature, pressure, chemical potential, etc. In this case, combining two systems will *not* yield an overall intensive quantity equal to the sum of its components. For example, two identical subsystems do not have a total temperature or pressure twice those of its components. A sum over intensive variables carries no physical meaning. Dividing meaningless totals by the number of components cannot reverse this outcome. In special circumstances averaging might approximate the equilibrium temperature after mixing, but this is irrelevant to the analysis of an out-of-equilibrium case like the Earth's climate.

Setting mixing aside, consider two disjoint isolated equilibrium systems, a and b , with functions of state $U^a = f^a(X_1^a, X_2^a, \dots, X_n^a)$ and $U^b = f^b(X_1^b, X_2^b, \dots, X_n^b)$ respectively, where U^a and U^b are the respective internal energies. X_j^a and X_j^b represent the corresponding extensive variables in systems a and b , respectively. Obviously, given f^a , the extensive variables for system a completely define the thermodynamic state of system a . Similarly for system b .

The partial derivatives, $\partial U^a / \partial X_j^a$ and $\partial U^b / \partial X_j^b$, are the j th intensive variables for the respective systems. The temperature of each system is, of course, the particular partial derivative with respect to the system's entropy,

$$T = \partial U / \partial S. \quad (1)$$

Together, the intensive variables form the tangent spaces of the respective functions of state. As such, they are local properties of the state space and are thus independent of the scale of the system.

Let us propose that an average over temperatures from both systems is required to be a temperature. This proposition produces a contradiction. The state, and the temperature, of system a , say, is completely determined by the variables $\{X_i^a\}$ and does not change in response to a change only in $\{X_i^b\}$. But any average is a function of both temperatures. Thus, while each temperature is a function of the extensive variables in its own system only, the average must depend explicitly on both sets of extensive variables, $\{X_i^a\}$ and $\{X_k^b\}$. That is it must depend on *both* states and it can change as a result of a change in either one. Then the average cannot be a temperature for system a , because system a is mathematically and thermodynamically independent of system b by assumption. Similarly, it cannot be a temperature for system b . Consequently, the average is not a temperature anywhere in the system, which contradicts the proposition that the average is a temperature.

While it is thus simple, obvious, and unavoidable that there is no one physically defined temperature for the combined system, the example illustrates the contradiction that arises in requiring an average over a local equilibrium temperature field to be itself a temperature of anything.

With reference to global temperature statistics, consider a projection of the full temperature field, $T(\mathbf{r}, t)$, onto a single real value, $\Xi(t)$, at time t , usually the present. For concreteness, consider an average of temperatures:

$$\Xi(t_M) = \frac{1}{N} \sum_{i=1}^N \left[T(\mathbf{r}_i, t_M) - \frac{1}{M} \sum_{j=1}^M T(\mathbf{r}_i, t_j) \right]. \quad (2)$$

Outside of small details, such as mappings of actual measurements onto a uniform grid on the surface, Ξ is a common statistical measure actually used in discussions of global warming, intended to quantify temperature changes over the entire globe at time t_M , by deviations from a sliding average in the past. This statistical measure is also known as the global temperature anomaly. The summation over i estimates a spatial integration of the field over the surface of the Earth, where $T(\mathbf{r}_i, t_M)$ is the air temperature at location \mathbf{r}_i and time t_M , and the summation over j accounts for a customary interval of time from the past to establish a local time mean (30 years is traditional).

Ξ is explicitly a function of $T(\mathbf{r}_k, t)$ for all \mathbf{r}_k (i.e., $\Xi = f(T(\mathbf{r}_1, t_1), T(\mathbf{r}_2, t_1), \dots, T(\mathbf{r}_1, t_2), T(\mathbf{r}_2, t_2), \dots, T(\mathbf{r}_N, t_M))$). In contrast, local equilibrium states in a field are defined at a particular location, \mathbf{r} . Other locations not in an infinitesimal neighborhood are independent, just as systems a and b are independent of each other. The statistical measure Ξ , on the other hand, exhibits the same property as an average of temperatures between the systems a and b : it depends on the states of all the local subsystems, not just the state at any one location.

Thus the same paradox arises as for systems a and b . If Ξ is supposed to be important at \mathbf{r} , it also depends explicitly on states at all remote locations $\{\mathbf{r}_k\}$. However, chemical, physical, or biological processes governed by temperature at \mathbf{r} are not functions of temperature other than at \mathbf{r} . So if one insists on Ξ defining processes like the melting of glaciers at some location, say, the melting is forced into being a function of temperatures elsewhere on the planet, too, which is physically untenable. Claiming otherwise is tantamount to temperature at a distance! While forces at a distance are considered an acceptable concept for classical gravity and electromagnetism, they are not suitable for thermodynamic processes.

It is worth noting in passing that Ξ suffers from the additional problem that it also depends explicitly on temperatures from the past. So in the picture implied by them, not only are the fate of glaciers today explicitly dependent on temperatures in a remote desert, say, but also on temperatures in that desert from twenty years ago.

On top of this problem, even if we let the two systems a and b come to equilibrium so that the temperature is the same everywhere and all possible definitions of average are equivalent (see Section 3), this common temperature will not be uniquely specified by the original states of a and b . It will also depend on the *path* taken by the process. Suppose two thermal systems a and b have equal heat capacities but initially different temperatures T^a and T^b . After putting them in thermal contact, they will equilibrate at the common temperature $(T^a + T^b)/2$. If on the other hand they equilibrate reversibly, i.e., while producing work, their common final temperature will be $\sqrt{T^a T^b}$.

Even in this simplest case, “averaging” between these two systems is not unique.

2.2. Average over temperature field as index or proxy

The argument is often heard that the global temperature, however it is calculated, while not being really what is physically driving the climate, is a good index or proxy for whatever does drive it. Science and engineering are used to such indirect measurements. Temperature itself, for example, is almost never measured according to its thermodynamic definition, $T = (\partial U / \partial S)_{V, N_1, N_2, N_3, \dots}$, but by measuring a volume, bending, electrical conductance, eigenfrequency of a crystal, radiation spectrum, etc. All such measurements rely on different assumptions being met. Physicists and chemists are well aware of these restrictions. However, no such physical arguments have been made for using a statistic as an index for climate driving force. Not even a statistical correlation has been forwarded for such a connection. ISO standards for measuring and calculating such an index were optimistically

promised after the Johannesburg climate meeting [8], and standards have been published regarding measurement of greenhouse gases, but the ISO has been conspicuously quiet on the measurement of a global temperature.

If temperature field averages are not temperatures, can they function as an index or as a proxy for climate dynamics? While an index need not be a physical variable explicitly, there are nonetheless unavoidable requirements before a statistic can be regarded as an index. An index is a statistic produced by a consistently applied rule that characterizes some, usually complex, process usefully, even though the precise connections between it and the process may not be fully known or understood. As an index, the statistic usefully encapsulates a possibly large set of data into one dimension for a specific purpose.

Proposing a statistic to be an index can be risky. What is regarded as “useful” may prove to be quite imprecise and subjective. Therefore, the burden to demonstrate the utility of a statistic as an index must lay with those who propose it as one. Similarly, the burden should never be to prove a statistic is not an index.

However, a peculiar circumstance exists for the case of averages over the temperature field. False, but common, impressions that such averages are actual physical variables generally, and temperatures in particular, have left the task of demonstrating the utility of such statistics undone. The apparent physical origins have led to the most uncritical thinking about what trends in such statistical measures or statistics might imply.

While it is impossible to prove that such averages are not useful in any manner whatsoever, it can be shown that temperature field averages cannot be useful in certain specific roles in which they are cited. For example, upward trends in Ξ are said, uncritically, to indicate trends in underlying dynamics, such as changes in the numbers or severity of storms, and trends in rainfall, not to mention melting of glaciers or many other local physical dynamical processes.

What connection exists between these dynamical processes and such averages? No physically precise reasoning has been proposed as the basis of such a connection. Indeed, it is unlikely that there is any such connection because intensive variables, like temperature, do not generally drive dynamics, and in particular do not drive the dynamics of the atmosphere and oceans, whose dynamics are caused by gradients, or differences in thermodynamic intensities, rather than the intensities themselves.

To be more explicit, consider a free thermodynamic mixing process between two subsystems a and b , isolated from everything except each other. In the absence of sources and sinks due to chemical reactions, extensive quantities

summed over subsystems a and b , except entropy, will not change. The change in entropy is equivalent to the existence of the process:

$$dS = \left(\frac{1}{T^a} - \frac{1}{T^b} \right) dU^a + \left(\frac{P^a}{T^a} - \frac{P^b}{T^b} \right) dV^a - \left(\frac{\mu_1^a}{T^a} - \frac{\mu_1^b}{T^b} \right) dN_1^a + \dots, \quad (3)$$

where T is the temperature, P the pressure, and μ_i is the chemical potential of species i . Superscripts refer to the subsystems a and b . When the subsystems are in equilibrium with each other, all intensities are the same in both systems ($T^a = T^b$, $P^a = P^b$, etc.). In that case dS vanishes and nothing happens. Nowhere in Eq. (3) do intensities enter in isolation. Differences cause processes.

It is true that phase changes and rates of chemical and biological processes depend parametrically on (local) absolute temperature. However, if there are no differences in intensities, like differences in chemical potential for chemical reactions, thermodynamic processes would not happen in the first place. Even the sensation of warming or cooling is not a consequence of a single temperature, but rather because of the thermodynamic forces that are set up between the intensities in our bodies and those of our surroundings. Thus two people can disagree on whether a given room is cold or hot.

If absolute intensities do not drive climate dynamics, there is no reason to expect that averages over them will. Moreover, at least for small changes, an average over the temperature field can be insensitive to climate change. Because of the local independence of functions from their gradients, an average can change in time without any appreciable changes in dynamics, while great climate changes can be envisioned under a constant average.

For example, simply rotating thermodynamic intensity fields with respect to the Earth's surface leaves field averages invariant but can imply significant climate change. In the special case of Ξ , invariance only requires changes that leave the sums in Eq. (2) unchanged. There is an infinite range of possibilities of such changes with Ξ invariant that would constitute significant climate change. Indeed, an overlooked type of climate modeling experiment is climate change under fixed temperature field averages.

Ultimately, it can be no surprise that local states are not well represented by a globally based statistic, casting further doubt on the role of Ξ as an index. It is also worth remembering that an invariant Ξ would not generally imply invariance for other temperature field statistics, from different averaging rules. As there are an infinite number of such statistics extractable from a field, there is also the question of whether the different statistics convey the same message or not, if any is conveyed at all. It is difficult to claim a statistic is a useful index when many very similar statistics seem to say something different.

Thus the proponents of Ξ as an index have been able to avoid two crucial tasks:

- i) They have not been required to justify or demonstrate the physical usefulness of a statistic extracted from the temperature field.
- ii) They have not answered why Ξ is the statistical measure of choice and not some other extract of the temperature field with potentially very different behavior.

The very different behaviors of other statistics are discussed in Section 3.

2.3. Field average as energy proxy

If a field average is not a temperature, and it is not an index, can it be simply a direct proxy for internal energy instead? This calls for statistics like Ξ to be given up as drivers of climate or as indicators of local dynamics. Instead, temperature is to be viewed as equivalent to energy, and thus the numerator of the right side of Eq. (1) is to represent changes in the global internal energy. Then any physical meaning is born by changes in energy instead of temperature.

This strategy has two problems. First, substituting temperature for internal energy amounts to extensifying an intensity, and second, the role for a global averaged internal energy in climate is unclear, especially given the few percent that we could expect it to vary.

Despite popular thinking otherwise, temperature and energy are not equivalent. Temperatures can be very high at very low energies. While *heat* is a form of energy, temperature is, fundamentally, a measure of how energy is spread over quantum states. For example, radiation from a small laser powered by flashlight batteries can have temperatures peaking as high as $\sim 10^{11}$ K. This is higher than many stellar interiors but one cannot even feel the heat of the beam on one's hand [6].

The myth of the equivalence of temperature and energy comes from the proportionality of internal energy U to temperature T , through a heat capacity. In the case of an ideal gas with no internal molecular structure, we have

$$U(N, T) = \frac{3}{2} NkT. \quad (4)$$

While Eq. (4) is not an equation of state in extensive variables alone, it is easily derived from one by differentiating it according to the definition Eq. (1). In the result, U is not only proportional to T but to N as well, so, being independent, changes in T are not equivalent to changes in U .

Introducing a local version of Eq. (4) by dividing through by volume, V ,

$$u(n, T) = \frac{3}{2}nkT, \quad (5)$$

where $u \equiv U/V$ and $n \equiv N/V$, we see the same issue holds locally: measuring changes in T cannot determine changes in u in environments where the number density is also variable.

But are changes in n large enough to matter for the processes in this case? Using Boyle's law and Eq. (5),

$$\frac{du}{u} = \frac{dP}{P} = \frac{dn}{n} + \frac{dT}{T}. \quad (6)$$

As dP/P is about 5% across the entire Earth's surface while dT/T would be closer to 20%, variation of n cannot be ignored, globally speaking, in comparison to variation in T for the processes actually in question. As T varies about four times as much as u over the Earth, temperature cannot be a proxy for internal energy in climate. In fact, these numbers suggest that places of high T tend to correspond to low n , which agrees with the well-known tendency of aircraft to have diminished takeoff performance at high air temperatures.

Even if trends in global U were measurable and known (they are not), the question remains what it would mean for climate dynamics. Like variations in Ξ , significance is simply presumed without further consideration.

3. Averaging without physics

The notion of averaging exists independently of physics. Loosely speaking, it is nothing more than holding one mathematical object to stand in for many. Usually we think of a single number standing in for a set of numbers or standing in for a more complex mathematical object. While having a single representative taken in place of many can be convenient, clearly something is lost in it. The representative may not exhibit what is physically relevant, as the selection rule need not be set by the physics.

Setting aside the statistical analysis of physical data, averages arise within physical theory in two basic ways. They can emerge, for example, as an interpretation of a previously existing structure rather than as an intrinsic property. In the case of quantum mechanics, the term *expectation value* is used for an eigenvalue that characterizes a mathematical operator and thus a quantum state. The expectation value has the character of, if not the name of, an average, but it need not be interpreted as one. The interpretation is definitely after the fact.

In other physical applications, averages arise in a somewhat *ad hoc* manner. They are simply sums over physical quantities that have been normalized. In principle, normalization is a convenience. It is not essential, and so neither is the notion of averaged quantities essential to the mathematics in such physical theories. What is essential is that the sums have meaning in the first instance.

As we saw in the preceding, sums are not generally meaningful over all variables. Temperature, pressure, and chemical potential are just a few of the exceptions. Nonetheless, it may happen that a variable is not additive, while a transformation on that variable is. For example, a variable defined by the square of internal energy is not additive, but its square root is. The physics thus presents a coordinate system for that quantity that is preferred over another. Similarly, normalization of the corresponding sums results in a preferred average.

However, for thermodynamic intensities there is no preferred coordinate system. The sum has little physical meaning, so all transformations are in play. Normalized sums are not invariant under all transformations. Thus more than one average is unavoidable. Statisticians have long known this. Unlike physicists, they work with data unconnected to underlying physical theory and have uncovered classes of averages that reflect this. One is faced with having an infinity of averages, having all possible values within the range of the data set.

The most well known of the class are the Hölder or power means [16]. These arise from transforming the mean under simple powers. Alternatively, they can be deduced by varying geometric assumptions about what constitutes the length of a vector. This method has pedagogical significance because it forces an enlightening conflict between conventional biases: one cannot have both the classical mean and Euclidean distance. This class of average is by no means exhaustive.

This section makes three contributions to a thermodynamic discussion of non-equilibrium fields:

1. A simple thermodynamic example is shown where different averages over given temperatures exhibit contradictory time trends. Thus the notions of warming and cooling are problematic for non-equilibrium temperature fields.
2. Global conditions are deduced for when it is possible to say, unequivocally, that one temperature field is warmer or cooler than another.
3. It is demonstrated that any value of a data set can represent the mean in some coordinate system. If there is no preferred coordinate system given for the average because the physics does not provide one, then any value in the data set can stand as an average of the temperature field.

3.1. Warming and cooling simultaneously?

3.1.1. The averaging operator and the origin of contradictory trends For the case of a finite data set, an averaging operator, \mathcal{A} , maps the set onto a real number within the domain of the original data. Clearly, a value that does not satisfy this requirement would be a pathological representative of the set. In the case of temperature data, the domain of \mathcal{A} is the range of data from the original field.

Thus the resulting real number is within the interval induced by the original set. This definition is still quite general. All of the well-known (standard) averages, such as the mean, the mode, the median, the root mean square, the harmonic mean, the geometric mean, or the infinity of nonstandard, unnamed others satisfy this requirement, without necessarily being equivalent to each other. It also follows from this definition that a set made up of identical data values can have only a single average value, independent of the type of averaging operation. That single value corresponds to the single repeated data value.

The aim of these mappings is of course beyond the mathematics and beyond any statistical theory. Neither mathematics nor statistics suffice to determine whether any result of such mappings achieve any aim. Moreover, the data and its underlying physics do not depend in any way on \mathcal{A} .

On the basis of the definition, \mathcal{A} is not a unique operation unless some additional rules, \mathcal{R} , beyond the fundamental definition are specified (cf. the importance of path followed as discussed at the end of Section 2.1). Thus if we average over a set ϕ to get x , we must specify the operation, $\mathcal{A}_{\mathcal{R}}$ in terms of \mathcal{R} ,

$$\mathcal{A}_{\mathcal{R}}(\phi) = x. \quad (7)$$

If there are two different rules, \mathcal{R}_α and \mathcal{R}_β , and one set, ϕ , then in general the averages are different,

$$\begin{aligned} \mathcal{A}_{\mathcal{R}_\alpha}(\phi) &= x_\alpha \\ \mathcal{A}_{\mathcal{R}_\beta}(\phi) &= x_\beta, \end{aligned} \quad (8)$$

where $x_\alpha \neq x_\beta$. Furthermore, if the ϕ is a function of t (i.e. $\phi(t)$), then $\dot{x}_\alpha(t)$ and $\dot{x}_\beta(t)$ can have opposite signs. This idea is illustrated in the following physical example.

3.1.2. A physical example of contradictory trends Let us consider a specific example involving temperature. A glass of ice water at 2°C is sitting beside a

cup of coffee at 33°C. The two remain isolated, but are allowed to relax to room temperature, which is 20°C, according to Newtonian cooling (heating). To complete the example, a plausible relaxation time of eight minutes for each container was set for the sake of this illustration, but the phenomenon we will find is not unique to this value. In this manner, the ice water is allowed to warm, while the coffee cools accordingly.

For this example, the two independent temperatures were averaged in four different ways. They are not exhaustive by any means. Furthermore, examples with other temperature units and other averages may be formulated, but these would not add materially to the value of the example.

The averages used are:

$$\begin{aligned}
 \mathcal{A}_{\mathcal{R}_{-1}}(T_1, T_2) &= \left(\frac{T_1^{-1} + T_2^{-1}}{2} \right)^{-1}, \\
 \mathcal{A}_{\mathcal{R}_1}(T_1, T_2) &= \left(\frac{T_1 + T_2}{2} \right), \\
 \mathcal{A}_{\mathcal{R}_2}(T_1, T_2) &= \left(\frac{T_1^2 + T_2^2}{2} \right)^{1/2}, \\
 \mathcal{A}_{\mathcal{R}_4}(T_1, T_2) &= \left(\frac{T_1^4 + T_2^4}{2} \right)^{1/4}.
 \end{aligned} \tag{9}$$

\mathcal{R}_{-1} is the harmonic mean. It is preceded, for example, in the case of temperature in connection with minimum entropy production and the radiation field [17]. It also appears in connection with average resistance in a parallel circuit and in average travel times on a road network with varying speed limits.

\mathcal{R}_1 is the simple mean, often used over small temperature ranges and in connection with simple Newtonian heat exchange.

\mathcal{R}_2 is the root mean square, which is well preceded in statistics and statistical mechanics. In the latter case, it appears particularly in connection with kinetic energy. It can also emerge in connection with the potential energy of a spring.

\mathcal{R}_4 would appear in connection with black body radiation.

The results of the Newtonian cooling calculations are shown in Figure 1. It is clear that the starting “system temperature” varies widely depending on the averaging formula chosen. Because of the property that all averages must be the same for a set composed of equal values, all averages must approach the room temperature. All of these averages clearly exhibit this proper behavior.

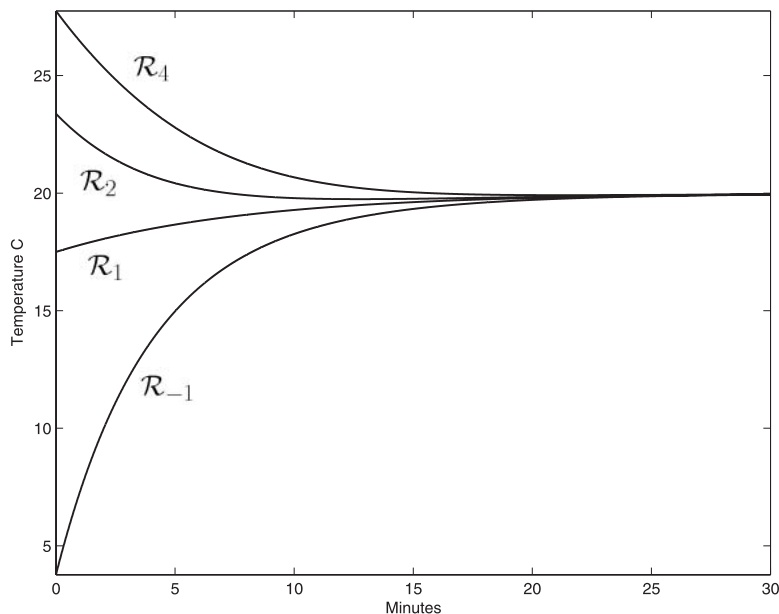


Figure 1 Four averages over one thermodynamic system. The rules for each, denoted by \mathcal{R}_{-1} , \mathcal{R}_1 , \mathcal{R}_2 and \mathcal{R}_4 are defined in Eq. (9).

Clearly, whether the system seems “warmer” or “cooler” at time $t = 5$, say, in relation to $t = 0$ depends on the average chosen. But the data are independent of the averaging rule used, therefore the sign of the derivatives is not intrinsic to the data, but a property of the averaging rule selected.

If the physics does not prescribe one rule to be used over another, as it does not for temperature, we may use any rule. If one interpreter of the data chooses one rule, while another chooses a different rule, there is no way to settle a disagreement as to whether the system is getting “warmer” or “cooler” with time.

Alternatively, if one has a change of opinion about what rule to use, the ranking of the two systems may be instantaneously reversed. In a mathematical sense, this would just be a quantitative alteration in the time derivative. However, “warming” versus “cooling” is a qualitative distinction for thermodynamics. Changing the sign of the derivative reverses the ranking of the two states catastrophically, in the sense of a qualitative change that is extrinsic to the system studied. That is, rank order is catastrophically reversed by simply changing how the data are interpreted – a rank order catastrophe.

There are only two options: admit the possibility that non-equilibrium systems can simultaneously warm and cool or take the position that these terms

have no meaning in such cases. We adopt the latter convention. In either case, the ice water is warming and the coffee is cooling, and while that is happening they are not in thermodynamic equilibrium with one another or with the room. There is no physically coherent way to reduce this dynamic to changes in a single temperature number for the system as a whole.

Arbitrary averages can and do give contradictory behaviors in more complex cases, too, as we shall see below. Averaging does not represent any means of avoiding the fact that a system is not in global thermodynamic equilibrium. Thus, in the case of non-equilibrium thermodynamics, temperature averages fail in the most basic role of an average, which is for one value to represent many.

3.2. Range overlap and global ranking of fields

In Sections 3.1.1 and 3.1.2 it was shown that objects out of equilibrium cannot necessarily be compared to each other as being “hotter” or “cooler”. It makes problematic the claim that Earth’s temperature field is warmer or cooler today than it was a hundred years ago, or that one century is hotter than another century.

In contrast, it is obviously valid to make the observation that the temperature field of the Sun is hotter than the temperature field of Pluto, yet each of these bodies have non-equilibrium temperature fields. What makes this comparison different? The simple answer is that there are no common values in the respective temperature fields. This was not the case for the example in the preceding section, where the interval spanned by the range of temperature at time $t = 0$ contained the interval spanned at later times.

In the case of a temperature field on Pluto, the Sun, or the Earth, we work with a subset of the field which constitutes a finite number of values. This is a practical necessity even if it is not a mathematical one. The range of values from the field induces the domain of the averaging operation. This domain is an interval, which according to the definition in 3.1.1 limits what an average can be. Consider the interval operation, \mathcal{I} , which defines the domain of the data set ϕ . For two non-overlapping data sets ϕ_1 and ϕ_2 ,

$$\mathcal{I}(\phi_1) \cap \mathcal{I}(\phi_2) = \emptyset. \quad (10)$$

This condition is global in that it is entirely independent of specific locations in the original field. It follows that

$$\mathcal{A}_{\mathcal{R}}(\phi_1) - \mathcal{A}_{\mathcal{R}}(\phi_2) \quad (11)$$

has no zeros over all possible rules \mathcal{R} satisfying the definition in Section 3.1.1. No zeros in Eq. (11) means it never changes sign, so there cannot be a rank order catastrophe as \mathcal{R} is varied over all rules. If the global condition Eq. (10) holds, then it is not ambiguous to say that one set is warmer or cooler than the other. In the case of the Sun and Pluto, it is reasonable to expect that Eq. (11) holds, so that we may say that the Sun is hotter than Pluto without contradicting the preceding.

But if

$$\mathcal{I}(\phi_1) \cap \mathcal{I}(\phi_2) \neq \emptyset, \quad (12)$$

we will say that the fields yielding ϕ_1 and ϕ_2 have *range overlap*. In the case of range overlap, we cannot guarantee that there are no zeros in Eq. (11) and therefore there can be rank order catastrophes, which means that the ranking of temperature fields as “warmer” or “cooler” is fundamentally problematic. It is possible for Eq. (11) to have no zeros over all reasonable \mathcal{R} in cases where Eq. (12) holds if a *local* condition is applied such that every location in the field is increasing or decreasing. This condition does not occur in realistic fields such as that of the Earth, and it is not global.

In the case of the Earth’s temperature field, the temperatures form a set of values, ϕ_{\oplus} . The field at time t_1 is being compared to the field at time t_2 . In that case, Eq. (12) holds. In fact, this is among the most extreme cases of range overlap because,

$$\mathcal{I}(\phi_{\oplus}(t_1)) \approx \mathcal{I}(\phi_{\oplus}(t_2)), \quad (13)$$

from which we conclude that ranking the Earth’s temperature fields at two instants in time is highly problematic. It simply is not comparable to the case of Pluto versus the Sun.

3.3. Coordinate transformations and means

A naive view suggests that the simple mean is always the appropriate scalar summary for any set of data, including temperature data. But this is untenable for reasons more general than those related to the thermodynamic issues discussed above.

Data have no intrinsic meaning. The context establishes the meaning. It also establishes whether the numbers themselves are used or some other numbers derived from the raw data. “Derived” data always raises the issue of coordinate transformations on the raw data. As noted above, temperature data used in climate measurement are inherently derived, since air temperature is never

measured using the thermodynamic definition Eq. (1), but using instrument-based proxies such as length, volume, electrical conductance, etc., not to mention natural proxies such as tree ring widths and wood density, ice core layer isotope ratios, and so forth.

Even in the absence of a distinction between raw and derived data, questions about the natural form of an average emerge. A context for a data set may take the form of an auxiliary structure on the data that induces a geometry. The geometry can imply the natural form of an average that can differ from the mean. In the absence of a context, different norms on vectors formed from data can induce a class of averages, which have the same mathematics as a range of transformations between raw and derived data. The change from raw to derived data inevitably involves mappings and coordinate transformations.

One physical quantity is measured, but another physical property is sought. The two quantities are linked by a relationship that amounts to a coordinate transformation. Consider a case where basic observational data must be transformed to acquire the quantity for which an average value is sought. Suppose that we measure a raw set $\{y_i\}$ with N positive elements from which we derive a set $\{x_i\}$ by a nonsingular monotone transformation.

In examples such as kinetic speed or radiation energy as raw values, an average over derived kinetic energies or derived temperatures would be expressed in terms of a power law. The raw and derived data are related by $y = x^r$ for positive r , then the mean of $\{y_i\}$ is equivalent to the mean over $\{x_i^r\}$,

$$\mathcal{A}_{R_1}(\{y_i\}) = \mathcal{A}_{R_1}(\{x_i^r\}) = \frac{1}{N} \sum_i x_i^r. \quad (14)$$

But the mean over the derived data is

$$\mathcal{A}_{R_1}(\{x_i\}) = \frac{1}{N} \sum_i x_i. \quad (15)$$

Thus, in either case there are two data sets and two simple means – one for each of the raw and derived sets. For the radiation example, we essentially measure $\{T_i^4\}$, ignoring constants of the classical black body law. To get temperature, we must transform the measurements by taking the 1/4 power. That is $r = 4$. Similarly, for the speed problem we measure $\{|v_i|\}$ and take the square, which means that $r = 1/2$.

But what is not mandated by the naive view is which mean to work with. That is, do we use the transformation to compute $(\mathcal{A}_{R_1}(\{y_i\}))^{1/r}$ or do we

compute $\mathcal{A}_{R_1}(\{y_i^{1/r}\})$ instead? Do we take the mean first and then do the transformation, or do we do the transformation and then take the mean? Statistical theory does not provide an answer.

Outside of exceptional cases,

$$(\mathcal{A}_{R_1}(\{y_i\}))^{1/r} \neq \mathcal{A}_{R_1}(\{y_i^{1/r}\}), \quad (16)$$

i.e., the mean operation does not generally commute with the powers. As a consequence, the mean is not a coordinate system invariant.

3.3.1. Vector norms, geometry, and averaging The most important exception to Eq. (16) is when all the elements of $\{y_i\}$ are identical to each other. In this case, all acceptable definitions of averages must yield the same value. Accordingly, for the degenerate class of sets where all elements are identical, averages must also be invariants under transformations. This degenerate class turns out to tell us about the geometry of averages.

Suppose one aims to find the average length of streets in a city. Let us use $\{x_i\}$ to denote the set of lengths of the city's streets. Form a vector from this set, $\langle x_1, x_2, \dots, x_N \rangle$. Vectors can have lengths in their own right. Seek another vector with the same length but having all elements being equal. The equicomponent vector that results is $\langle \mathcal{A}_{R_1}(\{x_i\}), \mathcal{A}_{R_1}(\{x_i\}), \dots, \mathcal{A}_{R_1}(\{x_i\}) \rangle$, provided the vector's length is computed as the sum of the (positive) elements (e.g. $x_1 + x_2 + \dots + x_N$), which is the sum of the individual lengths of the streets. In this case, the traditional mean emerges as the average, and the length of the vector is nothing more than the summation of lengths of all of the city streets, which seems to make particular sense for this case.

It is notable here that the naive view of the mean runs afoul of the naive view of length. The latter would hold that the vector's length ought to be Euclidean (i.e., $(x_1^2 + x_2^2 + \dots + x_N^2)^{1/2}$). But an equicomponent vector of equal Euclidean length is $\langle \mathcal{A}_{R_2}(\{x_i\}), \mathcal{A}_{R_2}(\{x_i\}), \dots, \mathcal{A}_{R_2}(\{x_i\}) \rangle$ or $\langle (\mathcal{A}_{R_1}(\{x_i^2\}))^{1/2}, (\mathcal{A}_{R_1}(\{x_i^2\}))^{1/2}, \dots, (\mathcal{A}_{R_1}(\{x_i^2\}))^{1/2} \rangle$, which does not yield the mean. It yields the root mean square. One cannot normally have the mean and an Euclidean geometry for data.

Let us suppose, without loss of generality, that some data set, $\{x_i\}$, has positive elements. Just as this data does not necessarily have a context, vectors like $\langle x_1, x_2, \dots, x_N \rangle$ need not possess a mathematical geometry (an inner product) and hence any particular length. In choosing a length, one induces a geometry and an average. The vector lengths or norms, discussed above, both belong to a class of norms known as ℓ -norms, ℓ_r , which depend on a parameter, r ,

$$(\ell_r)^r \equiv \sum_i x_i^r = N \mathcal{A}_{R_1}(\{x_i^r\}). \quad (17)$$

This class of norms is not exhaustive, but it is sufficient for our discussion. For a data set $\{x_i\}$ the equicomponent vector arising from any such ℓ -norm is

$$\left\langle \frac{\ell_r}{N^{1/r}}, \frac{\ell_r}{N^{1/r}}, \dots, \frac{\ell_r}{N^{1/r}} \right\rangle. \quad (18)$$

A component of Eq. (18) is the *Hölder mean*. Note that for the city streets example the length is ℓ_1 , which is known, appropriately, as the “taxicab” or “city streets” norm. The Euclidean case was ℓ_2 . All other positive values of r induce different geometries.

From this we conclude

$$(\mathcal{A}_{R_1}(\{y_i\}))^{1/r} = \frac{\ell_r}{N^{1/r}} \neq \mathcal{A}_{R_1}(\{x_i\}), \quad (19)$$

where $y_i = x_i^r$ was substituted so that we could observe that Eq. (19) is the same equation as Eq. (16). However, in this section we have not presumed a distinction between raw and derived data. The issue is the underlying geometry, which brings exactly the same transformation group into play even if transformations are not introduced at the outset. Considering the geometric context generates the same infinite family of distinct averages as emerged in the previous section.

3.3.2. Averages in general This infinite family of averages agrees with the basic definition from Section 3.1.1. If we take the limit as $r \rightarrow \infty$ then $\ell_\infty = \max\{x_i\} \equiv x_{max}$, which is known as the ℓ_∞ -norm or max-norm. According to Eq. (18), the average becomes x_{max} , too. Thus it is not only the largest element of the set, but it follows that it is the largest of the family of averages, as r is varied over positive values:

$$\frac{\ell_r}{N^{1/r}} = \left(\frac{\sum_i x_i^r}{N} \right)^{1/r} \leq \left(\frac{N x_{max}^r}{N} \right)^{1/r} = x_{max}. \quad (20)$$

For $r = 0$ we find $\ell_r/N^{1/r}$ becomes the geometric mean in the limit: $\exp[(\sum_i \ln x_i)/N] \leq \exp[(N \ln x_{max})/N] = x_{max}$.

Negative r is easily treated by forming the set $\{w_i\}$ where $w_i \equiv x_i^{-1}$. Then w_{max} will be the largest average over $\{w_i\}$ for positive r . As $\min\{x_i\} = 1/w_{max}$, then $x_{min} \equiv \min\{x_i\}$ is the smallest average for negative r . As it is also smaller

than the geometric mean, because $\exp[(\sum_i \ln x_i)/N] \geq \exp[(N \ln x_{min})/N] = x_{min}$, it is the smallest average for all r .

Thus any value from the entire interval, $\mathcal{I}(\{x_i\})$, is admissible as “the” average, even values not in the original set but within the interval. This agrees with the basic notion of the average from Section 3.1.1. Any member of the original set can be representative of the whole set – very democratic.

Even though we have resurrected the entire data domain of the data set to be the range of possible averages, we have not encompassed all averaging rules. Potential averaging rules do not end with simple powers. With the harmonic mean as a clue, any non-singular monotone function, $z(\cdot)$, will do as long as it has an inverse, $z^{-1}(\cdot)$,

$$z^{-1}(\mathcal{A}_{\mathcal{R}_1}(\{z(x_i)\})) \equiv \mathcal{A}_z(\{x_i\}). \quad (21)$$

Even this cannot be said to be comprehensive in itself as it does not yet include weighted averages, nor have we extended the notion to a continuum, which we must face for the atmosphere, at least at a formal level. A little bit of imagination will widen the possibilities substantially.

Nonetheless, this class is already more than broad enough for our purposes. In fact, we will reduce it slightly to use families of averages that include the ℓ -norm averages, which are sometimes called the “mean of order r ” family, hereinafter “ r -means” for convenience. We can also introduce the “mean of order s ” family, hereinafter “ s -means” [18], where these families of symmetric means are important in the theory underlying index numbers in economics. The s -means are given by $z(\cdot) = \exp(s(\cdot))$, or the resulting average is

$$\frac{1}{s} \ln(\mathcal{A}_{\mathcal{R}_1}(\{e^{sx_i}\})) \quad (22)$$

or

$$\begin{aligned} r\text{-mean: } \mathcal{A}_{\mathcal{R}_r} &\equiv \left(\frac{1}{N} (x_1^r + \dots + x_N^r) \right)^{1/r} \\ s\text{-mean: } \mathcal{A}_{\mathcal{R}_s} &\equiv \frac{1}{s} \ln \left(\frac{1}{N} (e^{sx_1} + \dots + e^{sx_N}) \right). \end{aligned} \quad (23)$$

These means will be used in the next section.

4. Averages and actual atmospheric data

As discussed previously, the temperatures of the Earth form a continuous field, $T(\mathbf{r}, t)$, which varies in time. However, observations can only be made

at specific locations, $\{\mathbf{r}_i\}$, and times $\{t_j\}$. While the desideratum of the global averaging operation would seem to be a component of an equicomponent vector that has the same length as the temperature vector extracted from the field itself as per Section 3.3.1, it is still an intensive field that provides no definite length. One might naively expect something like $\mathcal{A}_{\mathcal{H}_1}(T(\mathbf{r}_i, t_j))$ to be the average used. However, what is actually done is much more arcane than any example of averaging operation discussed in the preceding sections [1, 2, 4, 7, 19].

The averaging rule varies in time. For example, there has been a substantial drop in the number of available weather stations since the 1970s, and especially since 1990, when the number of sampling points around the world fell roughly in half over four years ([7], Figure 2). The resulting land-based average is then combined (in some cases) with a series constructed using samples of seawater temperature collected during the twentieth century by ships, using a combination of bucket-and-thermometer observations and automated engine cooling-water observations.

Observations are not made at a common instant, t_j , for the whole field, as one might naively expect in the first instance. Instead they have a common day, d , and measurements are made in terms of local time of day time producing a set,

$$\{T_{\mathbf{r}_i}^{\tau, d}\} = \mathcal{O}_t(T(\mathbf{r}, t)), \quad (24)$$

where \mathbf{r}_i is place and τ is time on day, d . The operation \mathcal{O}_t is the time-dependent operator that takes the full field to the set.

Land-based meteorological stations around the world provide records of daily maximum and daily minimum temperatures, denoted $\max(T_{\mathbf{r}_i}^{\tau, d})$ and $\min(T_{\mathbf{r}_i}^{\tau, d})$. The daily average is defined as $(\max(T_{\mathbf{r}_i}^{\tau, d}) + \min(T_{\mathbf{r}_i}^{\tau, d}))/2$. Most land-based stations only report this type of average, although some provide the min/max.

From Section 3.3.1, this average is the mean of the ℓ_∞ and $\ell_{-\infty}$ Hölder means over daily temperatures. This combination is a new type of average in terms of this paper, which naturally can be expected to have distinctive behaviors. We introduce a new averaging operator $\mathcal{A}_{\mathcal{H}}^x$, which only averages over x holding other variables fixed, thus producing a set and not just a real number as output. In this case, \mathcal{A}_∞^τ indicates the average is over τ and we denote the mean of the Hölder means by ∞ .

Clearly, based on Section 3 and particularly 3.1.2, it may be problematic to determine whether one location is warmer or colder than another on any particular day with this statistic. Thus the observational network actually provides

$$\mathcal{A}_\infty^\tau(\mathcal{O}_t(T(\mathbf{r}, t))). \quad (25)$$

At every \mathbf{r}_i the mean of the daily values is taken over a month. The resulting series of monthly values is translated by subtracting the mean of the monthly means themselves further averaged over a thirty-year interval (nowadays 1961–1990 is used). These reduced values, known as anomalies, are then subjected to *ad hoc* adjustments which speculatively reduce the effects attributable to any land-use changes since before human settlement in the area, effects due to changes in equipment or measurement site, missing stretches of data, etc.

The adjusted anomalies are then subjected to a mapping to yield estimates of what the anomalies might be if the actual observation sites were distributed with a uniform density over the surface of Earth. The average series are grouped into latitude–longitude boxes and averaged up to a “gridcell” level, with weights applied so that the variance is not a function of the number of individual stations in the gridcell. In this way, records from isolated stations count more heavily than records in more densely sampled locations.

These translations to “anomaly” coordinates and the other mappings that speculate on what the values of observations might be at different locations not measured, or under different conditions not known, are denoted by \mathcal{M} ,

$$\mathcal{M}(\mathcal{A}_\infty^\tau(\mathcal{O}_t(T(\mathbf{r}, t)))). \quad (26)$$

Finally, after taking the simple mean, we arrive at the global temperature statistic Ξ that is actually in use,

$$\Xi = \mathcal{A}_{\mathcal{R}_1}(\mathcal{M}(\mathcal{A}_\infty^\tau(\mathcal{O}_t(T(\mathbf{r}, t)))).) \quad (27)$$

This statistic is clearly more complex than the Ξ introduced in Eq. (2). It is far removed from $\mathcal{A}_{\mathcal{R}_1}(T(\mathbf{r}, t))$, which Ξ is widely confused with. There is no evidence that trends in $\mathcal{A}_{\mathcal{R}_1}(T(\mathbf{r}, t))$ need agree with that of Ξ , nor is there a physical reason to prefer one over the other.

Debates about the difference between trends in Ξ and Y , where Y is the output of an entirely different averaging map of microwave fluxes, \mathbf{F} , received from satellites, are also problematic. There is no physical reason why Ξ must equal Y or even have the same trends. Ξ and Y are entirely different statistics in a mathematical sense, even if they each are mapped from the same $T(\mathbf{r}, t)$ for the Earth.

So far we have shown that different averages exist, they are used, and that contradictory trends can emerge between them. We have shown that the conditions exist in the atmosphere where such paradoxical behavior can be expected to be found. In the following we show that it does in fact appear in real atmospheric data.

4.1. Contradictory trends in global temperature averages

Ranking a particular type of field average computed over a sequence of times amounts to determining a trend in that average. Here we show that the sign and size of such a trend computed statistically is dependent on the choice of averaging rules, which will suffice to demonstrate both that the “global temperature trend” is not a unique physical variable, and that ranking this or that year as the “warmest of the millennium” is not possible, since other averages will give other results with no grounds for choosing among them.

To illustrate with actual temperature data, we computed averages of temperatures over twelve sites (see Table 1) and computed a linear trend in each case. The trends through the 1979–2000 period were computed with r -means and s -means. The raw data are themselves averaged (simple monthly means: $r = 1$) smoothing out some variability, but this could not be avoided, and does not affect the main results below. Stations were selected to give reasonable geographic variation, but whether it is a “global” sample or not is secondary for the purpose of the example. Stations had to be in continuous use during the 1979–2000 interval. Missing months were interpolated linearly as long as there was no more than one missing month in sequence, and it was not at the start or finish of the sample.

For each value of r , s (cf. Eq. (23)), the monthly r , s -means across the stations were computed, then a linear trend was fitted using ordinary least squares after deleting rows with missing data. The trend values are plotted in Figures 2 and 3. For the simple mean ($r = 1, s = 0$) the decadal “warming” trend was $0.06^\circ\text{C}/\text{decade}$. This turns out to be the peak value of the trend: for most values of r and s the trends are negative, indicating “cooling” across the 1979 to 2000 interval.

It might seem contradictory that the same data show “global warming” of about $0.02^\circ\text{C}/\text{decade}$ for $s = 0.04$, but “global cooling” of $-0.04^\circ\text{C}/\text{decade}$ for $s = -0.04$. But there is no contradiction in the data: They do not show “global” anything. The data are local. The interpretation of “global” warming or cooling is an artificial imposition on the data achieved by attaching a label to, respectively, a positive or negative trend in one particular average.

Table 1 Twelve weather station records used for Figures 2 and 3. The data are monthly averages from the Goddard Institute of Space Studies (GISS) [3].

Phoenix, Arizona	Sioux Falls, South Dakota
Cartagena, Colombia	Egedesminde, Greenland
Dublin, Ireland	Salehard, Russia
Chiang Mai, Thailand	Ceduna, Australia
Jan Smuts, South Africa	Halley, Antarctica
Honolulu, Hawaii	Souda, India

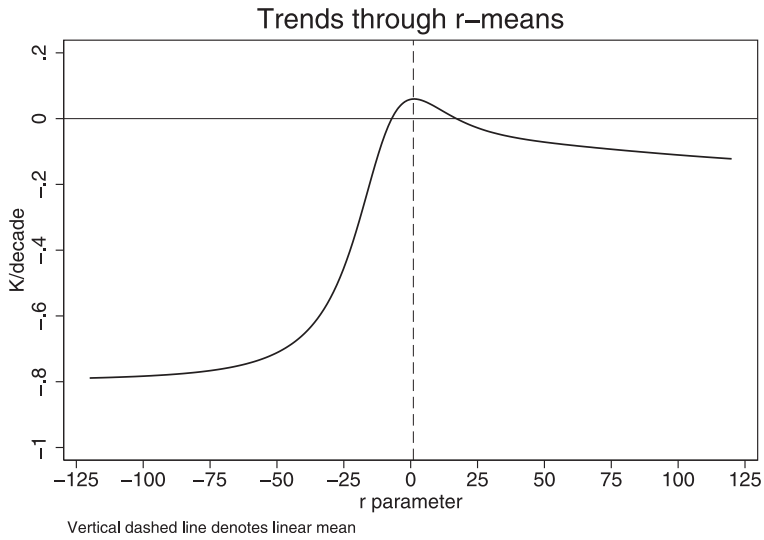


Figure 2 Trends (in K/decade) through r -means from 12-station sample.

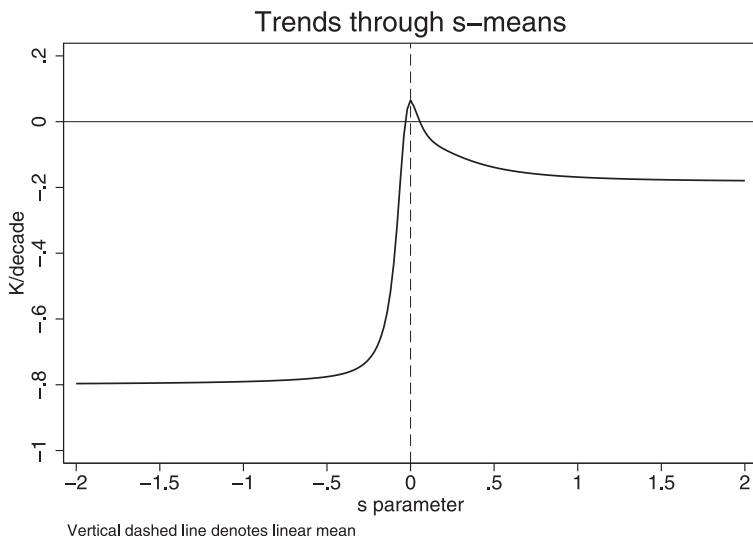


Figure 3 Trends (in K/decade) through s -means from 12-station sample.

5. Conclusion

There is no global temperature. The reasons lie in the properties of the equation of state governing local thermodynamic equilibrium, and the implications cannot be avoided by substituting statistics for physics.

Since temperature is an intensive variable, the *total* temperature is meaningless in terms of the system being measured, and hence any one simple average has no necessary meaning. Neither does temperature have a constant proportional relationship with energy or other extensive thermodynamic properties.

Averages of the Earth's temperature field are thus devoid of a physical context that would indicate how they are to be interpreted, or what meaning can be attached to changes in their levels, up or down. Statistics cannot stand in as a replacement for the missing physics because data alone are context-free. Assuming a context only leads to paradoxes such as simultaneous warming and cooling in the same system based on arbitrary choice in some free parameter. Considering even a restrictive class of admissible coordinate transformations yields families of averaging rules that likewise generate opposite trends in the same data, and by implication indicating contradictory rankings of years in terms of warmth.

The physics provides no guidance as to which interpretation of the data is warranted. Since arbitrary indexes are being used to measure a physically non-existent quantity, it is not surprising that different formulae yield different results with no apparent way to select among them.

The purpose of this paper was to explain the fundamental meaninglessness of so-called global temperature data. The problem can be (and has been) happily ignored in the name of the empirical study of climate. But nature is not obliged to respect our statistical conventions and conceptual shortcuts. Debates over the levels and trends in so-called global temperatures will continue interminably, as will disputes over the significance of these things for the human experience of climate, until some physical basis is established for the meaningful measurement of climate variables, if indeed that is even possible.

It may happen that one particular average will one day prove to stand out with some special physical significance. However, that is not so today. The burden rests with those who calculate these statistics to prove their logic and value in terms of the governing dynamical equations, let alone the wider, less technical, contexts in which they are commonly encountered.

Acknowledgement

Thanks to Kirill Kondratyev for encouraging the publication of this work.

References

- [1] Global Temperature Record, Climatic Research Unit Web site, <http://www.cru.uea.ac.uk/cru/info/warming/>, accessed December 2005.

- [2] Global Temperature (Meteorological Stations), Goddard Institute of Space Studies Web site, <http://data.giss.nasa.gov/gistemp/graphs/>, accessed December 2005.
- [3] Goddard Institute of Space Studies (GISS), Surface temperature: Station data, http://www.giss.nasa.gov/data/update/gistemp/station_data/.
- [4] Intergovernmental Panel on Climate Change, *Climate Change 2001: The Scientific Basis*, Cambridge University Press, Cambridge, 2001.
- [5] <http://cdiac.esd.ornl.gov/trends/temp/jonescru/jones.html>.
- [6] Essex, C., Kennedy, D., Berry, R.S., How hot is radiation?, *Am. J. Phys.*, 71 (2003), 969–978.
- [7] Peterson, T.C., Vose, R.S., An overview of the global historical climatology network temperature database, *Bull. Am. Meteorol. Soc.*, 78 (1997), 2837–2849.
- [8] Boehmer, K., ISO and climate change: Delivering on the promise, *ISO Bulletin*, August 2002, <http://www.iso.org/iso/en/commcentre/isobulletin/comment/2002/August2002.html>.
- [9] NRC (National Research Council) Panel on Reconciling Temperature Observations, *Reconciling Observations of Global Temperature Change*, National Academy Press, Washington DC, 2000.
- [10] Christy, J.R., Norris, W.B., What may we conclude about global tropospheric temperature trends?, *Geophys. Res. Lett.*, 31 (2004), L06211, doi 10.1029/2003GL019361.
- [11] Fu, Q., Johanson, C., Warren, S., Seidel, D., Contribution of stratospheric cooling to satellite-inferred tropospheric temperature trends, *Nature*, 429 (2004), 55–58.
- [12] Esper, J., Wilson, R.S.J., Frank, D.C., Moberg, A., Wanner, H., Luterbacher, J., Climate: past ranges and future changes, *Quaternary Sci. Rev.*, 24 (2005), 2164–2166.
- [13] Trenberth, K., Uncertainty in hurricanes and global warming, *Science*, 308 (2005), 1753–1754.
- [14] Pounds, J., Bustamante, M., Coloma, L., Consuegra, J., Fogden, M., Widespread amphibian extinctions from epidemic disease driven by global warming, *Nature*, 439 (2006), 161–167.
- [15] Yoganathan, D., Rom, W.N., Medical aspects of global warming, *Am. J. Ind. Med.* 40 (2001), 199–210.
- [16] Bullen, P.S., *Handbook of Means and Their Inequalities*, 2nd ed., p. 537, *Mathematics and Its Applications*, Vol. 560, Springer, New York, 2003.
- [17] Essex, C., Kennedy, D., Bludman, S.A., The non-equilibrium thermodynamics of radiation, Part II, Ch. 12, in: *Variational and Extremum Principles in Macroscopic Systems*, Eds. S. Sieniutycz, H. Farkas, pp. 603–626. Elsevier, Amsterdam, 2005.
- [18] Diewert, W.E., Symmetric means and choice under uncertainty, *Essays in Index Number Theory*, Vol. 1, Eds. W.E. Diewert and A.O. Nakamura, pp. 355–433, North Holland, Amsterdam, 1993.
- [19] Hansen, J., Ruedy, R., Sato, M., Imhoff, M., Lawrence, W., Easterling, D., Peterson, T., Karl, T., A closer look at United States and global surface temperature change, *J. Geophys. Res.*, 106 Issue D20 (2001), 23947–23964.

Paper received: 2006-05-03

Paper accepted: 2006-06-26

