

Planets, life and the production of entropy

Ralph D. Lorenz

Lunar and Planetary Laboratory, University of Arizona, 1629 E. University Blvd, Tucson, AZ 85721, USA
e-mail: rlorenz@lpl.arizona.edu

Abstract: Two thermodynamic principles offer considerable insight into the climatic and geological settings for life on other planets, namely (1) that natural systems tend to actually achieve the ideal ('Carnot') limit of conversion of heat into work and (2) if a fluid system such as an atmosphere has sufficient degrees of freedom, it will choose the degree of heat transport that maximizes the generation of work (equivalently, that which offers maximum entropy production). The first principle agrees well with results on terrestrial cumulus convection, and the mechanical energy released by tectonic activity. The second principle agrees with the observed zonal climates of Earth, Mars and Titan, and shows promise for planetary interiors too; I discuss applications in the investigation of paleoclimates and habitability. I compare the work performed by planetary atmospheres and interiors on the terrestrial planets and thereby predict a weakly eroded landscape on Titan. The association of life with the production of entropy is also noted, and the possibility of evaluating planetary entropy production by telescopic observation is discussed.

Received 25 September 2001

Key words: atmospheres, climate, complex systems, entropy and non-equilibrium thermodynamics.

Introduction

Ever since Rudolf Clausius defined the quantity 'entropy' and articulated the law of thermodynamics, which states that in a closed universe the total entropy must increase, prospects for life in the universe have been given finite and immutable bounds. However, while life is associated with a local decrease in entropy, to maintain the structures and gradients of temperature and concentration to sustain life *requires* the external generation of entropy. As an analogy, to run a refrigerator to generate a local decrease in temperature requires someone, somewhere to be burning fuel in a power station. Thus the progressive degradation of energy from nucleosynthesis to sunlight, to heat and motion, and ultimately to thermal radiation lost to space is the only setting in which life can take place. Even on planets where life does not exist, it is the generation of entropy that is the fundamental driver of interesting phenomena such as the tectonic forces that move mountains and the generation of weather systems that erode them.

While the second law of thermodynamics is possibly the most robust and succinct result in physics, it says nothing about how rapidly entropy is generated. Classical thermodynamics considers only equilibrium, the heat-death state at which motion and evolution of the system has ceased. Yet, planets are essentially non-equilibrium places, continuously but unevenly supplied with low-entropy energy from a parent star. It has been realized only recently that fluid motions on planets may organize themselves (i.e. select the amount of heat transport by convection or advection) to *produce entropy at*

the maximum rate possible. I discuss this maximum entropy production (MEP) principle at some length in Section 3 of this paper.

First, however, I discuss a related topic. While engineers usually express the efficiency of a system in terms of work output divided by heat supplied, and struggle to attain the ideal ('Carnot') limit, this view is somewhat misleading in a planetary context. Where engineers fail to meet the Carnot limit, it is because some heat is escaping before it reaches the cold end of the engine (the temperature contrast between this and the hot end defines the Carnot limit) or because some of the work produced is lost by friction. Both are effects that an engineer might strive to mitigate. In considering the work produced by geophysical heat flows, these considerations are moot – the heat 'leaks' are determined by the radiative and conductive settings of the planet, and 'friction' is the work we are interested in anyway – and thus the Carnot limit is actually achieved.

These two principles are of considerable astrobiological interest. The MEP conjecture allows the zonal climates (i.e. the gradient of temperature with latitude) to be estimated, without detailed knowledge of the circulation patterns in the ocean and atmosphere. Habitability in the conventional definition relies on planetary surface temperatures lying within a specified range, usually estimated as a global average. Latitudinal gradients in temperature may, however, make some parts of a surface perpetually uninhabitable, or others habitable while on a global average basis the planet is not. The ideal heat-to-work conversion considers how heat flows, whether selected by MEP or not, are expressed as work on a planet

and give some insight into how interesting (or dangerous) that planet is. Comparing the work produced by atmospheric heat flows with that by interior convection yields some insight into the planetary landscape.

Heat-to-work conversion

When heat is supplied to a thermodynamic engine, it can only perform work via flowing down a temperature gradient. If the temperature difference across the engine is ΔT , and the temperature of the hot end of the engine is T_h , the efficiency of the conversion of heat into work is $\eta = \Delta T/T_h$. The only way of realizing complete heat-to-work conversion is if the hot end is at an infinite temperature (which poses some difficulties in material selection for the engine) or if the cold end is at absolute zero (which would require the engine to operate with an infinitesimal heat flow and thus zero power output).

Natural systems (e.g. hurricanes, Emanuel 1999) operate as heat engines. The heat fluxes and the end-point temperatures are largely set by the radiative balance at the hot and cold ends of the system: to a first order the vertical structure of the Earth's atmosphere may be considered in this way. Of the 340 W m^{-2} of heat supplied to the top of the atmosphere on an annually averaged basis, about $80\text{--}100 \text{ W m}^{-2}$ reaches the surface and is not reflected or radiated away. This heat is transported vertically, to colder levels of the atmosphere where it is radiated thermally to space. Thus we have a heat engine, with a Carnot efficiency defined by the average surface temperature of 288 K at the hot end, and the effective temperature of 250 K defining the cold end. (Arguably the tropopause temperature of 215 K might be used – the difference is not especially important. I consider the effective temperature to be more appropriate, since some of the heat must be radiatively rejected before reaching the top of the troposphere.) Hence the Carnot efficiency is $(288\text{--}250)/288 \sim 13\%$. Thus, the work that can be performed by the atmosphere is about 0.13 times 90 W m^{-2} , or 12 W m^{-2} or so.

Meteorological texts tend to give a figure for the frictional dissipation by the atmosphere of 2 W m^{-2} , an estimate first obtained by Brunt (see, e.g., Brunt 1952), although often attributed to E. Lorenz (1960) or Peixoto and Oort (Peixoto *et al.* 1991). However, it has been realized only recently that this is probably a significant underestimate (a point also raised by Michaud (1995)), in that much dissipation is due to violent but rare events such as hurricanes (since dissipation relates to the cube of wind speed, ignoring the high-end tail of the wind-speed distribution by using a simple average ignores much of the dissipation.) Further, additional sources of dissipation, such as the friction around falling raindrops – which originally were lofted by the convective heat flow – are being recognized.

Lorenz and Renno (unpublished results) note that assuming an ideal heat-to-work conversion predicts that vertical convection in the Earth's atmosphere requires somewhat rare convective plumes (occupying around 10^{-4} of the surface), which are rather energetic (with updrafts of some 30 m s^{-1}) – in rather good agreement with observations. If convection were to occur with a much lower efficiency, then

there would be rather more widespread, but weaker, convection: a similar result has been obtained in simulations of Martian convection by Odaka (2001). Of course, the kinetic energy realized in these plumes is ultimately degraded, as momentum is shared with a larger and larger mass of air (i.e. entrainment in the plume, by viscosity) to achieve somewhat the same effect, but the observed energetic plumes can only be created with an initially ideal conversion. The diffusion of momentum is, incidentally, a degradation that is exactly analogous to the degradation of heat by conduction: energy is conserved, of course, but the energy is made less useful and entropy is produced.

Perhaps surprisingly, a similar ideal heat-to-work conversion is found in the Earth's interior. Around 80 m W m^{-2} of heat flows out of the Earth – a continuous release by radioisotopes together with a small component of secular cooling of the Earth's hot interior. Of that heat flow, around two-thirds is due to radioisotopes in the rigid crust, and this heat is principally leaking out by conduction. One-third, however, is transported from the fluid mantle. The dominant heat transport process there is by fluid motion. The temperature difference across the upper mantle is some 1500 K or more, and taking the cold-end temperature as the base of the crust at around 1500 K gives a Carnot efficiency of about 50%. Thus mantle convection yields some $\sim 20 \text{ m W m}^{-2}$ of available work, or around $3 \times 10^{20} \text{ J yr}^{-1}$. This is a striking number, in that it is not much more than the energy dissipated annually as earthquakes (somewhat more than $10^{20} \text{ J yr}^{-1}$ (Gutenberg & Richter 1954). Stacey (1967) makes some similar arguments. Allowing for the considerable uncertainty in all of the above quantities (not without close analogy to the discussion of the atmosphere above, where crude estimates of dissipation lead to a similar factor of ~ 5 underestimation), and dissipations other than those owing purely to earthquakes, the assumption of an ideal heat-to-work conversion appears to be a remarkably useful one. As an aside, it is interesting to note that Sadi Carnot himself noted the, far from obvious, connection between earthquakes and heat in the opening paragraphs of his treatise on steam engines: 'We must attribute to heat the great movements that we observe all around us on the earth. Heat is the cause too of earthquakes and volcanic eruptions' (Carnot 1824).

The Earth's annually averaged climate

The problem of latitudinal heat transport on a spherical planet was perhaps first discussed at length by James Croll in the context of the cause of the ice ages (e.g. Croll 1890). Because the planet is spherical, high-latitude regions receive less sunlight than the tropics (except at high obliquities, when the opposite is the case). On an airless body, this is the whole story and there is a substantial equator-to-pole temperature gradient and the planet is everywhere at radiative equilibrium.

Fortunately, for the inhabitants of the Earth, planets with significant atmospheres (and oceans) permit significant heat flows that follow, and therefore mitigate, the temperature gradient. Were this heat flow performed with a near-infinite

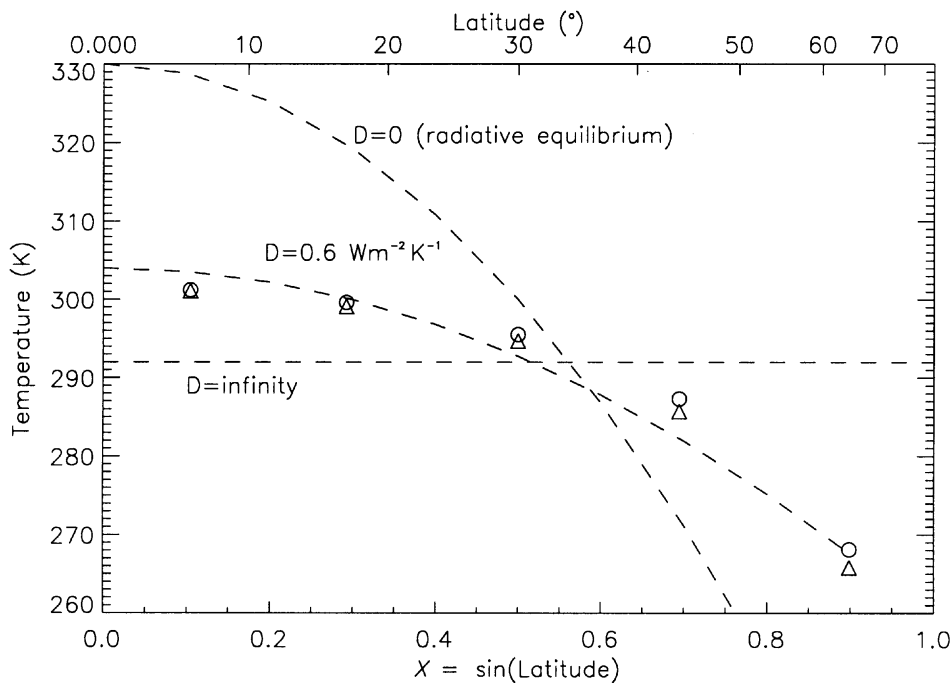
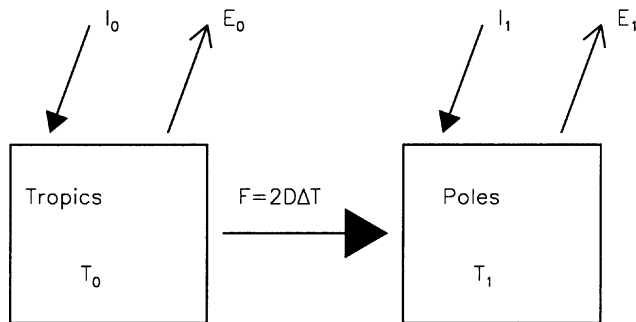


Fig. 1. Model temperature distributions on the Earth, showing the influence of changes in the heat transfer parameter D . Observed annual-mean temperatures (Paltridge 1975) are shown as triangles (Northern hemisphere) and circles (Southern hemisphere). The terrestrial climate is reasonably recovered for $D \sim 0.6 \text{ Wm}^{-2} \text{ K}^{-1}$, although see the text for discussion of the residual discrepancy. Note the ample scope for maintaining low-latitude temperatures above freezing by reducing D .



$$I_0 - E_0 - F = 0$$

$$I_1 - E_1 + F = 0$$

Fig. 2. Box model. The boxes are of equal surface area, one corresponding to equatorial regions (bounded by $\pm 30^\circ$ latitude) and the other the lumped ‘polar’ regions. Each box has an associated absorbed solar flux I_0 and I_1 , with I_0 corresponding to the equatorial region being larger. Between the boxes, there is a latitudinal heat transport term F , which is assumed to be proportional to the temperature difference between the boxes, i.e. $F = 2D\Delta T$, where $\Delta T = (T_0 - T_1)$. Each box also has an outgoing thermal flux, E_0, E_1 .

conductivity (or diffusivity) the planet would be isothermal. The distribution is indicated schematically in Fig. 1.

It was noted by Paltridge (1975) that the Earth’s zonal climate can be reproduced remarkably well by assuming that these heat flows maximize the production of entropy (in fact, his original paper defined the state as one of minimum entropy exchange). Simply put, if the heat flow F from tropical regions

at temperature T_0 to cooler polar regions at T_1 , the quantity $dS/dt = F/T_1 - F/T_0$ is maximized. Lorenz (1960) had previously noted a related and essentially equivalent observation, that the Earth’s climate maximizes the rate of production of available potential energy (APE).

Fig. 2 shows a simple two-box model (Lorenz *et al.* 2001) that captures the essence of the situation. For the present-day Earth, the insolation I_0 in the low-latitude box (taking into account the albedo) is about 240 Wm^{-2} , while at higher latitudes we have $I_1 = 140 \text{ Wm}^{-2}$. Heat leaves the planet as thermal radiation from the two zones, with the outgoing emission being related to temperature $E_x = A + B T_x$. In the grey approximation, $B \sim 4\sigma T^3 / (1 + 0.75\tau)$, where τ is the infrared optical depth, σ the Stefan–Boltzmann constant and T is the temperature about which we are linearizing. Earth has $\tau \sim 0.9$ and $T \sim 290 \text{ K}$, so $B \sim 3 \text{ Wm}^{-2} \text{ K}^{-1}$, close to the value ($\sim 2.1 \text{ Wm}^{-2} \text{ K}^{-1}$) obtained empirically from satellite observations (e.g. North 1975; North *et al.* 1981). If heat is transferred between the boxes at a rate $F = 2D(T_0 - T_1)$, then it is easy to show that the temperature difference $\Delta T = T_0 - T_1 = (I_0 - I_1) / (B + 4D)$. It can be shown in this formulation that the entropy production dS/dT has a maximum for $D = B/4$, i.e. $\sim 0.5\text{--}0.8 \text{ Wm}^{-2} \text{ K}^{-1}$ – just as required (North 1975; Wyant *et al.* 1988). The dependence of the entropy production and the two temperatures on D is shown in Fig. 3: to reproduce the present conditions indeed corresponds to a maximum entropy production value of D .

This model captures the essence of the principle in a succinct, albeit crude, form, one in which it has been applied to Mars and Titan (see below). Since Paltridge’s original paper, many

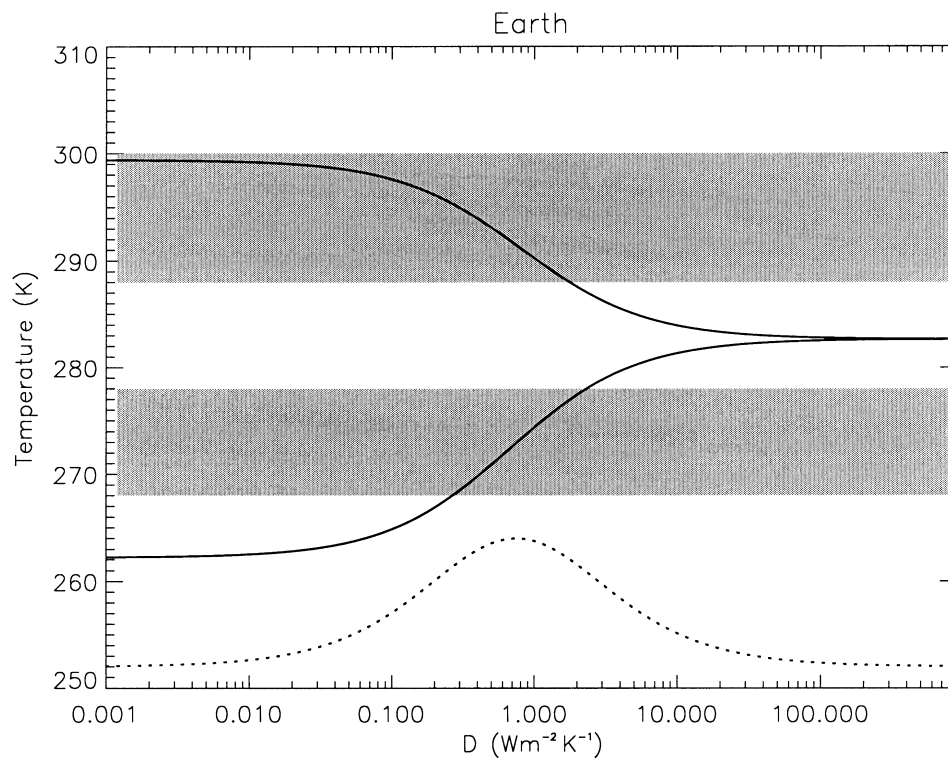


Fig. 3. Low- and high-latitude surface temperatures on Earth, computed using the simple two-box model. The upper curve corresponds to tropical temperatures, while the lower one is for high latitudes: at high D the curves meet. Shaded areas denote approximate measured temperatures. The dashed curve at the bottom denotes entropy production, dS/dT – note that it peaks where the temperature curves have their steepest slopes, and where model temperatures are as observed.

others have confirmed the basic result that the Earth's observed climate appears to be in an MEP state. Most of these papers use zonal energy balance models: examples are those by Grassl (1981), and Sohn & Smith (1993). Sadly, much of the literature discussing the principle is mired in forbidding algebra, which has not helped in communicating its importance. One of the crispest discussions is that due to Lin (1982), and ironically Rodgers' criticism (1976) of Paltridge's original paper expresses the idea more succinctly than did Paltridge himself.

An interesting point to note in considering zonal energy balance models is that a single, global D value (whether chosen empirically or via MEP) underestimates the temperature gradient slightly (North 1975) at lower latitudes and overestimates the gradient at higher latitudes (see Fig. 1). However, it follows from the dependence of D on T that D should be higher at the warmer, low latitudes. This effect flattens the temperature distribution at low latitudes, just as observed.

Application of MEP to zonal climate on other planets

Many of the papers discussing the application of MEP note that the generality of the principle should allow it to apply to other planetary bodies, and indeed since the coincidence of the observed terrestrial climate state with one of MEP might be just that, coincidence, an evaluation of the principle on other planets would be very useful.

Recently, Lorenz *et al.* (2001) noted that the observed

brightness temperature contrast (Samuelson *et al.* 1997) on Saturn's cold atmosphere-shrouded moon Titan, while surprisingly large (3 K, compared with an equatorial temperature of about 93 K) considering the thick, slowly rotating atmosphere, is exactly what would be predicted from MEP. While zonal models using a pressure-scaled D of 10^2 – 10^3 $\text{Wm}^{-2} \text{K}^{-1}$ would predict contrasts of around 0.01 K, the MEP principle mandates a D value rather lower than that of Earth, $\sim 4 \sigma(93)^3/(1+0.75\tau)$ or around $0.02 \text{ Wm}^{-2} \text{K}^{-1}$, leading to temperature contrast of a few K.

It is not clear how Titan's apparent heat transport is so inefficient – it may be that the latitudinal winds are suppressed by the strong zonal wind field, or that a condensation/evaporation phenomenon (Stevenson & Potter 1986) pins the polar temperatures at a low value. However, the *prima facie* agreement of the principle with the observed temperatures (although some uncertainty regarding a possible stratospheric contribution to the brightness temperatures exists) on Titan lends strong support to the MEP principle.

Lorenz *et al.* (2001) also consider Mars. At first look, Mars does not obey MEP – its climate can be largely reproduced with the very small heat transport expected in a thin atmosphere (i.e. $D < 0.01 \text{ Wm}^{-2} \text{K}^{-1}$). However, its winter poles would get too cold, and models are forced to pin them at the CO_2 condensation temperature of around 150 K. This is a quite reasonable 'fix' to the models, given that we can observe this process in action with the seasonal growth and decay of CO_2 frost caps. However, when the latitudinal heat transport

implicit in this process (since latent heat is being transported from the condensing cap to the atmosphere and thence to the subliming cap) is taken into account, the heat flow is entirely consistent with what would be predicted from MEP, of around $4\sigma(200)^3/(1+0.75\tau)$ or $\sim 0.4 \text{ W m}^{-2} \text{ K}^{-1}$. Were meteorologists unaware of the CO_2 condensation, the predictions of GCMs or zonal models would be gravely in error, yet only knowledge of the basic radiative setting of the planet (insolation, obliquity, albedo and infrared opacity) allows MEP to correctly predict that the atmosphere does *something* to ensure a heat transport of $\sim 0.4 \text{ W m}^{-2} \text{ K}^{-1}$, corresponding to $\sim 1 \text{ m}$ of seasonal frost and winds of several m s^{-1} , as observed. This underscores the potential utility of MEP for astrobiological studies in predicting the resultant climate state, even when all the contributing mechanisms are not known.

On hot Venus, temperature contrasts are believed to be quite low (although there are no accurate measures of surface temperature with latitude). However, both MEP and more conventional pressure scaling of heat transport coefficients give more-or-less equivalent results: while Venus may indeed obey MEP, it is not an ideal test case.

Application of MEP to vertical convection

Although the entropy production in the Earth's atmosphere by vertical transport of latent and sensible heat has been considered for a decade (Peixoto *et al.* 1991) and even half a century (Wulf & Davis 1952) ago, it remains an active topic. Ozawa & Ohmura (1997) developed a remarkably simple model to study the applicability of MEP to vertical convection. They first developed a radiative equilibrium of the Earth's atmosphere, simply assuming that the optical and thermal optical depths of the atmosphere were proportional to pressure. The more conventional approach adopted by radiative-convective models (Ramanathan & Cloakley 1978) is to check where the resultant radiative temperature profile exceeds the adiabatic lapse rate, and to fix it to that lapse rate where it does. Some iterative adjustment is usually then required to make the fluxes balance.

What Ohmura and Ozawa did instead is to impose an arbitrary convective flux profile. The temperature structure responds accordingly to restore the energy balance at each level by changing the thermal emission. There is a one-to-one correspondence between the imposed convection and the resultant temperature profile, but there is nothing to discriminate between different imposed profiles, except the entropy production by the convective flux. Interestingly, if the profile is allowed to vary, the flux profile that results in MEP is remarkably close to the observed flux in the Earth's atmosphere. It follows that the 'convective adjustment' and the adiabatic lapse rate are 'discovered' independently by MEP (at the risk of muddling a maximum entropy state with one of maximum entropy *production*, it is worth noting that the adiabatic lapse rate can be derived from entropy considerations alone (e.g. Bohren & Albrecht 1998)).

Thermodynamic subtleties aside, this result is important in two respects. First, it lends support for the potential universality of the MEP principle, operating as it does in a

completely different dimension from the zonal climate applications in the previous section. Secondly, it shows that the atmospheric temperature structure for an atmosphere can be constructed using only its opacity structure: no specific composition need be assumed to determine whether the profile is superadiabatic. This may be useful in applications to extra-solar planets where knowledge of the atmospheric composition is likely to be incomplete. There is an interesting parallel between the convective 'fix' applied to radiative models, and the frost-point pinning of Martian zonal models – in each case some 'correction' is being applied, yet the need for such a correction is predicted by the MEP principle.

Application to planetary interiors

The most striking discovery of two centuries of geological study of the Earth is the sheer impermanence of its features, and the fluid motion undergone by its interior. Despite the glacial pace of mantle motions, could these be sufficiently free to organize themselves to maximize entropy production? A simple calculation shows that they may, and therefore the MEP principle may make useful predictions: this discussion is reproduced below.

Entropy analysis has been applied quite successfully in the optimization of nuclear power plants (e.g. Bejan 1992, 1998), which are in fact a useful analogue of planetary heat flow. A reactor or radioisotope source generates heat. Kept in perfect isolation, it would become hot without limit. Realistically conductive loss paths to the low-temperature environment force a limiting temperature, with the difference between the limit temperature and the environmental temperature (assumed as a heat sink of infinite capacity) being proportional to the thermal resistance of the heat leak.

Now imagine we wish to generate power from this heat source. We operate some power converter between the hot and cold reservoirs, tapping some of the heat flow (i.e. it runs in parallel with the conductive heat leak). This converter, whether a thermoelectric device or a Stirling engine or whatever, will produce work or electricity or whatever with an efficiency constrained by the Carnot efficiency $(1 - T_c/T_h)$. As with the atmospheric case, there are obvious boundary conditions. The designer's free parameter is the thermal resistance of the converter. A high resistance will abstract only a small amount of the available heat flow: the hot-end temperature (and therefore the Carnot efficiency) will be maximized, but most of the available heat flow is wasted through the heat leak. On the other hand, a low converter resistance will tap off the lion's share of the available heat flow, but since the resistance is low, the hot-end temperature falls and the conversion efficiency is very low. In between, the heat flow through the converter has an intermediate value, and the power output is a maximum.

For the Earth we need a fractionally more complex model, since the convecting mantle (the 'converter') is also the source of heat. (We ignore the Earth's core, and the concentration of radiogenic heating in the crust from where it is readily conducted away.) We assume that the Earth's mantle (e.g. Davies 1999) conducts with a thermal conductivity of $3 \text{ W m}^{-1} \text{ K}^{-1}$,

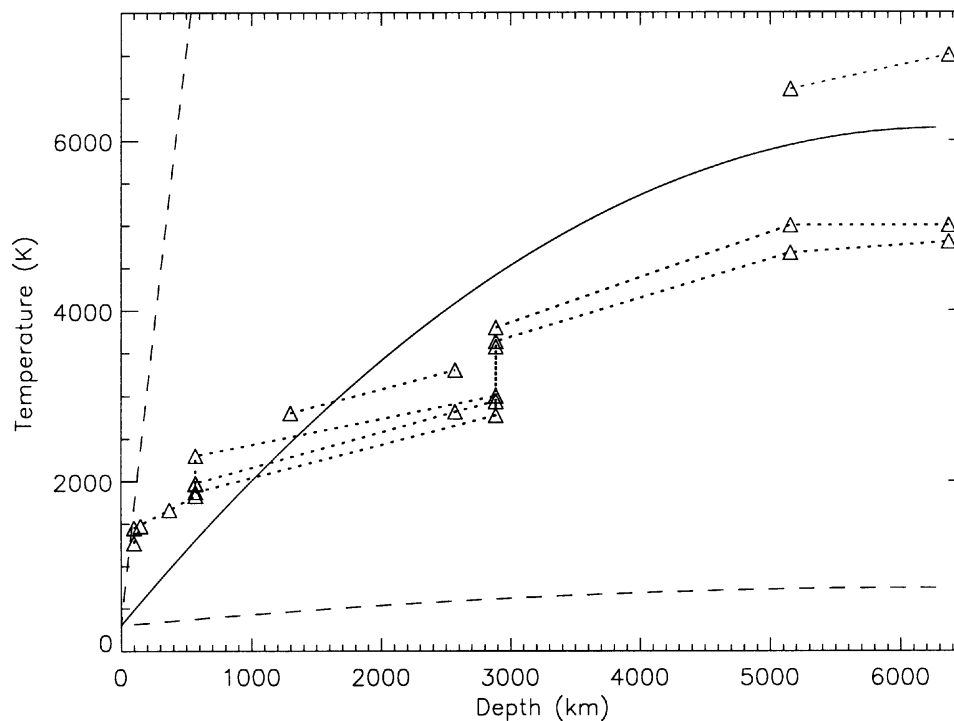


Fig. 4. Internal temperature distribution of the Earth. Dotted lines with triangles are from a compilation in Poirier (1991). Dashed lines are for fixed $Nu = 1$ (near-vertical) and 100 (near-horizontal). The solid curve corresponds to the MEP value of $Nu = 7.6$.

and generates heat at a rate of $2 \times 10^{-8} \text{ W m}^{-3}$ (this volumetric production would therefore account for about half of the observed terrestrial heat flow – the remainder owing to the concentration of radioisotopes in the crust and to secular cooling). The heat transfer through the interior may be expressed as the Nusselt number Nu , which is the ratio between the actual (convective + conductive) heat transfer and the conductive-only part. The temperature structure as a function of radius r is given by $dT/dr = -Hr/3kNu$. Given a surface temperature of 300 K and a radius of 6370 km, then if $Nu = 1$ (conductive-only) throughout, the temperature at the centre of the planet would be some 45000 K. If Nu is infinite then the temperature everywhere is the surface value of 300 K. Both of these cases are valid solutions of the problem, but are at odds with what we know about the behaviour of rocks, which do not remain rigid at 45000 K, nor convect vigorously at 300 K.

Conventional geophysical models have a variety of complications including compositional layering, but fundamentally (Turcotte & Schubert 1982) rely on estimates of mantle viscosity from post-glacial rebound and parametrization of convective heat transport (typically of the form $Nu \sim 0.1Ra^x$, where x is an exponent ~ 0.3 , and Ra is the Rayleigh number). Predictions from this class of models (Poirier 1991) of the Earth's internal structure are shown in Fig. 4, and yield central temperatures of around 4400–7000 K.

What profile do we obtain if Nu is allowed to be a free parameter, subject to the sole constraint that the entropy production by the convective part of the heat transfer is maximized? With the parameters above, MEP occurs for $Nu = 7.6$, and a central temperature of 5600 K. While again not an unassailable proof of MEP, the agreement of MEP (indeed,

the shape of the temperature curve with depth – Fig. 4 – not just the central temperature) using a minimum of assumptions with the results of vastly more elaborate (and ‘tuned’) models is striking.

The astute reader will see this problem as a direct analogue of the foregoing discussion on atmospheric convection (Ozawa & Ohmura 1997), with the only distinction being that radiogenic heat is substituted for solar energy deposition, and conduction ‘loss’ for outgoing thermal emission.

It should be emphasized that this trivially simplistic kind of steady-state calculation may not be applicable to all planets – evolutionary models may need to be applied. However, MEP formulations of the heat flow during planetary evolution are likely to be simpler but just as effective as conventional parametrizations on Rayleigh number, etc., and fidelity to MEP models may be added as desired (e.g. Nu may be allowed to vary throughout the planet). These models can almost certainly be applied to the evolution of icy satellites as well.

Lorenz (2001) has noted that the application of the MEP principle to geothermal heat flow in the Galilean satellites of Jupiter suggests that their interiors should well exceed the melting point of water. This agrees with the discovery of induced magnetic field signatures of conductive (presumably brine) layers beneath the surfaces of Ganymede and Callisto.

Application to planetary boundary layers

Another situation where MEP may be useful is in the heat transfer between the surface and atmosphere. On Earth, this is typically treated with highly empirical parametrizations (Sutton *et al.* 1978), the application of which to other planets may be of doubtful accuracy. Consider the diurnal temperature

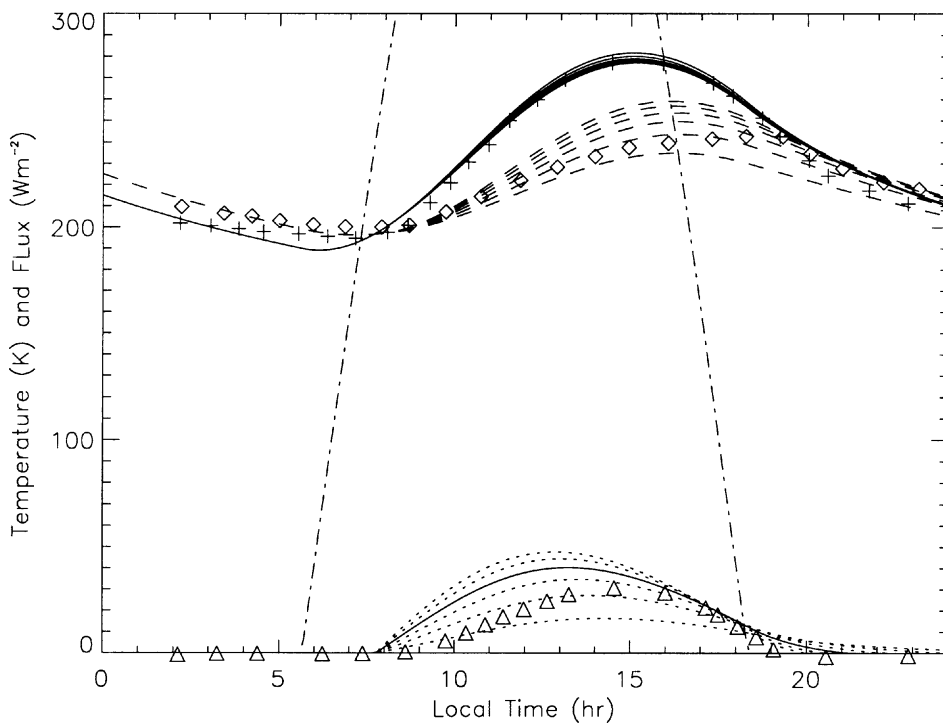


Fig. 5. Diurnal temperatures and fluxes at the surface of Mars. The family of solid lines are surface temperatures, while the dashed lines are for atmospheric temperatures. Different curves correspond to different values of the heat transfer coefficient K (see the text). Surface temperatures inferred from models based on thermal emission data are shown as crosses and atmospheric temperatures measured by Pathfinder are shown as diamonds. The chain line indicates insolation. Model boundary layer fluxes are shown as dashed lines at the base of the plot, except for that giving MEP, shown as a solid line. Estimated fluxes from Haberle *et al.* (1999) are shown as triangles.

evolution of the surface of Mars (the example that follows pertains to the landing site of the Sagan Memorial Station, also known as ‘Mars Pathfinder’): as the sun rises, the surface temperature increases, peaking a little after noon when the insolation peaks. Heat is lost (considering the surface as a single slab with a fixed heat capacity) by thermal radiation and by boundary layer convection, with the first term being rather more important. The latter term is often called ‘afternoon cooling’ since it is driven by the temperature difference between the temperature inversion between the warm surface and the cool atmosphere above, which is strongest in the afternoon.

The atmosphere, which is not optically thick either at solar or thermal wavelengths, varies in temperature rather less throughout the day than does the ground. It warms somewhat after sunrise owing to a small amount of solar absorption by dust, and by slight trapping of thermal radiation from the ground. Because these heat fluxes are fairly small, the input of heat from the ground by boundary layer convection is by comparison fairly significant. Fig. 5 shows how the surface and atmospheric temperatures vary as a function of the efficiency of boundary layer convection. The convective heat flux is expressed simply as $F = K(T_{\text{surf}} - T_{\text{atm}})$, where $(T_{\text{surf}} > T_{\text{atm}})$, otherwise zero. Conventional meteorological parametrizations (e.g. Sutton *et al.* 1978) have K depend on surface wind stress, roughness scales and the like. Here we leave it as a free parameter and explore the entropy production, $F/T_{\text{atm}} - F/T_{\text{surf}}$.

As K is increased, it is seen from the figure that the atmosphere is warmed more strongly. The effect on the surface temperature is fairly small, since the heat capacity of the surface is rather more than that of the boundary layer. As K is increased and the heat flux increases, the entropy production also increases. However, as it does so, the temperature difference over which the flux operates is reduced and so a maximum in entropy production is reached. As seen in the figure, the value of K ($\sim 1 \text{ Wm}^{-2} \text{ K}^{-1}$) that yields MEP appears to be quite consistent with the observed temperature evolution of the surface and atmosphere, and the corresponding surface heat flux is close to that predicted by the parametrization embedded in the Mars GCM (Haberle *et al.* 1999).

It is freely admitted that the choice of surface and atmospheric emissivities and heat capacities allow ‘tuning’ of the model. Again, this model is not intended as a robust proof of MEP on the Martian surface, but rather as a demonstration that this is the kind of problem to which MEP may be applied, and that it appears to yield results that are consistent with more empirical (but sophisticated) models.

Discussion

The MEP principle makes many atmospheric dynamicists uncomfortable, in that it appears to ignore parameters of cherished significance to conventional meteorology, such as rotation rate. General circulation models run with different planetary rotation rates yield different temperature contrasts between low and high latitudes, whereas MEP would predict

similar values; it is not clear whether MEP is inapplicable, or GCMs are being pushed further than they should. (While the basic physics such as conservation of momentum etc. that goes into such models is not in question, the same may not be said for many of the crude parametrizations of subscale heat transports in the models, which are of course tuned to present terrestrial conditions.)

One of the principal challenges is to define the applicability domain of MEP. As Rodgers (1976) argued immediately after the publication of Paltridge's original work, the thin atmosphere of Mercury cannot follow MEP (although he incorrectly said the same about Mars). Evidently there are physical limits (e.g. the speed of sound) to the amount of heat that can be transported by the atmosphere, such that the heat transport may be constrained to be below that for MEP: however, it is conjectured that natural systems will not have transfers higher than that for MEP.

Golitsyn (1970) developed a similarity theory of planetary circulation, which has thermodynamic (heat-work) arguments as its basis, but includes characteristic speeds such as sound velocity and planetary rotation rates. That theory may merit further examination in the context of MEP.

We have shown that the large-scale annually averaged temperatures of three planetary bodies seem to follow MEP, in the sense that the annually averaged large-scale entropy production is maximized. However, does the system maximize its instantaneous entropy production? Probably it cannot, since many of the heat transports (notably the ocean thermohaline circulations) have organization timescales ('spin-up times') that are much longer than 1 year.

The work of Malkus (1954) is relevant here. He showed with a simple heat transfer experiment that as the temperature gradient across a fluid is increased, the heat transfer increases too. This is of course well known (and usually expressed succinctly as the empirical dependence of the Nusselt number on the Rayleigh number), but what Malkus noted is that the heat transport increases in discrete steps as new convective modes 'kick in'. Presumably the smaller-scale modes, with shorter time constants, are invoked first by the system. As these fail to transport enough heat, the system invokes larger and larger modes until MEP is reached. Malkus makes a rather striking remark – '*Perhaps the most suggestive observation in these experiments was the apparent lack of dependence of the heat transport on the character of the horizontal motion; almost as though the heat transport was separately determined while the fluid motion adjusted itself to fit new boundary constraints*'.

As an aside, it has been noted (Bak 1998) that the state with the highest throughput that can be naturally attained is the so-called critical state, whether the system is a climate heat transport, traffic on a highway, a securities market or a seismic system. Self-organized critical states are often characterized by so-called $1/f$ statistics, where the frequency of events depends on their magnitude raised to some exponent (typically -1). Exploring the connection between MEP and self-organized criticality will be an interesting area of future investigation. $1/f$ statistics were noted by Gutenberg and Richter in

earthquake energies (which we have discussed in Section 2) and by Zipf (1949), in the rank-frequency distributions of words in many languages and in tonal jumps in music. The title of Zipf's book (*Human Behavior and the Principle of Least Effort*) suggests an extremal principle may be at work, and languages are after all self-organized, although see Mandelbrot (1961) for a note of caution.

Another challenge for future work is to determine exactly what is being maximized. For many purposes, a simple zonal or vertical model will do, and the results do not depend significantly on whether entropy production is maximized, or some other similar property. (Shutts (1981) found that a property related to entropy production was maximized in a wind field generated from vorticity transfer theory – an important supporting argument for MEP, and one of few that do not rely on empirical agreement with the observed terrestrial climate state.) However, the production of entropy by vertical convection (Li *et al.* 1994) is much larger on Earth than that by latitudinal heat transport, and two-dimensional models that consider both (Pujol & Llebot 2000) yield different answers (albeit only modestly different) if entropy production by only one heat flow is optimized. Similarly, some models (Paltridge 1975; Grassl 1981) allow the system to manipulate the cloud cover fraction (which changes both the incoming solar flux as well as the outgoing thermal flux) with good results, and again the models yield slightly different answers if this freedom is restricted.

A final, more general challenge, is to develop a convincing rationale as to why systems should be in an MEP state (there is often confusion from the fact that a totally constrained system, e.g. a conductive bar heated at one end, will evolve into a state of *minimum* entropy production). Some arguments have been made (Paltridge 1979) regarding non-linearities in meteorological processes tending to push the system via small perturbations to a state of maximum dissipation, but a robust theory remains to be developed.

Other applications of MEP

The most obvious application of MEP is perhaps to study the zonal climate of planets in other solar systems. The zonal climate is of interest not only because a range of conditions may be experienced over a planet's surface (especially in cases of high obliquity), but the evolution of significant mass (and importantly, opacity) reservoirs such as the Earth's water ice caps, and the CO₂ frost caps of Mars, are controlled not by the average planetary temperature, but by the coldest polar temperatures.

It has been noted that the runaway glaciation of the Earth into a possible 'snowball' state is contingent upon a sufficiently vigorous latitudinal heat transport (Endal & Schatten 1982) – if the low and high latitudes are sufficiently decoupled, then even with a lower insolation and greenhouse opacity, the tropics might always remain above freezing point. An implication of MEP (and the corollary of the principle, that heat transfer is more effective at higher temperatures, as supported by the previous paragraph) is that as the Earth cools, heat transport will be suppressed, and thus prevent complete

glaciation (Paltridge 1975; Gerard *et al.* 1990). Although paleoclimates have been explored with general circulation models (GCMs), it has been noted (Jenkins & Smith 1999) that the crude heat transport parametrizations in GCMs mean that their predictions should be regarded with caution.

To date, studies of extrasolar planets have assumed heat transport efficiencies that scale with planetary rotation rate and total pressure (Williams & Kasting 1997). It is likely that different results will be obtained using MEP models. Another habitability application is that of synchronously rotating planets (Joshi *et al.* 1997). The rotational regime here is far from the terrestrial one, so it is not clear how trustworthy GCMs using terrestrially tuned heat transfers are likely to be. Again, MEP offers some hope of being sufficiently general to apply.

There are likely to be applications in studying gas giant planets too. However, these are complicated by two factors. One is that there is no solid surface to form a useful reference level which bounds the fluid motions and therefore makes an important test site for MEP. The second complication is that there is a significant internal heat source, so that three heat flows have to be considered; that from low to high latitudes as before, and those from the interior to low and high latitudes, respectively.

More speculatively, as well as planetary atmospheres and interiors, MEP may find useful application in models of protoplanetary clouds and discs. Students of the remarkable morphology of planetary rings may even find the MEP principle of interest – rings are dissipative systems with many degrees of freedom, after all. A challenge in these more diverse systems will be to correctly define the sources and sinks of entropy, and the correct analogue property of thermodynamic temperature.

A synthesis: atmospheric and internal work

An interesting ratio to consider is that between the work performed on a planetary surface by its atmosphere and the work made available by mantle convection. For the Earth, the atmosphere performs around 12 Wm^{-2} of work, as discussed in Section 2, while internal heat flow generates around 0.03 Wm^{-2} . The ratio is thus ~ 500 , but would have been perhaps an order of magnitude lower in the Earth's past. The ratio is not itself readily interpretable as surface morphology, but indicates the relative roles of tectonics (which can be thought of as largely constructive of topography) and erosion, which tends to degrade topography.

For Mars, despite the thin atmosphere, the heat transport is comparable to the Earth's, as discussed in Section 3. Because of the substantial temperature contrasts, the conversion efficiency is even higher than for Earth, around 25%. The atmospheric work is therefore around the same, of order 10. Geothermal heat flow is not known, but is usually estimated at around $20\text{--}30 \text{ mWm}^{-2}$. Mars being a small planet has cooled more rapidly, and much of that heat flow is due to crustal production and so is not available for work – perhaps less than a quarter is associated with mantle convection. Taking a

Carnot efficiency of 20%, the likely internally generated work is probably of the order of 0.01 Wm^{-2} or less. Thus, present-day Mars has a ratio of > 1000 , it is a more atmosphere-worked planet than Earth at present (though as its huge volcanoes and rifts attest, its past probably had a much more significant internal contribution).

Venus might be considered, all things being equal, to have a similar internal heat production, with perhaps a similar or slightly lower thermodynamic efficiency (since its stiffer, dehydrated rocks will flow only at a higher temperature). However, despite the proximity to the Sun and its high surface temperature, both the heat flows across its surface due to the atmosphere, and the temperature gradients that enable the production of work, are small. Only a few per cent of the incident sunlight reaches the Venusian surface, and so heat flows of only a few Wm^{-2} can be supported. Estimates of surface temperature contrasts are at most a few K, so conversion efficiencies of $< 1\%$ follow. (Although isolated locations may have kilometre-scale topography that drives slope winds with temperature contrasts of the order of 10 K, and thus conversion efficiencies of a few per cent.) Thus the typical atmospheric work available is $< 0.01 \text{ Wm}^{-2}$. Comparing this with the terrestrial or assumed Venusian internal heat flows suggests that the internal heat flows are, in fact, dominant, i.e. the ratio is less than 1. The Venusian surface bears very few Aeolian or other erosive landforms, and is dominated by tectonism, together with a few scattered volcanoes and impact craters (the surface as a whole is only 500 Myr old, but this is ample time on Earth at least for significant erosion to occur). Venus might therefore be considered to have an internally worked surface.

As for Titan, MEP suggests atmospheric heat flows of $\sim 0.08 \text{ Wm}^{-2}$, driven by temperature contrasts of around 4 K, hence a conversion efficiency of $\sim 4\%$. Thus atmospheric work is performed at about 0.003 Wm^{-2} . Heat production by radiogenic isotopes in its rock core (all of which must be convected or conducted through its ice and water mantle) gives around 0.006 Wm^{-2} . Conservatively taking the efficiency as 5%, the tectonic work is 0.0003 Wm^{-2} . Hence for Titan, the ratio of atmospheric work to internal work is around $1/10$ – rather closer to Venus than to Earth or Mars, and hence its surface may be considered to be internally worked, despite the common association of a thick atmosphere with processes of erosion.

Although tidal heating on Titan makes a minimal (perhaps 10%) contribution to the heat flux, it must be remembered that this is tidal working, not just heating. Thus the energy supplied here is all available as work, without a conversion efficiency. Since the Carnot efficiency of the mantle is also of this order it follows that on Titan, body tides and internally driven tectonics may contribute about equally to surface deformation (tides in seas are another matter, however).

Io and Europa, and perhaps Ganymede, have internal workings that are dominated by tidal forcing rather than heat flow, although in the case of Io, the heat flows are themselves made manifest on the surface via widespread, energetic and continuing volcanism that largely dwarfs tectonism. These

bodies, and Callisto, however do not have atmospheres thick enough to transport significant amounts of heat, let alone choose an MEP level of heat transport. The atmospheric work performed on their surfaces is therefore negligible.

The mechanical energy administered to a planetary surface by the impact of a comet or asteroid is another source of deformation. Although a comparison of that energy source with the others listed above (perhaps requiring the location of different planets in a tetrahedral parameter space) would be interesting, it will not be attempted here.

The ratios discussed here depend to a first order on only a few parameters: the size and ice/rock fraction of the planet (driving the energy of formation and radiogenic heat production) and on the distance from the parent star and the strength of the greenhouse effect. These are also the parameters that affect habitability – it may be possible to make some informed speculation about the likely natures of different classes of habitable planets.

Life and entropy production

Those bodies on which MEP does appear to be attained (Earth, Mars, Titan) are all bodies of astrobiological interest. Bodies where MEP manifestly does not apply, such as Mercury, are of little interest.

Swenson (1989) has argued that the spontaneous and successive symmetry breaking leading to the emergence of organized structures in the universe (in hydrodynamics, astrophysics as well as biological evolution) is a natural result of systems seeking to maximize their production of entropy. Ulanowicz & Hannon (1987) observe that living communities tend to augment the rate of entropy production over what it would be in the absence of biota, which as Lotka (1922) pointed out is a natural consequence of evolution – a more rapidly growing (more dissipative) species will displace a slowly growing one, and suggest that the spectrum of outgoing thermal radiation from a planet should contain information on the entropy production thereon. This idea is examined in some detail by Lesins (1991) who notes that the entropy production on Earth is dominated by the general down conversion of sunlight to thermal emission: the differences owing to additional entropy production by biota are likely to be immeasurably small.

There is of course much scope for pseudoscientific nonsense in discussions along the lines of those above, but the links between entropy production and the evolution of complexity are significant, and deserve fuller exploration. To end on a slightly frivolous note, Aoki (1992) has observed that the calorimetrically determined entropy production of pigs increases from birth until the age of about 8 months and then declines until death. This might not sound surprising, but it should be noted that the animal continues to grow in size well beyond 8 months: but this is simply a mature pig getting bigger. A maximum in entropy production is reached as the pig completes its development, growing from a piglet into a mature animal. Aoki suggests that this entropy production pattern may be a characteristic of living things.

Conclusions

The wide applicability of a principle of maximum entropy production is becoming apparent. Although much further work remains, the initial indications are that most systems with sufficient degrees of freedom to challenge conventional analytic methods render themselves subject to the principle. While MEP is unlikely to supplant conventional methods, it makes a useful and convenient alternative where information is limited, such as on extrasolar planets.

The predictive abilities of MEP are such that one is reminded of Mark Twain's remark 'There is something wonderful about Science. One gets such wholesome returns of conjecture out of such trifling investment of fact.' However, it is clear that its full utility will only be realized when its limitations are robustly defined.

Some general aspects of entropy production in complex systems have been discussed, and a tentative link drawn between MEP and self-organized criticality has been drawn, although much work remains to explore this possibility. The connections between thermodynamic entropy and information entropy should be explored further.

The notion that convective flows tend to be ideally converted to work has also shown to be a useful predictive tool, and a tectonically deformed landscape on Titan is predicted as a result.

Acknowledgements

I acknowledge the support of NASA's Cassini and DS-2 programmes. Fruitful discussions with Chris McKay, Jonathan Lunine, Paul Withers, Pascal Rannou, Garth Paltridge, Toni Pujol, Hisashi Ozawa and others in the development of the ideas in this paper are gratefully acknowledged.

References

- Aoki, I. (1992). Entropy physiology of swine – a macroscopic viewpoint. *J. Theor. Biol.* **157**, 363–371.
- Bak, P. (1998). *How Nature Works*. Springer, New York.
- Bejan, A. (1992). *Entropy Generation through Heat and Fluid Flow*. Wiley-Interscience, New York.
- Bejan, A. (1998). Entropy generation minimization: the new thermodynamics of finite-size devices and finite-time processes. *J. Appl. Phys.* **79** (3), 1191–1217.
- Bohren, C. F. & Albrecht, B. A. (1998). *Atmospheric Thermodynamics*. Oxford University Press, Oxford.
- Brunst, D. (1952). *Physical and Dynamical Meteorology*, 2nd edn. Cambridge University Press, Cambridge.
- Carnot, S. (1824). *Reflexions on the Motive Power of Fire, and on Engines Fitted to Develop that Power*. Bachelier, Paris.
- Croll, J. (1890). *Climate and Time in their Geological Relations*. Appleton, New York.
- Davies, G.F. (1999). *Dynamic Earth: Plates, Plumes and Mantle Convection*. Cambridge University Press, Cambridge.
- Emanuel, K.A. (1999). Thermodynamic control of hurricane intensity. *Nature* **401**, 665–669.
- Endal, A.S. & Schatten, K.H.J. (1982). The faint young Sun – climate paradox: continental influences. *J. Geophys. Res.* **87**, 7295–7302.
- Gerard, J.-C., Delcourt, D. & Francois, L.M. (1990). The maximum entropy production principle in climate models: application to the faint young sun paradox. *Q. J. R. Met. Soc.* **116**, 1123–1132.

- Golitsyn, G.S. (1970). A similarity approach to the general circulation of planetary atmospheres, *Icarus* **13**, 1–24.
- Grassl, H. (1981). The climate at maximum entropy production by meridional atmospheric and ocean heat fluxes, *Q. J. R. Met. Soc.* **107**, 153–170.
- Gutenberg, B. & Richter, C.F. (1954). *Seismicity of the Earth*. Princeton University Press, Princeton, NJ.
- Haberle, R.M. *et al.* (1999). General circulation model simulations of the Mars Pathfinder atmospheric investigation/meteorology data. *J. Geophys. Res.* **104**, 8957–8974.
- Jenkins, G.S. & Smith, S. (1999). GCM simulations of snowball Earth conditions during the Late Proterozoic. *Geophys. Res. Lett.* **26**, 2263–2266.
- Joshi, M., Haberle, R.M. & Reynolds, R. (1997). Simulations of the atmospheres of synchronously rotating terrestrial planets orbiting M dwarfs: conditions for atmospheric collapse and the implications for habitability. *Icarus*, **129**, 450–465.
- Lesins, G.B. (1991). Radiative entropy as a measure of complexity. In *Scientists on Gaia*, eds Schneider, S. & Boston, P., pp. 121–127. American Geophysical Union.
- Li, J., Chylek, P. & Lesins, G.B. (1994). Entropy in climate models. I. Vertical structure of entropy production. *J. Atmos. Sci.* **51**, 1702–1708.
- Lin, C.A. (1982). An extremal principle for a 1-dimensional climate model. *Geophys. Res. Lett.* **9**, 716–718.
- Lorenz, E.N. (1960). Generation of available potential energy and the intensity of the general circulation. In *Dynamics of Climate*, ed. Pfeffer, R.L., pp. 86–92. Pergamon, Oxford.
- Lorenz, R.D., Lunine, J.I., McKay, C.P. & Withers, P.G. (2001). Titan, Mars and Earth: entropy production by latitudinal heat transport. *Geophys. Res. Lett.* **25**, 415–418.
- Lorenz, R.D. (2001). Of course Ganymede and Callisto have oceans: application of a principle of maximum entropy production to icy satellite convection. *Lunar and Planetary Science Conf.*, Houston, March 2001 (abstract 1043 on CD ROM).
- Lotka, A.J. (1922). Contribution to the energetics of evolution. *Proc. Natl. Acad. Sci.* **8**, 147–154.
- Malkus, W.V.R. (1954). Discrete transitions in turbulent convection. *Proc. R. Soc. A* **225**, 1856–195.
- Mandelbrot, B. (1961). On the theory of word frequencies and on related Markovian models of discourse. In *Structure of Language and its Mathematical Aspects*, vol. XII, *Proc. Symposia in Applied Mathematics*, pp. 190–219. American Mathematical Society, Providence, RI.
- Michaud, L.M., 1995. Heat to work conversion during upward convection Part 1: Carnot engine method. *Atmosph. Res.* **39**, 157–178.
- North, G., Cahalan, R.F. & Coakley J.A. (1981). Energy balance climate models. *Rev. Geophys. Space Phys.* **19**, 91–121.
- North, G. (1975). Theory of energy-balance climate models. *J. Atmos. Sci.* **32**, 2033–2043.
- Odaka, M. (2001). A numerical simulation of martian atmospheric convection with a two-dimensional anelastic model: a case of dust-free Mars. *Geophys. Res. Lett.* **28**, 895–898.
- Ozawa, H. & Ohmura, A. (1997). Thermodynamics of a global-mean state of the atmosphere – a state of maximum entropy increase. *J. Climate* **10**, 441–445.
- Paltridge, G.W. (1975). Global dynamics and climate change: a system of minimum entropy exchange. *Q. J. R. Met. Soc.* **101**, 475–484.
- Paltridge, G.W. (1979). Climate and thermodynamic systems of maximum dissipation. *Nature* **279**, 630–631.
- Peixoto, J.P., Oort, A.H., Almeida, M., & Tome, A. (1991). Entropy budget of the atmosphere, *J. Geophys. Res.* **96**, D6, 10981–10988.
- Poirier, J-P. (1991). *Introduction to the Physics of the Earth's Interior*. Cambridge University Press, Cambridge.
- Pujol, T. & Llebot J.E. 2000. Extremal climatic states simulated by a 2-dimensional model Part 1: sensitivity of the model and present state. *Tellus* **52A**, 422–439.
- Ramanathan, V. & Cloakley, J.A. (1978). Climate modeling through radiative-convective models. *Rev. Geophys. Space Phys.* **16**, 4, 465–489.
- Rodgers, C.D. (1976). Comments on Paltridge's 'minimum entropy exchange' principle. *Q. J. R. Met. Soc.* **102**, 455–457.
- Samuelson, R.E., Nath, N.R. & Borysow, A. (1997). Gaseous abundances and methane supersaturation in Titan's troposphere. *Planet. Space Sci.* **45**, 959–980.
- Shutts, G.J. (1981). Maximum entropy production states in quasi-geostrophic dynamical models, *Q. J. R. Met. Soc.* **107**, 503–520.
- Sohn, B. & Smith, E.A. (1993). Energy transport by ocean and atmosphere based on an entropy extremum principle. Part 1: zonal averaged transports. *J. Climate* **6**, 886–899.
- Stacey, F.D. (1967). Convecting mantle as a thermodynamic engine. *Nature* **214**, 476–477.
- Stevenson, D.J. & Potter, A.E. (1986). Titan's latitudinal temperature distribution and seasonal cycle. *Geophys. Res. Lett.* **13**, 93–96.
- Sutton, J.L., Leovy, C.B. & Tillman, J.E. (1978). Diurnal variations of the martian surface layer meteorological parameters during the first 45 Sols at two Viking Lander sites. *J. Atmos. Sci.* **35**, 2346–2355.
- Swenson, R. (1989). Emergent attractors and a law of maximum entropy production: foundations to a theory of general evolution. *Syst. Res.* **6**, 187–197.
- Turcotte, D.L. & Schubert, G. (1982). *Geodynamics: Application of Continuum Physics to Geological Problems*. Wiley, New York.
- Ulanowicz, R.E. & Hannon, R.M. (1987). Life and the production of entropy. *Proc. R. Soc. London B* **232**, 181–192.
- Williams, D. & Kasting, J.F. (1997). Habitable planets with high obliquities. *Icarus* **129**, 254–267.
- Wulf, O.R. & Davis, L. Jr. (1952). On the efficiency of the engine driving the atmospheric circulation. *J. Meteorology* **9**, 2, 7982.
- Wyant, P.H., Mongroo, A. & Hameed, S. (1988). Determination of the heat-transport coefficient in energy-balance climate models by extremization of entropy production. *J. Atmos. Sci.* **45**, 189–193.
- Zipf, G.K. (1949). *Human Behavior and the Principle of Least Effort*. Addison-Wesley, Cambridge, MA.