

How does the Earth system generate and maintain thermodynamic disequilibrium and what does it imply for the future of the planet?

BY AXEL KLEIDON*

*Biospheric Theory and Modelling Group, Max-Planck-Institut für
Biogeochemie, Hans-Knöll-Strasse 10, 07745 Jena, Germany*

The Earth's chemical composition far from chemical equilibrium is unique in our Solar System, and this uniqueness has been attributed to the presence of widespread life on the planet. Here, I show how this notion can be quantified using non-equilibrium thermodynamics. Generating and maintaining disequilibrium in a thermodynamic variable requires the extraction of power from another thermodynamic gradient, and the second law of thermodynamics imposes fundamental limits on how much power can be extracted. With this approach and associated limits, I show that the ability of abiotic processes to generate geochemical free energy that can be used to transform the surface–atmosphere environment is strongly limited to less than 1 TW. Photosynthetic life generates more than 200 TW by performing photochemistry, thereby substantiating the notion that a geochemical composition far from equilibrium can be a sign for strong biotic activity. Present-day free energy consumption by human activity in the form of industrial activity and human appropriated net primary productivity is of the order of 50 TW and therefore constitutes a considerable term in the free energy budget of the planet. When aiming to predict the future of the planet, we first note that since global changes are closely related to this consumption of free energy, and the demands for free energy by human activity are anticipated to increase substantially in the future, the central question in the context of predicting future global change is then how human free energy demands can increase sustainably without negatively impacting the ability of the Earth system to generate free energy. This question could be evaluated with climate models, and the potential deficiencies in these models to adequately represent the thermodynamics of the Earth system are discussed. Then, I illustrate the implications of this thermodynamic perspective by discussing the forms of renewable energy and planetary engineering that would enhance the overall free energy generation and, thereby ‘empower’ the future of the planet.

Keywords: habitability; free energy; thermodynamics; global change; geoengineering

*akleidon@bgc-jena.mpg.de

One contribution of 13 to a Theme Issue ‘Climate predictions: the influence of nonlinearity and randomness’.

1. Thermodynamic disequilibrium as a sign of a habitable planet

In the search for easily recognizable signs of planetary habitability, Lovelock [1] suggested the use of the chemical disequilibrium associated with the composition of a planetary atmosphere as a sign for presence of widespread life on a planet. He argued that the Earth's high concentration of oxygen in combination with other gases, particularly methane, constitutes substantial chemical disequilibrium that would quickly be dissipated by chemical reactions if it were not continuously replenished by some processes. Since life dominates these exchange fluxes, he argued that life is the primary driver that generates and maintains this state of chemical disequilibrium in the Earth's atmosphere (see also Lovelock [2]).

Atmospheric composition is the only one aspect of the Earth system that is maintained far from a state of equilibrium. Another example of disequilibrium is the atmospheric water vapour content, which is mostly far from being saturated. If it were not for the continuous work being performed in the form of dehumidification by the atmospheric circulation [3,4], then the atmosphere would gain moisture until it is completely saturated, as this is the state of thermodynamic equilibrium between the liquid and gaseous state of water. Topographical gradients on land also reflect disequilibrium as erosion would act to deplete these gradients into a uniform equilibrium state, and life plays an important role in the processes that shape topographical gradients [5,6].

The maintenance of these disequilibrium states seems to contradict the fundamental trend towards states of thermodynamic equilibrium, as formulated by the second law of thermodynamics. To understand why this is fully consistent and even to be expected from the laws of thermodynamics, we need to resort to the formulations of non-equilibrium thermodynamics and formulate Earth system processes and their interactions on this basis. While Lovelock's further research, e.g. on the Gaia hypothesis [7], the Daisyworld model [8] has contributed substantially to the emergence of Earth system science that considers the functioning of the Earth as one, interconnected system [9–11], the use of thermodynamics as a basis for the holistic integration of processes of the Earth system is practically absent from mainstream Earth system science. If we had such a basis, then Lovelock's conjecture could easily be evaluated and we could evaluate how human activity alters the fundamental nature of the planet. Such a basis would seem critical to have to guide in the process of managing the impacts of human activity within the Earth system in the future.

This paper lays out how non-equilibrium thermodynamics can be used to develop a holistic view of how disequilibrium is generated and maintained within the Earth system, what this view would imply for the effects of human activities on the Earth system, and what potential deficits there are in the numerical models that we use to assess the Earth system change. To do so, the paper is structured along a series of questions. First, a brief overview of non-equilibrium thermodynamics is provided to address the question of how disequilibrium is generated and maintained without violating the second law of thermodynamics. In essence, it is shown how free energy is generated from one thermodynamic gradient and transferred to another, causing disequilibrium in thermodynamic variables that are not directly related to heat and entropy. More specifically, it is shown how work is derived (or power generated) from the planet's external forcing and initial conditions that then fuels a hierarchy

and cascade of free energy generation, transfer and dissipation. One can imagine this to be similar to an engine that is fuelled by heating and that can drive a series of belts and wheels, resulting in motion and cycling of mass. Then, the fundamental constraints that limit the extent of disequilibrium are addressed, which leads to the Carnot limit, its limited applicability to Earth system processes, and its extension to the maximum power principle and the proposed principle of maximum entropy production (MEP). (For a reader familiar with the thermodynamic basis, §§2–4 can easily be skipped.) This is followed by an evaluation of generation rates of free energy of the present-day Earth system to estimate the relative importance of the drivers of present-day disequilibrium states, particularly regarding chemical disequilibrium. The impacts of human activities are then discussed in this context and their likely consequences for planetary disequilibrium and free energy generation. After explicitly discussing some potential deficits in the present-day Earth system models regarding the dynamics of free energy, the paper closes with a brief summary and conclusions.

2. How is disequilibrium generated and maintained?

Typically, textbook teaching of thermodynamics mostly deals with isolated systems that are maintained in a state of thermodynamic equilibrium, characterized by a maximum in entropy. This is the consequence of the second law because processes can increase only entropy in an isolated system. To understand the generation and maintenance of disequilibrium—and the associated low entropy within the system—we note that the Earth system is not an isolated system, so that the nature of its thermodynamic state is substantially different. A state of disequilibrium does not violate the laws of thermodynamics, but is rather a consequence of these. In the following, a brief overview of the thermodynamics away from equilibrium is given to provide the basics of generating and maintaining disequilibrium and its relation to free energy generation within a system. The following overview is not meant to be exact and encompassing, but rather illustrative and explanatory. For more details, the reader is referred to textbooks on non-equilibrium thermodynamics [12,13].

(a) *Defining the Earth as a thermodynamic system*

Even though it seems somewhat formal, we first need to define the boundaries of the Earth system. The way we choose the boundaries is, in theory, arbitrary. Depending on how we choose the boundary, we may get different types of thermodynamic systems in that we need to consider different types of exchange fluxes through the boundaries. By choosing the boundaries well, we can make our description of the thermodynamic system a lot easier because we may need to consider fewer exchange fluxes to describe the system.

Three types of thermodynamic systems exist: (i) isolated systems are systems in which no exchange of energy and mass takes place with the surroundings, (ii) closed systems exchange energy, but no mass with the surroundings, and (iii) open systems that exchange both energy and mass.

When we deal with the Earth system, a good choice for the boundary is the top of the atmosphere. There, the dominant exchange is radiative, with low entropy solar radiation—in terms of its photon composition as well as its confinement to

a narrow solid angle—entering the Earth system and terrestrial radiation with some scattered solar radiation being returned to space. With this choice of boundary, the Earth is almost a closed system (ignoring the relatively small exchange owing to gravity and mass, such as hydrogen escape to space that could have played an important role in the Earth's past).

Note that subsystems of the Earth are typically placed at some thermodynamic meaningful boundaries as well. For instance, the separation of the climate system into the atmosphere, ocean and land follows the boundaries defined by the different states—gaseous, liquid and solid. When we deal with the boundary of the biosphere, we would place it at the interface of organic, living matter to its inorganic, non-living surroundings. The shape of this boundary would be rather complex. Because the exchange of mass between these subsystems is substantial, these subsystems are examples of open thermodynamic systems.

(b) *The first and second law of thermodynamics*

Once the boundary is defined, we need rules to determine the limits on how exchange fluxes at the boundary can be altered. This leads us to the first and second laws of thermodynamics. These laws express the constraints on the rate at which work can be extracted from a heat gradient to generate free energy, and it provides the direction of natural processes towards states of higher entropy. These two laws form the basis to understand what is needed to drive and maintain a state of thermodynamic disequilibrium.

The first law balances the change in internal energy dU within a system with the amount of heat exchange dQ with the surroundings and the amount of work performed by the system dW ,

$$dU = dQ - dW. \quad (2.1)$$

In principle, the internal energy U of a system represents mostly the amount of stored heat, but other types of energy also contribute to this term (see below; these contributions are generally small). When we deal with the Earth system, however, work is not extracted from the system, but is performed within the system itself and, hence, dW is not taken away from U . This work is, for instance, reflected in the acceleration or lifting of mass, i.e. it is taken away from the heat content to generate a non-heat form of free energy. So, when we want to use the first law to quantify generation rates of free energy, we first need to distinguish between heat versus non-heat contributions to U (with U_{heat} referring to the contribution of heat to U). For completeness, we also need to consider that the free energy generated by the system is also eventually dissipated, so that an additional, dissipative heating term D needs to be considered in the balance of U_{heat} . So, when we consider changes through time, we need to balance the changes in heat dU_{heat}/dt with the rates of net heating of the system owing to external forcings, $J_h = dQ/dt$, the extracted power $P = dW/dt$ to forms of free energy and the dissipative heating D resulting from irreversible processes within the system that dissipate free energy,

$$\frac{dU_{\text{heat}}}{dt} = J_h - P + D. \quad (2.2)$$

While equation (2.2) looks like a typical climatological energy balance, it is more than that because it includes the transfer and depletion rates of free energy related to P and D associated with other forms of free energy. The magnitudes of P and D are typically quite small in comparison with J_h and therefore not important for the heat balance that determines the changes in temperature. However, P and D are critical for the dynamics of the planet as these are driven by free energy generation and dissipation, and are intimately linked to the maintenance of the disequilibrium state, as we will see further below.

The second law states that the entropy S of an isolated system can only increase, i.e. $dS \geq 0$. The change in entropy dS is defined as the amount of heat dQ added or removed at the temperature T of the system,

$$dS = \frac{dQ}{T}. \quad (2.3)$$

To demonstrate how the diffusion of heat is a manifestation of the second law, let us consider two reservoirs of heat with different temperatures T_h and T_c with $T_h > T_c$ within a system. The difference in temperatures would drive a diffusive heat flux that would remove some amount of heat dQ from the warmer reservoir and add it to the colder reservoir. The entropy of the warmer reservoir would decrease by dQ/T_h , whereas the entropy of the colder reservoir would, once dQ is fully mixed, increase by dQ/T_c . The entropy of the whole system would change by $dS = -dQ/T_h + dQ/T_c$, and $dS > 0$ because $T_h > T_c$. When this heat exchange is sustained through time, resulting in a diffusive heat flux $J = dQ/dt$, then the diffusive heat exchange produces entropy σ at a rate of

$$\sigma = \frac{dS}{dt} = J \times \left(\frac{1}{T_c} - \frac{1}{T_h} \right). \quad (2.4)$$

When we deal with a non-isolated system, we also need to consider the entropy changes owing to the exchange fluxes of energy and mass with the surroundings. Then, equation (2.3) becomes a balance equation for the entropy of the system that relates the changes in entropy dS/dt with the entropy production σ by irreversible processes within the system, e.g. heat diffusion, and the net exchange of entropy across the system boundary, $J_{s,\text{net}}$. The net entropy exchange captures the fact that the energy which is exchanged with the surroundings is not of the same quality. For instance, it captures the fact that solar, shortwave radiation is of different composition to terrestrial, longwave radiation that leaves the Earth. Hence, the entropy balance is written as

$$\frac{dS}{dt} = \sigma - J_{s,\text{net}}, \quad (2.5)$$

with an isolated system represented by $J_{s,\text{net}} = 0$.

Note that when we deal with the system and its surroundings together, e.g. Earth and space, then we still deal with an isolated system. The total entropy of the system plus its surroundings then increases by $J_{s,\text{net}}$, which is consistent with the second law of thermodynamics [14,15].

(c) Quantifying thermodynamic disequilibrium

Thermodynamic disequilibrium and the balance equation for free energy A are directly linked to the first and second law (equations (2.2) and (2.5)). Free energy is generated when the power extracted from heating (P in equation (2.2)) is used to perform the work of building up some other gradient, which then stores free energy in some other form. The change of free energy dA is given by the balance of added power P , which adds free energy, and the rate of dissipation D by some irreversible process, which depletes free energy,

$$\frac{dA}{dt} = P - D. \quad (2.6)$$

Dissipation is linked to both: to the energy balance through the dissipated heat, and to the entropy balance by the entropy that this dissipation produced (D is thus linked to σ in equation (2.5)). Not considered in equation (2.6) is the potential transfer of free energy from one type to generate free energy of another type. This aspect is dealt with further below when discussing free energy generation and transfer within the Earth system.

The free energy A in a system is directly related to the distance to thermodynamic equilibrium. To relate the two, we consider the first law (equation (2.1)) and note that the added heat dQ either contributes to the internal energy of the system dU or to the free energy generated dA by performing the work dW . We can then express the change of entropy dS (as in equation (2.3)) as the sum of two contributions, $dS = dS_{\text{heat}} + dS_{\text{diseq}}$, relating to the change in internal energy dU/T and to the change in free energy dA/T within the system. Hence, the disequilibrium of a system, as expressed by S_{diseq} , is directly related to the free energy content A by $dS_{\text{diseq}} = dA/T$.

In steady state, $P = D$. When the rate of dissipation is expressed as $D = A/\tau$, then the steady-state free energy within a system is $A_{\text{ss}} = P \times \tau$ and the associated disequilibrium $S_{\text{diseq,ss}} = P \times \tau/T$. The extent of disequilibrium (and mean free energy) within a system is hence directly related to how much power P is contained in the processes that generate free energy. Since the disequilibrium is also related to the time scale of its depletion τ , the extent of observed disequilibrium does not necessarily tell us how active the system is in generating and dissipating free energy.

(d) Illustration with a simple model

The concepts of free energy and disequilibrium for two types of thermodynamic systems are demonstrated using the two simple systems shown in figure 1. The two systems consider two heat reservoirs that are linked by a heat flux and have initially the same, uneven distribution of heat. The only difference between the system is that the system on the right column of figure 1 exchanges heat with its surroundings. For the details of the model, see Kleidon [16,17].

The system on the left of figure 1 is an isolated system (figure 1(i)). In this system, the difference in heat content is dissipated through time, as shown by the equilibration of the temperatures (figure 1b(i)). The heat flux decreases with the declining temperature difference, as does the associated entropy production (figure 1c(i)). While the entropy of the colder box increases, the entropy of the

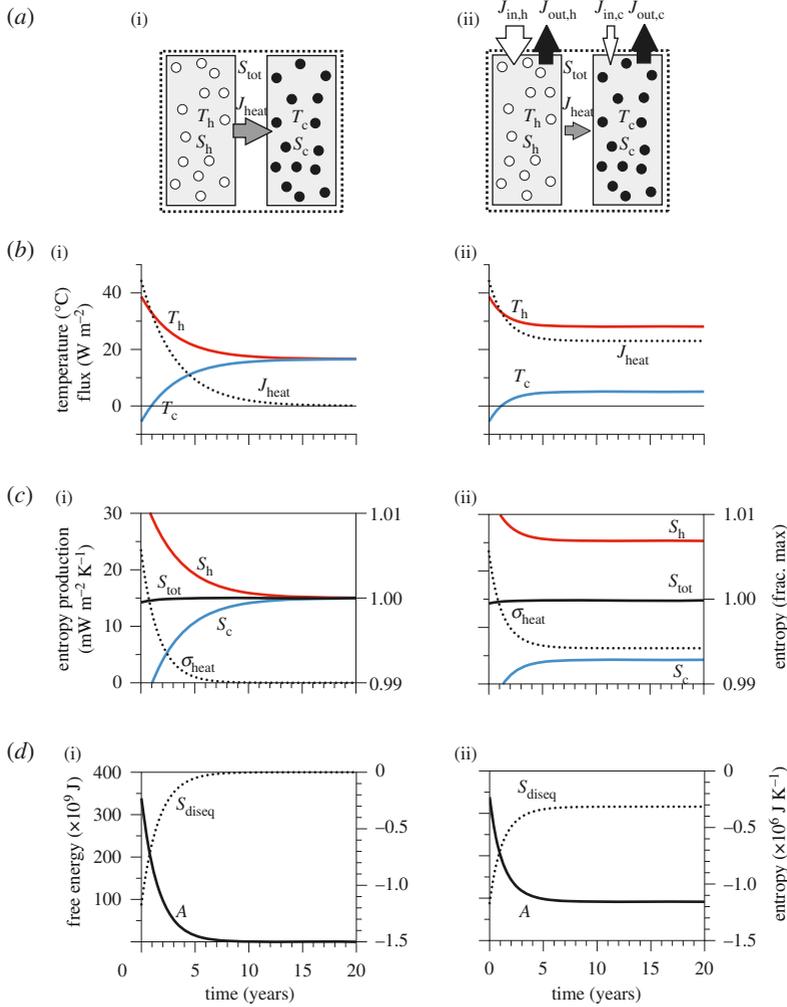


Figure 1. Simple box models to demonstrate the time evolution of (i) an isolated and (ii) non-isolated thermodynamic system. (a) Layout of the box model with T_h and T_c temperatures of the hot and cold reservoir, respectively; J_{heat} , heat flux; S_h , S_c and S_{tot} , entropy of the hot, cold reservoir and the whole system, respectively; $J_{in,h}$, $J_{in,c}$, $J_{out,h}$ and $J_{out,c}$, heat exchange fluxes with the surroundings (subscripts refer to: in, heating from the surroundings; out, cooling to the surroundings; h, hot reservoir; c, cold reservoir). (b) Time evolution of temperatures and heat flux J_{heat} between the boxes. (c) Entropies S_h , S_c and S_{tot} and entropy production σ by heat exchange. (d) Free energy A and disequilibrium entropy S_{diseq} . Adapted from Kleidon [17]. (Online version in colour.)

warmer box decreases owing to the removal of heat. The entropy of the whole system increases owing to the equilibration, which is clearly seen in the depletion of A and S_{diseq} (figure 1d(i)).

The system on the right of figure 1 is a closed system, in which both heat reservoirs exchange heat with the surroundings. The initial difference in heat content also declines in this system, but it reaches a steady state away from

equilibrium with a non-vanishing difference owing to the differential heating that the two reservoirs receive (figure 1*b*(ii)). We observe a steady-state heat flux J_{heat} that aims to deplete the temperature difference, but cannot deplete it to zero owing to external, differential heating of the system. Hence, the fluxes through the boundaries play a critical role for the maintenance of disequilibrium within the system, as reflected by a difference in the entropies between the reservoirs (figure 1*c*(ii)), non-zero free energy A and disequilibrium S_{diseq} (figure 1*d*(ii)). Feedbacks of the dynamics of the system onto the exchange fluxes at the boundaries and constraints that shape the flexibility of these determine how far a system can evolve and maintain itself away from a state of thermodynamic equilibrium.

3. How is disequilibrium and free energy generated within the Earth system?

We now take the simple formulations of §2 and apply them to the Earth system. As discussed earlier, the exchange of the Earth system with its surroundings is accomplished primarily by the exchange fluxes of solar and terrestrial radiation. These radiative fluxes result in radiative heating and cooling when absorbed or emitted. The associated radiative heating and cooling fluxes, J_{heat} and J_{cool} , and the energy balances that reflect the temperatures at which this heating and cooling takes place, T_{heat} and T_{cool} , form the constraints for physical forms of free energy generation by heat engines, as formulated by the first and second law of thermodynamics.

Note that these heat engine constraints do not apply directly to photochemical means of free energy generation. Solar radiation contains a significant amount of free energy that is embedded in the photon composition, which reflects the high emission temperature of the Sun. When solar radiation enters the Earth system, its flux density is much diluted compared with when it was emitted, so that its photon composition is far from equilibrium with its respective radiative equilibrium temperature and equilibrium photon composition. A simple way to illustrate the free energy in solar radiation is to consider the temperature gradient between the emission temperatures of the Sun and the Earth, which would yield a Carnot efficiency of $\Delta T/T \approx (5760 - 255 \text{ K})/5760 \text{ K} \approx 95\%$ (which is only an approximation, as detailed analyses would need to consider the entropy of radiation; for the explanation of the Carnot limit, see below). This disequilibrium, however, is not provided in the form of a temperature gradient, but rather in the form of the associated photon composition. Hence, it can only be exploited by processes that involve photon transformations, such as absorption, but not by heat engines.

Photochemistry is one way that uses the free energy of solar radiation. It uses excited electrons that have absorbed solar photons in the visible range. Electronic absorption can generate electronic free energy and thereby avoid that the energy of solar photons is directly converted into heat after absorption. Photochemistry can, in principle, yield substantially more free energy than heat engines driven by the same radiative fluxes, but it requires suitable photochemical mechanisms to exploit this free energy. By performing photosynthesis, life does exactly this and generates substantial amounts of chemical free energy at the Earth's surface. Note that the photochemistry taking place in the Earth's stratosphere also generates free energy by converting molecular oxygen into ozone, but this free energy is

rapidly dissipated, and is therefore unlikely to provide substantial rates of free energy that are available to drive geochemical reactions at the surface.

We deal with this contribution later and focus first on the constraints imposed by physical transfer processes related to heat engines and what these imply for the functioning of the planet.

(a) *Planetary balance equations*

The dynamics of the planet are constrained by the associated global balance equations for total internal energy, U_{planet} ; heat, $U_{\text{planet,heat}}$; free energy, A_{planet} (with $U_{\text{planet}} = U_{\text{planet,heat}} + A_{\text{planet}}$) and entropy, $S_{\text{planet}} = S_{\text{heat}} + S_{\text{diseq}}$,

$$\frac{dU_{\text{planet}}}{dt} = J_{\text{heat}} - J_{\text{cool}}, \quad (3.1)$$

$$\frac{dU_{\text{planet,heat}}}{dt} = J_{\text{heat}} - J_{\text{cool}} - P_{\text{total}} + D_{\text{total}}, \quad (3.2)$$

$$\frac{dA_{\text{planet}}}{dt} = P_{\text{total}} - D_{\text{total}} \quad (3.3)$$

and
$$\frac{dS_{\text{planet}}}{dt} = \frac{dS_{\text{heat}}}{dt} + \frac{dS_{\text{diseq}}}{dt} = \frac{J_{\text{heat}}}{T_{\text{heat}}} - \frac{J_{\text{cool}}}{T_{\text{cool}}} + \sigma_{\text{total}}, \quad (3.4)$$

where P_{total} is the total power (or generation rate of free energy) of the Earth system, D_{total} is the total dissipation of free energy and σ_{total} the total entropy production by irreversible processes. Note that the change in disequilibrium, dS_{diseq}/dt , is directly related to the change in free energy, dA_{planet}/dt , by $dS_{\text{diseq}}/dt = -1/T_{\text{planet}} \times dA_{\text{planet}}/dt$ (as above, with a representative temperature T_{planet}), and the total dissipation D_{total} contributes to the total entropy production σ_{total} (although entropy is also produced by diffusive processes that are not related to the dissipation of free energy, so that $\sigma_{\text{total}} \geq D_{\text{total}}/T$).

Disequilibrium and free energy of the planet can result from the depletion of planetary initial conditions and by exploiting the fluxes at the planetary boundary. Both of these have, of course, to be consistent with the second law, which requires $\sigma_{\text{total}} \geq 0$ in the global balance equations above. To understand how disequilibrium and free energy in other forms of energy than heat is generated, we need to include these in our considerations of the total internal energy U_{planet} . Other forms of energy, such as gravitational energy, kinetic energy, binding energy, chemical energy and so on, can be expressed in terms of conjugate variables, pairs of thermodynamic variables that taken together express different forms of energy. Changes in the total internal energy dU_{planet} can then be expressed as the sum of changes in heat, and changes in the energy of the other forms,

$$dU_{\text{planet}} = d(TS) + \sum_i d(p_i v_i) + \sum_i d(\phi_i M_i) + \sum_i d(\mu_i M_i) + \sum_i d(A_i \xi_i) + \dots \quad (3.5)$$

Here, the first term on the right side expresses the contribution by the heat content of the system (temperature T and entropy S , yielding $U_{\text{planet,heat}}$), by kinetic energy (momentum p_i and velocity v_i), by gravitational energy (gravitational potential ϕ_i and mass M_i), by binding energies (chemical potential

μ_i and mass M_i) and chemical energy (affinities of the reactions A_i and extent of the chemical reactions ξ_i). The sums in equation (3.5) run over the different contributions to a particular form of free energy, e.g. the contributions of air flow, water flow (oceans and rivers), mass flow of plate tectonics and mantle convection and so on to the total kinetic energy within the system. Other forms of energy, e.g. electric or magnetic energy, can be expressed in a similar fashion [12,18], but are neglected here for simplicity.

(b) *Dynamics of free energy and disequilibrium*

The change in total internal energy dU_{planet} does not tell us much about the dynamics that take place within the Earth system and the associated disequilibrium. When one form of energy is converted into another, the total internal energy dU_{planet} does not necessarily change because this is just a conversion between different contributions to dU_{planet} . When, for instance, free energy in some chemical compound is depleted by an exothermic chemical reaction, then the free energy associated with chemical compounds is reduced (i.e. $d(A_i\xi_i) < 0$), which is balanced by the increase in the heat content (i.e. $d(TS) > 0$). Likewise, when a gradient in heat results in the generation of kinetic energy, then the term $d(TS)$ decreases and the term $d(p_iv_i)$ increases, but dU_{planet} is unaffected. When atmospheric motion lifts dust, the kinetic energy fuels the generation of potential energy of the dust grains, i.e. $d(p_iv_i) < 0$ to result in $d(\phi_j m_j) > 0$. Ultimately, the kinetic energy is dissipated by friction, i.e. $d(p_iv_i) < 0$ and $d(TS) > 0$, which again is not associated with a change dU_{planet} . Hence, the energy balance does not inform us about how much free energy is generated, dissipated and transferred between conjugate variables and how much disequilibrium is being maintained.

What is not captured by the energy balance is that work is required to generate free energy and drive-associated dynamics. For instance, work is needed to accelerate mass to convert a gradient in heat into motion. Or, more generally, the dynamics of conversions are driven by the generation and depletion of gradients of the different sets of conjugate variables, where the depletion of one gradient fuels the generation of another gradient. These dynamics are described by the generation and depletion rates of the different forms of free energy that all contribute to A_{planet} .

The conversions among different forms of energy cannot take place arbitrarily at any rate or direction. The second law requires that $\sigma_{\text{total}} \geq 0$. It thereby constrains the entropy balance, and imposes limits on the direction and the rates of conversions. In steady state, the entropy balance yields $J_{s,\text{net}} = \sigma_{\text{total}}$, i.e. the total rate of entropy production by all irreversible processes within the Earth system is constrained by the entropy exchange at the system boundary. For the Earth system, these are the exchange fluxes of solar and terrestrial radiation. Hence, the entropy balance is intimately linked to the energy balance, specifically to the net entropy exchange associated with J_{heat} and J_{cool} .

Furthermore, the entropy balance is linked to the free energy balance. First, the rates of free energy generation (P) are restricted by the entropy balance, specifically by the requirement of the second law of $J_{s,\text{net}} = \sigma_{\text{total}} \geq 0$. This aspect will be dealt with in more depth in the next section in connection with how the entropy balance limits power generation to a characteristic maximum possible

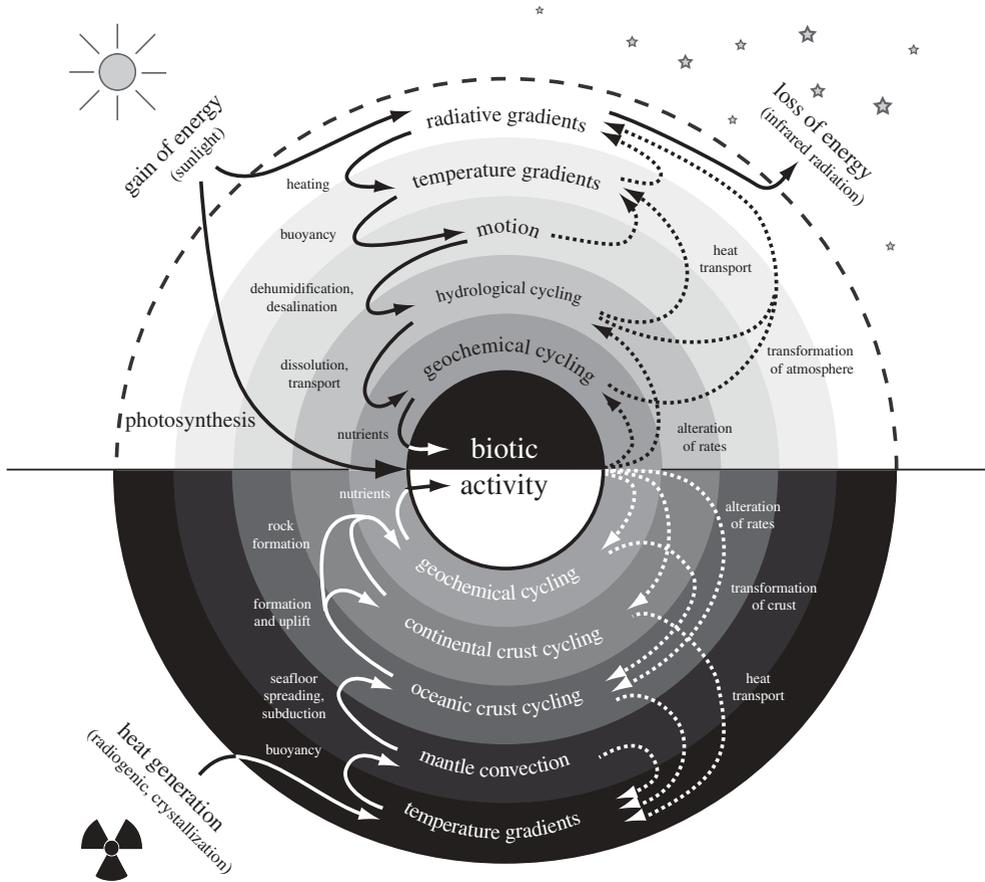


Figure 2. Schematic of the planetary hierarchy of free energy generation, transfer and dissipation (solid lines), and associated effects (dotted lines). The different layers are associated with different forms of free energy and gradients associated with disequilibrium. For instance, motion is associated with gradients in momentum and represents kinetic energy. Hydrological cycling is associated with gradients in chemical potential, and geopotential is associated with potential and chemical free energy. Adapted from Kleidon [17].

rate. Second, the dissipated free energy D is connected to the total entropy production within the system σ_{total} . Hence, the overall dynamics of free energy generation and the resulting disequilibrium state are closely connected to the net entropy exchange at the system boundary.

(c) *Planetary hierarchy of free energy generation, transfer and dissipation*

When we consider most of the abiotic processes within the Earth system, these are ultimately driven by the free energy generated by heat engines driven by external forcings or initial conditions. The power extracted from the differential heating is then converted further into some other forms. Hence, understanding the maintenance of disequilibrium within the Earth system needs to be viewed in an encompassing, holistic perspective of free energy generation and transfer among different processes, as shown in simplified form in figure 2.

This holistic view, which was developed in Kleidon [19,20], Dyke *et al.* [6] and Kleidon [16,17], shows two major drivers for heat engines that fuel abiotic free energy generation. The first driver results from the spatial and temporal variation in solar radiation at the system boundary. This flux generates gradients in temperature, which causes changes in air density, pressure gradients and results in motion. Atmospheric motion powers the dehumidification of the atmosphere, which generates the free energy that drives evaporation and desalination of seawater at the surface and condensation aloft, and hence the global cycling of water. The transport of water on land provides the free energy in the form of potential free energy to drive river runoff and sediment transport, and in the form of chemical free energy associated with freshwater to chemically dissolve the continental crust. Hence, we get a sequence of transformations from a radiative energy flux that results in a small fraction of free energy generation that can alter the geochemical nature of the surface.

The second driver is associated with the depletion of the initial conditions of the planet at formation, in the form of secular cooling of the interior, heating by radioactive decay and crystallization of the core. The differential heating results in a similar sequence of free energy generation that results in plate tectonics, uplift of continental crust and generation of geochemical free energy at the surface [6].

The distribution of free energy generated by the heat engines across different types of free energy results in a different way to look at interactions and feedbacks within the Earth system (dotted lines in figure 2). The extent to which power is transferred from one form of energy to another obviously affects the dynamics associated with both forms of energy. For instance, atmospheric motion is driven by pressure and temperature gradients, but this motion results in the transport of mass and heat that depletes these gradients. When power is removed from motion, e.g. to lift moisture, then it strengthens water cycling at the expense of slowing down motion. Large-scale hydrological cycling also depletes the driving force of motion through the transport of latent heat. Hence, this holistic view places interactions and feedbacks into the context of how these affect the rates of power generation as the primary driver for Earth system dynamics. Given that there are characteristic maximum rates of power generation for many processes, as we will see in the following, these place upper bounds on to the strength of interactions within the Earth system.

4. What are the limits to the generation of disequilibrium and free energy?

The maximum of free energy generation in classical thermodynamics is characterized by the Carnot limit. In the following, I briefly explain how the Carnot limit is derived in order to understand the implicit assumptions being made. Then, the Carnot limit is applied to a slightly different setting more representative of natural processes to show that this yields the maximum power principle and/or MEP for complex, non-equilibrium systems such as the Earth system.

(a) *The Carnot limit*

The Carnot limit is derived directly from the first and second laws of thermodynamics. It considers a system as shown in figure 3a, where a heat flux

J_{in} adds heat to the system, whereas the heat flux J_{out} removes heat from the system. The first law in steady state then tells us that

$$0 = J_{\text{in}} - J_{\text{out}} - J_{\text{ex}}, \quad (4.1)$$

where $J_{\text{ex}} = P_{\text{ex}}$ is the extracted power from the system. The constraint on the maximum value of P_{ex} , and consequently the Carnot limit, originates from the entropy balance of the system. The entropy exchange of the system is set by a fixed temperature gradient, with the influx of heat taking place at a corresponding temperature of T_{in} , and the outflux at a temperature of T_{out} . Hence, the entropy balance of the system in steady state is given by

$$0 = \sigma + \frac{J_{\text{in}}}{T_{\text{in}}} - \frac{J_{\text{out}}}{T_{\text{out}}}. \quad (4.2)$$

Because P_{ex} is free energy, it is not associated with an entropy, i.e. it can be completely used for performing work, and therefore does not show up as a term in equation (4.2). The best case for extracting work is when no irreversible process takes place within the system and the entropy production within the system is zero (i.e. $\sigma = 0$). This condition ($\sigma = 0$) is the limit of what is permitted by the second law. In this case, the entropy balance can be used to express J_{out} as a function of J_{in} , T_{in} and T_{out} , and the first law (equation (4.1)) yields an expression for the maximum rate $P_{\text{ex,max}}$ at which work can be extracted from the system,

$$P_{\text{ex,max}} = \frac{(T_{\text{in}} - T_{\text{out}})}{T_{\text{in}}} J_{\text{in}}. \quad (4.3)$$

This is the well-known Carnot limit, with an associated thermodynamic efficiency defined as $\eta = (T_{\text{in}} - T_{\text{out}})/T_{\text{in}}$.

Note that the derivation of the Carnot limit contains important assumptions. First, it assumes that the temperature gradient is fixed, i.e. there is no effect of the rate of extracted work on the temperatures at the system boundary. Second, it assumes that no other irreversible processes take place. If $\sigma > 0$ because of some unavoidable irreversible process taking place within the system, then the extractable work would need to be less than the Carnot limit. Third, the balance of free energy is not in a steady state because the free energy generated by the extracted work is not dissipated, and the related waste heat is not added back to the system. These assumptions cannot be made for Earth system processes. Interactions play a critical role, and heat fluxes often deplete the gradients by which these are driven. Also, processes compete for the same driving gradient, e.g. convection and conduction, hence the assumption of no other irreversible processes cannot be made. Finally, the Earth system is closed in that extracted work is dissipated in steady state.

We next consider what the maximum power limit is when these assumptions are relaxed.

(b) *The maximum power limit*

The assumptions of the Carnot limit can be relaxed in a slightly more complex setting as shown in figure 3a(ii). In this setting, free energy generation competes with diffusive dissipation of the temperature gradient. The extracted work is

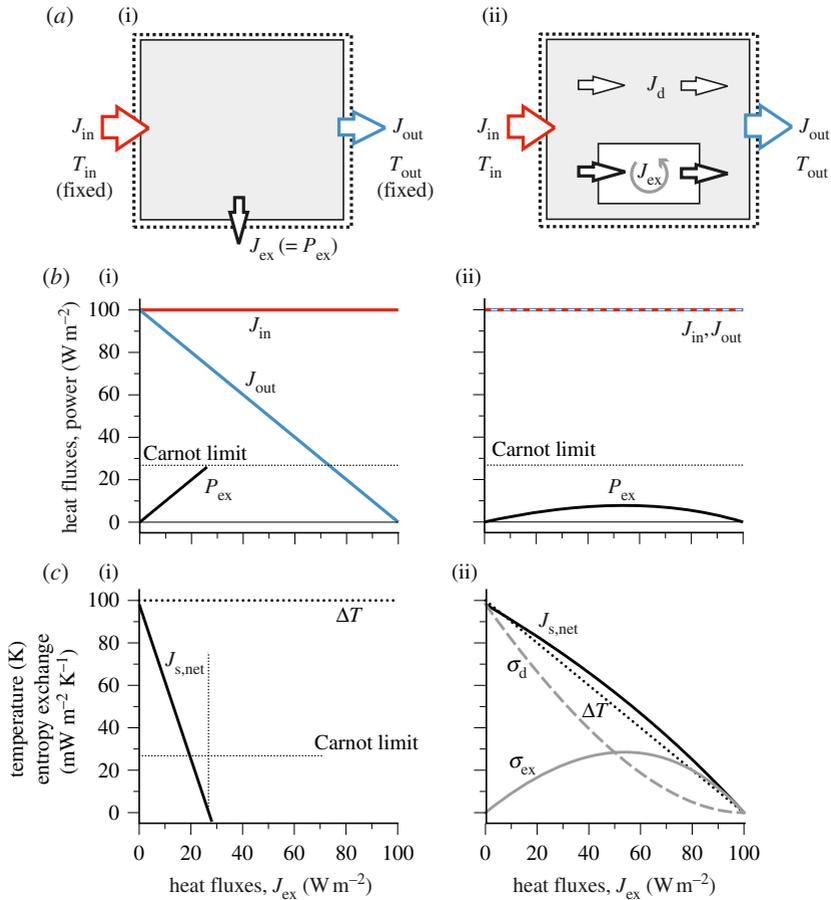


Figure 3. Illustration of the maximum extractable power in (i) the common Carnot limit and in (ii) a nonlinear system in which power extraction competes with another dissipative process (e.g. diffusion and radiative exchange) and the temperature gradient is affected by the extracted power and the associated dynamics. (a) Definition of system boundary and fluxes. In addition to an influx of heat J_{in} , an outflux J_{out} and an extracted heat flux J_{ex} , the right system also has a diffusive heat flux J_d that competes with J_{ex} for the depletion of the temperature gradient. (b) Heat fluxes and power versus extracted heat flux J_{ex} . The Carnot limit is shown by the dashed lines for comparison. (c) Temperature difference ΔT and net entropy exchange $J_{s,net}$ versus extracted heat flux J_{ex} . The Carnot limit corresponds to the extracted heat flux J_{ex} at which $J_{s,net} = 0$ (i). In the case of the system shown in (ii), the maximum in extracted power corresponds closely to a maximum in entropy production σ_{ex} associated with power extraction. See Kleidon [17] for model details. (Online version in colour.)

dissipated and the resulting heat is added back into the system. When expressing the diffusive process as $J_d = k_d(T_{in} - T_{out})$, and applying the Carnot limit to the extracted heat flux J_{ex} , we obtain for the work extraction (P_{ex})

$$P_{ex} = \frac{(T_{in} - T_{out})}{T_{in}} J_{ex} = \frac{(J_{in} - J_{ex}) \times J_{ex}}{k_d T_{in}}, \quad (4.4)$$

where the expression for J_d and the energy balance constraint $J_{in} = J_d + J_{ex}$ were used to substitute $(T_{in} - T_{out})$ and J_d . It is easy to show that this expression has a maximum for a heat flux $J_{ex,opt} = J_{in}/2$ of (if the dependency of the T_{in} in the denominator is neglected, see also Kleidon [17])

$$P_{ex,max} = \frac{1}{4} \frac{(T_{in,0} - T_{out,0})}{T_{in,0}} J_{in}, \quad (4.5)$$

where $T_{in,0}$ and $T_{out,0}$ are the temperatures at the system boundary in the absence of work extraction (i.e. $J_{ex} = 0$). The reduction to one-quarter compared with the Carnot efficiency if it were applied directly to J_{in} can be understood as follows: at the state of maximum power, the temperature gradient is depleted to half its value because the extracted work enhances the overall heat flux. The competition with diffusive loss yields another reduction by a factor of two. This results in the maximum power limit of equation (4.5), which is equivalent to the well-known maximum power principle in electrical engineering.

The state of maximum power extraction coincides very closely to the state at which the entropy production is at a maximum (figure 3). This maximum in entropy production relates to the proposed principle of MEP [14,21–23], which states that complex systems with sufficient degrees of freedom maintain a steady state at which entropy production is maximized. There are some theoretical developments to support the general nature of this principle and relate it to the maximum entropy formalism of equilibrium thermodynamics [24–27]. While the derivation of the maximum power limit above merely establishes an upper bound—just as the Carnot limit—the proposed MEP principle would imply that complex systems would actually evolve to and maintain such a steady state—just like an engineer would work towards achieving the Carnot limit when designing an engine.

There are several indications that natural processes, such as turbulence [28], convection [29] and the atmospheric circulation on the Earth [30–33] and other planetary systems [34,35], operate close to states of maximum entropy production and/or maximum power/dissipation (see also reviews by Ozawa *et al.* [14], Kleidon & Lorenz [21], Martyushev & Seleznev [22] and Kleidon [36]). This would suggest that the dynamics of complex systems are indeed characterized by maximization of power and dissipation.

(c) *Maximum power of material processes*

The above derivation of the maximum power limit is based on the extraction of work from a heating gradient. Equivalent derivations can be made for material fluxes and associated forms of energy. In order to find states of maximum power, we need to identify a trade-off between the force that drives the flux, and the flux that depletes the force. In the above example of maximum power derived from a heating gradient, the force is associated with the temperature gradient that drives the convective heat flux. This flux turn depletes the force by depleting the temperature gradient, resulting in the limit of extractable power.

The maximum power limit is well established within electrical engineering. Here, the trade-off exists between the resistance associated with a load that is connected to a generator with an inherent, internal resistance. The greater the resistance of the load, the greater the gradient in electric potential across the load,

but because of the higher overall resistance, the current is reduced. Because the electric power drawn by the load is the product of current and potential gradient across the load, a maximum exists for the power that can be drawn from the electric generator.

To demonstrate an equivalent trade-off that is more relevant for the Earth system, take the transfer of power contained in some flow (e.g. air flow and river flow) to perform work on a material flux (e.g. dust transport and sediment transport). The momentum gradient between the flow and the surface exerts a drag force that performs the work of lifting and accelerating particulates to the speed of the flow. The greater the removal of power from the flow, the more work can be performed on the particulates, but at the cost of a reduced flow velocity and thereby a reduced ability to transport particulates. Hence, a trade-off exists between the force, which is proportional to the momentum gradient, and the resulting flux of particulates, which depletes the force. Consequently, maximum power limits should be a rather general feature of Earth system processes. These limits constrain the extent to which power can be transferred within the Earth system as shown in figure 2, and to the extent to which thermodynamic variables can be maintained at states of thermodynamic disequilibrium.

(d) *Maximum power and feedbacks*

These upper bounds and the associated state of maximum power are relevant for how the Earth system responds to perturbations. If the emergent dynamics were organized in such a way that they maximize power generation, then these dynamics would evolve towards the maximum power state after any perturbation. Hence, maximization is inherently associated with the system reacting to perturbations with negative feedbacks [14].

If the nature of the boundary conditions change, then this would alter the conditions to which the maximization is subjected. In this case, the emergent dynamics would evolve in such a way as to maximize power under the altered boundary conditions.

(e) *Maximum power and disequilibrium*

When we want to relate these limits in power generation to the original motivation of understanding the drivers for disequilibrium within the Earth system, we note first that the state of maximum power depends on boundary conditions only, and not on the material properties of the process under consideration. This can be seen in equation (4.5), which depends on the temperatures and the heat flux, both of which merely describe the conditions at the boundary. It does not depend on the density or viscosity of the fluid. The associated disequilibrium, however, does depend on material properties. The disequilibrium and the associated free energy content results from the balance of generation and dissipation of free energy. While the maximum state in generation is described by the maximum power state of the boundary conditions, the dissipation is intimately linked to material properties such as density and viscosity of the fluid in the case of motion. Hence, a state of maximum power is not necessarily equivalent with states of maximum disequilibrium or maximum free energy content. As it would seem that the maximization of power (or entropy

production) is based on a more general and better justified basis, I focus on power and free energy generation rather than disequilibrium and free energy content in the following sections.

5. What are the generation rates of free energy for the present-day Earth?

To understand the drivers for present-day disequilibrium, we need to estimate the generation rates of free energy within the Earth system. These can be estimated by considering the primary drivers that supply free energy from external sources and that feed the hierarchy of transfer shown in figure 2. Using these drivers, a global free energy budget is derived and shown in figure 4. In contrast to the well-established global energy balance, the free energy balance emphasizes the importance of the biota in the planetary free energy generation (in particular, in the form of chemical free energy) and highlights the magnitude of human activity in dissipating free energy. The estimates are based on Kleidon [17] and are described in the following sections.

(a) Drivers of free energy generation

Four major drivers are responsible for free energy generation within the Earth system.

- *Solar heat engines.* Incident solar radiation is associated with spatial and temporal variations that maintain temperature gradients and fuel heat engines. These gradients are the main driver for climate system processes. The total free energy generated from this source includes the generation of potential free energy associated with air, water vapour and aerosols, the kinetic energy associated with motion in the atmosphere, oceans and river flow, the chemical free energy generated by dehumidification and desalination of sea water, the electric free energy generated by thunderstorms and so on. All of these are fuelled by uneven heating and cooling, resulting in vertical and horizontal gradients in density and pressure. We can estimate the maximum power available to maintain these types of free energy by considering the radiative forcing at the boundary using simple considerations [17]. Absorption of solar radiation at the surface and cooling by emission of radiation aloft generates a vertical gradient in heating that can be converted into free energy. Using a mean surface heating of 170 W m^{-2} and typical temperatures of 288 and 255 K, Kleidon [17] estimates that free energy generation from this vertical gradient is less than 5000 TW (note that much of this power is used for vertical mixing and transport and is likely to contribute relatively little to large-scale cycling and transport). Owing to the planet's geometry and rotation, absorption of incident solar radiation results in horizontal gradients. Using a mean difference in solar radiation of about 40 per cent between the tropics and the poles yields an upper limit of about 900 TW. The temporal variation of heating in time yields an additional power of about 170 TW at maximum efficiency, so that the total power generated from radiative heating gradients is of the order of 6000 TW.

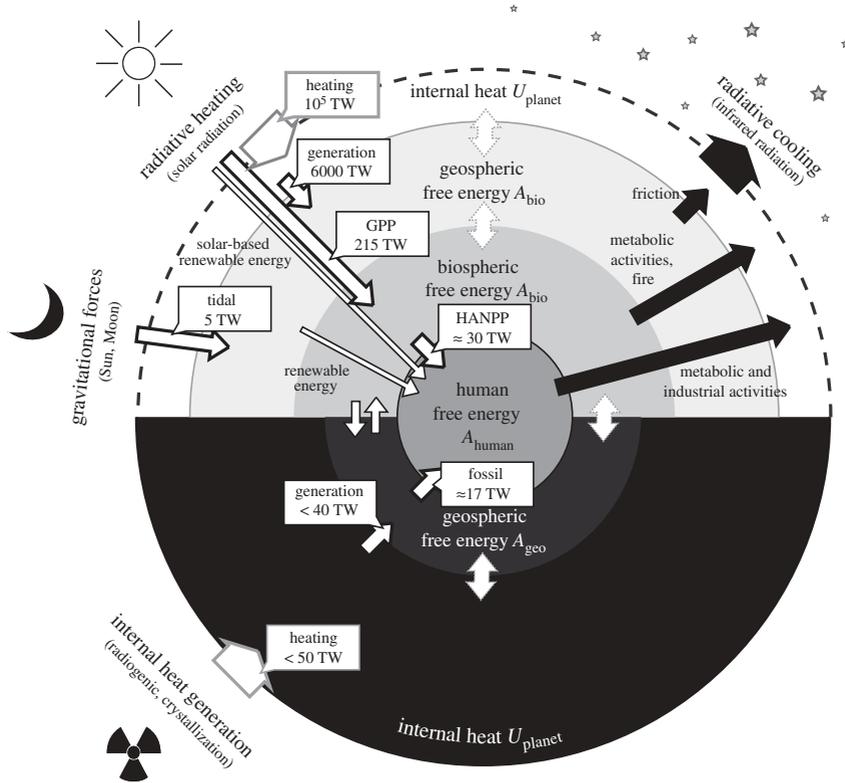


Figure 4. Estimates for global free energy generation rates at the planetary scale. Heating rates by solar radiation and interior processes are shown in boxes with grey borders. Free energy generation processes are shown by white boxes and arrows with black borders ('tidal', free energy generation by tidal forces; 'generation', free energy generation by heat engine processes; 'GPP', gross primary productivity, i.e. chemical free energy generation by photosynthesis). Human activity is driven by about 30 TW of human appropriation of net primary productivity of the biosphere ('HANPP') and the consumption of fossil fuels ('fossil'). $1 \text{ TW} = 10^{12} \text{ W}$. Adapted from Kleidon [17].

- *Solar photochemical engines.* Incident solar radiation contains wavelengths that can be used to generate chemical free energy when visible or ultraviolet radiation is absorbed by electronic absorption or photodissociation. Photodissociation can, in principle, generate radicals that are associated with free energy, but it is omitted here because those compounds have very short residence times and therefore unlikely to result in sustained free energy generation of significant magnitude. Photosynthesis is able to generate longer lasting free energy using complex photochemistry that prevents rapid dissipation. Using typical values for global gross primary productivity and typical free energy content of carbohydrates yields a generation rate of chemical free energy of about 215 TW [6].
- *Gravitational engines.* Gravitational forces by the Moon and the Sun provide some additional free energy by generating potential free energy mostly in the ocean in the form of tides. Estimates place the total generation rate at around 5 TW [37].

- *Interior heat engines.* Radioactive decay, crystallization of the core and secular cooling of the interior provide means to generate free energy within the interior. This free energy is associated with the kinetic energy of mantle convection and plate tectonics, with potential free energy generation associated with plate tectonics, with magnetic free energy generation associated with the maintenance of the Earth's magnetic field, and with geochemical free energy generation associated with metamorphism and other geochemical transformations. Given that the heat flux from the interior is less than 0.1 W m^{-2} at the Earth's surface, maximum efficiency estimates by Dyke *et al.* [6] yield a maximum generation rate of free energy in the various forms of about 40 TW.

In summary, the total free energy generation for current conditions yield about $P_{\text{geo},a} \approx 6070 \text{ TW}$ of power by physical processes within the atmosphere from the exchange fluxes at the Earth–space boundary, about $P_{\text{bio}} \approx 215 \text{ TW}$ of chemical free energy by photosynthesis, and about $P_{\text{geo},b} \approx 40 \text{ TW}$ driven by the depletion of initial conditions in the interior. Hence, the total power generation by the planet is about $P_{\text{planet}} = P_{\text{geo},a} + P_{\text{geo},b} + P_{\text{bio}} \approx 6325 \text{ TW}$.

(b) *Free energy transformations and planetary geochemical disequilibrium*

Only a small fraction of the total generated power P_{planet} is available for geochemical transformations. Of the $\approx 6070 \text{ TW}$ of geophysical free energy generated within the Earth's atmosphere, most is likely to be dissipated by atmospheric convection. More certain are large-scale dissipative terms. About 1000 TW are dissipated by frictional dissipation by the large-scale atmospheric circulation [38], 65 TW are transferred into the oceans to generate waves and maintain the wind-driven circulation [37], about 560 TW are associated with lifting water to the height at which it condenses and precipitates to the ground [17,39], and about 27 TW are associated with desalinating seawater.

Geochemical free energy is generated by abiotic means mostly by hydrological cycling. Of the 560 TW involved in hydrological cycling, about 13 TW are involved in maintaining streamflow. Of these 13 TW, only some fraction can be used to mechanically transform the continental crust by transporting sediments to the ocean. Precipitation on land also yields chemical free energy associated with disequilibrium of freshwater and the Earth's crust. This chemical free energy can be used to dissolve minerals of the continental crust. The power generated by precipitation is about 0.15 TW. The contribution by interior processes to the generation of geochemical free energy is likely to be very small because most of the 40 TW of power are involved in the generation of kinetic energy associated with mantle convection and plate tectonics.

In contrast to these very small generation terms of chemical free energy, biotic productivity generates 215 TW of chemical free energy. Not all of this free energy is available for geochemical transformations of the environment, as the metabolic activity of organisms consumes about half of the generated free energy. This contribution is nevertheless likely to be 1–2 orders of magnitude larger than abiotic means of geochemical free energy generation. Hence, this estimate substantiates the suggestion given by Lovelock [1] that the Earth's planetary geochemical disequilibrium is mostly attributable to the presence of widespread life on the planet.

(c) Free energy appropriation and dissipation by human activity

It is well recognized that human activity substantially alters the planet, as reflected by the suggestion to refer to the current geological era as the ‘anthropocene’ [40]. The additional heating by the burning of fossil fuels is, however, minute with its approximately 17 TW of primary energy consumption [41] compared with the solar heating of the order of 10^5 TW. The planetary impact of human activity is much easier to notice when considering the free energy appropriation related to human activity. Free energy is needed for maintaining human activity as it fuels the basic metabolic requirements (mostly food production) as well as the industrial activities (mostly fossil fuel consumption).

The basic requirement for metabolic activity is met by food production and the associated human appropriation of net primary production (HANPP) [42–45] from the biosphere. The concept of HANPP is more encompassing than merely the metabolic needs of humans, as it also considers contributions such as grazing by domesticated animals and the use of firewood. It is used here as a first-order approximation for the human demands for meeting metabolic energy needs. This need for free energy is estimated to be about 10–55% of the net primary productivity on land [43]. Using the estimated 40 per cent as a best guess [42] and converting the annual net primary productivity into units of free energy, this yields a free energy appropriation of $P_{\text{human,meta}} \approx 30$ TW. Note that this number is quite uncertain, not all of it is associated with food production, and 30 TW are not needed to directly meet human metabolic needs. With a human population of about five billion, and a basal metabolic rate of 100 W, this merely yields 0.5 TW of dissipation. The difference between 0.5 TW and 30 TW could be explained by a relatively large inefficiency of converting biotic productivity into food. If we, for instance, assume that the conversion efficiency is about 10 per cent for converting the productivity of grass to the maintenance of livestock and for the conversion of livestock into meat production, then the energetic needs to meet human metabolic requirements would translate into an equivalent chemical free energy generation by biotic productivity of the order of 50 TW, so that the estimate of 30 TW seems reasonable, but would require a more careful analysis. In addition, the industrial use of free energy, mostly from fossil fuels, is associated with $P_{\text{human,ind}} \approx 17$ TW of free energy.

In total, human activity, therefore, consumes about $P_{\text{human}} = P_{\text{human,meta}} + P_{\text{human,ind}} \approx 47$ TW. When compared with the planetary budget of free energy generation, especially regarding abiotic means of geochemical free energy generation of less than 40 TW, human energy consumption is a substantial term in the budget. The 47 TW of human consumption exceeds all free energy generated and consumed by geological processes of less than 40 TW in the Earth’s interior.

6. How does human activity change planetary free energy generation?

The free energy used for human activities are, of course, drawn out of the Earth system and thereby affect its state. At present, much of the free energy needs for industrial use are met by depleting a stock of geological free energy (in the form of fossil fuels) and this results in global climatic change owing to higher concentrations of carbon dioxide in the Earth’s atmosphere. If this depletion is

going to be replaced by renewable sources of free energy—as commonly suggested to avoid emissions of carbon dioxide—then this is going to leave an impact on the free energy balance of the planet. Other impacts of human activity, such as the emission of methane, nitrous oxide or soot, also relate, directly or indirectly, to the combustion of fuels or to food production, and should also relate to the Earth’s free energy balance. Hence, it would seem appropriate to relate human activity as well as its impacts on the Earth system to its basic driver, the need for free energy. This need for free energy would seem to be the most important metric to measure the impact of humans on the planet and would seem to serve to be a highly useful metric to evaluate potential future impacts.

As we have already seen in §5*c*, human activity already consumes a considerable share of the free energy in relation to how much is generated within the Earth system. When we think about the future state of the planet, it would seem almost inevitable that human activity will increase further, in terms of population size and standard of living, among others. Both of these will require more free energy to sustain. Then, the central question is going to be whether this increase in human activity is going to be met by degrading the ability of the Earth system to generate free energy or whether these demands will be met by enhancing the ability of the Earth system to generate free energy. If we think of the free energy budget shown in figure 4 as a pie, then these questions amount to the issue of whether an increase in human activity in the future is going to decrease or increase the planetary pie of free energy generation, thereby depleting or enhancing the planetary disequilibrium. This concept is illustrated in figure 5, and its application is illustrated qualitatively in the following section.

The need for meeting our metabolic demands results in the conversion of natural lands into agricultural use. Because humans appropriate productivity (i.e. $\Delta P_{\text{human,meta}} > 0$), the associated changes in land cover are likely to result in less productivity available for the biota (i.e. $\Delta P_{\text{bio}} < 0$), therefore resulting in less free energy generation to drive biotic activity. The associated land-cover changes result in different functioning of the land surface with consequences for the atmosphere. A shift from forests to agricultural lands is usually accompanied with a higher reflectivity and a reduced ability to recycle water back into the atmosphere [46]. When we consider the extreme scenario of all vegetation being removed from land, the associated climatic conditions would likely be less favourable to biospheric free energy generation, as demonstrated by extreme climate model simulations [47,48]. It would hence seem that an increase in agricultural activity ($\Delta P_{\text{human,meta}} > 0$), in principle, would result in shrinking the pie of biotic free energy generation ($\Delta P_{\text{bio}} < 0$). This decrease was also shown in sensitivity simulations with a coupled vegetation–climate model to the magnitude of human appropriation by Kleidon [45].

The need for heating and industrial activity associated with human activity, $P_{\text{human,ind}}$, is currently fuelled to a large extent by fossil fuels. Even though the fossil free energy is not taken away from an active, geological process that affects the Earth system at present, its consumption results in carbon dioxide emissions into the atmosphere. The expected increase in energy demands $\Delta P_{\text{human,ind}} > 0$ as well as the necessary shift towards sustainable sources of free energy, such as wind power, hydropower, tidal power and so on, the appropriation of such sources of free energy by humans would obviously reduce the free energy within the system, i.e. $\Delta A_{\text{planet}} < 0$. Because these processes are likely to operate already at maximum

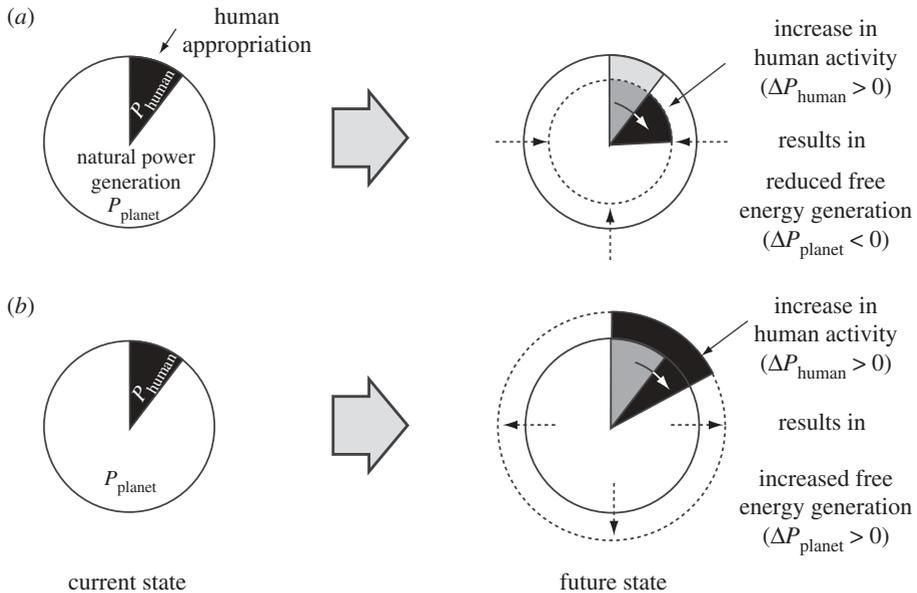


Figure 5. Schematic of possible contrasting effects of an increase in human activity, corresponding to an increase in free energy consumption, $\Delta P_{\text{human}} > 0$, on the free energy generation capability of the biosphere and the Earth system. (a) Effects of $\Delta P_{\text{human}} > 0$ result in a reduction of free energy generation by natural processes, i.e. $\Delta P_{\text{planet}} < 0$. (b) Alternative case in which the effects of $\Delta P_{\text{human}} > 0$ result in an overall enhancement of free energy generation by natural processes, i.e. $\Delta P_{\text{planet}} > 0$. For a sustainable future with inevitable increases in energy demands by human activity, it is argued in the text that human effects should follow (b).

efficiency, the associated impacts are likely going to be such that the power of these processes is going to decrease as a result of human appropriation (i.e. $\Delta P_{\text{planet}} < 0$). That this is indeed the case has been demonstrated with sensitivity simulations with a climate model to the magnitude of wind power extraction at the surface [49].

Both examples of meeting the human demands for free energy suggest that human activity will result in detrimental effects in terms of the ability of the Earth system to generate free energy. We can, however, also imagine another scenario. If human activity is directed to have impacts of the sort that would act to enhance free energy generation within the Earth system, as shown in figure 5b, then this could have beneficial effects on the overall system in that the ability to generate free energy within the Earth system is enhanced. Two examples are given below to illustrate such potentially positive impacts owing to human activity.

To meet an increase in demands for metabolic free energy generation, one could imagine that if currently unproductive lands are used and converted into agricultural lands with the use of technology (e.g. by irrigation with desalinated seawater), then this could result in enhanced free energy generation by the biota. Then, an increase $\Delta P_{\text{human,meta}} > 0$ would be met with a change $\Delta P_{\text{bio}} > 0$. This could, for instance, be accomplished using technology to ‘green’ the desert (see also Ornstein *et al.* [50]). If the resulting gain in productivity is larger than the

technological needs for free energy to enable desert greening, then we would have a win–win situation of overall enhanced free energy generation by the Earth system. If ‘desert greening’ would indeed yield overall such an energetic gain, however, would need to be carefully evaluated in greater detail.

The increase in demands for industrial free energy could be met by more efficient use of solar radiation. Currently, most of the solar radiation is absorbed and immediately converted into heat. Differential heating is, however, very inefficient in converting solar radiation into free energy. Photovoltaic cells or the use of direct solar radiation are two examples of technologies that could be used to enhance the efficiency of converting solar radiation into free energy. Their use in currently unvegetated regions, such as deserts, could result in an overall enhanced ability of the Earth system to generate free energy with the use of human technology—just as in the example of ‘desert greening’.

Such conscious planning of human effects to enhance the power generation by the Earth system would constitute a form of ‘geoengineering’ or ‘Earth system engineering’. Currently proposed measures of ‘geoengineering’ focus on active measures to ‘undo’ the effects of human activities. In particular, they focus on counteracting the warming induced by higher greenhouse gas concentrations [51]. What seems like a logical consequence of these free energy considerations is that we first need to acknowledge that human activity will result in unavoidable impacts (see also Allenby [52]), but that the resulting impacts should be ‘engineered’ towards enhancing the ability of the planet to generate free energy. Given the sheer magnitude of human activity in terms of free energy consumption and its expected increase in the future, it would seem that such a form of the Earth system engineering is the only sustainable way to support such an increase in human activity without depleting the sources of free energy generation within the Earth system. Such active and conscious intervention would, of course, require careful analysis.

7. What is needed in Earth system models to adequately represent free energy generation, dissipation and transfer?

To evaluate the effect of the different options of how humans appropriate free energy from the Earth system, one would need Earth system models that capture the dynamics of free energy generation, transfer and dissipation. Such an analysis, however, is practically absent in current applications of Earth system models. While the energetics of some Earth system processes are diagnosed, for instance, the kinetic energy generation in the atmosphere [53] or the ocean [37], most analyses focus on the balance of heating and cooling terms and associated temperature differences (e.g. for the case of global warming, see IPCC [54]). For differences in heat content and temperature, this is justified, because the magnitude of free energy generation, transfer and dissipation is much smaller than the mean heating terms, so that their effect in the energy balance can be neglected. However, it is the dynamics of free energy generation, transfer and dissipation that shapes disequilibrium, and it is this disequilibrium that seems an appropriate way to characterize a habitable planet. So, it would seem that we miss a critical aspect when diagnosing Earth system model simulations of global change.

To allow for the analysis of the free energy balance, it would seem that some aspects are being missed in current Earth system models regarding the diagnostics, and, more importantly, regarding the adequate representation of free energy dynamics. In terms of process representation, some processes are treated as diffusive, even though they are not. In contrast to a diffusive process, a non-diffusive process is associated with free energy generation, dissipation and, potentially, transfer to another process. This is, for instance, the case for preferential flow of water in soils. Preferential flow describes the rapid flow of water along spatially connected flow paths of minimum flow resistance within the soil, and this results in the faster depletion of gradients than would be predicted using common, diffusion-based models [55]. Missing the dynamics of kinetic energy generation associated with the rapid flow does not allow for work possibly being drawn from the flow to transport material that could generate and maintain these connected flow paths within the soil. Overall, without the dynamics of free energy generation, transfer and dissipation, the dynamics of water movement would likely be slower, less dissipative and hence misrepresented.

Thermodynamics and thermodynamic maximization can potentially also help us to address the representation of heterogeneity and its effects on the functioning at larger scales in climate models. Heterogeneity inherently reflects gradients, and as such reflects thermodynamic disequilibrium. One example where thermodynamic maximization has been used to parametrize the effect of subgrid-scale heterogeneity is the case of pattern formation of vegetation in semiarid regions. Schymanski *et al.* [56] used a simple model based on a reaction–diffusion equation to show that the formation of patterns results in a thermodynamic maximization associated with water redistribution and biotic productivity. It has also been hypothesized that the patterns found in sand dunes reflect maximum aeolian transport [57], i.e. the maximum transfer of kinetic energy of boundary-layer winds to perform work on sand transport. These studies provide a clue of the potential use of thermodynamic maximization to characterize the non-randomness of subgrid-scale heterogeneity and how it could possibly be described at a larger scale. This would need further evaluations.

Another aspect that is likely absent in some model implementations is the adequate representation of free energy transfer. When, for instance, air flows over an open water surface, some of the momentum gradient between the airflow and the water surface is transferred to generate water flow. This momentum is not dissipated by turbulence, but rather transferred into another form of free energy in the form of kinetic energy of the moving water. Hence, the acceleration of the water surface is another form of reducing the momentum gradient between the airflow and the water surface. The same example would also apply to the surface–atmosphere exchange on land, where momentum from the air flow is transferred to lift dust or to sway canopies. When this form of momentum reduction is not accounted for and momentum gradients are only reduced by turbulent dissipation, this should result in an overestimation of the associated turbulent fluxes that exchange heat and matter across the surface.

Apart from ensuring thermodynamic consistency, adequately representing the dynamics of free energy generation, transfer and dissipation in Earth system models would allow us to do the following.

- Test how close natural processes operate near-maximum efficiency of free energy generation and transfer. This could potentially result in simpler and better founded model parametrizations.
- Phrase interactions and feedbacks within the Earth system in the context of free energy generation and transfer. This would then help us to better understand the processes and dynamics that are involved in maintaining maximum efficiency.
- Quantify the role of the biota in the generation of free energy by the Earth system. This would allow for a better quantification of how much biotic activity contributes to the planetary disequilibrium of the Earth.
- Quantify the impact of human activity on the generation of free energy by Earth system processes. This would potentially yield a better way to measure the impact of human activity than changes in surface temperature.
- Better quantify the free energy budget. Because this budget includes the generation rates of free energy, it would provide a baseline budget for the availability of different forms of renewable energy.
- Determine strategies for future free energy appropriation by humans with minimum impact on free energy generation by Earth system processes. This would ensure that the impacts of human activity on the Earth system have no detrimental effects on free energy generation by Earth system processes.

It would thus seem that an adequate representation of free energy dynamics in Earth system models is not just a matter of evaluating the theory proposed here. It would rather seem that free energy dynamics, especially regarding the chains of conversions of free energy from one form to another—as in the example of momentum transfer at the surface above—are a fundamental aspect that needs to be properly accounted for in models that aim to represent the non-equilibrium dynamics of the Earth system and its response to change.

8. Summary and conclusions

I have provided a holistic description of the functioning of the whole Earth system that is grounded in the generation, transfer and dissipation of free energy from external forcings to geochemical cycling and the associated fundamental limits to these rates. Because free energy generation is needed to maintain a disequilibrium state, this description allows us to understand why the Earth system is maintained far from equilibrium without violating the second law of thermodynamics. I showed how biotic activity generates substantial amounts of chemical free energy by exploiting free energy in solar photons that is not accessible to purely physical heat engines. Hence, Lovelock's notion of chemical disequilibrium within the Earth's atmosphere as a sign for widespread life can be substantiated and quantified. This paper can hence be seen as a direct continuation of the work by Lovelock [1,2] on understanding the Earth as a single system that is strongly shaped by life.

The relevance of this holistic description of the Earth system becomes apparent when the impacts of human activity are evaluated from this perspective. When using free energy consumption as a measure for human activity, it is evident that human activity as an Earth system process is far greater and significant in comparison with natural processes than what it would seem using other, more traditional measures. An increase in human activity in the future, e.g. in terms of population size and standard of living, would inevitably result in greater needs for free energy generation. If these needs are met by appropriating renewable sources of free energy within the Earth system, then this is inevitably going to leave its impact in that the processes associated with this form of free energy will become less intense and ‘slow down’. The only sustainable way to meet the increasing needs for free energy by human activity would seem to use human technology in such a way that it would enhance the overall ability of the Earth system to generate free energy. This was illustrated using the two examples of ‘desert greening’ and the direct use of solar energy by photovoltaics or by heat engines using direct solar radiation in deserts. Even though this would require careful analysis and planning of potential, detrimental side effects, it would seem that it is only through the large-scale use of human technology that the Earth system could sustainably generate more free energy, yielding a more prosperous and empowered future of the planet.

The evaluation of such potential scenarios of the planetary future associated with growing human activity with respect to their impact on the ability of the Earth system to generate free energy is at this point qualitative and would need to be more rigorously evaluated. This can, in principle, be performed by climate models. However, while these models are routinely used to predict typical meteorological properties such as temperature and precipitation rates, the evaluation in terms of rates of free energy generation and transfer is rare. Apart from providing fundamental insights, thermodynamic maximization could provide a useful approach to capture subgrid-scale heterogeneity in global scale models. The transfer of free energy among different variables would need to be adequately represented in these models to allow for the thermodynamic consistency of the dynamics. Such thermodynamic consistency would, of course, be a first requirement for an improved prediction of human impacts and the planetary future.

The author acknowledges funding from the Helmholtz Alliance Planetary Evolution and Life. This manuscript resulted from a presentation and subsequent discussions at the Newton Institute in August 2010. The author thanks the organizers for the stimulating meeting. Fruitful discussions with members of the Biospheric Theory and Modelling group are also acknowledged. The author thanks two reviewers for their constructive and stimulating comments.

References

- 1 Lovelock, J. E. 1965 A physical basis for life detection experiments. *Nature* **207**, 568–570. (doi:10.1038/207568a0)
- 2 Lovelock, J. E. 1975 Thermodynamics and the recognition of alien biospheres [and Discussion]. *Proc. R. Soc. Lond. B* **189**, 167–181. (doi:10.1098/rspb.1975.0051)
- 3 Pauluis, O. & Held, I. M. 2002 Entropy budget of an atmosphere in radiative convective equilibrium. Part I: maximum work and frictional dissipation. *J. Atmos. Sci.* **59**, 126–139. (doi:10.1175/1520-0469(2002)059<0125:EBOAAI>2.0.CO;2)

- 4 Pauluis, O. & Held, I. M. 2002 Entropy budget of an atmosphere in radiative convective equilibrium. Part II: latent heat transport and moist processes. *J. Atmos. Sci.* **59**, 140–149. (doi:10.1175/1520-0469(2002)059<0140:EBOAAI>2.0.CO;2)
- 5 Dietrich, W. E. & Perron, J. T. 2006 The search for a topographic signature of life. *Nature* **439**, 411–418. (doi:10.1038/nature04452)
- 6 Dyke, J. G., Gans, F. & Kleidon, A. 2010 Assessing life's effects on the interior dynamics of planet Earth using non-equilibrium thermodynamics. *Earth Syst. Dyn. Discuss.* **1**, 191–246. (doi:10.5194/esdd-1-191-2010)
- 7 Lovelock, J. E. & Margulis, L. 1974 Atmospheric homeostasis by and for the biosphere: the Gaia hypothesis. *Tellus* **26**, 2–10. (doi:10.1111/j.2153-3490.1974.tb01946.x)
- 8 Watson, A. J. & Lovelock, J. E. 1983 Biological homeostasis of the global environment: the parable of Daisyworld. *Tellus* **35B**, 284–289. (doi:10.1111/j.1600-0889.1983.tb00031.x)
- 9 Lovelock, J. E. 2003 Gaia: the living Earth. *Nature* **426**, 769–770. (doi:10.1038/426769a)
- 10 Schneider, S. H. & Boston, P. J. (eds) 1991 *Scientists on Gaia*. Cambridge, MA: MIT Press.
- 11 Schneider, S. H., Miller, J. R., Crist, E. & Boston, P. J. (eds) 2004 *Scientists debate Gaia: the next century*. Cambridge, MA: MIT Press.
- 12 Kondepudi, D. & Prigogine, I. 1998 *Modern thermodynamics: from heat engines to dissipative structures*. Chichester, UK: Wiley.
- 13 Lebon, G., Jou, D. & Casas-Vazquez, J. 2008 *Understanding non-equilibrium thermodynamics: foundations, applications, frontiers*. Berlin, Germany: Springer.
- 14 Ozawa, H., Ohmura, A., Lorenz, R. D. & Pujol, T. 2003 The second law of thermodynamics and the global climate system: a review of the maximum entropy production principle. *Rev. Geophys.* **41**, 1018. (doi:10.1029/2002RG000113)
- 15 Lineweaver, C. H. & Egan, C. A. 2008 Life, gravity and the second law of thermodynamics. *Phys. Life Rev.* **5**, 225–242. (doi:10.1016/j.plrev.2008.08.002)
- 16 Kleidon, A. 2010 Life as the major driver of planetary geochemical disequilibrium: reply to comments on 'Life, hierarchy, and the thermodynamic machinery of planet Earth'. *Phys. Life Rev.* **7**, 473–476. (doi:10.1016/j.plrev.2010.11.001)
- 17 Kleidon, A. 2010 Life, hierarchy, and the thermodynamic machinery of planet Earth. *Phys. Life Rev.* **7**, 424–460. (doi:10.1016/j.plrev.2010.10.002)
- 18 Alberty, R. A. 2001 Use of Legendre transforms in chemical thermodynamics (IUPAC technical report). *Pure Appl. Chem.* **73**, 1349–1380. (doi:10.1351/pac200173081349)
- 19 Kleidon, A. 2009 Maximum entropy production and general trends in biospheric evolution. *Paleontol. J.* **43**, 130–135. (doi:10.1134/S0031030109080164)
- 20 Kleidon, A. 2010 Non-equilibrium thermodynamics, maximum entropy production and Earth-system evolution. *Phil. Trans. R. Soc. A* **368**, 181–196. (doi:10.1098/rsta.2009.0188)
- 21 Kleidon, A. & Lorenz, R. D. (eds) 2005 *Non-equilibrium thermodynamics and the production of entropy: life, earth, and beyond*. Heidelberg, Germany: Springer.
- 22 Martyushev, L. M. & Seleznev, V. D. 2006 Maximum entropy production principle in physics, chemistry, and biology. *Phys. Rep.* **426**, 1–45. (doi:10.1016/j.physrep.2005.12.001)
- 23 Kleidon, A., Malhi, Y. & Cox, P. M. 2010 Maximum entropy production in environmental and ecological systems. *Phil. Trans. R. Soc. B* **365**, 1297–1302. (doi:10.1098/rstb.2010.0018)
- 24 Dewar, R. C. 2003 Information theory explanation of the fluctuation theorem, maximum entropy production, and self-organized criticality in non-equilibrium stationary states. *J. Phys. A* **36**, 631–641. (doi:10.1088/0305-4470/36/3/303)
- 25 Dewar, R. C. 2005 Maximum entropy production and non-equilibrium statistical mechanics. In *Non-equilibrium thermodynamics and the production of entropy: life, earth, and beyond* (eds A. Kleidon & R. D. Lorenz), pp. 41–56. Heidelberg, Germany: Springer.
- 26 Dewar, R. C. 2005 Maximum entropy production and the fluctuation theorem. *J. Phys. A* **38**, L371–L381. (doi:10.1088/0305-4470/38/21/L01)
- 27 Niven, R. K. 2009 Steady state of a dissipative flow-controlled system and the maximum entropy production principle. *Phys. Rev. E* **80**, 021113. (doi:10.1103/PhysRevE.80.021113)
- 28 Ozawa, H., Shimokawa, S. & Sakuma, H. 2001 Thermodynamics of fluid turbulence: a unified approach to the maximum transport properties. *Phys. Rev. E* **64**, 026303. (doi:10.1103/PhysRevE.64.026303)

- 29 Ozawa, H. & Ohmura, A. 1997 Thermodynamics of a global-mean state of the atmosphere: a state of maximum entropy increase. *J. Clim.* **10**, 441–445. (doi:10.1175/1520-0442(1997)010<0441:TOAGMS>2.0.CO;2)
- 30 Paltridge, G. W. 1975 Global dynamics and climate: a system of minimum entropy exchange. *Q. J. R. Meteorol. Soc.* **101**, 475–484. (doi:10.1002/qj.49710142906)
- 31 Paltridge, G. W. 1978 The steady-state format of global climate. *Q. J. R. Meteorol. Soc.* **104**, 927–945. (doi:10.1002/qj.49710444206)
- 32 Kleidon, A., Fraedrich, K., Kunz, T. & Lunkeit, F. 2003 The atmospheric circulation and states of maximum entropy production. *Geophys. Res. Lett.* **30**, 2223. (doi:10.1029/2003GL018363)
- 33 Kleidon, A., Fraedrich, K., Kirk, E. & Lunkeit, F. 2006 Maximum entropy production and the strength of boundary layer exchange in an atmospheric general circulation model. *Geophys. Res. Lett.* **33**, L06706. (doi:10.1029/2005GL025373)
- 34 Lorenz, R. D., Lunine, J. I., Withers, P. G. & McKay, C. P. 2001 Titan, mars and earth: entropy production by latitudinal heat transport. *Geophys. Res. Lett.* **28**, 415–418. (doi:10.1029/2000GL012336)
- 35 Lorenz, R. D. 2010 The two-box model of climate: limitations and applications to planetary habitability and maximum entropy production studies. *Phil. Trans. R. Soc. B* **365**, 1349–1354. (doi:10.1098/rstb.2009.0312)
- 36 Kleidon, A. 2009b Non-equilibrium thermodynamics and maximum entropy production in the Earth system: applications and implications. *Naturwissenschaften* **96**, 653–677. (doi:10.1007/s00114-009-0509-x)
- 37 Ferrari, R. & Wunsch, C. 2009 Ocean circulation kinetic energy: reservoirs, sources, and sinks. *Annu. Rev. Fluid Mech.* **41**, 253–282. (doi:10.1146/annurev.fluid.40.111406.102139)
- 38 Li, L., Ingersoll, A. P., Jiang, X., Feldman, D. & Yung, Y. L. 2007 Lorenz energy cycle of the global atmosphere based on reanalysis datasets. *Geophys. Res. Lett.* **34**, L16813. (doi:10.1029/2007GL029985)
- 39 Pauluis, O., Balaji, V. & Held, I. M. 2000 Frictional dissipation in a precipitating atmosphere. *J. Atmos. Sci.* **57**, 987–994. (doi:10.1175/1520-0469(2000)057<0989:FDIAPA>2.0.CO;2)
- 40 Crutzen, P. J. 2002 Geology of mankind. *Nature* **415**, 23. (doi:10.1038/415023a)
- 41 International Energy Outlook 2009 Technical Report. DOE/EIA 0484(2009), US Department of Energy, Energy Information Administration, Washington, DC.
- 42 Vitousek, P. M., Ehrlich, P. R., Ehrlich, A. H. & Matson, P. A. 1986 Human appropriation of the products of photosynthesis. *Bioscience* **36**, 368–373. (doi:10.2307/1310258)
- 43 Rojstaczer, S., Sterling, S. M. & Moore, N. J. 2001 Human appropriation of photosynthesis products. *Science* **294**, 2549–2552. (doi:10.1126/science.1064375)
- 44 Imhoff, M. L., Bounoua, L., Ricketts, T., Loucks, C., Harriss, R. & Lawrence, W. T. 2004 Global patterns in human consumption of net primary production. *Nature* **429**, 870–873. (doi:10.1038/nature02619)
- 45 Kleidon, A. 2006 The climate sensitivity to human appropriation of vegetation productivity and its thermodynamic characterization. *Glob. Planet. Change* **54**, 109–127. (doi:10.1016/j.gloplacha.2006.01.016)
- 46 Bonan, G. B. 2008 Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *Science* **320**, 1444–1449. (doi:10.1126/science.1155121)
- 47 Kleidon, A., Fraedrich, K. & Heimann, M. 2000 A green planet versus a desert world: estimating the maximum effect of vegetation on land surface climate. *Clim. Change* **44**, 471–493. (doi:10.1023/A:1005559518889)
- 48 Kleidon, A. 2002 Testing the effect of life on Earth’s functioning: how Gaian is the Earth system? *Clim. Change* **66**, 271–319. (doi:10.1023/B:CLIM.0000044616.34867.ec)
- 49 Miller, L. M., Gans, F. & Kleidon, A. 2011 Estimating maximum global land surface wind power extractability and associated climatic consequences. *Earth Syst. Dyn.* **2**, 1–12. (doi:10.5194/esd-2-1-2011)
- 50 Ornstein, L., Aleinov, I. & Rind, D. 2009 Irrigated afforestation of the Sahara and Australian outback to end global warming. *Clim. Change* **97**, 409–437. (doi:10.1007/s10584-009-9626-y)
- 51 Lenton, T. M. & Vaughan, N. E. 2009 The radiative forcing potential of different climate geoengineering options. *Atmos. Chem. Phys.* **9**, 5539–5561. (doi:10.5194/acp-9-5539-2009)

- 52 Allenby, B. 2000 Earth systems engineering and management. *IEEE Technol. Soc. Mag.* **19**, 10–24. (doi:10.1109/44.890078)
- 53 Lorenz, E. N. 1955 Available potential energy and the maintenance of the general circulation. *Tellus* **7**, 157–167. (doi:10.1111/j.2153-3490.1955.tb01148.x)
- 54 Intergovernmental Panel on Climate Change (IPCC) 2001 Climate change 2001: the scientific basis. Contribution of Working Group I to the third assessment report of the IPCC. Cambridge, UK: Cambridge University Press.
- 55 Zehe, E., Blume, T. & Blöschl, G. 2010 The principle of ‘maximum energy dissipation’: a novel thermodynamic perspective on rapid water flow in connected soil structures. *Phil. Trans. R. Soc. B* **365**, 1377–1386. (doi:10.1098/rstb.2009.0308)
- 56 Schymanski, S. J., Kleidon, A., Stieglitz, M. & Narula, J. 2010 Maximum entropy production allows a simple representation of heterogeneity in semiarid ecosystems. *Phil. Trans. R. Soc. B* **365**, 1499–1455. (doi:10.1098/rstb.2009.0309)
- 57 Rubin, D. M. & Hunter, R. E. 1987 Bedform alignment in directionally varying flows. *Science* **237**, 276–278. (doi:10.1126/science.237.4812.276)