Chapter 3

Accounting scales of motion

Fluids are fascinating because they contain a lot of different motions at the same time, evocative of the potentialities of mind. Motions of different *sizes* are especially intriguing, like vigorous little scrambles (a dragon! no, a peacock!) caught up in sky-sized swirls, all drifting by on continental or planet-sized currents. These raise profound and important questions about predictability and knowability. To address them rationally, we need bookkeeping tools to decompose motion into scale components. Chapter 1 reviewed how to account for stuff (including momentum and kinetic energy) moving in space. Now we wish to keep track of motions in the abstract domain of *size*, whose vastness requires us to use logarithmic *scale*. The tools are crude, which will entice us into entity characterization in Part II, but valuable ideas exist along the way.

3.1 One size cut: molecular vs. macroscopic

Borders between regions of space are crossed by individual careening molecules, and also by coherent masses of them moving together. These cases should be distinguished, because the continuum limit of calculus only applies safely to the latter. In addition to contributing discrete little fluxes of mass, molecules carry specific stuff like momentum and energy from conditions at their place of 'origin' (last contact with other matter) to the place of their next collision, where that stuff is rapidly shared out in a new *neighborhood*, a term that properly implies the spatial averaging inherent in continuum reasoning.

Logic (not detailed here; think) tells us that the net flux of stuff by numerous random bidirectional exchanges of molecules act as a *flux down the gradient, proportional to the strength of the gradient.* Such a flux is called *diffusive*. Treating the proportionality coefficient as constant, the time-tendency called *diffusion* (the *convergence of diffusive flux*) is thus positively proportional to the *Laplacian operator*, $\boldsymbol{\nabla}^2 = \boldsymbol{\nabla} \cdot \boldsymbol{\nabla}$, which we also met in the pressure equation (problem 1.4.4 and solution 1.5). This diffusive molecular flux (called *viscosity* in the case of momentum) is separated from the flux by fluid motions in the *Navier-Stokes equations* of fluid dynamics. Specifically, $\nu\boldsymbol{\nabla}^2 u,\ \nu\boldsymbol{\nabla}^2 v,\ \nu\boldsymbol{\nabla}^2 w$ get pulled out of the flux terms in equations (1.4 b,c,d), respectively, and added to the RHS, leaving the flux by coherent groups of molecules to be treated as advection in equation sets like (2.10) and (2.15).

3.2 Another cut: large-scale flow vs. small "eddies"

Do *fluid blobs smaller than the flow we care about* (that is, the "eddies" in whatever stream we are pondering) act like molecules? That is, as they cross a border we care about, between regions we distinguish -- both ways, because of mass continuity -- can their effect be represented as a *diffusive flux* of specific stuff? If so, hooray! We simply reuse the argument of section 3.1, enhancing[a] molecular viscosity ν with an *eddy diffusivity,* called *eddy viscosity* when momentum is the stuff being transported. In that case, the Laplacian's curvaceousness-feeling derivatives must be redefined to have a broader span, based on *the size of those flow features we care about*. These personal words emphasize that scale separation is a tool for our reasoning, not a property of nature.

Specifically, the words *mean conditions* and *neighborhood* from section 3.1 should be redefined to denote a broader averaging operation or *spatial smoothing filter* (denoted by overbars $\bar{u}, \bar{v}, \bar{w}$) that just preserves, with acceptable blurring, the features we care about. We then write eddy

---

[a] or *replacing,* since eddy flux is orders of magnitude larger for Earthly convection.

quantities, representing all the smaller scales removed by the filter, with a prime[b]:

$$u' \equiv u - \bar{u} \qquad (3.1)$$

Such a definitional equation for all variables can be substituted into the equations in flux form, and the terms involving products of primes (sub-filter-scale *eddy flux* terms) moved to the right hand side. For instance, for u, (2.15b) can be elaborated in this way as:

$$\begin{aligned}\bar{u}_t = -\bar{u}\,\overline{u_x} - \bar{v}\,\overline{u_y} - \bar{w}\,\overline{u_z} - \overline{\pi_x} + \overline{fv} \\ +\overline{(u'u')_x} \;\; + \overline{(v'u')_y} \;\; + \overline{(w'u')_z} \qquad (3.2)\end{aligned}$$

Since the powerful Law of mass continuity applies to both large and small scales, the overbar terms (transport by smooth, large-scale, *filtered* motions) can be converted from flux form back to the *apparent advection* form in (3.2), as in problem 3.10.1[c].

What has been gained from this exercise? Nothing yet, until the eddy terms on the RHS are treated somehow, but one rigorous scale cut has been made, and the messy eddy flux terms are its notational cost. Only an intrepid few feel inspired to write comparable bookkeping for *higher order* expansions of RHS terms (e.g. André et al. 1976), or make another size cut with a second smoothing filter operation (e.g. Tan et al. 2018).

For those interested in planetary-scale flows, a typical horizontal filter scale is Earth's size divided by about 1000 (for our visual screens) or 100 (fitting our brain's social hardware[d]) or 10 (for counting on our fingers). Such a filter means that all of the white parts of several or many

---

[b] The prime was also used in chapter 2 where we only *cared about* the whole-fluid horizontal mean denoted by the special averaging operators $\bar{\bar{\rho}}$ and [KE] in problem 2.4.1.

[c] Averaged prime terms like $\overline{\bar{u}u'}$ vanish only if the averaging is done over *fixed limits of integration*, not for a sliding boxcar or convolution smoother that leaves $\bar{u}$ a continuous function. For this reason, x,y,z differentiation subscripts are kept under the bar.

[d] https://en.wikipedia.org/wiki/Dunbars_number.

convective clouds are implied in the definition of "eddy", from which we then want to know only a *net effect:* the flux of stuff across the borders of regions of sizes we care about.

The atmosphere invites a peculiar application of this averaging nomenclature, in which the bar reflects only a *horizontal* averaging filter, retaining full dependence on vertical position. Models built around this mathematics retain far more vertical resolution than most viewers of the fields "care about" examining. But users want skillful horizontal flow evolution over time, which requires that filter-scale motions and physics (including sub-*horizontal*-filter-scale eddy fluxes) be computed accurately in the vertical. This is the strange realm of *convection parameterization* (confronted more in Part III), an engineering challenge that does also stand as an important and genuine test of our scientific understanding of convection (Arakawa 2004).

3.3 On deviations, anomalies, eddies, perturbations, etc.

Primes and overbars like (3.2) are sprinkled throughout the literature of fluid science, with various meanings. Customarily these are used for *time* averages and *deviations* from them, called *anomalies* or sometimes *fluctuations.* Square brackets and asterisks like $[v^*u^*]$ represent *zonal* (east-west around the globe) spatial averages and *zonal eddy* deviations therefrom, in planetary-scale atmospheric general circulation work. In the *Reynolds average* interpretation, bars and primes are viewed as an *ensemble* mean plus individual *member or realization* deviations from that mean. Sometimes deviations are vaguely called *perturbations,* which makes the most sense if some agent changes or perturbs something else, like in an *experiment minus control* setting. In *time series* data analysis, primes and bars are sometimes interpreted as a proxy for spatial deviations or for Reynolds realizations, a view whose validity hinges on caveats like *stationarity* or its mathematical rabbit-hole cousin *ergodicity*. Wikipedia pages and literatures abound for all these terms and concepts, but here we merely emphasize that readers of the diverse literature should allocate enough mental capacity for thinking clearly about these conceptual distinctions, since the terms are sometimes used carelessly or interchangeably.

3.4 Fourier (spectral) decomposition and logarithmic *scale*

To make more than one or two size distinctions, we can leap to another kind of accounting: a *spectrum*[e] of size components such as sine and cosine waves (*Fourier harmonics*). Fourier analysis is one particular type of *Galerkin decomposition,* which divides complex structure into a weighted sum over a *basis set* of complete and orthogonal (mutually exclusive) basis functions. The advanced student should explore *wavelet analysis,* which illuminates the fundamental tradeoff or uncertainty principle between accounting for location vs. for scale.

To appreciate Fourier spectra clearly, the familiar words *wavelength* in space and *period* in time must be rethought as two pi (2 x 3.141...) times their inverses, called *wavenumber* in space and *frequency* in time. The abstractness of wavenumber (frequency) is lessened by measuring them in the units *number of cycles over a given distance* (or *cycles over a given time interval*), using our brain's counting faculty for wave crests or troughs. The word *amplitude* refers to the coefficient of each term, and its square (double the *variance* of each wavenumber component) is often called *power* (or *power spectral density*) for reasons that may be more historical than sensible: The units are variance (kinetic energy, in the case of velocity variance) per frequency bin width, not energy per time (true power from physics).

For finite data series (which Fourier analysis necessarily treats as repeating or periodic), the set of possible wavenumbers is discrete. Since these wavenumbers are "bins" over which we distribute conserved stuff like energy (variance)[f], the amount of stuff falling in each bin depends on the widths (that is, the spacings) of those bins. The inverse dependence of wavenumber on wavelength (or frequency on period) makes the bins

---

[e] The spooky revelation of secret colors in white light evoked *spectres* or apparitions.

[f] See https://en.wikipedia.org/wiki/Parsevals_theorem

.

wider for long waves and narrower for short waves. For instance, when the fluctuations in a mesh of 4km discrete spatial grid boxes around the Earth's equator (10,000 points) are expressed as a power spectrum, all the variance in the wavelength range from 4000-40,000 km falls into just 10 wavenumber bins, while energy in the wavelength range 40-400 km is divided over 1000 bins. Partly for this reason, "power" spectra almost invariably tend to have much more "energy" in long wave bins.

Mathematically, the Fourier Transform and its inverse are linear and reversible, allowing us to jump back and forth between the *physical domain* (space and time) and *spectral domain* (wavenumber and frequency components) without any loss of information, with just a few pencil strokes, or now keystrokes. This is an essential power tool every science student should learn to appreciate and apply, if not to fully understand to its depths[g]. Computer exercises in section 3.9 encourage readers to play with Fourier analysis of imagery and sound.

To tame the vastness of the size or wavenumber domain, the *log* function suggests itself. In fact, logarithms are fundamentally what we mean by the word *scale* as opposed to *size*. We speak of *meter-scale eddies* whose size is about 1-10m but not 100m, while *kilometer-scale* clouds are about 1-10km across but not 100km. Scale is measured in *decades* (factors of 10), or sometimes more finely in *octaves* (factors of 2; but named for the mysteriously pleasing brainfeel of tones 1/8 of an octave apart in the ancient art of music on stringed instruments).

Any spatial or temporal data series can be *decomposed into* a sum of Fourier components. But does this mean that nature is truly, deeply, secretly *composed of* those components? Consider a short wave (with large wavenumber *K*), $\psi = A\ sin(Kt)$, whose amplitude A is *modulated by* a long wave with small wavenmber *k*: $A = \ C\ sin(kt)$ so that $\psi = C\ sin(kt)\ sin(Kt)$. This "modulation" of a small thing by a large thing

---

[g] The world-changing Fast Fourier Transform (FFT) is a masterpiece of mid-20th century matrix factorization, where *NlogN* computations give the answer to a formerly N^2 sized question, *N* being the number of points in a series resolving *N/2* frequencies.

might be quite physical and real, but Fourier analysis decomposes it into an *interference pattern* or *beating pattern* between two high wavenumbers near K, $\psi = C/2\,[cos(kt - Kt) - cos(kt + Kt)]$, with no long-wavelength energy involved at all. Which is the truer picture: the true scale *interaction* of "modulation", or the mere additive superposition of wave "interference"? AM and FM radio show us that both are useful descriptions, and their majestic equality means that mathematics is silent on which is really happening in the physical processes giving rise to the data. The lesson is that the categories of bookkeeping systems must not be mistaken for the realities they describe *even a complete and correct description may still be non-unique.*

3.5 Eddies, shear, and energy transfer across scale

3.5.1 *Downscale energy transfer: shear instability*

Whatever the interpretation of Fourier spectra, Parseval's theorem does give us an accounting framework for distributing energy over size or scale bins. Consider *shear instability* (the roll-up of sheets or filaments of vorticity into ball-like vortices). This process transfers conserved energy *downscale,* from larger to smaller scales, with no trace in the global kinetic energy budget's ultimate source term [*wb*] (Problem 2.4.1), at least for horizontal roll-ups.

3.5.2 *Upscale energy flux and convection-LS interaction*

Fluid eddies, unlike diffusive molecules, can carry momentum *up its mean gradient (up-shear)*, acting as a *negative viscosity* (Starr 1968). From a kinetic energy perspective, such eddies transfer energy from their own scale into the shear. This situation is common, since shear's effect on an eddy structure is to tilt it in such a way that the $\overline{(v'u')_y}$ or $\overline{(w'u')_z}$ terms in (3.2) then transfer momentum upshear. As long as there is some ongoing energy source for eddies (like buoyant convection), postulating this upscale flux process is not as exotic as it may seem at first blush. That local shear is always part of some larger-

.

scale flow feature, but it may be of a vastly larger scale (like a planetary-scale jet stream), or simply a swirl one octave or decade larger than the eddy. In other words, this up-shear momentum flux need not be local in the scale domain.

3.5.3 *Triads and tunneling*

Detailed bookkeeping of kinetic energy transfers by the advection terms (substituting Fourier forms into the governing equations and rearranging) reveals that energy transfers across scale may occur in *resonant triads* of wavenumber (e.g. Bretherton 1969), based on frequency sum-and-difference formulas like in the beating vs. modulation example examined above and in problem 3.9.4. These "tunnels" for energy flux across the scale domain, transferring energy among well-separated wavenumbers, are one more way in which the *cascade* presumption about of multiscale flow may be full of holes.

3.6 The cascade fallacy in spectral energetics

The classical theory of turbulence (Kolmogorov 1941) shows that *if* a turbulent flow is driven at large scale (for instance, stirred by a huge spoon), and *if* the only dissipation is by diffusion at the molecular scale, and *if* the energy transfer in the intervening "inertial subrange" of scales is local (with only eddies of comparable size exchanging energy), *then* by purely dimensional analysis a logarithmic graph of the kinetic energy power spectrum must have a slope of energy vs. wavenumber of -5/3. This paradigm is called a *cascade*[h] because of its local-in-scale energy transfer assumption.

A cascade is only one particular type of waterfall, but such paradigms are sticky in the brain, especially in the absence of specific alternatives. When a gross statistic of observations (the approximate -5/3 slope) agrees with a mechanistically hazy (purely dimensional) prediction

---

[h] A *cascading* waterfall is one with many short drops from one pool to the next.

arising from a sweeping assumption, what is the weight of that evidence for the assumption?

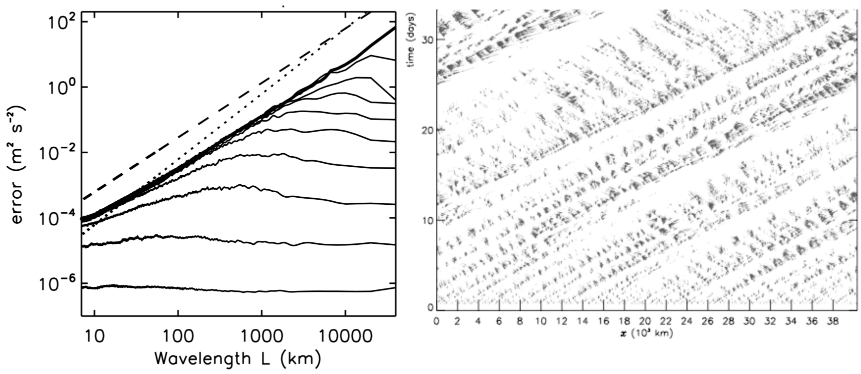

Fig. 3.1. Kinetic energy spectrum of a 2D (x-z plane) cloud model, with -5/3 reference slope (dashed) and -2 (dotted). The heavy line is the time-averaged spectrum, while lighter lines show the upward progression of "errors" (differences between two simulations after seeding the model with tiny uncorrelated random perturbations). Adapted from Fig. 5 of Mapes et al. (2008). Right panel: Surface precipitation time section for the first (spinup) month of the 40,000 km simulations analyzed at left.

Explicit simulations of deep moist convection, within the wavy stratified environment it creates and inhabits, exhibit a nearly -5/3 logarithmic slope to their energy spectrum at equilibrium with steady destabilization forcing. While global 2D wavenumber spectra on the sphere are hard to interpret, simulations with a more affordable 2D cloud model in the equator-like *x*-z plane fall extremely, compellingly close to that -5/3 slope across 5 decades of scale when multiscale structure is fully developed (Mapes et al. 2008; heavy line in Figure 3.1a).

However, the spectrally resolved kinetic energy budget of these simulations shows nothing like a simple cascade from a large forcing scale to a small dissipation scale. Instead, buoyant KE generation [*wb*] is important at every wavenumber. The kinetic energy generated by this gravity work is transferred from vertical to horizontal wind by pressure, and the horizontal wind energy is then lost to drag. The same is true in the spectral energy budget of a 3D Large Eddy Simulation (LES) of a cloud-topped boundary layer (Harm Jonker 2018, *pers. comm.*). The mechanisms of dissipation of kinetic energy differ somewhat between those two models, but the point is that source term *[wb]* it is broadband (active at many wavenumbers).

All scales convect is the overarching lesson of such spectral analyses of the [*bw*] term. This is true both for *initial energy growth* from small random fluctuations in a quiescent but unstable atmosphere, and for *difference-energy or difference variance growth* of initially small perturbations in predictability experiments on fully developed, steadily convecting flow. However, the low-wavenumber or long-wave energy (or squared difference magnitude in the predictability experiment) takes longer to *saturate* than the energy in high-wavenumber motions, because of the sloping ceiling of the -5/3 spectrum. For the initial growth problem, "saturation" means that the winds strengthen until dissipative processes match the buoyant generation. For difference growth, "saturation" means that predictability is lost: the difference energy has become as great as the background equilibrium spectrum of energy.

The word *growth* above meant the increase of energy *magnitude* at a given wavenumber; what Mapes et al. (2008) called *up-amplitude* growth. But there is also *up-scale* growth, meaning that *an increasing fraction of the energy is in the low-wavenumber end of the spectrum.* This up-scale energy progression is shown for difference growth in the upward-moving sequence of difference-energy spectra (thin lines) in Fig. 3.1.

The convenience of the cascade paradigm has led many into an under-appreciated cascade fallacy. Up-scale difference energy growth occurs in Fig. 3.1 not because energy flows through the smaller scales to reach the larger scales, but rather because the small scales simply saturate sooner, and at a lower energy level. Low-wavenumber motions (or differences thereof) can increase in magnitude for longer, and accumulate more energy (or difference energy), simply because their saturation 'ceiling' (heavy line in Fig. 3.1a) is higher. But why was that again? We have no adequate answer except the inadequate cascade paradigm of local-in-space energy transfer, as the foregoing explained in several ways.

For multi-scale convection in its stratified and thus wavy environment, the basic tenets of a cascade are simply inapplicable. Might there be other lines of reasoning, perhaps based on deeper unarticulated principles, behind compellingly clean (if rather abstract) spectral plots like figure 3.1 above, Fig. 4 of Wikle et al. (1999), and others? As lamented by Sun et al. (2017), "the authors have been unable to develop a simple explanation for why a −5/3 slope develops." The straightness in Figure 3.1 is surely no coincidence, but for all the formidable breadth and depth of modern turbulence theory (Alexakis and Biferale 2018, with the word *cascade* elevated to its title), and long-standing wave-spectrum saturation ideas in oceanography that might seem to be relevant (Munk 1980), we still lack a compelling paradigm or model for the multiscale *coupled* wave-convection problem (Bretherton 1969, Kiladis 2009).

3.7 Macro-description, entropy, information, and "systems"

The molecular vs. macroscale separation in section 3.1 expresses ideas from the field of *statistical thermodynamics*. Historically, that discipline matured only after the confirmation of atoms and molecules, greatly illuminating the older, more empirically grounded science of thermodynamics (e.g. Jaynes 1957). If temperature is sensibly translated into units expressing the energy of molecules (rather than arbitrary decimal "degrees" based on water's properties), then entropy is dimensionless, as a measure of pattern or probability properly should be. This dimensionless entropy is best translated as *missing information* (Ben-Naim 2008), in the Shannon (1948) sense of the word *information*, with units of *bits* that most 21st century reader will find intuitive. Specifically, S is a *functional* (an integrated, summarizing scalar) of the distribution of probabilities p_i that a set of molecules is in *microscopic state i*:

$$S = -\sum_i p_i log(p_i) \quad (3.3)$$

.

An example from Ben-Naim (2008) in exercise 0.8 makes this formula intuitive.

The words *missing information* indicate that entropy is a property of *our description* of nature -- that is, it is a conceptual tool, a tactic of scale separation between the "macro" scales we choose to represent explicitly and the "micro" scales we choose to describe only statistically, discarding information that is redundant or uninteresting, or simply assuming the least we can when information is unavailable. Might similar tools of description be borrowed for a statistical treatment (Part III) of large numbers of small[i] "entities" in a fluid (Part II), just as eddy viscosity of section 3.2 extended the concept of true molecular viscosity from section 3.1?

Crises of description can spotlight real scientific issues. For instance, Fourier analysis leads to a crisis at small scales. Even in one dimension, there are ten times more Fourier wavenumber bins in the 10s of km scale range than the 100s of km scale range, and so on to smaller scales. Each bin is a *degree of freedom* (*DOF*), a container for information (the drawing of distinctions, measurable in units of *bits*) as well as for energy as discussed above. The limitlessly growing multiplicity of DOFs with smallness of scale presented a philosophical crisis for sensible thinking about macroscopic phenomena in early 20th century physics. This *ultraviolet catastrophe* of infinite energy in the infinity of small scales was an early clue to the existence of irreducible *quanta* as a floor to the bottomless pit of microstructure. Classical philosophy's *atomism* was finally proven out.

Above quantum theory's lowest basement floor of physicalism, there remains a philosophical (or descriptive) crisis in multiscale (micro to macro) reasoning. A strict version of reductionism insists that the most fundamental or real causality must lie at the smallest scale, with all larger-scale entities being mere *epiphenomena* which are *supervenient* on the underlying microscale. This view leaves no role for macroscopic

---

[i] Much smaller than convecting flows, albeit much larger than molecules.

identity or agency, from convective cells to biological cells, neurons to behavior. While rather abstruse, supervenience subtly undermines causality discourse in all the macro-sciences, privileging the hardest of hard sciences (particle physics), which turns around and hands off its ultimate causality to quantum indeterminacy. There is no bottom, it is turtles[j] all the way down!

Of course, this is a crisis only of description: nature has no problem operating sensibly at macroscales. Fluid flow usefully challenges any and all attempts at defining tidy bounds of macroscale sense-making, with its scale-dependent time horizons of predictability and even knowability. Might our humble little corner of science offer a unique vantage point and testing ground for bulk accountings of microstructure in macroscopic weather? Exploring this possibility may require statistical accounting schemes in the middle ground between rigorous statistical thermodynamics with its practical infinity of profoundly indistinguishable molecules, and strained analogies in fields like biology where individuals have identites rooted in both life history and very long-term heritable structure.

Information theory might offer the necessary toolkit for this project of creating accounts of multiscale phenomena, with philosophy-of-science platforms that could span the supervenience abyss at levels usefully high above the molecular or quantum basement (e.g. Hoel 2018). Neuroscience seems to be at the leading edge of this exciting intellectual project, because its practitioners have all the ingredients: an interest in complex structure (biology), rooted in chemistry's utilitarian understanding of *entropy*, but also a functional interest (the teleological purpose of neurons) in processing of entropy's mathematical close cousin *information* (Ben-Naim 2008).

The core concept of statistical thermodynamics is the *system.* An *isolated system* exchanges no matter or energy across its *boundaries*, and the

[j] evidently a Hindu image, https://en.wikipedia.org/wiki/Turtles_all_the_way_down

.

Second Law says that its *state* tends toward *equilibrium*. A *closed* system is a fixed set of matter which exchanges energy with its surroundings. An *open* system exchanges both matter and energy. By these exchanges it may exist in a state that remains far from equilibrium, indefinitely. At first blush, the *nonequilibrium thermodynamics of open systems* may seem almost like a post-modern parody of science. Since the only thing that really defines such an open "system" is provisionally labeling it as one, this sounds like another supervenience crisis: the study of arbitrary human-assigned labels, not of anything fundamentally real in nature. Until we realize that our very bodies and brains are precisely such systems! Might we usefully draw on the theory of such open systems as a bulk description basis for convective "systems" in the atmosphere?

The power of systems descriptions often lies in *integral principles* those systems putatively obey. For instance, the *principle of least time* or of *least action* in physics allow the computation of wave refraction and other complex behaviors, sidestepping the need for detailed local mechanistic descriptions. In thermodynamics, the Second Law's mere inequality that entropy does not decrease with time ($dS/dt \geq 0$) can be given real teeth if it is strengthened it to a principle of *maximum entropy production rate.* Might such a principle, applied to a descriptive treatment of small-scale motions about which our "missing information" is very great, gain us some useful traction on the problem of convection?

The principle of maximum entropy production rate has been invoked as an approach to atmospheric dynamics, as a direct *extension* of statistical thermodynamics to convective cells (Asai and Kasahara 1971) or the planetary scale mean atmospheric circulation (Paltridge 1978). The word "extension" emphasizes that these theories maximize the production rate of *thermodynamic entropy* -- the missing information in the zillions of redundant *molecules* in the atmosphere. Despite the fascination of applying a principle across such vast scale separations (see review in Liu 2011, inspirational mention in Palmer 2019), this elegant approach has arguably not yet unlocked the compelling predictive or even explanatory power one might hope for.

Perhaps an *analogy* to statistical thermodynamics, rather than such a direct *extension*, could be more fruitful. Such an analogy might invoke entropic assumptions (such as maximum missing information; e.g. Jaynes 1957) about statistically redundant but not identical puffs of air, without adding to it the vast underlying abyss of molecular entropy. Borrowing an analogy rather than extending thermophysics theory would embody the view that such principles are really tools of our description, not expressions of fundamental physics obeyed by convecting flow (and every other system) supervening upon an ultimate "basement" causality of molecular interactions.

Biological systems reasoning is sometimes of this analogy type. For instance, *ecosystem succession* is a familiar and powerful concept in that field, and might seem to be a relevant analogue for atmospheric convection developing and organizing (or 'evolving') over land on a summer day. Both can be viewed as problems in which an available energy resource (an energy *flow*, ultimately sunshine) can be processed (assisted, but also lived off of) by open "systems" (entities) which compete in a trade space of energy inefficiency vs. structural complexity.

In this trade space, "organized" systems may be optimal for exploiting their specific niche, but are complicated (unlikely to form spontaneously) and perhaps highly contingent on other systems (trophic levels). These compete for existence against simpler, general-purpose structures that can spontaneously or more readily develop (i.e., more "cheaply" in terms of some required resource), but may be less efficient at the "job" of consuming or transmuting energy. It is doubtful that a whole ecosystem has a teleological "job"[k] that can be expressed as an optimum principle around some simple scalar like energy: the diversity of ecosystems shows that at the very least such an optimum must be wildly non-unique. But the "job" of convection in lowering the center of gravity of a bottom-heated body of fluid seems clearer, and more likely to be relatable to the key local force (buoyancy) that drives the process.

.

These ideas will be revisited in Part III, in an attempt to apply them to the larger-than-molecular convective flow entities of Part II. While no great breakthrough in cumulus parameterization emerges straightforwardly from the pure ideas and definitions of information theory, these ideas comprise a distinct facet of this chapter's topic of "accounting scales of motion", and point to incompletely explored and possibly promising avenues for progress or greater appreciation.

3.8 Problems and solutions

3.8.1 *Scale separation (large scale vs. eddy)*

Derive (3.2) from (3.1), noting the footnote there.

3.8.2 *Spectra of spatial data (your photograph)*

View the Jupyter notebook Spectra.ipynb from this Github page: (*URL*). Using Jupyter-Python (easily installed as explained at *Unidata's python page URL*), operate the notebook to replicate its figures. Replace the photograph with your own, and adjust code there to explore how Fourier analysis in 1 or 2 horizontal dimension works to decompose an energy-like quantity (*variance of brightness* in an image domain).

3.8.3 *Modulation vs. beating*

a. In a Jupyter notebook or other coding and exposition environment, illustrate the modulation vs. beating (interference) interpretations at the end of section 3.4.

b. Read about and explain in your own words the difference between AM and FM radio.

3.8.4 *Multiscale solutions to fluid equations*

Show and explain the truth of this statement from Palmer (2019):

"Although the Navier–Stokes equations cannot be solved directly, they have certain symmetry properties ... One of these is a scaling symmetry: if u(x,t) is the velocity field and p(x,t) is the pressure field associated

$$u_\tau(x,t) = \tau^{-\frac{1}{2}} u\left(\frac{x}{\tau^{1/2}}, \frac{t}{\tau}\right) \qquad (1)$$

and

$$p_\tau(x,t) = \tau^{-1} p\left(\frac{x}{\tau^{1/2}}, \frac{t}{\tau}\right) \qquad (2)$$

with a solution to the Navier–Stokes equations, then so are:
For another approach, see section 2.2 of Lovejoy and Schertzer (2014).

3.8.5 *Missing information, measured in bits*

Shannon (1948) information, the negative of the dimensionless entropy, has units of *bits*. This unit should feel familiar or even visceral in the digital age, where most of us pay money for data plans. To illustrate it, consider the Missing Information denoted by H in Ben-Naim (2008), a book which also offers many additional exercises and deeper discussion. H is a *functional*, a sum or integral over an underlying distribution, in this case a probability distribution *P(situation)*. Consider this example from Ben-Naim (2013) section 3.1:

Situation: One stone is in one of 8 identical boxes arranged in a line.

Question: How many bits of information are missing from this incomplete description of the situation?

.

a. Compute the missing information H from the formula
$H = -\sum_i p_i \log_2 p_i$ for the 8 equal probabilities. The unit is bits, because the log is base 2.

b. Explain an optimum strategy to learn which box contains the stone from someone who knows, by asking the smallest possible number of yes-no questions (so that each answer is 1 bit of information). Can you do it in any less than the number of bits from a? Does the inefficiency of other possible but nonoptimal strategies change what your intuition would properly call the "missing information" (MI) in the description of the *situation*? Should it? Is the term MI well chosen?