Weather vs. Climate

DEEPS, Brown University, Providence, RI, USA

ABSTRACT

This is a tale involving partial differential equations, chaos, predictability, ergodicity, models, and statistics. We will be discussing for about 2 weeks. There are homework problems at the end which we will present in class when the lectures are complete.

1. Introduction

We hear a lot in the news about climate change, weather events, and weather or climate disasters. What is only generally covered is exactly what we mean by weather and climate and how this conception came to be accepted.

Defining what constitutes the climate is a matter of law and policy. According to the UN FCCC, "The Paris Agreement is a legally binding international treaty on climate change. It was adopted by 196 Parties at the UN Climate Change Conference (COP21) in Paris, France, on 12 December 2015. It entered into force on 4 November 2016. Its overarching goal is to hold 'the increase in the global average temperature to well below 2 °C above pre-industrial levels' and pursue efforts 'to limit the temperature increase to 1.5 °C above pre-industrial levels.' " This quotation suggests that the global average temperature is the measure of the climate, or at least the measure consistently enough defined to be counted. By global average temperature, they mean the area-weighted surface temperature (GMST) which is easy to observe from satellites or sometimes the air temperature at 2 m above the surface (GSAT), which is a little easier to calculate in models.

As you may have noticed this past July, 2023–the warmest July recorded thus far–it is tricky to define precisely what we mean. The average global surface temperature in July was 1.12°C above average, ranking it as the warmest July in NOAA's 174-year record. Earth's average land and ocean surface temperature in 2022 was 0.86 °C above the 20th-century average of 13.9 °C — the sixth highest among all years in the 1880-2022 record. The IPCC (2021a) reported that " Global surface temperature was 1.09 [0.95 to 1.20] °C higher in 2011–2020 than 1850–1900, with larger increases over land (1.59 [1.34 to 1.83] °C) than over the ocean (0.88 [0.68 to 1.01] °C)." In the IPCC version, brackets are used to denote a range of estimates (90% confidence window, or the *very likely* range).

These statements reveal many things, including:

1. It matters what background temperature you are comparing to. It is not clear from these definitions alone how they relate to "pre-industrial" conditions per the Paris Agreement.

2. It matters whether you are using a month's temperature, or a decade, or longer.

3. There is a difference in magnitude between the warming of the land and the warming of the ocean.

4. None of these mention the warmest day or warmest temperature ever recorded.

In fact, there is some dispute about what temperature captures the warmest recorded temperature (`https://en.wikipedia.org/wiki/Highest_temperature_recorded_on_Earth`) with the present World Meteorological Organization-certified one being back in 1913. Wouldn't we think that climate change would have made recent temperatures warmer? Apparently, weather has something to do with these more transient measurements, while the longer decadal averages are taken to be indicative of the climate state.

2. Climate is what you expect, Weather is what you get

The title of this section is a quotation attributed to Mark Twain. It has the sense of a statistical claim, where we are used to distinguishing between measured values and expectation values (i.e., the average of all possible outcomes weighted by their likelihood).

a. The Mean, Histogram, and Expectation

There is another way to calculate the mean of a random variable which is closely related to the mean over a volume or area. Suppose we have a (not normalized) histogram $h(x)$ over a set of n different values of x over N different experimental results. We could sum up all of the columns of the histogram like this

$$N = \sum_{j=1}^{n} h(x_j). \tag{1}$$

to arrive at the total number of experiments. Or, we could form a normalized histogram like this

$$1 = \sum_{j=1}^{n} \frac{h(x_j)}{N}. \tag{2}$$

Note that here j runs over all of the possible outcomes of x, hitting each only once. Since each value of the histogram is associated with one particular value of the variable x_i, we can compare this to the ordinary average, which runs over all N measurements,

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^{N} x_i. \tag{3}$$

Suppose we rearranged the order of the sum in the average, so that we first summed all of the occurrences that equal the first possible value x_1, then the occurrences of the second value x_2, etc. The number of terms of each type in the sum would just be the *histogram*! Thus,

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^{N} x_i = \frac{1}{N} \sum_{j=1}^{n} x_j h(x_j). \tag{4}$$

The first expression shows the expected value is just the mean over all of the experiments, and the final one shows that this is equal to the likelihood-weighted expectation value from the histogram summed over all of its bins. We can think of a few ways of expressing the same idea,

$$\langle x \rangle = \frac{1}{N} \sum_{j=1}^{n} x_j h(x_j) = \sum_{j=1}^{n} \frac{x_j h(x_j)}{N} = \frac{\sum_{j=1}^{n} x_j h(x_j)}{\sum_{j=1}^{n} h(x_j)}. \tag{5}$$

Each of these is exactly equal to the sample average over the experiments included in the histogram.

Taking this idea to the limit of infinite experiments, where our histogram approximates a probability distribution, we see that we can write the expected value as

$$\langle x \rangle = \sum_{j=1}^{n} x_j p(x_j). \tag{6}$$

For a probability density function of a continuous variable x, the equivalent form is

$$\langle x \rangle = \int_{-\infty}^{\infty} x \rho(x) \mathrm{d}x. \tag{7}$$

Notice that the normalization of the probabilities and probability density means that we do not need the denominator if we sum or integrate over all possible values.

b. Climate as an expectation

Metrics, or measures, of climate describe the statistics of atmospheric and oceanic variables at a given location and time of year. Weather on the other hand is the specific outcome one gets at a given time and location.

Which statistics make up the climate? Well, expectation value for one, but also variation about the expected value, which could be described by "variability" or "extremes" or other words capturing not just the center of the probability distributions of atmospheric and oceanic variables, but also their tails. One way to measure these tails that's just an extension of the expected value is through the *moments* of probability distributions.

The term *moment* is used to describe an operation like the average above over a distribution or density function. Thus, the first moment of the probability density function is the average:

$$\mu = \langle x \rangle = \int_{-\infty}^{\infty} x \rho(x) x. \tag{8}$$

The second moment is

$$\langle x^2 \rangle = \int_{-\infty}^{\infty} x^2 \rho(x) x. \tag{9}$$

The m^{th} moment is

$$\langle x^m \rangle = \int_{-\infty}^{\infty} x^m \rho(x) x. \tag{10}$$

This name comes from the close relationship between this form and moments in physics, such as the moment of inertia.

The moments are sometimes centralized and normalized. The central moments are the moments about the mean. The m^{th} centralized moment is

$$\mu_m = \langle (x - \langle x \rangle)^m \rangle = \int_{-\infty}^{\infty} (x - \langle x \rangle)^m \rho(x) x. \tag{11}$$

The moments can be normalized using the standard deviation $\sigma_x = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}$. The m^{th} normalized moment is

$$\frac{\langle x^m \rangle}{\sigma_x^m} = \frac{\langle x^m \rangle}{\left(\sqrt{\langle x^2 \rangle - \langle x \rangle^2} \right)^m}. \tag{12}$$

The standardized moment, which is just the normalized, centralized moment is

$$\frac{\mu_m}{\sigma_x^m} = \frac{\langle (x - \langle x \rangle)^m \rangle}{\left(\sqrt{\langle x^2 \rangle - \langle x \rangle^2} \right)^m}. \tag{13}$$

In this language, the first moment is the mean, the second centralized moment is the variance, the third centralized, normalized moment is the skewness, and the fourth centralized, normalized moment is the kurtosis.

Climate certainly contains expectations about the mean and variance of environmental variables, but sometimes also the higher moments as well. Generally, the higher you go the more data you need to measure moments and the better your model needs to be to predict them.

c. Direct statistical simulation

There are ways to directly simulate the evolving probability distribution of the climate (e.g., Li et al. 2021; Allawala et al. 2020) or the evolving expectation value and variance (e.g., Nicklas et al. 2023). However, it is far more common to use a *deterministic* model to simulate an example climate *trajectory*. Often an ensemble of trajectories is considered together, each with different initial conditions but with the same climate forcing. This ensemble is thought to map out the probability space of the range of possible trajectories.

One other important limitation of simulations should be mentioned: limited resolution. The Earth is very big, and we know that in order to directly numerically simulate fluid motion we have to resolve down to the scales where dissipation and viscosity dominate. On Earth, the dissipation scale for the atmosphere and oceans are tiny–just cubic centimeter or millimeter scales–so we'd need many, many gridpoints to simulate at this scale. The volume of the atmosphere is $4.2\times10^{18}\,\mathrm{m}^3$ and of the ocean is $1.4\times10^{18}\,\mathrm{m}^3$. Assuming we need $1\,\mathrm{mm}^3$ grid cells to reach direct numerical simulation, this means our models would have 4.2×10^{36} and 1.4×10^{36} grid cells in the atmosphere and ocean. Present computing power barely reaches kilometer-scale modeling (Hewitt et al. 2022), which have about 4×10^{16} times fewer gridpoints that global direct numerical simulation would require. Even with Moore's law doubling computing power every 18 months, it will still be another 105 years before we'll be able to carry out these simulations.

Thus, global models tend to be coarser than the fluid dynamics of the real world requires. This means that some of the variability in the real world on small space and time scales is just missing from these simulations. Does that mean that we are only simulating the large-scale "climate" and not the small-scale "weather"? Not quite, because even our weather models are limited in resolution. We also parameterize some of the key unresolved effects on resolved scales, because there is no guarantee that just resolving the largest scales does a good job of capturing the key processes important for climate and weather. Sometimes, *stochastic* variability, or random noise, is added to the equations to resemble the variability that would have been stimulated by the unresolved scales. In these stochastic models, it is expected that one would run an ensemble of trajectories with different noise each time to find the average result across these simulations.

3. Boundary vs. Initial: The heat equation

We are beginning to gather a statistical sense of the meaning of weather vs. climate, but there is also a mathematical/numerical sense of the distinction based on the kinds of equations that are solved by climate and weather models. Both the atmosphere and the ocean are fluids, and they also carry climatically important tracers such as thermal energy (equivalent to measures of temperature), humidity, salt, and trace gasses such as carbon dioxide or the many forms of carbon in the ocean. All of these processes are simulated using partial differential equations, specifically those of fluid mechanics. Other important climate changes, in sea ice, ice sheets, glaciers, seasonal snow and ice on land, and the hydrological cycle are also simulated in climate models or in conjunction with climate models in offline calculations; these phenomena are also simulated using partial differential equations, but not of the fluid equations, instead solid mechanics, percolation theory, and other hybrids of different variants of continuum mechanics are used. Finally, the changes to biology and the land surface, such as forests changing to grasslands, fire, and sometimes ocean biology, are simulated with different approaches, such as agent-based models, lookup tables, data-driven approaches, and other ways to generate outcomes from the changing weather and climate conditions that may feed back onto the energy cycle through albedo, the water cycle through transpiration, or the carbon cycle through growth for example. The combined equation set including all of these different aspects is far beyond our consideration here, but we can examine one equation that is suitable for illustrative purposes.

The thermal equation in the ocean is

$$\partial_t\Theta + \mathbf{v}\cdot\nabla\Theta = \nabla\cdot\kappa\cdot\nabla\Theta \tag{14}$$

where Θ is a measure of the temperature and $\mathbf{v}$ is the velocity of the seawater (including a boost from unresolved motions similar to Stokes drift). The diffusivity κ is an anisotropic tensor that varies in space and time with the flow to parameterize the kinds of turbulent motion that are expected. The right side has the divergence of $\kappa\cdot\nabla\Theta$, because that is just the divergence of the diffusive heat flux, which will heat or cool at the location of its divergence. Thus both the advective term (the one with the velocity) and the diffusive term (the one with the κ) are nonlinear, as they depend on nonlinear products of the predicted variables $(\mathbf{v},\kappa,\Theta)$.

a. Solving the heat equation

A related equation is the heat equation, which we can solve analytically and understand a bit better. Suppose we neglect the advective term and just take κ to be a constant, then

$$\partial_t \Theta = \kappa \nabla^2 \Theta \tag{15}$$

Let's make the added simplification that we'll neglect the spherical nature of the Earth and just consider depth (z) and meridional location (y). Finally, notice that any constant added to Θ doesn't change the equation, so we can think of Θ as just a temperature anomaly from a background temperature.

One way to attack this problem is by separation of variables.

$$\Theta = \sum_k \phi_k(y,z) T_k(t), \tag{16}$$

$$\kappa T_k(t) \nabla^2 \phi_k(y,z) = \phi_k(y,z) \partial_t T_k(t), \tag{17}$$

$$\frac{\kappa \nabla^2 \phi_k(y,z)}{\phi_k(y,z)} = -\kappa k^2 = \frac{\partial_t T_k(t)}{T_k(t)}. \tag{18}$$

The last step introduces a new constant, k, chosen to have convenient units by multiplication with κ. We know this must be a constant because the left side of the equation depends on space only and the right side of the equation depends only on time. We can exploit this to split apart the equation into two separate ones,

$$\nabla^2 \phi_k(y,z) + k^2 \phi_k(y,z) = 0 \tag{19}$$

$$\text{and } \partial_t T_k(t) + \kappa k^2 T_k(t) = 0. \tag{20}$$

So, we have two equations left for any chosen constant k. Since we picked this constant somewhat randomly, there are likely to be many of them, and there will be a different solution for each of them (hence the subscript k on each of the functions). In the end, we'll sum up all of these to arrive at a complete solution.

The first of the separated equations is just the Helmholtz equation for ϕ, which we can solve by a second round of separation of variables (left for homework). The other is a first-order, homogeneous, constant-coefficient, linear differential equation for the time variation, which is just an exponential.

In both cases, we can assume a solution of the form

$$\Theta(y,z,t) = \sum_{m,n,\sigma} T_{mn\sigma} = \sum_{m,n,\sigma} A_{mn\sigma} e^{my} e^{nz} e^{\sigma t}, \tag{21}$$

$$\kappa(m^2 T_{mn\sigma} + n^2 T_{mn\sigma}) = \sigma T_{mn\sigma}, \tag{22}$$

$$\therefore \kappa(m^2 + n^2) = -\kappa k^2 = \sigma. \tag{23}$$

We now see that we chose the sign of k^2 for two reasons. We could have had either exponential growth $(\sigma > 0)$ or exponential decay $(\sigma < 0)$ in time, and we are more interested in the latter. Also, it means that $m^2 + n^2 = -k^2$, so at least one of m, n will be imaginary (that is, the exponential associated with it will become sinusoidal).

It's clear that initial conditions can be met by setting $t = 0$ and then evaluating $A_{mn\sigma}$ over different m, n combinations to match their spatial pattern. Similarly, boundary conditions can be met by matching simple boundary conditions such as "no anomaly along the boundaries" ($\Theta = 0$ on the boundaries) or "no heat transfer through the boundaries", which is $\mathbf{n} \cdot \kappa \nabla \Theta = 0$ with $\mathbf{n}$ as an outward normal vector.

b. Heat equation behavior

So, now that we have a solution form, let's consider some applications that are climate and weather relevant. Suppose we start out with a temperature anomaly within the ocean. For example, if we take increasingly negative z to be the bottom of the ocean and $y = \pm L$ to be its meridional limits, then we might have a center-intensified blob of temperature anomaly like

$$\Theta(y,z,0) = -A \sin(\pi z/H) \sin(\pi y/L) \tag{24}$$

We can see that this anomaly already might be consistent with no-anomaly, Dirichlet boundary conditions ($\Theta = 0$) on the bottom ($z = 0, z = -H$) and coasts ($y = \pm L$). We can also see that if we choose $m^2 = -\pi^2/L^2$ and $n^2 = -1/H^2$ it will be easy to fit the general solution form (21). What about its temporal behavior?

$$n = \pm i\pi/H, \tag{25}$$

$$m = \pm i\pi/L \tag{26}$$

$$\kappa(m^2 + n^2) = -\frac{\kappa\pi^2}{H^2} - \frac{\kappa\pi^2}{L^2} = \sigma. \tag{27}$$

We see that this anomaly will decay in time with a timescale that depends on the lengthscales of the initial disturbance and the rate of diffusion. The smaller the basin ($L, H \longrightarrow 0$) or the faster the diffusion ($\kappa \longrightarrow \infty$), the quicker the decay.

Without changing the size of the basin, if the initial disturbance had been smaller, then the decay would also have been faster.

$$\Theta(y,z,0) = -A \sin(2\pi y/L) \sin(2\pi z/H), \tag{28}$$

$$n = \pm 2i\pi/H, \tag{29}$$

$$m = \pm 2i\pi/L \tag{30}$$

$$\kappa(m^2 + n^2) = -\frac{4\kappa\pi^2}{H^2} - \frac{4\kappa\pi^2}{L^2} = \sigma. \tag{31}$$

How do the boundary conditions come into play? As mentioned, the anomaly has no anomaly ($\Theta = 0$) at $z = 0, -H$, and $y = \pm L$. We could instead have chosen a specified flux condition, ($\mathbf{n} \cdot \kappa \nabla \Theta = \mathcal{F}(y,z,t)$), such as no flux through the basin walls and a global warming flux through the surface. In this case we'd expect the average temperature to rise, probably in a surface-intensified way. However, this approach uses mixed boundary conditions (Neumann and Dirichlet).

A simpler way to illustrate this effect is to chose a different surface boundary condition. We would then have a *boundary* source of anomaly. In order to match the surface boundary condition, many of the vertical modes are needed to work together to form a jump initially, which can be superimposed on the previous initial condition. At the beginning the added terms would cancel out everywhere except right at the wall–thus satisfying the initial conditions (with $\Theta = 0$ on all four boundaries) and the nonzero surface boundary condition together. Over time, first the smaller scales then the larger scales would come to reflect the steady boundary condition instead of the decaying anomaly from the initial conditions–a heat anomaly would spread from the top into the domain. Thus, over time, the impact of the initial conditions will decay and the impact of the boundary conditions would grow and fill the domain. Eventually, after the decay of the smallest σ (the one associated with the largest-scale sinusoids in the sum, the same one to do with the initial conditions in (24)), the initial conditions would be forgotten and the boundary would be the whole of the solution. Note that we could use a variety of different choices for A in (24), and the final answer would not vary–the initial conditions are forgotten while the surface boundary anomaly would remain. Thus, the initial conditions in this system have a finite duration impact on the solution.

4. Boundary vs. Initial: Predictability

In realistic models, a similar takeover of initial conditions by boundary conditions as in the heat equation tends to occur. Consider the following example from Sane et al. (2021): a coastal ocean model of Narragansett Bay and nearby waterways driven by surface forcing (winds, precipitation, heating and cooling), river freshwater, and offshore forcing of oceanic temperature, salinity, velocities, and tides.

The model is initialized simply, with uniform temperature and salinity and zero velocity. Over time, the model approaches the observations (Fig. 1). This effect indicates not only that the model works fairly well eventually, but it does so as the initial conditions are forgotten. Just as in the heat equation, initial conditions become increasingly unimportant after some time and the boundary forcing (constructed to be as close as possible to observed boundary conditions) takes over to make the model temperature converge on the observed temperature.

Are the initial condition effects guaranteed to eventually decay while the boundary conditions take over in any dynamical system? Not necessarily. Consider if the initial conditions produced a wave that just sloshed back and forth over and over–in some non-dissipative models such a wave could persist forever so the initial conditions would never be forgotten. Forgetting the initial conditions is a feature of the way weather, climate, and ocean models behave, and

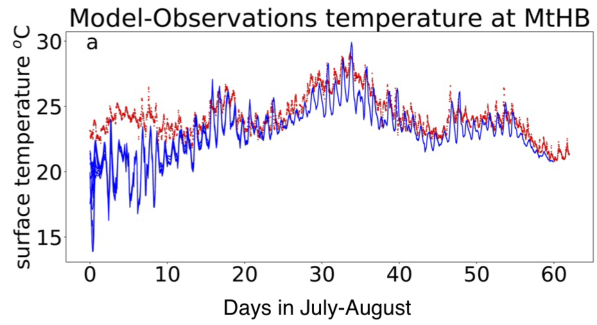

FIG. 1. A comparison between modeled temperature (blue) and observed temperature (red) at one location in Mt. Hope Bay. Figure from Sane et al. (2021).

evidence suggests that this reflects a real behavior of the earth system due to the chaotic behavior.

The timescale over which the initial conditions are forgotten is called the *predictability* timescale. Fig. 2 shows a schematic of how this timescale can be measured. An

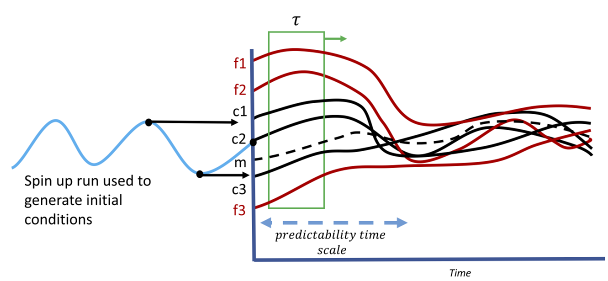

FIG. 2. A schematic of how a spin-up run (blue) is used to produce an ensemble of realistic initial conditions drawn from states visited at different times (black) to map the climatology and its uncertainty. The red lines indicate an initially amplified anomaly which convergences into the climatological ensemble over the predictability timescale. Figure from Sane et al. (2021).

initial run is carried out to provide realistic states for the model over a given time of year, and an ensemble of simulations each with initial conditions drawn from a different day of the initial run are used to map out the *climatology*. For reasons we'll explain below to do with chaos, the climatology is not taken to be a single prediction, but instead taken as an ensemble of predictions that stay close to one another but not exactly equal.

To measure the timescale over which initial conditions are forgotten, a second ensemble is created with larger initial anomalies than the climatology ensemble's initial spread. The upper panel of Fig. 3 shows how the perturbation ensemble temperature converges to within the climatology ensemble spread.

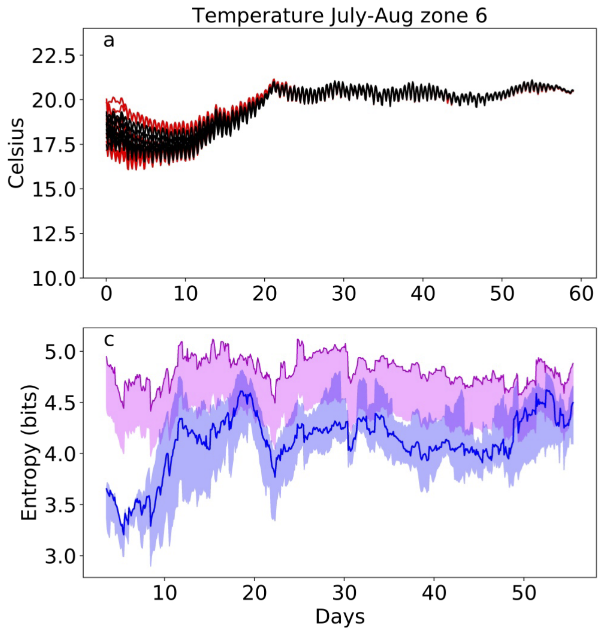

FIG. 3. ((Upper) An example of how the perturbation ensemble (red) converges onto the climatology ensemble (black) over time. (Lower) the Shannon information entropy and mutual information between these ensembles. Figure from Sane et al. (2021).

a. Information theory

However, an even clearer picture can be seen if instead of raw temperatures, the similarities and distinctions in the temperature trends across the ensemble are highlighted. Sane et al. (2021) propose using metrics from information theory to zoom in on the ensemble behaviors. The Shannon (1948) entropy is defined for a random variable (X) with a histogram providing estimates of probability ($p(x_i)$) of that variable being in bin x_i for i from 1 to M as

$$H(X) = -\sum_{i=1}^{M} p(x_i)\log_2 p(x_i). \qquad (32)$$

Shannon famously named this "entropy" after a conversation with Von Neumann who said, "No one really understands entropy. Therefore, if you know what you mean by it and you use it when you are in an argument, you will win every time." Surely it was also not lost on Shannon that the Boltzmann expression for thermodynamic entropy of a physical system that can visit many microstates has the same formula as (32) multiplied by the Boltzmann constant k_b.

Here we need the entropy because it measures how many states are possible and what their probability is. The units of measure of Shannon entropy are bits–i.e., the number of binary numbers needed to count, or quantify, all of the states visited by the system. If a system varies a lot, it needs more bits to capture that variability, if it stays constant, then only one bin in the histogram is filled and the entropy is zero bits.

Related to entropy is the mutual information of two random variables (X, y),

$$I(X) = -\sum_{i=1}^{M}\sum_{i=1}^{N} p(x_i, y_j)\log_2 \frac{p(x_i, y_j)}{p(x_i)p(y_j)}. \qquad (33)$$

where $p(x_i, y_j)$ is the joint probability of both X being in bin x_i and Y being in bin y_j. If these variables are unconnected, then they are independent and their joint probability is just the product $p(x_i)p(y_j)$. If $X = Y$, then $p(x_i, y_j) = p(y_j) = p(x_i)$ and the mutual information equals the Shannon entropy (32). In the heat equation discussion of the last session, as the initial conditions are forgotten then the solution approaches the boundary-condition dominated solution. In the system considered by Sane et al. (2021), the intrinsic variability of the system prevents an asymptotic convergence but reaches a statistical convergence, so the mutual information stays somewhere between zero and the Shannon entropy.

Thus, the lower panel of Fig. 3 shows the Shannon entropy range for all of the climatology ensemble members (pink) versus the ensemble of mutual information values when taken between the perturbation ensemble members and the climatology ensemble members. When these two distributions overlap, one can no longer distinguish between the two ensembles and convergence has occurred. In this specific case, this happens just after 10 days, the predictability timescale, for the first time and holds for much of the longer timescales as well.

b. Forecasts, predictions, and projections

Now that we have the idea of ensembles of predictions with different initial conditions and the predictability timescale, we can contrast predictions of the weather from projections of the climate. The glossary of the IPCC (2021b) contains the following definitions:

Climate prediction: A climate prediction or climate forecast is the result of an attempt to produce (starting from a particular state of the climate system) an estimate of the actual evolution of the climate in the future, for example, at seasonal, interannual or decadal time scales. Because the future evolution of the climate system may be highly sensitive to initial conditions, has chaotic elements and is subject to natural variability, such predictions are usually probabilistic in nature.

Climate projection: Simulated response of the climate system to a scenario of future emissions or concentrations of greenhouse gases (GHGs) and aerosols and changes in land use, generally derived using climate models. Climate projections are distinguished from climate predictions by their dependence on the emission/concentration/radiative forcing scenario used, which is in turn based on assumptions concerning, for example, future socio-economic and technological developments that may or may not be realized.

Based on our previous section, we can then categorize predictions as forecasts that are sensitive to initial conditions, whereas projections provide a range of outcomes under the effects of different forcing (i.e., the changes that are under the control of boundary conditions). Often, when we think about "predictions" or "forecasts" we really mean of the weather. Weather models tend to have a predictability timescale of less than two weeks. Beyond this window, we just have projections of the climate–which might be estimated by the span of an ensemble of many weather outcomes describing the spread after the predictability timescale. The fact that a projection of the possible outcomes doesn't just continue to spread out wider and wider means that there is something else to describe beyond the weather. The projection describing the bounds, mean, or other statistics of multiple predictions beyond the predictability timescale is a projection of the climate.

Things get a bit confusing because for some variables in some models, the predictability timescale is so long that it reaches levels we normally think of as governed by *climate* rather than weather. Often the long predictability stems from the slower but still predictable evolution of the oceans, biology, or cryosphere in comparison to the faster evolving atmosphere which becomes chaotic in only a few weeks. Climate variability (e.g., El Nino) and climate predictions blur these lines, as phenomena such as El Nino may be predictable as much as 18 months ahead, even though the atmospheric weather associated with them is not. So, the terminology of projection and prediction helps further distinguish.

c. Sources of uncertainty in climate projections

Hawkins and Sutton (2009) describe a method to categorize the three different types of uncertainty in climate projections: internal variability, scenario uncertainty, and model error. Internal variability is just the intrinsic error due to different initial conditions leading to different, but neighboring, outcomes. Scenario uncertainty stems from the fact that we're not sure what humans will do in the future, so the boundary condition contributions which represent the effects of human emissions on the energy budget of Earth are uncertain. Finally, model error reflects the fact that models' limited resolution and imperfect parameterizations of unresolved phenomena introduce uncertainty. We cannot directly estimate the error in any one given model, but we can analyze models developed by different groups making different choices. Thus, the differences between the models samples across the spread of these choices and quantifies our ignorance in building models.

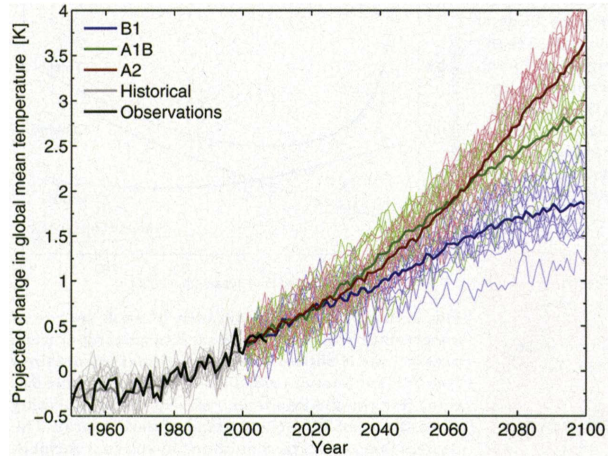

FIG. 4. (Timeseries of the CMIP3 ensemble climate predictions (thin lines). Different scenarios are in different colors, and observations are in black. Historical simulations where there is no scenario uncertainty are in grey. Figure from Hawkins and Sutton (2009).

Fig. 4 shows examples of the timeseries that go into the error decomposition method. These Climate Model Intercomparision version 3 (CMIP3) models were run at modeling centers around the world and collected together by Hawkins and Sutton (2009). Internal variability is measured by the temperature anomalies of each model from a smooth fit to itself. Scenario uncertainty is measured by the variance between the multi-model means for each scenario. Model uncertainty is measured for each scenario by the variance among the fits to each model for each scenario. The total is just the mean temperature change of all predictions. All temperatures are measured relative to the 1971-2000 mean of that model, and the means and variances were calculated with weighted averages based on model performance during the historical period.

There are a few things to note about the estimates in Fig. 5. For these decadal averages, the internal variability is lower than the combined model plus scenario uncertainty for the global temperature, but not so for the regional temperature for short timescale projections. That is, the initial conditions matter a lot for regional temperature projections soon after they are initialized. As time goes on, the model uncertainty shrinks as the climate takes over from the weather and these climate change signal gets large (i.e., the models agree more for larger change). Finally, the impacts of scenario uncertainty increase through time, which reflects both that the emission scenarios differ more at later times (Fig. 4) and the fact that cumulative emissions control the temperature change, not instantaneous emissions.

5. Chaos and ergodicity

One remaining question has to do with how different the initial conditions need to be for the models in an ensemble

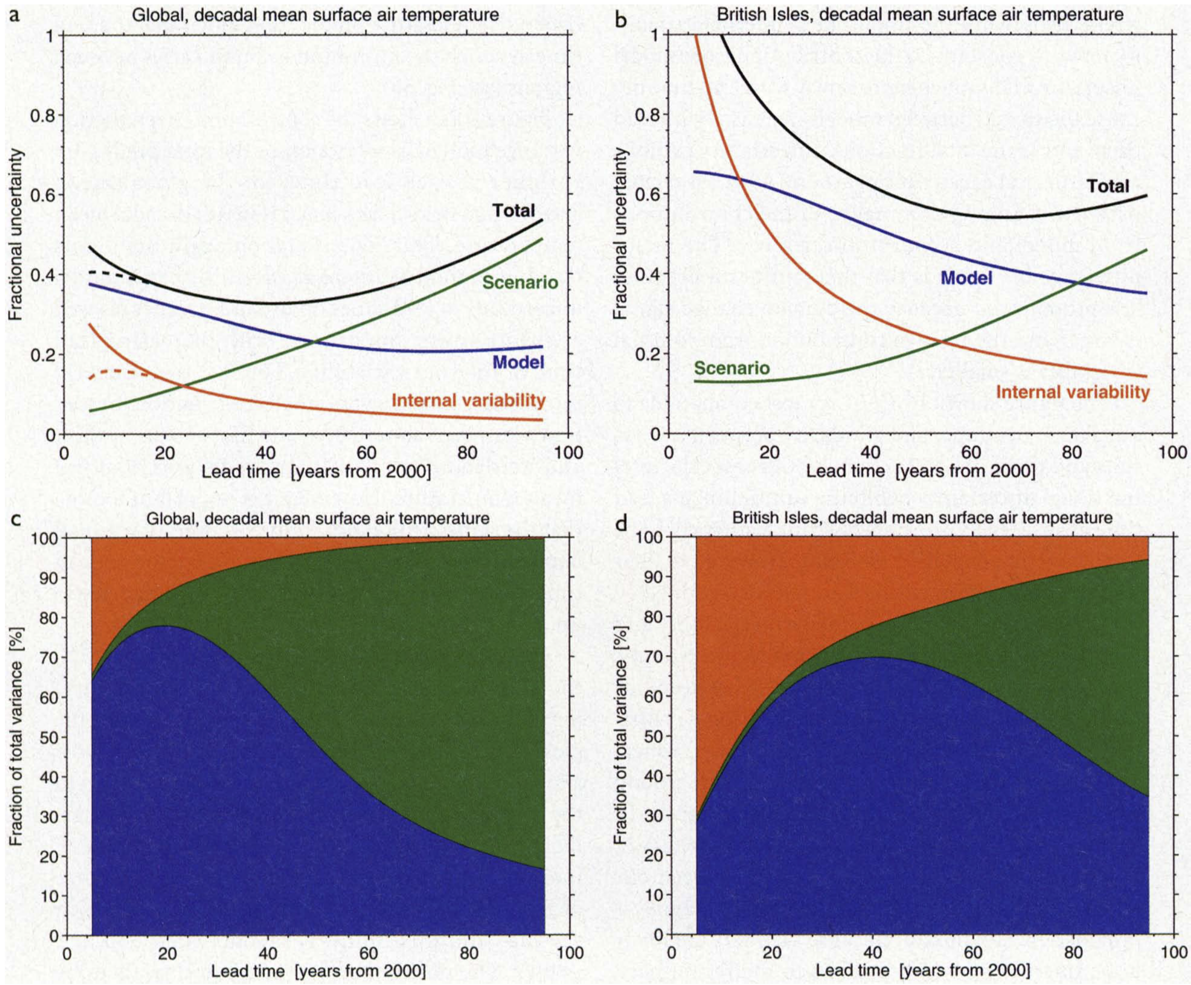

FIG. 5. (Decompositions of the CMIP3 ensemble climate projection uncertainties for global (left) and regional (right; British Isles only) decadal temperature averages. Upper panels show uncertainty and lower panels show ratios of uncertainty sources (labelled colors) with the total (black). Dashed lines illustrate the effects of optimizing ocean initial conditions. Figure from Hawkins and Sutton (2009).

to differ. This has to do with the chaotic nature of fluids and the climate.

Since it was used by Lorenz (1963) to describe the results of simulations carried out by Ellen Fetter, chaos has become a central theme in weather and climate science. Chaotic behavior is not just disordered behavior, it is something quite a bit more specific.

In a deterministic dynamical system, everything is known and in principle can be calculated to arbitrary accuracy. Deterministic climate and weather models used to make predictions are good examples: you get the same answer over and over so long as you start with the bit-by-bit same initial conditions and use the exact same boundary conditions. However, in the real world, the initial and boundary conditions are subject to some measurement uncertainty; instruments are imperfect and not everywhere is measured at the same time.

In nonchaotic dynamical systems, the closer your initial conditions are, the closer your trajectories remain. Consider a forecast function F that predicts the location at a future time for the distance between two initial locations.

$$F(\mathbf{x}_{0,1}, \mathbf{x}_{0,2}, t) = |\mathbf{x}_1(t) - \mathbf{x}_2(t)|. \tag{34}$$

We can consider starting the initial locations arbitrarily oriented but close to one another,

$$F(\mathbf{x}_0, \mathbf{x}_0 + \delta\hat{\mathbf{x}}, t) = |\mathbf{x}_1(t) - \mathbf{x}_2(t)|, \tag{35}$$

where δ is a small scalar and $\hat{\mathbf{x}}$ is a unit vector in any arbitrary direction. Putting in this way, we can make this system resemble the tools from analysis, given a tolerance ϵ

and a time t, for what δ can we guarantee that the following is true?

$$F(\mathbf{x}_0, \mathbf{x}_0 + \delta\hat{\mathbf{x}}, t) = |\mathbf{x}_1(t) - \mathbf{x}_2(t)| < \epsilon. \quad (36)$$

If we can prove this is true, then

$$\lim_{\delta\to 0} F(\mathbf{x}_0, \mathbf{x}_0 + \delta\hat{\mathbf{x}}, t) = 0. \quad (37)$$

This kind of behavior is typically assumed for continuum mechanics!

In chaotic systems, nearby initial conditions tend to diverge from one another quickly. In Figs. 6-7, we see the complex behavior in time and in a projection of phase space of the Lorenz (1963) system. Over time, trajectories loop back and forth over the "butterfly" shape of the phase space. Sometimes the trajectory loops repeatedly on one side and sometimes it crosses between the two sides. It is difficult to predict which will occur. Indeed, neighboring initial conditions soon make different choices at the branch and end up far apart from one another. However, the trajectory never flies away to infinity, it stays near an *attractor* which is s a structure in phase space. Our climate model similarly was chaotic, but bounded in that its solutions stay within the climatological range.

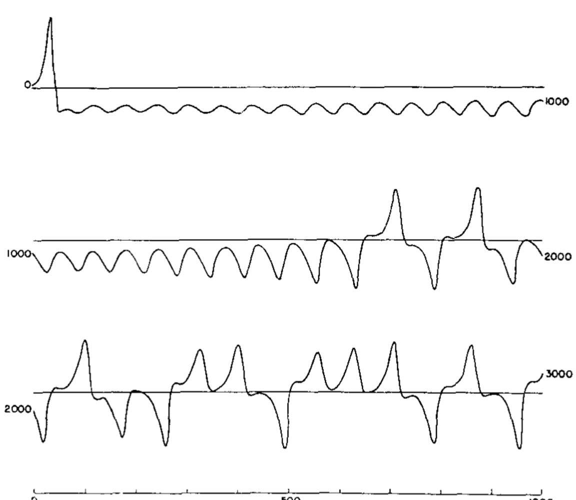

FIG. 6. Timeseries of the trajectory of 1 of three variables in the chaotic model of Lorenz (1963). Figure from Hawkins and Sutton (2009).

The Lyapunov exponent is a measure of how quickly initial conditions separate from one another in a chaotic system. Applying this notion to our system above, the Lyapunov exponent λ would appear as

$$F(\mathbf{x}_0, \mathbf{x}_0 + \delta\hat{\mathbf{x}}, t) = |\mathbf{x}_1(t) - \mathbf{x}_2(t)| \approx e^{\lambda t}\delta. \quad (38)$$

Does a limit exist in this case?

$$F(\mathbf{x}_0, \mathbf{x}_0 + \delta\hat{\mathbf{x}}, t) = |\mathbf{x}_1(t) - \mathbf{x}_2(t)| \approx e^{\lambda t}\delta < \epsilon, \quad (39)$$

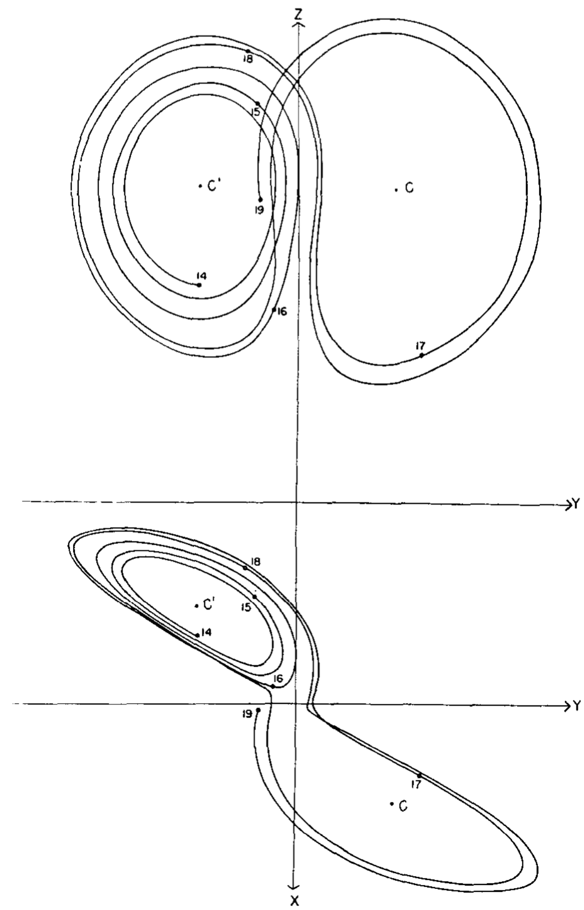

FIG. 7. Projections of trajectories in 3D of the chaotic model of Lorenz (1963). Figure from Hawkins and Sutton (2009).

So, for a given time t and a tolerance ϵ there is a δ small enough that the separation stays together under a growing Lyapunov exponent, but it is very small–exponentially small–for moderate times t,

$$\delta < \epsilon e^{-\lambda t} \quad (40)$$

For example, if $\lambda t = 1$, then δ must be a little less than $\epsilon/3$. If $\lambda t = 3$, then δ must be a bit less than $\epsilon/20$. If $\lambda t = 10$, then δ must be a bit less than $\epsilon/20000$.

One further consideration is whether there are underlying patterns, such as the Lorenz attractor, that can be understood. If so, then aspects of the climate system, such as the bounds of the attractor, can be used to improve climate projections. The Earth system has many wobbles, such as El Nino and the Madden-Julian Oscillation or the Quasi-Biennial Oscillation and variability of the Atlantic Meridional Overturning and Antarctic Circumpolar Current transport. It is unclear if these wobbles are forced oscillations that are just weakly damped, chaotic variability on some undescribed underlying attractor, nonlinear or linear resonances with random weather forcing, or some other mathematical description connecting the underlying processes. An important consequence distinguishing among these descriptions pragmatically is that applications could take advantage of the underlying system rather than assuming it is all just random.

Ergodicity is the property that any sizeable sample of a process is representative, statistically speaking, of the whole. We often assume that weather is ergodic, which means that we can take averages of it over time, space, or multiple initial condition ensembles and arrive at much the same climate. This is implicit in the approaches of Hawkins and Sutton (2009) and Sane et al. (2021). But, chaos and dynamical systems can be much more interesting than this, featuring *strange attractors* such as the one in the Lorenz (1963) system–it appears to be just two connected loops but later study has shown that it is in fact a fractal-like shape, with underlying complexity. The Lorenz (1963) produces quasi-periodic variability with deep underlying structure, not just white noise or other ergodic signals. Other classical nonlinear systems also do not reach ergodicity (e.g. Fermi et al. 1955). Oftentimes, we can use systems that are much simpler than we know the dynamics to be, such as Linear Inverse Models with added noise to predict El Nino or other climate variability (e.g. Penland and Magorian 1993; Weiss et al. 2019), in the hopes that the distinctions between random noise and the deep underlying structure is not consequential. On the other hand, decadal predictions (e.g. Meehl et al. 2009; Lovenduski et al. 2019) may have some support from the strange attractors of the system.

Putting all of these pieces together, when we estimate a predictability time τ in a chaotic system, we are estimating something similar to when the Lyapunov exponent obeys $\lambda\tau \leq 1$. For a more chaotic system, λ is bigger and τ is smaller. Beyond this predictability time the behavior is inherently chaotic, but bounded in that solutions stay within a climatological range. The study of climate is better understanding what processes and conditions govern this climatological range, while the study of weather seeks to understand what processes and conditions govern the predictable early phase and to extend where possible to the farthest possible forecast window.

6. Homework

a. Computing Cost

Revisit the calculation of global Direct Numerical Simulation cost from the *Direct statistical simulation* section, but calculate it if a grid that is $1\,\mathrm{cm}^3$ instead of $1\,\mathrm{mm}^3$ is adequate to resolve the atmospheric and oceanic fluid mechanics.

b. Helmholtz

Use separation of variables on the Helmholtz equation ($\nabla^2\phi_k(y,z) + k^2\phi_k(y,z) = 0$) to find two governing equations for a mode in y and a mode in z. Show that if the solutions to these are written in a separation of variables combined form $\sum_m \sum_n \phi_{mn} e^{ny} e^{mz}$, the whole system works if a particular relationship between m, n, k holds.

c. Entropy

Consider rolling 1 die (6 equally likely states) and 2 dice (36 equally likely states, but only 11 unequally likely sums from 2 to 12). Calculate the Shannon entropy of 1 die, the mutual information of 2 dice, and the Shannon entropy of the sum of 2 dice.

References

Allawala, A., S. Tobias, and J. Marston, 2020: Dimensional reduction of direct statistical simulation. *Journal of Fluid Mechanics*, **898**, A21.

Fermi, E., P. Pasta, S. Ulam, and M. Tsingou, 1955: Studies of the nonlinear problems. Tech. rep. doi:10.2172/4376203, URL https://www.osti.gov/biblio/4376203.

Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, **90 (8)**, 1095–1108, doi:10.1175/2009BAMS2607.1, URL https://doi.org/10.1175/2009BAMS2607.1.

Hewitt, H., B. Fox-Kemper, B. Pearson, M. Roberts, and D. Klocke, 2022: The small scales of the ocean may hold the key to surprises. *Nature Climate Change*, doi:10.1038/s41558-022-01386-6, URL https://rdcu.be/cPNdr.

IPCC, 2021a: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, chap. Summary for Policymakers, 3–32. Cambridge University Press, United Kingdom and New York, NY, USA, doi:10.1017/9781009157896.001, URL https://www.ipcc.ch/report/ar6/wg1/#SPM.

IPCC, 2021b: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, United Kingdom and New York, NY, USA, doi:10.1017/9781009157896, URL https://www.ipcc.ch/report/ar6/wg1/.

Li, K., J. Marston, and S. M. Tobias, 2021: Direct statistical simulation of low-order dynamosystems. *Proceedings of the Royal Society A*, **477 (2254)**, 20210427.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *Journal of atmospheric sciences*, **20 (2)**, 130–141, doi:10.1175/1520-0469(1963)020%3C0130:DNF%3E2.0.CO;2, URL https://doi.org/10.1175/1520-0469(1963)020%3C0130:DNF%3E2.0.CO;2.

Lovenduski, N. S., G. B. Bonan, S. G. Yeager, K. Lindsay, and D. L. Lombardozzi, 2019: High predictability of terrestrial carbon fluxes from an initialized decadal prediction system. *Environmental Research Letters*, **14 (12)**, 124074, doi:10.1088/1748-9326/ab5c55, URL https://doi.org/10.1088/1748-9326/ab5c55.

Meehl, G. A., and Coauthors, 2009: Decadal prediction: can it be skillful? *Bulletin of the American Meteorological Society*, **90 (10)**, 1467–1486, doi:10.1175/2009BAMS2778.1, URL https://doi.org/10.1175/2009BAMS2778.1.

Nicklas, J. M., B. Fox-Kemper, and C. Lawrence, 2023: Efficient estimation of climate state and its uncertainty using Kalman filtering with application to policy thresholds and volcanism. *Journal of Climate*, doi:10.31223/X5FH2C, URL https://doi.org/10.31223/X5FH2C, submitted.

Penland, C., and T. Magorian, 1993: Prediction of niño 3 sea surface temperatures using linear inverse modeling. *Journal of Climate*, **6 (6)**, 1067–1076.

Sane, A., B. Fox-Kemper, D. Ullman, C. Kincaid, and L. Rothstein, 2021: Consistent predictability of the Ocean State Ocean Model (OSOM) using information theory and flushing timescales. *Journal of Geophysical Research – Oceans*, **126 (7)**, e2020JC016 875, doi:10.1029/2020JC016875, URL http://doi.org/10.1029/2020JC016875.

Shannon, C. E., 1948: A mathematical theory of communication. *The Bell system technical journal*, **27 (3)**, 379–423.

Weiss, J. B., B. Fox-Kemper, D. Mandal, A. D. Nelson, and R. K. P. Zia, 2019: Nonequilibrium oscillations, probability angular momentum, and the climate system. *Journal of Statistical Physics*, 1–18, doi:10.1007/s10955-019-02394-1, URL https://doi.org/10.1007/s10955-019-02394-1.