Lecture 10: Multiple Equilibria of the THC

10.1 Introduction – 2-box model

We have, in the previous lectures, laid the groundwork for this one, which is arguably the central lecture concerning a conceptual understanding of the role of ocean circulation in climate dynamics. We introduce a box model, which represents the North Atlantic thermohaline circulation (THC) in its simplest possible form: The entirety of the low latitudes is represented by a single, well mixed box, as are the entire high latitudes. The model was introduced by Stommel over forty years ago (Stommel (1961)); we use here the simplification of Marotzke (1990). Despite its simplicity, the model displays an astonishing range of phenomena, many of which are central to a general theoretical understanding of dynamical systems.[1] All aspects of this model can be calculated analytically, and exactly, with the exception of the explicit time-dependent behaviour under time-varying forcing.

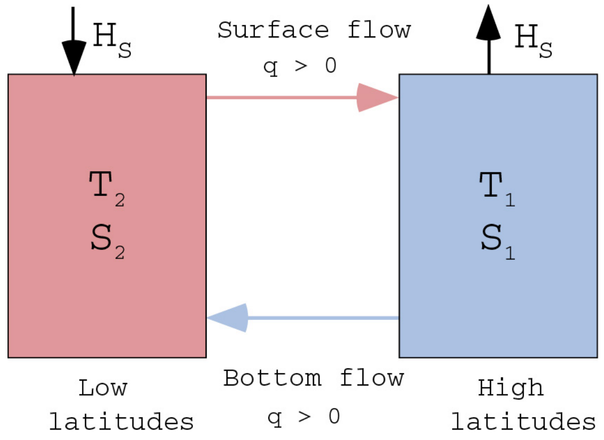

Figure 10.1: Stommel’s box model

[1] I think that this model plays – or should play – a role in understanding the THC and some classes of complex systems that is comparable to the role of the linear harmonic oscillator in basic physics. To compare it to another important paradigm: When I was a beginning graduate student, a very wise lecturer, Ulf Larsen, told us over and over again how important it was to illustrate the principles of quantum statistical mechanics with the simplest example, an isolated spin-½ particle, having just two quantum states (spin up and spin down). Being young and foolish, we used to chuckle, feeling that we wanted hard problems, not simple ones. I have since come to recognise our folly for what it was.

Heuristically, we assume that the atmosphere controls the ocean temperature and the surface fresh-water loss or gain, E (in m/s). Using the preceding lecture, we saw that this approximation is equivalent to assuming a Haney restoring law for heat flux with infinitely strong coupling; or, we use the extreme case of mixed thermohaline boundary conditions. We will see in the lectures on the coupled box model, later in this course, how to view this approximation as the limiting case of the coupled system. For now, let us proceed with the assumption that T_1 , T_2 , and E are prescribed as external parameters[2].

Again, as in the preceding lecture, we will use a virtual surface salinity flux, H_S:

$$H_S = S_0\, E/D \qquad (10.1)$$

where D is depth and S_0 a reference salinity.

The boxes are connected by pipes near the surface and the bottom; the pipes are assumed to have vanishing volume but are conduits for the flow. The thermohaline circulation strength is denoted by q (strictly speaking, q represents THC/Volume; q has units of s^{-1}). We use the sign convention that q>0 denotes poleward surface flow, implying equatorward bottom flow and, conceptually, sinking at high latitudes. This is the picture that we are used to when thinking about the North Atlantic THC. Conversely, q<0 means equatorward surface flow and poleward bottom flow. We assume a very simple flow law for q, namely, that it depends linearly on the density difference between high and low latitudes:

$$q = \frac{k}{\rho_0}\left[\rho_1 - \rho_2\right] \qquad (10.2)$$

where ρ_0 is a reference density and k is a hydraulic constant, which contains all dynamics, that is, the connection between density and the flow field. The equation of state is

$$\rho_i = \rho_0\left(1 - \alpha T_i + \beta S_i\right);\ \ i = 1, 2\ , \qquad (10.3)$$

where α and β are, respectively, the thermal and haline expansion coefficients,

$$\alpha \equiv -\left(1/\rho_0\right)\partial_T \rho \quad \beta \equiv \left(1/\rho_0\right)\partial_S \rho\ . \qquad (10.4)$$

[2] This is where we depart from Stommel (1961) and instead follow Marotzke (1990). Stommel (1961) used Haney-type conditions for both temperature and salinity, but with a longer restoring timescale for salinity. As a consequence, the original Stommel box model cannot readily be solved analytically.

For simplicity, we employ a linear equation of state; that is, both α and β are assumed constant. The flow law, (10.2), thus becomes, using (10.3),

$$q = k\left[\alpha\left(T_2 - T_1\right) - \beta\left(S_2 - S_1\right)\right] \qquad (10.5)$$

As we assume that the temperatures are fixed by the atmosphere and enter the problem as external parameters, we need not formulate a heat conservation equation. The salt conservation equations for the Stommel model are

$$\dot{S}_1 = -H_S + |q|\left(S_2 - S_1\right), \qquad (10.6)$$

$$\dot{S}_2 = H_S - |q|\left(S_2 - S_1\right), \qquad (10.7)$$

which may require a little explanation. We postulate that flow into a box carries with it the properties, in particular the salinity, of the originating box. (We note in passing that this is equivalent to “upstream differencing”). So, if q>0, the upper pipe brings water with salinity S_2 into Box 1, while the lower pipe takes water with S_1 out of Box 1. If q<0, it is the lower pipe that imports S_2 into Box 1, while the upper pipe exports S_1 out of Box 1. In either case, S_2 is imported into Box 1, while S_1 is exported out of Box 1, both at a rate given by the modulus of q. This is what (10.6) expresses. *Mutatis mutandis*, the same holds for Box 2 and (10.7).

We introduce the following abbreviations for meridional differences of temperature, salinity, and density:

$$T \equiv T_2 - T_1; \quad S \equiv S_2 - S_1; \quad \rho \equiv \rho_1 - \rho_2, \qquad (10.8)$$

which implies that

$$q = \frac{k}{\rho_0}\rho = k\left[\alpha T - \beta S\right]. \qquad (10.9)$$

Under normal conditions, net evaporation occurs at the warmer low latitudes and net precipitation at the colder high latitudes; in other words, temperature and salinity are expected to be both high at low latitudes and both low at high latitudes. In their influence on the THC, two cases can be distinguished. When the temperature difference dominates the salinity difference in their influence on density, high-latitude density is greater than the low-latitude density. Therefore, q>0, and the surface flow is poleward. One can say that the temperature difference, T, drives the THC and the salinity difference, S, brakes the THC, as seen from

$$q > 0 : |q| = q = k\left[\alpha T - \beta S\right]. \qquad (10.10)$$

Conversely, when the salinity difference dominates the temperature difference, high-latitude density is lower than the low-latitude density, q<0, the surface flow is equatorward. Now, S drives the THC, and T brakes it:

$$q < 0 : |q| = -q = k\left[\beta S - \alpha T\right] \qquad (10.11)$$

The sum of the salt conservation equations (10.6) and (10.7) gives

$$\dot{S}_1 + \dot{S}_2 = 0 \, , \qquad (10.12)$$

reflecting that total salt mass is conserved. (One consequence of this simplification is that we cannot determine the mean salinity from the set of equations we use here. Processes other than evaporation, precipitation, and oceanic transport of salinity must be invoked for the determination of the total oceanic salt content.) Because of the constancy of total salt mass, (10.12), equivalent to the constancy of global mean salinity, we need only consider the difference, S, between S_2 and S_1. The difference of the salt conservation equations (10.6) and (10.7) gives an equation for S:

$$\dot{S}_2 - \dot{S}_1 = \dot{S} = 2H_S - 2|q|S \, , \qquad (10.13)$$

or, using the flow law (10.9),

$$\dot{S} = 2H_S - 2k\left|\alpha T - \beta S\right|S \, , \qquad (10.14)$$

which completes the formulation of the model – its behaviour is completely characterised by (10.14) .

10.2 Equilibrium solutions

As the first step in our analysis of (10.14), governing the evolution of the salinity difference between the low and high latitude boxes, we look for steady-state or equilibrium solutions, defined by a vanishing of the time derivative:

$$H_S - k\left|\alpha T - \beta\overline{S}\right|\overline{S} = 0 \, , \qquad (10.15)$$

where the overbar marks a steady-state quantity. We must consider separately the cases where the argument of the modulus is positive or negative.

Case I:

$$\bar{q} > 0, \quad \alpha T > \beta \bar{S} \tag{10.16}$$

We can simply replace the modulus signs by brackets, giving

$$H_S - k\left(\alpha T - \beta \bar{S}\right)\bar{S} = 0\,, \tag{10.17}$$

or

$$\left(\beta \bar{S}\right)^2 - \left(\beta \bar{S}\right)\left(\alpha T\right) + \beta H_S / k = 0\,, \tag{10.18}$$

which has the roots

$$\left(\beta \bar{S}\right)_{1/2} = \left(\alpha T\right)\left\{\tfrac{1}{2} \pm \sqrt{\tfrac{1}{4} - \frac{\beta H_S}{k\left(\alpha T\right)^2}}\right\}. \tag{10.19}$$

For a positive radicand, defined by

$$\frac{\beta H_S}{k\left(\alpha T\right)^2} < \tfrac{1}{4}\,, \tag{10.20}$$

the model has two equilibrium solutions for poleward near-surface flow. These solutions can also be characterised as thermally dominated or, in the language of atmospheric science, “thermally direct” (meaning that rising motion occurs at the location of heating, and subsidence at the location of cooling). If the freshwater flux forcing exceeds the threshold defined by (10.20), no thermally-driven equilibrium exists.

Case II:

$$\bar{q} < 0, \quad \alpha T < \beta \bar{S} \tag{10.21}$$

Now, we must insert a minus sign when replacing the modulus signs by brackets,

$$H_S + k\left(\alpha T - \beta \bar{S}\right)\bar{S} = 0\,, \tag{10.22}$$

which gives

$$\left(\beta \bar{S}\right)^2 - \left(\beta \bar{S}\right)\left(\alpha T\right) - \beta H_S / k = 0\,, \tag{10.23}$$

and the single root

$$\left(\beta \bar{S}\right)_3 = \left(\alpha T\right)\left\{\tfrac{1}{2} + \sqrt{\tfrac{1}{4} + \frac{\beta H_S}{k\left(\alpha T\right)^2}}\right\} \tag{10.24}$$

Notice that we must discard the negative root; the radicand is greater than ¼, so that the negative root would imply $\overline{S} < 0$, in contradiction to the condition (10.21). The solution (10.24) has equatorward near-surface flow and can be characterised as salinity dominated or "thermally indirect". It exists for all (positive) values of the freshwater flux forcing.

Figure 10.2 shows the equilibrium solutions as a function of the freshwater flux forcing. In summary, we find the remarkable result that this simplest non-trivial model of the THC, represented in steady state by the pair of quadratic equations, (10.15) and (10.22), has three steady state solutions, provided that the freshwater flux forcing is not too strong [cf., (10.20)]. Two equilibria have $\overline{q} > 0$ (poleward surface flow); they are characterised by either a small salinity contrast and strong flow ($\beta\overline{S} < \tfrac{1}{2}\alpha T, \overline{q} > \tfrac{1}{2}k\alpha T$), or by a large salinity contrast and weak flow ($\beta\overline{S} > \tfrac{1}{2}\alpha T, \overline{q} < \tfrac{1}{2}k\alpha T$). These steady states exist only if $\frac{\beta H_S}{k(\alpha T)^2} < \tfrac{1}{4}$. The model has one steady-state solution with $\overline{q} < 0$ (equatorward surface flow), characterised by a very large salinity contrast ($\beta\overline{S} > \alpha T, \overline{q} < 0$). This solution always exists, and is the only one if $\frac{\beta H_S}{k(\alpha T)^2} > \tfrac{1}{4}$.

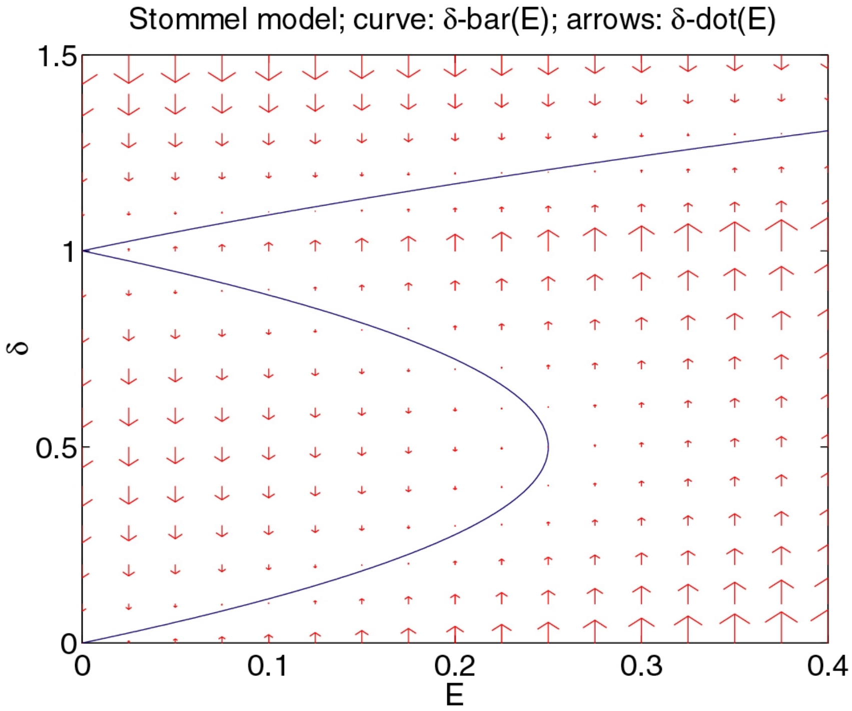

Figure 10.2 Solution portrait of the box model in phase space. Dimensionless salinity difference is denoted $\delta \equiv \beta S/\alpha T$; dimensionless surface salinity flux is $E \equiv \beta H_S / k(\alpha T)^2$. The curves mark the equilibrium solutions, $\overline{\delta}(E)$, while the arrows show the tendencies in phase space. Notice the existence of three steady states for E < ¼.

What is the physical reason behind the vanishing of the thermally direct solution if $\frac{\beta H_S}{k(\alpha T)^2} > 1/4$? Stronger surface salinity flux must by balanced by stronger salinity advection, qS. This can be accomplished either by increasing the salinity difference, S, between low and high latitudes, or by increasing the flow strength, q. But increasing S has the dynamical consequence of weakening the flow – (10.9) expresses that q decreases linearly with S. Obviously, the product, qS, is zero for either S = 0 or q = 0 (the latter implying $\beta\overline{S} = \alpha T$); qS is positive for intermediate values and attains a maximum at $\beta\overline{S} = 1/2\,\alpha T$ (see phase space diagram, Fig. 10.2). At this point, $\overline{q}\overline{S} = 1/4\,\frac{k}{\beta}(\alpha T)^2$, which marks the critical freshwater flux forcing, that is, the strongest forcing that can be balanced by salinity advection through thermally direct flow. For even greater H_S, balance is impossible.

An even deeper question than the one starting the preceding paragraph is, what makes the multiple equilibria possible in the first place? Two crucial ingredients are required. First is the advective nonlinearity: The flow advecting salinity is itself influenced by salinity gradients, through density. Without this nonlinearity the model would have a unique solution (or none at all). But there is a second requirement, that of different coupling of temperature and salinity to the atmosphere. We assume that the atmosphere controls temperature but the salinity *flux*. Imagine, instead, two extreme cases of equal coupling:

i. *Temperature **and** salinity prescribed:*

Then, density is prescribed as well, meaning that the flow prescribed. Trivially, no multiple equilibria are possible.

ii. *Heat **and** freshwater **flux** prescribed:*

Then, the surface density (or buoyancy) flux is prescribed and, hence, the steady-state horizontal density transport, $k\left|\overline{\rho}\right|\overline{\rho}$. As k and $\left|\overline{\rho}\right|$ are positive, the sign of $\overline{\rho}$ is uniquely determined by the sign of the surface buoyancy flux: If the low latitude box receives buoyancy from the atmosphere, it is less dense than the high latitude box, and $\overline{\rho}$ and $\overline{q}$ are both positive (thermally direct circulation). The converse is true for prescribed buoyancy loss at low latitudes. Hence, the steady-state circulation is uniquely determined.

Exercises:

1. *Loss of multiple steady states: What steady-state solutions are possible in the 2-box model if the flow field is given as an external parameter (that is, depends neither on temperature nor on salinity)?* Hint: *Plot the salinity difference as a function of freshwater forcing, with q given and constant.*
2. *Loss of multiple steady states: Prove the sequences i. and ii. outlined just above, by using the appropriate modifications of the equations for the Stommel box model, (10.2) - (10.7).*
3. *Loss of multiple steady states: Suppose that the surface heat and salt fluxes are formulated as restoring laws, as originally done by Stommel, i.e., the equations are*

$$\dot{T}_1 = \lambda_T \left(T_1^* - T_1\right) + |q|\left(T_2 - T_1\right) \qquad (10.25)$$

$$\dot{T}_2 = \lambda_T \left(T_2^* - T_2\right) - |q|\left(T_2 - T_1\right) \qquad (10.26)$$

$$\dot{S}_1 = \lambda_S \left(S_1^* - S_1\right) + |q|\left(S_2 - S_1\right) \qquad (10.27)$$

$$\dot{S}_2 = \lambda_S \left(S_2^* - S_2\right) - |q|\left(S_2 - S_1\right) \qquad (10.28)$$

where the starred quantities are the target values. Assume that $\lambda_T = \lambda_S$ *and construct a single ordinary differential equation for q. What are the physically meaningful steady-state solutions now? What would change if* $\lambda_T \neq \lambda_S$ *? N.B.: Do not solve the entire problem for* $\lambda_T \neq \lambda_S$ *.*

9.3 Stability

We have identified three equilibria of the 2-box model of the THC in a certain parameter range. Now, we concern ourselves with the stability of the equilibria – more precisely, with the "linear" stability. This means that we want to understand what happens if the equilibrium is perturbed by a tiny amount, either in the forcing, H_S, or in the solution, S. We will use a variety of techniques, each of which is important generally in the analysis of dynamical systems, and each of which illuminates one or several characteristics.

We start by investigating in more detail the equilibrium curves in phase space, Fig. 10.2. From the steady-state conditions, as expressed in eqs. (10.18) and (10.23), we obtain through a slight modification,

$$\bar{q} > 0, \frac{\beta\bar{S}}{\alpha T} < 1: \quad \frac{\beta H_S}{k(\alpha T)^2} = -\left(\frac{\beta\bar{S}}{\alpha T}\right)^2 + \left(\frac{\beta\bar{S}}{\alpha T}\right), \tag{10.29}$$

$$\bar{q} < 0, \frac{\beta\bar{S}}{\alpha T} > 1: \quad \frac{\beta H_S}{k(\alpha T)^2} = +\left(\frac{\beta\bar{S}}{\alpha T}\right)^2 - \left(\frac{\beta\bar{S}}{\alpha T}\right), \tag{10.30}$$

which expresses the dimensionless salinity gradient, $\delta \equiv \beta S / \alpha T$, as a function of the dimensionless surface salinity flux, $E \equiv \beta H_S / k(\alpha T)^2$. Thus, we can write (10.29) and (10.30)in dimensionless form as

$$\delta \le 1: \quad E = -\delta^2 + \delta = \delta(1-\delta) \tag{10.31}$$

$$\delta \ge 1: \quad E = \delta^2 - \delta = \delta(\delta - 1). \tag{10.32}$$

This pair of equations represents two sideways parabolas, with opposite orientation, intersecting at $\delta \equiv \beta S/\alpha T = 0$ (no salinity difference) and $\delta = 1$ ($\alpha T = \beta S$; no flow). In either case, the forcing must vanish ($E \equiv \beta H_S / k(\alpha T)^2 = 0$). The curves depicted in Fig. 10.2 are the zeros of the salinity conservation equation (10.14), rewritten in dimensionless form as

$$\frac{1}{2k\alpha T}\frac{d}{dt}\left(\frac{\beta S}{\alpha T}\right) = \frac{\beta H_S}{k(\alpha T)^2} - \left|1 - \left(\frac{\beta S}{\alpha T}\right)\right|\left(\frac{\beta S}{\alpha T}\right). \tag{10.33}$$

Notice that (10.33) implies an advective timescale, suitable for nondimensionalisation, of $(2k\alpha T)^{-1}$, and a nondimensional overturning strength of $\tilde{q} = 1 - \delta$. We can thus rewrite (10.33) as

$$\dot{\delta} = E - |1-\delta|\delta \tag{10.34}$$

Exercise:

4. *Prove the statement in the sentence following (10.33)* Hint:. *Write* $t = \hat{t}\,\tilde{t}$; $q = \hat{q}\,\tilde{q}$ *etc., where the caret denotes the scale and the tilde the non-dimensional quantity.*
5. *Find the steady-state solutions of (10.34), that is, perform the procedure leading to (10.19) and (10.24), but using non-dimensional quantities from the outset.*

From either (10.33) or (10.34), we can read off the following. On the equilibrium curve, the tendency (time rate of change) of the salinity difference between high and low latitudes vanishes. But to the left of the curve, E or H_S is smaller than required by the equilibrium condition. Hence, $\dot{S} < 0$, and S decreases, as indicated by the downward pointing arrows in Fig. 10.2. In fact, the arrows were calculated from the right-hand sides of (10.34). To the right of the curve, E or H_S is greater than required for equilibrium, hence $\dot{S} > 0$, and S increases. Notice that for every given $\overline{\delta}$ in Fig. 10.2, there belongs a unique E, so “left” and “right” of the equilibrium curve are unambiguously defined.

By visual inspection of Fig. 10.2, we can now read off the stability properties of the solutions. If, by any initial perturbation or change in forcing, we find ourselves to the left of the equilibrium curve, the evolution depends critically on which solution branch we started from. On the top $\left(\overline{\delta} > 1\right)$ and bottom $\left(\overline{\delta} < 1/2\right)$ branches in Fig. 10.2 (salinity dominated and thermally dominated-strong flow, respectively), the systems moves downward, back towards the equilibrium curve. But if one starts from the middle branch $\left(1/2 < \overline{\delta} < 1\right)$, which runs from top-left to bottom-right in Fig. 10.2, the system does not return, but instead undergoes a transition towards the lower, thermally dominated branch. If the initial perturbation or change in forcing leaves the system to the right of the equilibrium curve, the system moves upward, again back towards the equilibrium curve, if it started from the top or the bottom branch. But if it started from the middle branch, it would make a transition toward the salinity-dominated equilibrium. Hence we conclude that the salinity-dominated steady state is always stable, the strong-flow thermally dominated steady state is stable (if it exists), while the weak-flow thermally dominated steady state is unstable to infinitesimal perturbations. There exists a tell-tale sign allowing one to infer this instability even without investigating the full time-dependent equation. As one follows the unstable branch in Fig. 10.2 $\left(1/2 < \overline{\delta} < 1\right)$, from left to right, say, an increase in E implies a decrease in δ. Thus, an *increase* in forcing leads to a *decrease* in the steady-state response, which is, to my knowledge, an unfailing indication of instability.

Two points deserve special mention, since they are *semistable*, meaning that the system approaches them if it is on one side in phase space, but moves away from them if it is on the other side. These points are $\left(E = 0,\ \delta = 1\right)$, where the two parabolas meet, and $\left(E = 1/4,\ \delta = 1/2\right)$, the point beyond which no thermally direct steady state is possible. (In the language of dynamical systems, this is called a saddle node bifurcation.) Both these points show interesting mathematical behaviour, but they are not of great physical interest because this behaviour is not robust to small perturbations, such as a small amount of random noise.

Lyapunov potential

A powerful illustration of the stability properties discussed in the preceding paragraphs comes from a mathematical construct called the “Lyapunov potential”. In loose analogy to, say, the relationship between gravitational force and gravitational potential, the time rate of change of dimensionless salinity, $\dot{\delta}$, (cf., (10.34)), is written as the negative gradient of the Lyapunov potential, L, such that

$$-\frac{\partial L}{\partial \delta} = \dot{\delta} = E - |1-\delta|\,\delta\,. \tag{10.35}$$

By construction, the steady states of the system coincide with the extrema (maximum or minimum) of the Lyapunov potential. But we can say more: Plotting $L(\delta)$ immediately indicates the stability properties of the equilibria; indeed one can interpret the stability as if a bead was sliding on a wire under the influence of gravity: A minimum in L is a stable equilibrium, while a maximum is an unstable equilibrium. We first illustrate this graphically, before showing it mathematically.

It is readily shown that

$$\begin{aligned} L &= -E\delta - \tfrac{1}{3}\delta^3 + \tfrac{1}{2}\delta^2; \qquad && \delta \le 1 \\ L &= -E\delta + \tfrac{1}{3}\delta^3 - \tfrac{1}{2}\delta^2 + \tfrac{1}{3}; && \delta \ge 1 \end{aligned} \tag{10.36}$$

fulfils (10.35), including the (arbitrary) condition of $L(0)=0$ and the (non-arbitrary) condition of continuity at $\delta = 1$. Figure 10.3 shows the Lyapunov potential, as a function of δ, for a variety of choices for E. The case, $E = 0$, has one minimum at $\delta = 0$ and a double extremum (level turning point) at $\delta = 1$. The former is stable, according to Fig. 10.2, while the latter is semistable (approached from the right, moved away from on the left). Thus, we can visualise the evolution of the system as the inertia-less sliding of a bead on the “wire” $L(\delta)$. As E is nonzero but less than ¼, the minimum at the left moves from zero to higher values, while another minimum appears for $\delta > 1$ and growing. Since $L(\delta)$ is continuous, the two minima must be separated by a maximum. In other words, two stable equilibria must have an unstable equilibrium between them.

As E approaches ¼, the minimum at $\delta > 1$ becomes deeper than the one at $\delta < 1/2$, until, at $E = 1/4$, the two equilibria with $\delta < 1$ merge to form a level turning point. This is the second semistable point discussed in Fig. 10.2. For even greater E, the thermally dominated ($\delta < 1$) equilibrium vanishes altogether, although its vicinity can still be felt through the very small time rates of change nearby.

After gaining an intuitive understanding of how to interpret *L*, we can now derive mathematically how its shape reflects stability properties. At any point, if *L* increases with δ, the left-hand side of (10.35) is negative, $\dot{\delta} < 0$, and δ decreases. In the $L(\delta)$ phase plot, Fig. 10.3, one slides toward the left. The converse is true if L decreases with δ. In the vicinity of a minimum, hence, any deviation to the right (*L* increasing with δ) is followed by motion to the left, back toward the minimum. Likewise, any deviation to the left will be followed by motion back to the minimum. Near a maximum, instead, a deviation to the right, say, means that *L decreases* with δ, the left-hand side of (10.35) is positive, $\dot{\delta} > 0$, and δ increases further, that is, the system moves further to the right, away from the equilibrium. For deviations to the left of a maximum, $\dot{\delta} < 0$, and δ decreases further, again moving away from the equilibrium. Hence, if we can construct a Lyapunov potential as in (10.36), we can immediately read off the plot the stable and unstable steady states, in a completely intuitive manner.

Notice that in a case such as depicted in Fig. 10.3, sometimes the nomenclature is adopted to call the stable equilibrium with the shallower potential well “metastable”, reserving the term “stable” only for the steady state with the globally lowest potential. Here, we will largely only concern ourselves with distinguishing between stability and instability to infinitesimal perturbations.

Exercises

6. *Prove that (10.36) is the correct Lyapunov potential for the system described by (10.34).*
7. *At what value of δ are the two minima in Fig. 10.3 equally deep [$L(\delta)$ equal values]?*

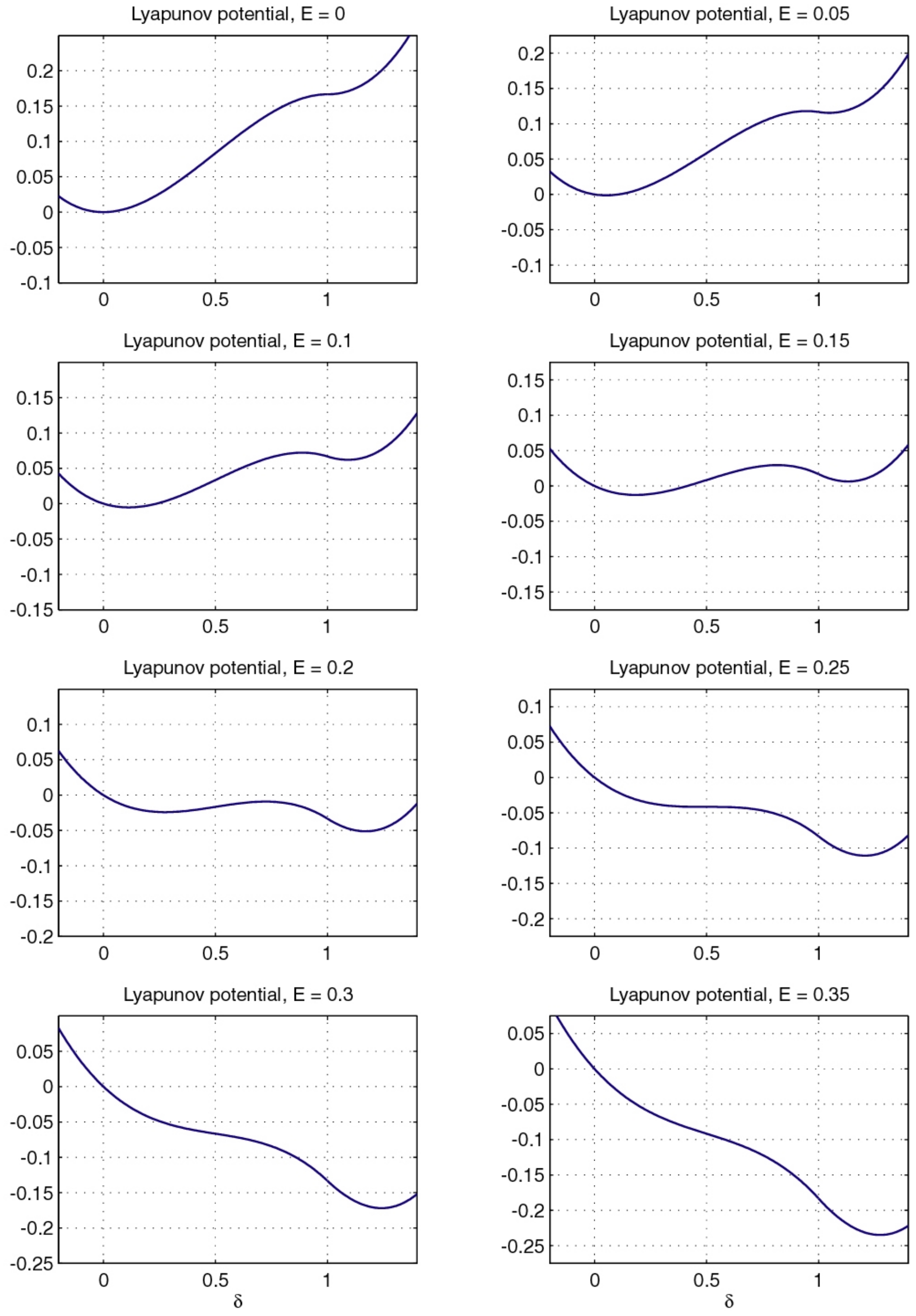

Figure 10.3: Lyapunov potential as defined by (10.36), for a variety of choices for E.

10.4 Feedbacks

We have found, from either the phase space plot, 10.2, or the Lyapunov potential, 10.3, how to characterise the multiple equilibria of the 2-box model as either stable or unstable. But what are the processes that lead to stability or instability? To this end, we now analyse the model equations in the vicinity of the steady states, employing a powerful technique applicable near any equilibrium state. The trick is to approximate the full, nonlinear equation through a linear one, such that the approximation ("*linearisation*") is good in the vicinity of the steady state (notice that one must linearise separately about every distinct equilibrium.

For this exercise, we return to the original, dimensional equation for the salinity difference between low and high latitudes, (10.13), and the flow law, (10.9). We write all quantities as the sum of the steady-state value, marked again by an overbar, and a deviation thereof, marked by a prime, such that

$$S = \overline{S} + S', \quad q = \overline{q} + q' \tag{10.37}$$

This separation is interesting in our case (or complicated, depending on taste), owing to the appearance of the modulus of q in the salinity advection. Care must be taken, and we again must distinguish between positive and negative *q*:

$$\begin{aligned} |q| = |\overline{q} + q'| &= \left| k\alpha T - k\beta\left(\overline{S} + S'\right)\right| \\ &= \left[k\alpha T - k\beta\left(\overline{S} + S'\right)\right] = |\overline{q}| - k\beta S'; \quad \overline{q} > 0\,, \\ &= \left[k\beta\left(\overline{S} + S'\right) - k\alpha T\right] = |\overline{q}| + k\beta S'; \quad \overline{q} < 0 \end{aligned} \tag{10.38}$$

where it has been used that

$$q' = -k\beta S' \tag{10.39}$$

because *T* is an external parameter. The salinity conservation equations, (10.13), is now written, using the expansion (10.37),

$$\dot{S} = \left(\dot{\overline{S}} + \dot{S}'\right) = \dot{S}' = 2H_S - 2|q|S = 2H_S - 2\left(|\overline{q}| \mp k\beta S'\right)\left(\overline{S} + S'\right); \quad - : \overline{q} > 0; \quad + : \overline{q} < 0 \;, \tag{10.40}$$

where we have used that the steady-state value does not change with time. We can subtract from (10.40) the steady-state condition, (10.15), leaving

$$\dot{S}' = -2\left|\overline{q}\right| S' \pm k\beta S'\left(\overline{S} + S'\right); \quad + : \overline{q} > 0; \quad - : \overline{q} < 0 \ , \qquad (10.41)$$

Notice that so far, we have not introduced any approximation yet, but merely rewritten the original equation in an inflated form. Now, however, we introduce the assumption that

$$S' \ll \overline{S};\ q' \ll \overline{q} \ , \qquad (10.42)$$

that is, the deviations from the equilibrium values are small compared to the equilibrium values themselves. In other words, we remain close to the steady-state. In this case, we can neglect the term containing the product of two perturbations quantities, leaving behind only terms that are linear in primed quantities (hence the term linearisation),

$$\dot{S}' \cong -2\left|\overline{q}\right| S' \pm 2k\beta S'\overline{S}; \quad + : \overline{q} > 0; \quad - : \overline{q} < 0 \ . \qquad (10.43)$$

On this approximated equation, or any other obtained through this approach, we can launch the full power of systematic solutions of linear differential equations. We know that if the coefficient multiplying *S'* on the right-hand side is negative, the perturbation *S'* is exponentially damped toward zero – the system returns to the steady state, which hence is stable. In contrast, if the coefficient multiplying *S'* on the right-hand side is positive, the perturbation *S'* grows exponentially, the system does not return to its equilibrium, which hence is unstable. Notice that *S'* does not go to infinity – as it grows too large, the assumption, (10.42), behind the linearisation breaks down, and one has to resort to the full nonlinear analysis.

Which processes determine whether the steady state is stable or unstable? We must analyse (10.43) to determine the contributors to the coefficient of *S'*. Each of the terms represents a *feedback*, meaning a contribution to a tendency in *S'* that is caused by *S'* itself. The first term represents the advection of an anomaly in salinity difference by the time-mean flow, and can hence be called the *mean flow feedback*: Assume that, from whatever cause, $S' > 0$. The first term on the right-hand side of (10.43) contributes negatively, $\dot{S}' < 0$, so *S'* is reduced by this term. In other words, the mean flow feedback works against the original anomaly, hence stabilises the equilibrium – which is the definition of a *negative feedback.* It is readily shown that negative anomalies (original $S' < 0$) are damped as well. Notice that the mean flow feedback works identically in the thermally dominated and haline dominated equilibria, though with different strengths.

The second term on the right-hand side of (10.43) represents the advection of mean salinity gradient by the perturbation flow, and can hence be called the *salinity transport feedback*. (Notice that advection of perturbation salinity gradient by perturbation flow is neglected in this linear approximation). The sign of the salinity transport feedback depends on the steady-state flow direction. If $\overline{q} > 0$, and $S' > 0$ (say), then $\dot{S}' > 0$, and the initial perturbation is further increased. Again, it is readily shown that this amplification is independent of the sign of the initial anomaly. If $\overline{q} > 0$, hence, the salinity transport feedback is a *positive feedback*, destabilising the equilibrium.

The situation is different for the haline dominated equilibrium, $\overline{q} < 0$. If $S' > 0$, then $\dot{S}' < 0$, from the contribution by the second term on the right-hand side of (10.43), and the initial perturbation is reduced. The salinity transport feedback is a negative, stabilising feedback.
In summary, we identify two negative feedbacks for the thermally indirect or haline dominated circulation, $\overline{q} < 0$. As all feedbacks are negative, this equilibrium is always stable to infinitesimal perturbations.

In contrast, the thermally direct circulation, $\overline{q} > 0$, has one positive feedback and one negative feedback. To determine the stability of the equilibrium, the relative strengths of the competing feedbacks must be evaluated. Using the dynamic flow law, (10.9), for $\overline{q}$ in the salinity perturbation equation (10.41), gives

$$\dot{S}' \cong -2\overline{q}S' + 2k\beta S'\overline{S} = -2k\left(\alpha T - 2\beta\overline{S}\right)S' . \qquad (10.44)$$

Hence, if $\beta\overline{S} < 1/2\,\alpha T$, the coefficient multiplying S' is negative, and the equilibrium is stable. In contrast, if $1/2\,\alpha T < \beta\overline{S} < \alpha T$, the coefficient multiplying S' is positive, and the equilibrium is unstable. In the former case, the stabilising mean flow feedback dominates, whereas in the latter, the destabilising salinity transport feedback dominates.

Exercise:

8. *Complete the discussion of feedback loops for all cases and show that the sign of the feedback is independent of the sign of the initial anomaly.*

10.5 Time-dependent solution

At the beginning of this lecture, I made a rather oblique remark concerning the exceptions to the statement that we can completely calculate the solution to the simplified Stommel model . Of course, one can always invent forcing histories, such as E(t) in the dimensionless salt conservation equation (10.34), that an analytical solution can only be given symbolically. But (10.34) permits the exact, and relatively simple, analytical solution of its full time-dependence. As of writing these notes (January 2002), I am unaware of any published account of this solution. And since the solution provides a perspective that cannot be obtained from the previous approaches, it is given here.

With some help from Matlab's ® Symbolic Math toolbox, one readily finds as the solution to (10.34):

$$\delta(t) = \tfrac{1}{2} - \sqrt{\tfrac{1}{4} - E}\ \tanh\left\{ t\sqrt{\tfrac{1}{4} - E} + \operatorname{atanh}\frac{\tfrac{1}{2} - \delta(0)}{\sqrt{\tfrac{1}{4} - E}} \right\};\ \delta \le 1, \qquad (10.45)$$

$$\delta(t) = \tfrac{1}{2} + \sqrt{\tfrac{1}{4} + E}\ \tanh\left\{ t\sqrt{\tfrac{1}{4} + E} + \operatorname{atanh}\frac{-\tfrac{1}{2} + \delta(0)}{\sqrt{\tfrac{1}{4} + E}} \right\};\ \delta \ge 1, \qquad (10.46)$$

where *atanh* is the inverse of the hyperbolic tangent, *tanh*, and $\delta(0)$ is the initial condition. Using $\frac{d}{dx}\tanh x = 1 - \tanh^2 x$ and noticing that the derivative of the argument of the tanh gives an additional factor of $\sqrt{\tfrac{1}{4} + E}$, we obtain from (10.45) that

$$\dot{\delta}(t) = -\left(\tfrac{1}{4} - E\right)\left(1 - \tanh^2\{\ldots\}\right);\ \delta \le 1. \qquad (10.47)$$

The validity of (10.34) is then readily shown by substitution of $\delta(t)$ and $\delta^2(t)$. That . (10.45) is valid for $t = 0$ is almost trivial.

Exercise

9. *Prove that (10.45) and (10.46) are the correct solutions of (10.34).*

In addition to showing mathematical validity, (10.45) and (10.46) offer other interesting aspects. The long-term behaviour is very simple; for large *t*, the first term dominates the argument of the *tanh* (the initial condition is forgotten), and since *tanh* approximates 1 for large argument, we recover the two stable equilibria,

$$t \to \infty: \;\; \delta(t) \to \tfrac{1}{2} - \sqrt{\tfrac{1}{4} - E} \;\; ; \;\; E < \tfrac{1}{4}, \qquad (10.48)$$

$$t \to \infty: \;\; \delta(t) \to \tfrac{1}{2} + \sqrt{\tfrac{1}{4} + E} \;\; ; \;\; E > \tfrac{1}{4}. \qquad (10.49)$$

Notice that there is no trace of the unstable equilibrium left in the time-dependent solution, reflecting the fact that time evolution is always *away* from the unstable steady state. [Writing $\delta(t) = \tfrac{1}{2} + \sqrt{\tfrac{1}{4} - E}\;\tanh\{\ldots\}$ etc. in (10.45) would not fulfil (10.34) – try it!]. Notice, further, that (10.45) is perfectly valid even for $E > \tfrac{1}{4}$; indeed, using that $\tanh ix = i \tan x$ etc., indicates that if $E > \tfrac{1}{4}$, $\delta(t)$ grows until it becomes greater than one, and (10.46) must be used.

Figure 10.4 shows evaluations of the full time-dependent solutions to the 2-box model, (10.45) and (10.46), as functions of initial conditions and time. Notice that, if the solutions crosses the $\delta(t) = 1$ threshold from below, at time t_c, use of (10.45) must be discontinued and (10.46) must be used instead, with initial condition $\delta(t_c) = 1$. The first row shows the solutions for E = 0.2. Three types of behaviour are discernible in Fig. 10.4a. Low and high initial conditions lead to rapid convergence to the stable thermally and haline dominated equilibria, respectively. Intermediate-size initial conditions mean that the solutions hover near the unstable equilibrium for a while, before departing from it and approaching one of the stable steady states. Fig. 10.4b illustrates this behaviour in a contour plot. Moving horizontally to the right indicates the solution changing in time as one crosses colour separations. For long times, the two stable equilibria fill out the entire phase space, as witnessed by the ever expanding areas of orange and blue. The transition between the two values becomes sharper as time progresses and indicates the ever shrinking region in phase space from where the system has not yet exited to one of the stable equilibria ("attractors"). The case, E = 0.24, close to the bifurcation point, shows this general behaviour in more pronounced form. (It is readily shown that the equilibria are $\delta = 0.4,\ 0.6,\ \text{and}\ 1.2$, which means that they fall on the boundaries between colours in the intervals chosen). Finally, if E = 0.26, and there is no thermally dominated equilibrium any more, some of the trajectories approach the (now unique) equilibrium quickly, while those starting from a small initial value hover near the (now vanished) steady state, its influence still there. But one by one, the trajectories undergo a rapid transition (Fig. 10.4e). The transition region between red and blue colours is not horizontal any more, as it was for E < 0.25, indicating that sooner or later, all initial conditions lead to the haline dominated equilibrium (Fig. 10.4f).

References

Marotzke, J., 1990: Instabilities and multiple equilibria of the thermohaline circulation. Ph.D. thesis. *Berichte aus dem Institut für Meereskunde Kiel*, **94**, 126pp.

Stommel, H., 1961: Thermohaline convection with two stable regimes of flow. *Tellus*, **13,** 224-230.

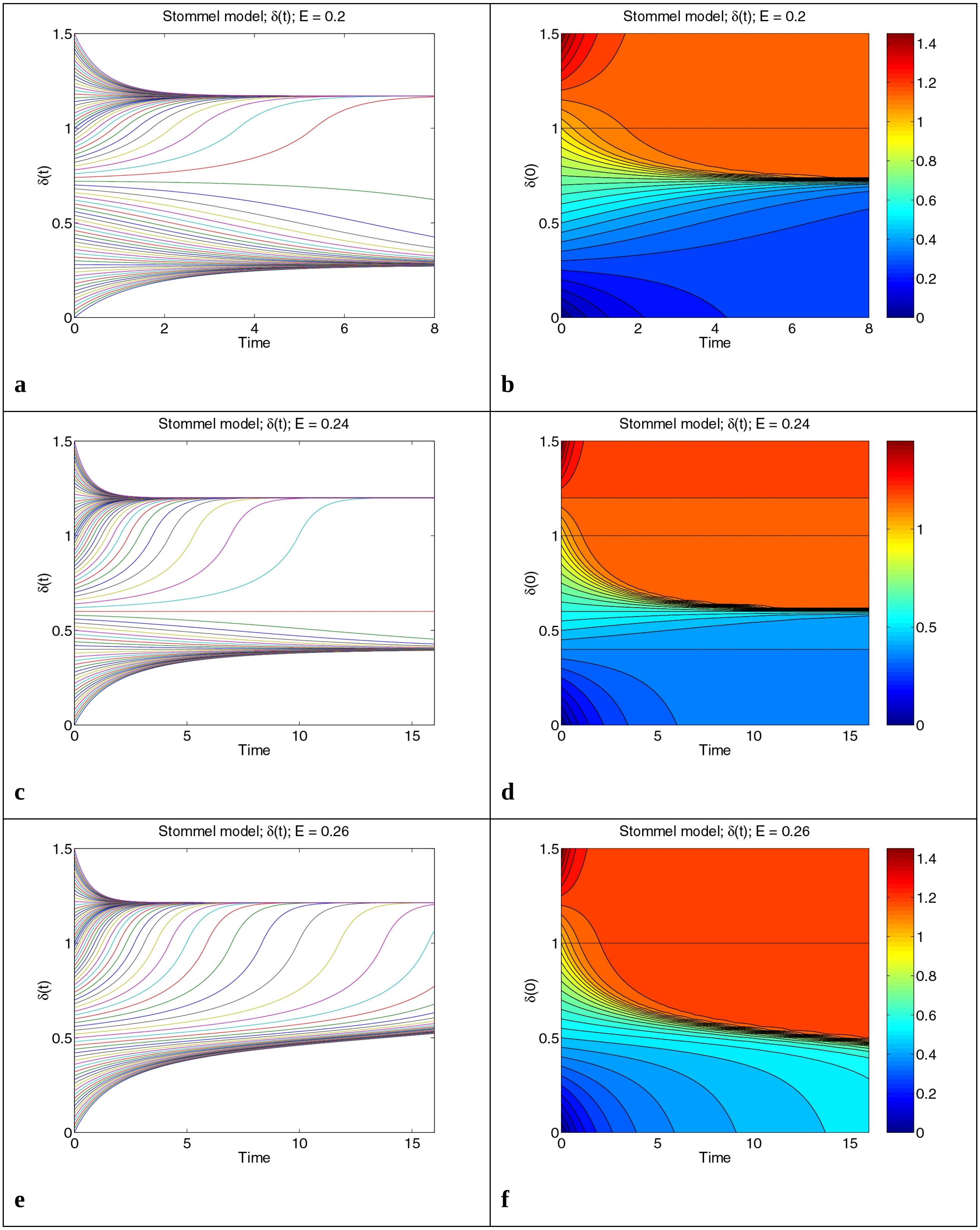

Figure 10.4. Solutions of 2-box model, as a function of dimensionless time and initial conditions. Left column: Time series of solutions. Right column: Contour plot of solutions.