12.306/12.806 - Computational Methods Exercise

Author / TA - Daniel Rothenberg (darothen@mit.edu)

Date - February, 2015

Introduction

Contemporary research in atmospheric physics and chemistry often relies on the use of both detailed observations from the field or laboratory as well as a wide array of computational tools of varying complexity. For instance, researchers who attempt to understand future climate change may use complex ``Earth system models'' which include many interlocking components which detail the dynamics and evolution of the atmosphere, ocean, ice sheets, and biosphere in three dimensions. On the other hand, some research may focus on detailed chemical transport of pollution due to local- and synoptic-scale meteorology at the spatial scale of a continent. And of course, there continues to be significant interest in detailed modeling of idealized chemical and physical sub-systems of the atmospheres - fogs, urban plumes, and even model reproductions of experimental setups are often invaluable for gleaning a deeper insight into things.

Why have models become ubiquitous in atmospheric physics and chemistry? There are a many potential reasons, but it's useful to focus for a moment on a few:

1. **We usually only get one realization of the real world**. Say you're a researcher studying the stratospheric aerosol layer, and you wish to understand how that layer is affected by volcanoes. There is really only one well-observed such case in the modern, satellite era - the eruption of Pinatubo in the early 1990's. Is one test case enough data to tease out the validity of your hypothesis on how volcanoes affect stratospheric aerosol?

2. **We can not always directly affect the system we wish to study**. It's probably not very easy to go out and cause a volcano like Pinatubo erupt when you're ready for it and have your observation system in place (it's probably not ethical, either). But with the right tools, maybe an ensemble of simulated eruptions could yield the insight you need to understand a particular volcanic effect on climate.

3. **We can't always observe things directly**. For instance, over the next decade many nations may choose to begin curbing emissions of carbon dioxide and other greenhouse gases from power plants and other stationary sources. Since emissions reductions are a contentious issue, there will likely need to be an independent check on *emissions inventories* - what is being emitted from where - for a variety of greenhouse gases. One (expensive) way to accomplish this might be to install new equipment to observe emissions at smokestack of every coal power plant. However, through inverse modeling, the same sort of information (with detailed uncertainty estimates) could be derived far more cheaply with the existing observation network.

In this class, we will study the basics of atmospheric chemistry and consider how the phenomena we examine have been incorporated into sophisticated tools for prediction and analysis. We do not assume any student has a detailed background in computational and numerical methods, let alone familiarity with numerical tools or scripting for scientific research. This **Computational Methods Exercise** is meant to be a guided introduction which will walk you through a simple observation-vs-modeling project and point out tips and tricks which might be useful during the coursework in 12.806 and beyond.

This project consists of several parts. First, you will grab data from an online source and use it to generate several plots. Second, you will build a multi-compartment model of the global carbon cycle and use the data you downloaded to both run and evalute the model. Finally, you will use your model and additional downloaded data to analyze future climate change scenarios with respect to the global carbon cycle.

In all parts of the exercise, try to rely first on built-in toolkits and packages in your analysis software or language of choice unless explicitly asked to implement a method. Often times, existing numerical packages are already more than sophisticated enough for general research problems, and will have both large user and experience bases which minimizes the risk of bugs. We haven't included any parts in the exercise that are designed to blow-up on you, since it's anticipated that this will happen while you practice coding and solving the problems. **Start early! Please contact the TA if you get stuck or need help!**

Part 1 | Contemporary Trends in Atmospheric CO_2

The National Oceanic Atmospheric Administration (NOAA) operates a global monitoring network for greenhouse gases. Critically, this network has measured and recorded carbon dioxide measurements at the Mauna Loa Observatory in Hawaii since the late 1950's; monthly mean data from these observations is freely available[1].

1. Download the monthly mean CO_2 data. Generate a figure which illustrates the contemporary trend in the monthly data. Additionally, compute yearly-mean data from this data and plot it on the same figure.

2. Generate polynomial fits (at least first and second order) and an exponential fit (of the form $y(t) = a + be^{-ct}$) to your analyzed yearly-mean CO_2. Do higher-order polynomials necessarily produce `better' fits to the observational data?

3. Use one of your polynomial fits to de-trend the monthly data, leaving the seasonal cycle. Comment on the structure of the seasonal cycle, noting changes over time. Is the cycle perfectly sinusoidal? If not, why? You may wish to average the seasonal cycle over each decade of data in the observations and plot them simultaneously, although other creative analyses could be pursued.

4. Often times, periodic or cyclical phenomena can be decomposed into a sum of harmonics represented by basic trigonometric functions, such as

$$y(t) \approx \sum_{k=1}^{n} A_k \sin \frac{2\pi k}{T_k}(t + \phi_k)$$

---

[1] ftp://aftp.cmdl.noaa.gov/products/trends/co2/co2_mm_mlo.txt

for some finite n. Model the seasonal cycle using at least one harmonic, and compare this result to the average seasonal cycle you get. You may want to consider what happens if you use all the data simultaneously, or attempt to fit the average cycle with your harmonics. Include estimated parameters for amplitude (A_k), period (T_k), and phase (ϕ_k) in your fits.

5. Finally, consider that we've decomposed the CO_2 record into two components - a long-term, slowly increasing part and a cyclical, seasonal anomaly. Model the CO_2 record as a linear combination of these two components, using your results form the previous section. Use the exponential form for the long-term trend a seasonal cycle based on at least the lowest two harmonics. Ultimately, you should have an equation - built piecewise - that looks something like

$$y(t) \approx a + be^{-ct} + \sum_{k=1}^{2} A_k \sin \frac{2\pi k}{T_k}(t + \phi_k)$$

Comment on differences between your modeled CO_2 curve and the Mauna Loa data. Is there a danger of *overfitting* the model we have built? Put another way, can you justify, using evidence we have discussed in class, the inclusion of each component in your model?

Part 2 | Compartmental Model of the Global Carbon Cycle

Please refer to Chapter 22, Section 22.2.2 - ``Compartmental Model of the Global Carbon Cycle''[2] of Seinfeld and Pandis which describes this model in more detail. Specifically, Figure 22.6 sketches an outline of the model and Table 22.1 (reproduced below) gives the model governing equations and parameter set.

2.1 Model Basics

The model introduced here approximates the exchange between several different `reservoirs' of carbon within the climate system:

1. atmosphere
2. warm surface ocean water
3. cool ocean surface water
4. deep ocean water
5. terrestrial biosphere
6. soil and detritus

It also considers a reservoir of carbon sequestered in the Earth's crust which can be unleashed via fossil fuel burning (M_7); G accounts for permanent changes in the terrestrial biosphere due to de/reforestation. The set of equations $\frac{dM_i}{dt}$ and $\frac{dG}{dt}$ represent a system of coupled ordinary differential equations which can be integrated together simultaneously. While you're welcome to attempt to derive an analytical solution to this system, we will proceed by building a numerical version of this model to approximate its time-dependent solution.

Before implementing the model, answer the following questions:

---

[2]Available on MIT's campus or off-campus.

Governing Equations	Parameters		
	Symbol	**Value**	**Units**
$\frac{dM_1}{dt} = -(k_{12}+k_{13})M_1 - k_{15}G\frac{M_1-\gamma}{M_1+\Gamma} + k_{21}M_2^{\beta_2}$	k_{12}	0.0931	yr^{-1}
$+k_{31}M_3^{\beta_3} + k_{51}M_5 + k_{61}M_6 + F_f(t) + F_d(t) - F_r(t)$	k_{13}	0.0311	yr^{-1}
$\frac{dM_2}{dt} = k_{12}M_1 - (k_{23}+k_{24})M_2 - k_{21}M_2^{\beta_2} + k_{42}M_4$	k_{15}	147	yr^{-1}
$\frac{dM_3}{dt} = k_{13}M_1 + k_{23}M_2 - k_{34}M_3 - k_{31}M_3^{\beta_3} + k_{43}M_4$	k_{21}	$58\times 730^{-\beta_2}$	$\text{PgC}^{(1-\beta_2)}\text{yr}^{-1}$
$\frac{dM_4}{dt} = k_{24}M_2 + k_{34}M_3 - (k_{42}+k_{43})M_4$	k_{23}	0.0781	yr^{-1}
$\frac{dM_5}{dt} = k_{15}G\frac{M_1-\gamma}{M_1+\Gamma} - (k_{51}+k_{56})M_5 - F_d(t) + F_r(t)$	k_{24}	0.0164	yr^{-1}
$\frac{dM_6}{dt} = k_{56}M_5 - k_{61}M_6$	k_{31}	$18\times 140^{-\beta_3}$	$\text{PgC}^{(1-\beta_3)}\text{yr}^{-1}$
$\frac{dM_7}{dt} = -F_f(t)$	k_{34}	0.714	yr^{-1}
$\frac{dG}{dt} = -\left[\frac{a_d F_d(t) - a_r F_r(t)}{M_5(t=0)}\right]$	k_{42}	0.00189	yr^{-1}
	k_{43}	0.00114	yr^{-1}
	k_{51}	0.0862	yr^{-1}
	k_{56}	0.0862	yr^{-1}
	k_{61}	0.0333	yr^{-1}
	β_2	9.4	
	β_3	10.2	
	γ	62.0	PgC
	Γ	198	PgC
	a_d	0.230	
	a_r	1.0	

Table 1: Governing Equations for Compartmental Model of Carbon Cycle

1. Using the estimates of the pre-industrial flux of carbon between each connecting reservoir (F_{ij}), estimate the residence time of a molecule of CO_2 in each reservoir. *Hint: the mean residence time is given as the ratio of its total abundance (in mass) to the total rate of removal from the reservoir (in mass per unit time).*

2. Why are two different surface ocean carbon reservoirs represented in the model?

3. Reservoir M_1 represents the atmospheric carbon burden. Assuming all of this carbon resides in gaseous carbon dioxide, derive an expression which allows you to convert the mass of carbon in the atmospheric reservoir to an equivalent mixing ratio of CO_2. If the pre-industrial atmospheric burden of carbon was 612 Pg, what was the atmospheric concentration of CO_2 *gas*? Note the difference between carbon and carbon dioxide here.

2.2 Model Implementation

We can now implement the model. Many systems of time-dependent, coupled ordinary differential equations which appear in the physical sciences can be represented as a simple initial value problem of the form

$$\dot{\mathbf{y}}(t) = f(t, \mathbf{y}(t)), \qquad \mathbf{y}(t_0) = \mathbf{y}_0 \tag{1}$$

where $\mathbf{y}(t)$ is a multi-component state vector representing the physical system, and $\dot{\mathbf{y}} = \frac{d\mathbf{y}}{dt}$ is the vector representing the time derivative of each component of the state vector.

The simplest way to integrate a coupled system of ODEs forward in time is to apply a finite-difference approximation to the right-hand side of Equation 1 (henceforth referred to as `RHS'),

$$\dot{\mathbf{y}}(t) \approx \frac{\mathbf{y}(t + \Delta t) - \mathbf{y}(t)}{\Delta t}$$

which can be re-arranged as

$$\mathbf{y}(t + \Delta t) \approx \mathbf{y}(t) + \Delta t \cdot \dot{\mathbf{y}}(t)$$
$$\approx \mathbf{y}(t) + \Delta t \cdot f(t, \dot{\mathbf{y}}(t)) \tag{2}$$

This yields ``Euler's Method'', a very simple way to march the system of ODEs forward in time. Starting from $\mathbf{y}_0$ and with a prescribed Δt, you can evaluate the RHS and apply it to the system's current state-vector to approximate to update the state vector at the new time. Since the state-vector at a future time is always generated from known, prior state-vectors, Euler's Method is called *explicit*. It's easy to code, but has undesirable properties (in terms of order and stability) which are beyond the focus of this exercise. *Implicit* schemes also exist, and trade-off enhanced stability (meaning you can use a larger Δt without fear of the numerical solution exploding) for the burden of solving a non-linear system. For *stiff* systems - systems where different components respond on very different timescales, such as ones typical in atmospheric chemistry - this trade-off is extremely favorable.

One way to improve the performance of explicit methods is to apply higher-order approximations to the derivative in the RHS equation. For instance, consider ``Heun's Rule'' (Equation 3), which essentially applies the trapezoidal rule to produce the finite difference equation:

$$\tilde{\mathbf{y}}(t) = \mathbf{y}(t) + \Delta t \cdot f(t, \mathbf{y})$$
$$\mathbf{y}(t + \Delta t) = \mathbf{y}(t) + \frac{\Delta t}{2} \cdot [f(t, \mathbf{y}(t)) + f(t + \Delta t, \tilde{\mathbf{y}}(t))] \tag{3}$$

One additional multi-step method which is commonly used is the ``Classical Runge-Kutta Method'', sometimes referred to as RK4 (Equation 4):

$$\mathbf{y}(t + \Delta t) = \mathbf{y}(t) + \frac{\Delta t}{6}(k_1 + 2k_2 + 2k_3 + k_4) \tag{4}$$
$$k_1 = f(t, \mathbf{y}(t))$$
$$k_2 = f(t + \frac{\Delta t}{2}, \mathbf{y}(t) + \frac{1}{2}k_1\Delta t)$$
$$k_3 = f(t + \frac{\Delta t}{2}, \mathbf{y}(t) + \frac{1}{2}k_2\Delta t)$$
$$k_4 = f(t + \Delta t, \mathbf{y}(t) + k_3\Delta t)$$

In modern practice, these simple methods have been superseded by more-sophisticated techniques which employ adaptive timesteps and other tricks to (a) reduce the computational complexity of integrating the system, (b) preserve the accuracy of the solution, and (c) derive useful additional information such as roots, zeros, or sensitivities. Such schemes include CVODE[3] and LSOD(E/A)[4]. These can be trivially deployed in most scientific programming projects, since they all extend a simple interface, requiring only the definition of the RHS equation and the initial conditions (some methods are optimized in cases where the ODE has an analytical Jacobian which the user can supply).

1. Assume the emissions from re-/deforestation and fossil fuel burning over the time-periods 1850 to 1990 can be modeled by the equations

$$
\begin{aligned}
F_r(t) &= 0 \\
F_d(t) &= 0.3 + 0.01t, \ (t = 0 \text{ for } 1850) \\
F_f(t) &= \begin{cases} 0.014t & \text{from 1850 to 1950 } (t = 0 \text{ for } 1850) \\ 1.4 + (4.6/40)t & \text{from 1950 to 1990 } (t = 0 \text{ for } 1950) \end{cases}
\end{aligned}
$$

   Implement Euler's, Heun's, and the Classical Runge-Kutta scheme and integrate the compartmental model from 1850 to 1990. Additionally, identify a numerical toolkit for your programming langauge of choice and use a high-performance solver from that toolkit as a reference solution. Use a timestep of 1 year and the initial conditions from Table 22.2 of Seinfeld and Pandis. Compare the performance of each method, and comment on the difficulty of implementation (or application of the toolkit). Be sure to research and writeup a few sentences on the details of that solver - you shouldn't *ever* use a numerical method as a blackbox and put blind faith that it's the right tool for the job!

You only ever need to plot your model's output for atmospheric carbon (M_1), and please use the conversion factor from the previous part to express your answer in parts per million.

2. Run your simulations again with each method, but allow your simulation to run until the concentration of atmospheric CO_2 doubles from its pre-industrial value. You can assume a ``business-as-usual'' scenario where the emissions increase at the same rate beyond 1990 (i.e. 0.01 PgC yr^{-1} and 4.6/40 PgC yr^{-1} for deforestation and fossil fuel burning, respectively) When does this doubling occur?

3. Suppose that in 2050, the world commits to decreasing its carbon dioxide emissions (both from fossil fuels and deforestation) such that the peak value following 2050 decays with an e-folding of 20 years. Write the mathematical form of $F_f(t)$ and $F_d(t)$ in this case, extending the definition from the first part of this sub-problem. When will the atmospheric CO_2 concentration peak and at what value? When will it return (if ever) to pre-industrial levels?

Part 3 | Model Application to Observed Atmospheric CO_2 Concentrations

Now, we'll deploy the compartmental model to simulate the observed and future evolution of atmospheric CO_2. Note that the sub-problems here are not numbered; please refer to the complete text below for instructions on how to run your simulations.

---

[3] http://robotics.stanford.edu/~scohen/cvode.paper.pdf
[4] http://people.sc.fsu.edu/~jburkardt/f77_src/odepack/odepack.html

3.1 20th Century Emissions

The Carbon Dioxide Information Analysis Center at Oak Ridge National Laboratory (CDIAC) reports global carbon emissions inventories from various sources. As in the past sub-problem, assume that emissions from re-forestation are negligible at all times. Then, use the land-use[5] and fossil-fuel[6] emissions data from the CDIAC to run a simulation from 1850-1950, to spin-up your model to pre-Mauna Loa observations. Use the initial conditions from the Table 22.2 of Seinfeld and Pandis.

To accomplish this, you'll need to consider how to incorporate yearly, tabular emissions data into your model. That's easier said than done! There are three ways you could do this:

- assume the emissions rate is constant for each year
- linearly interpolate the emissions rate to produce a time-varying emissions function
- filter the data and apply a higher-order local or global interpolation method

The second approach is the most common, since it assumes little knowledge about between-year values. The easiest way to implement the emissions function using the data would be to independently read in the data as some sort of table, and write a function which accepts time t as an argument, performs the interpolation on that data, and returns the emissions value. Before you run your spin-up simulation, prepare a plot of the emissions from deforestation and fossil-fuel burning over time. Plot the emissions curves for each of the methods suggested above. You can choose any technique for the final method - one possibility might be to apply a 1-2-1 filter (for each value, compute the average of twice that value plus each of its neighbors) and then a spline or piecewise-polynomial interpolating function (e.g. `pchip` in MATLAB or a `PchipInterpolator` in SciPy). Lean on whatever numerical packages are available for your coding software, but be sure to look into how the interpolator works and explain it in your commentary.

Another thing to note: unless provided in a universal, well-documented (with respect to metadata) format such as netCDF or HDF, you often need to pre-process the data you download. In general, it's easier to tweak the data in a text editor and spreadsheet software to conform to some standard (such as CSV or fixed-width tabular format) and use the input/output functions built into your programming toolkit than to write your own input routine from scratch. That's particularly true in this exercise, since you do not need to do any significant analysis on the emissions data once you have read it in. For this project, I'd strongly recommend importing the downloaded data into Excel, deleting all the meta-data and superfluous columns, and writing out an intermediate dataset which you'll read into MATLAB, Python, Fortran, etc.

Use the second option (linear interpolation) for your spin-up simulation, and save the output for the year 1950. We'll use that to re-initialize the model for the next part.

From the spun-up state, we'll now try to model the observed increase in atmospheric CO_2 over the past 50 years. Each year, the CDIAC compiles a Global Carbon Budget[7], which breaks down emissions by source and further by region. Using their fossil-fuel emissions and land-use change emissions (which we'll assume to be deforestation), run your model from 1960 to 2013. Plot the modeled atmospheric CO_2 concentrations against the yearly-averaged data from the very first part of this exercise. Comment on any differences, including suggestions for why they might exist in the first place.

One natural thing to do with the model and dataset we've developed in this exercise is to use the two in synergy. Suppose we had very high confidence in the accuracy of the observed CO_2 record and our model's

---

[5] http://cdiac.ornl.gov/trends/landuse/houghton/houghton.html
[6] http://cdiac.ornl.gov/trends/emis/tre_glob_2010.html
[7] http://cdiac.ornl.gov/GCP/carbonbudget/2014/

ability to matriculate emissions into carbon dioxide observations, but we had large uncertainty in our emissions inventory. Propose a methodology for using the model and your observational data to constrain the emissions (*hint: you may wish to research 'inverse modeling' - and we'll actually tackle just this problem later in the course!*).

3.2 Looking Towards the Future

Finally, let's use our model to investigate future climate change scenarios. In parallel with the Intergovernmental Panel on Climate Change's Assessment Report schedule, modeling groups from across the world collaborate on a series of experiments aimed at helping elucidate the potential global warming due to anthropogenic emissions of greenhouse gases. Since we do not have detailed foreknowledge of how society's emissions will unfold over the next few hundred yeras, a series of `scenarios' - in their most recent iteration, known as ``Representative Concentration Pathways" or ``RCP" for short - are constructed and used as input to global climate models.

Let's see what information about future climate change we can glean from our carbon cycle model. From the RCP data portal[8], download the historical 20th century emissions data (**20c3m**). Using the initial conditions again from Table 22.2 of Seinfeld and Pandis, simulate the change in atmospheric CO_2 between 1850 and 2005 using the historical data. Compare with your previous analysis and observations (preferably on the same plot). Save the result for year 2005.

Now, download the emissions data for each of the scenarios **RCP3-PD**, **RCP4.5**, **RCP6**, and **RCP8.5**. Using the spun-up model state from 2005, simulate the emissions scenario for each RCP, and produce a plot which shows the change in atmospheric CO_2 between 1850-2500 for each scenario (you'll want to extend your previous figure). Using the features of your plotting software, annotate this figure as thoroughly as possible. Some things you may wish to consider:

- different colors or line styles for the various RCP's
- something visual marker denoting where ``historical" emissions end and RCPs begin
- informative labels of the x- and y-axes.
- a companion plot which translates the ΔCO_2 into units of equivalent radiative forcing

Comment on your results. What are the implications of these various scenarios on global climate change? Assuming an equilibrium climate sensitivity of 3 degrees C per doubling of atmospheric CO_2, how might you expect global average temperature to evolve in each of these scenarios? Recall that our model is a simplified representation of the global carbon cycle. What factors in a warming climate might feedback on the global carbon cycle which we do not account for in our model, and how might that affect your interpretation of the emissions scenario results?

---

[8]http://www.pik-potsdam.de/~mmalte/rcps/