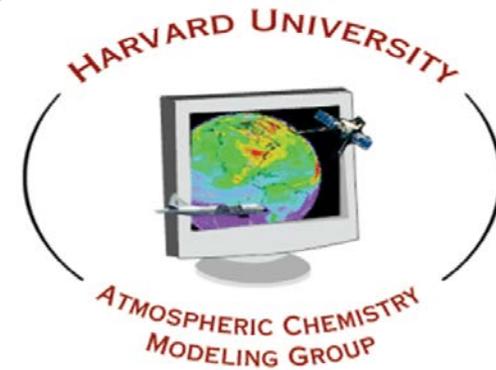


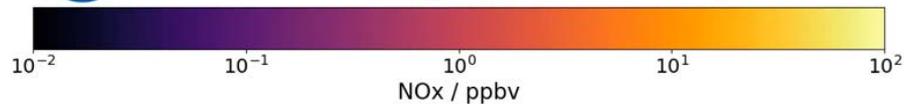
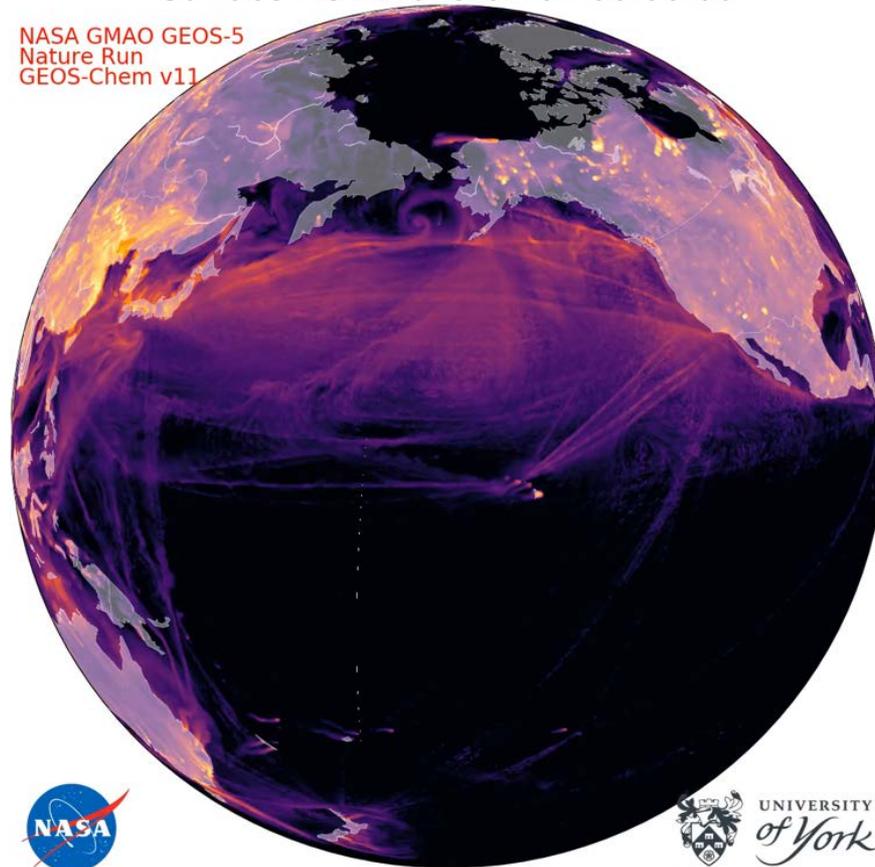
Definition, objectives, and types of models

Daniel Jacob, Harvard University



Surface NO_x: 2013-07-01 00:00:00

NASA GMAO GEOS-5
Nature Run
GEOS-Chem v11

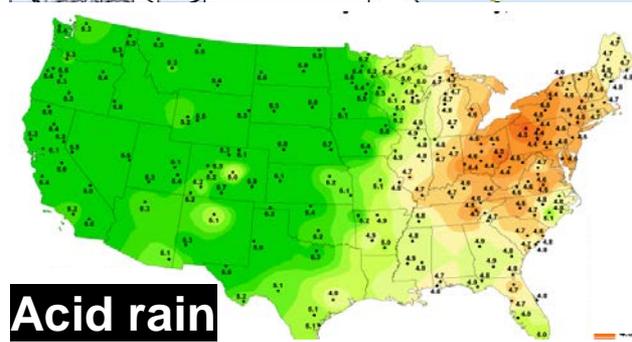
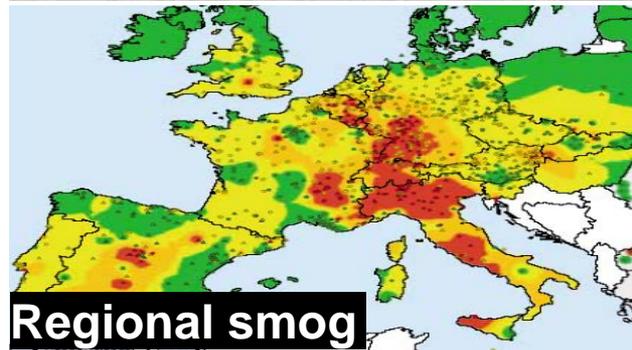


Atmospheric chemists are interested in a wide range of issues

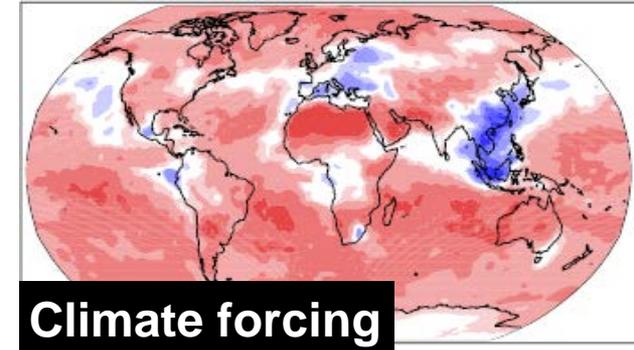
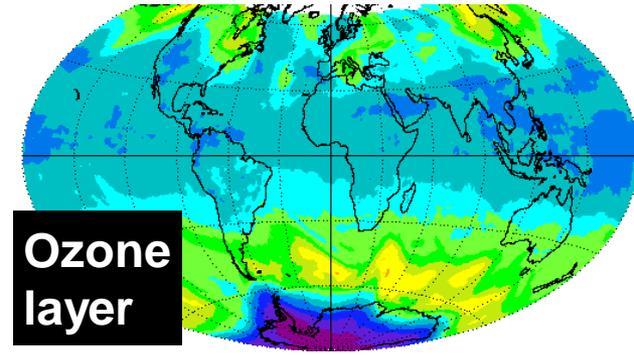
LOCAL



REGIONAL



GLOBAL



To address these issues we need models:

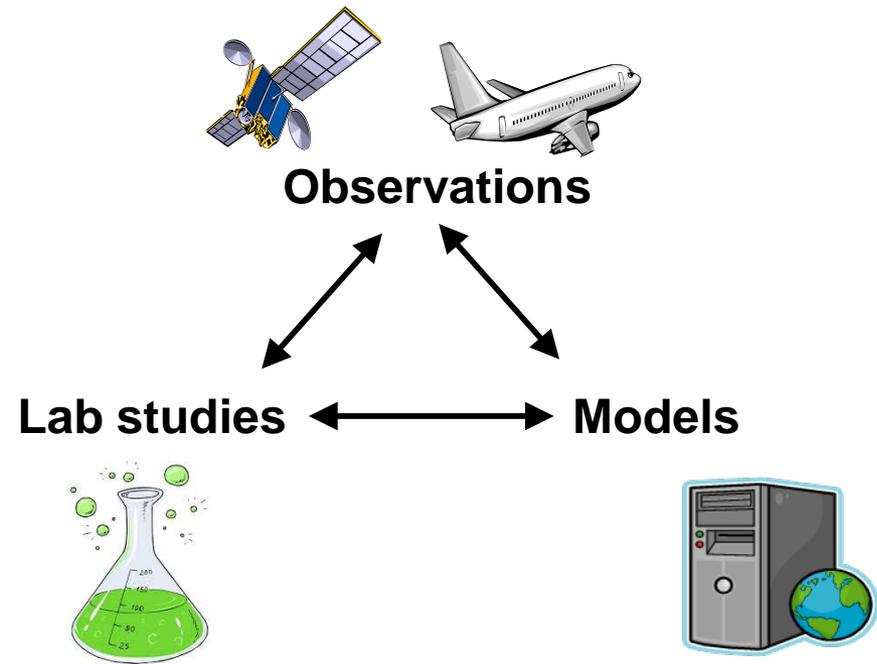
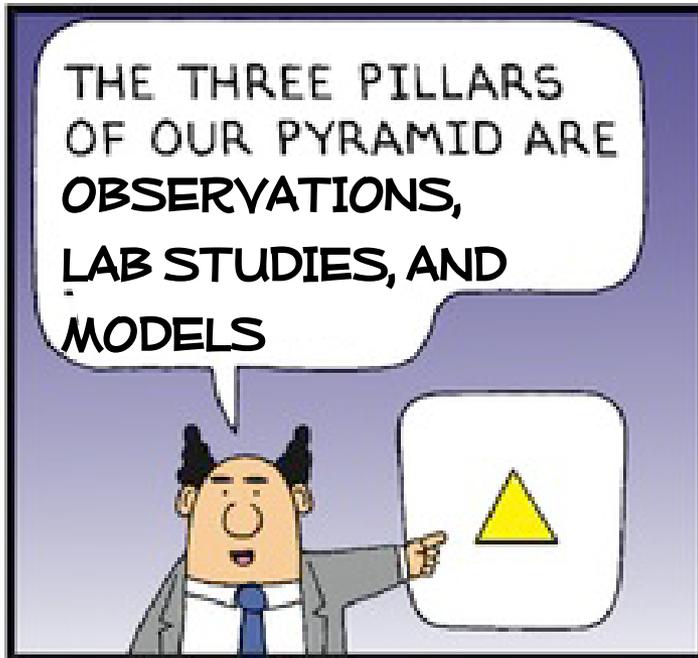
simplified representations of complex system amenable to analysis

- Minimize number of variables
- Ignore minor processes
- Draw on empirical relationships



- Interpret observations
- Gain understanding of processes
- Make future projections

“All models are wrong, but some are useful” (G. Box, 1976)

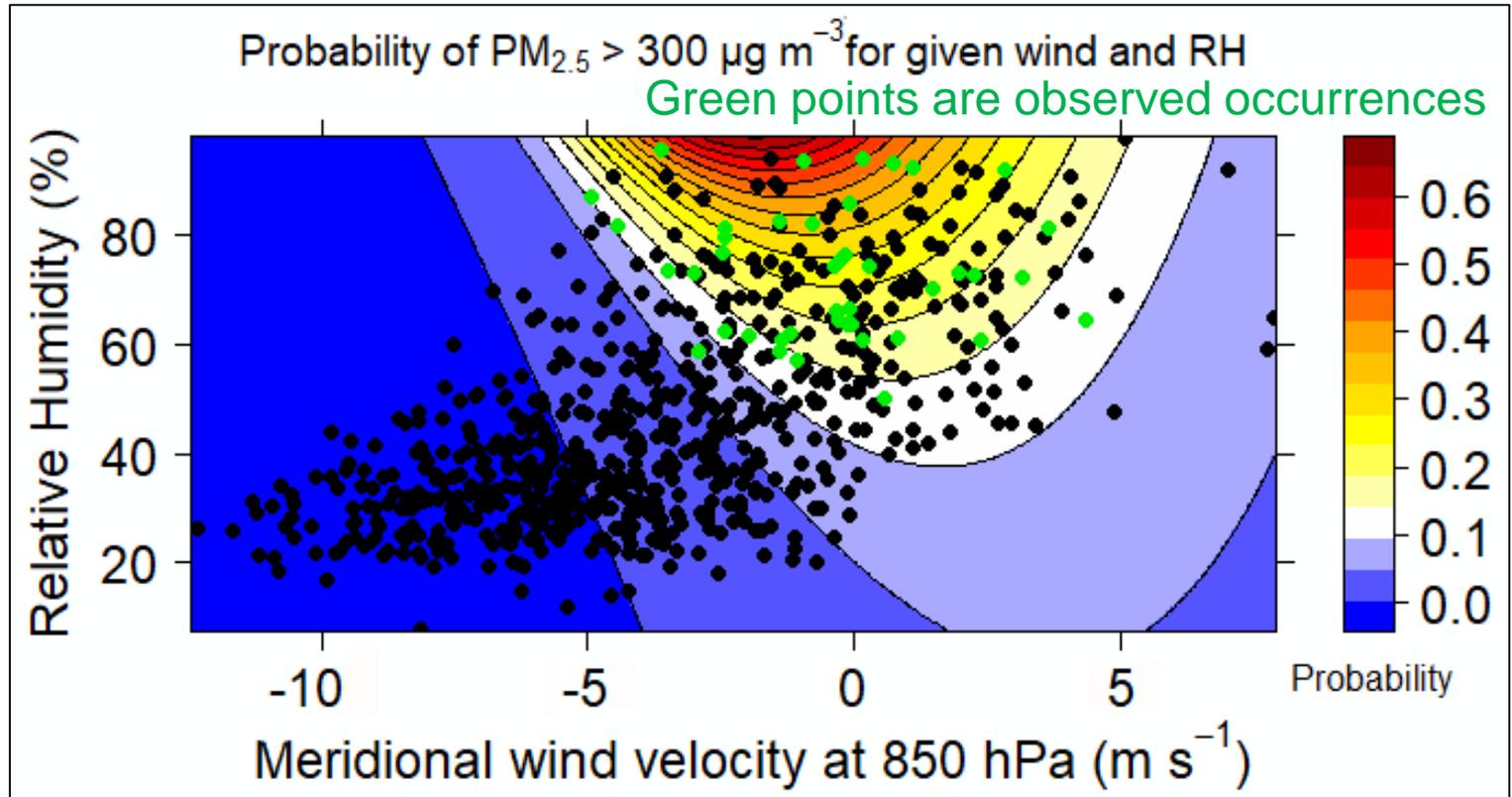


What kind of models?

- **Physical models:** simplify the physical equations describing the system
 - The most fundamental quantitative approach
- **Heuristic models:** draw mental inferences based on knowledge/experience
 - Your brain is a powerful computer. Use it!
- **Statistical models:** use empirical relationships to make projections
 - Need physical basis, but can be very powerful

Statistical model of extreme winter haze events in Beijing

Point process fit of DJF 2009-2017 daily $\text{PM}_{2.5}$ to wind velocity and relative humidity



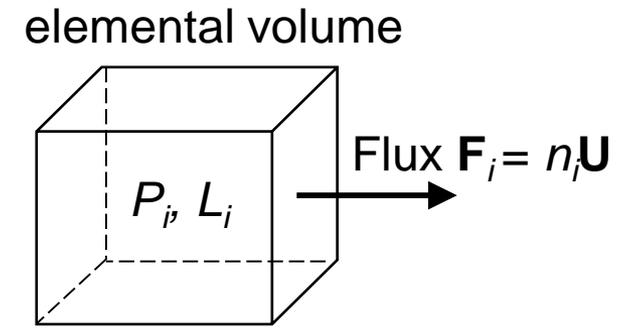
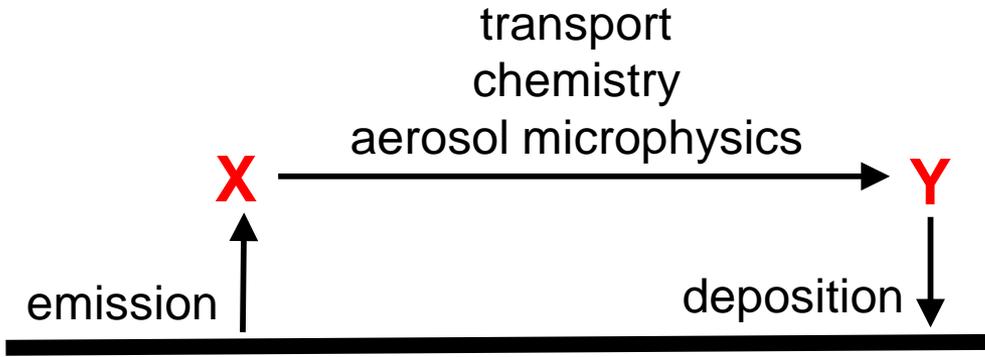
Expect lower RH over China in future climate
➔ less frequent extreme air pollution events



Drew Pendergrass

Pendergrass et al., in prep.

Physical modeling by solution of continuity equation



Solve continuity equations for number densities $\mathbf{n} = (n_1, \dots, n_K)$ of ensemble of K species:

Eulerian flux form

$$\frac{\partial n_i}{\partial t} = -\nabla \cdot (n_i \mathbf{U}) + P_i(\mathbf{n}) - L_i(\mathbf{n})$$

local trend in number density transport (flux divergence) emissions, deposition, chemical and aerosol processes

Eulerian advective form in terms of mixing ratios $\mathbf{C} = (C_1, \dots, C_K)$:

$$\frac{\partial C_i}{\partial t} = -\mathbf{U} \cdot \nabla C_i + P_i(\mathbf{C}) - L_i(\mathbf{C})$$

Lagrangian form (moving frame of reference):

$$\frac{dC_i}{dt} = P_i(\mathbf{C}) - L_i(\mathbf{C}) \quad \text{where} \quad \frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{U} \cdot \nabla$$

Break down dimensionality of continuity equation by operator splitting

Solve for transport and local terms separately over time steps Δt

$$\frac{\partial C_i}{\partial t} = -\mathbf{U} \cdot \nabla C_i + P_i(\mathbf{C}) - L_i(\mathbf{C})$$



Advection (and other transport):

Chemistry (and other local processes)

$$\frac{\partial C_i}{\partial t} = -\mathbf{U} \cdot \nabla C_i \quad (\text{Eulerian})$$

$$\frac{dC_i}{dt} = P_i(\mathbf{C}) - L_i(\mathbf{C})$$

$$\frac{d\mathbf{x}}{dt} = \mathbf{U}(\mathbf{x}, t) \quad (\text{Lagrangian})$$

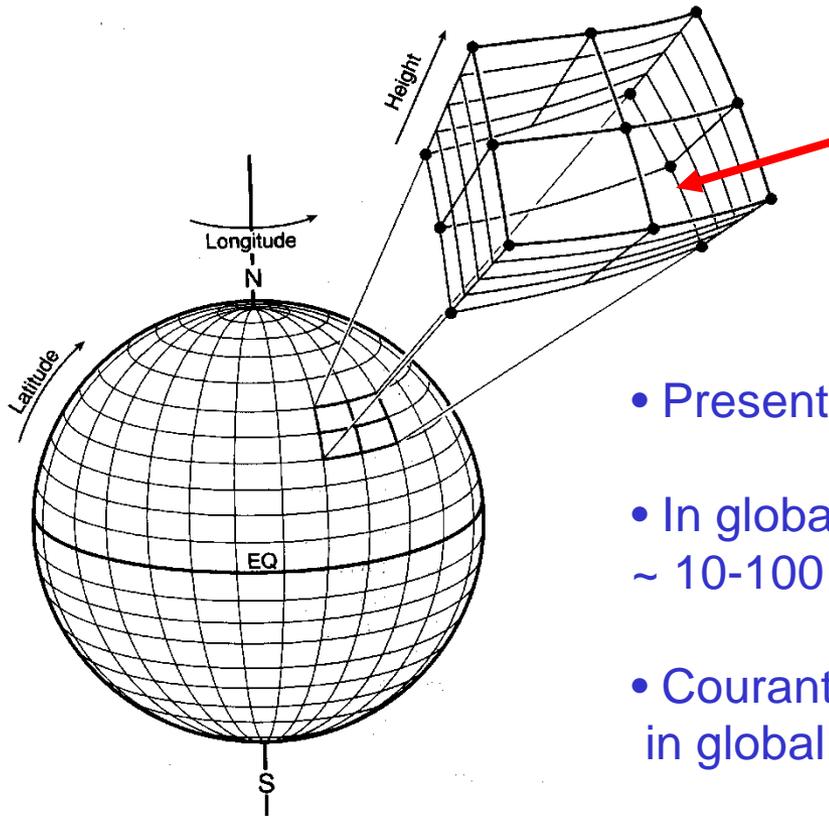


Advection equations:
PDEs with no coupling
between species

Chemical equations:
 K -dimensional system of ODEs

Eulerian models partition atmospheric domain into gridboxes

This discretizes the continuity equation in space



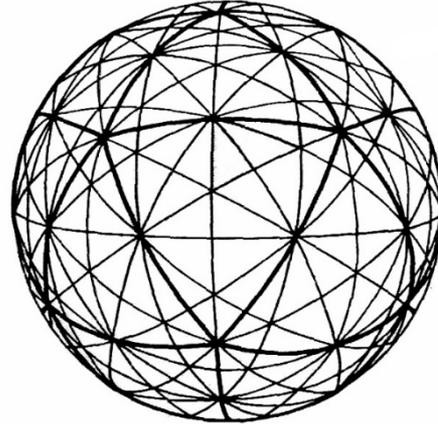
Solve continuity equation for individual gridboxes

- Present computational limit $\sim 10^8$ gridboxes
- In global models, this implies a grid resolution Δx of ~ 10 - 100 km in horizontal and 0.1 - 1 km in vertical
- Courant number limitation $u \Delta t / \Delta x \leq 1$; in global models, $\Delta t \sim 10^2$ - 10^3 s

Eulerian models often use equal-area or zoomed grids

Equal-area grids: avoid singularities at poles

icosahedral triangular



cubed-sphere

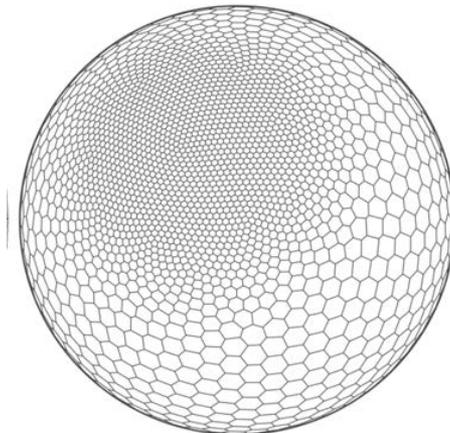


Zoomed grids: increase resolution where you need it (or when, in an *adaptive* grid)

nested

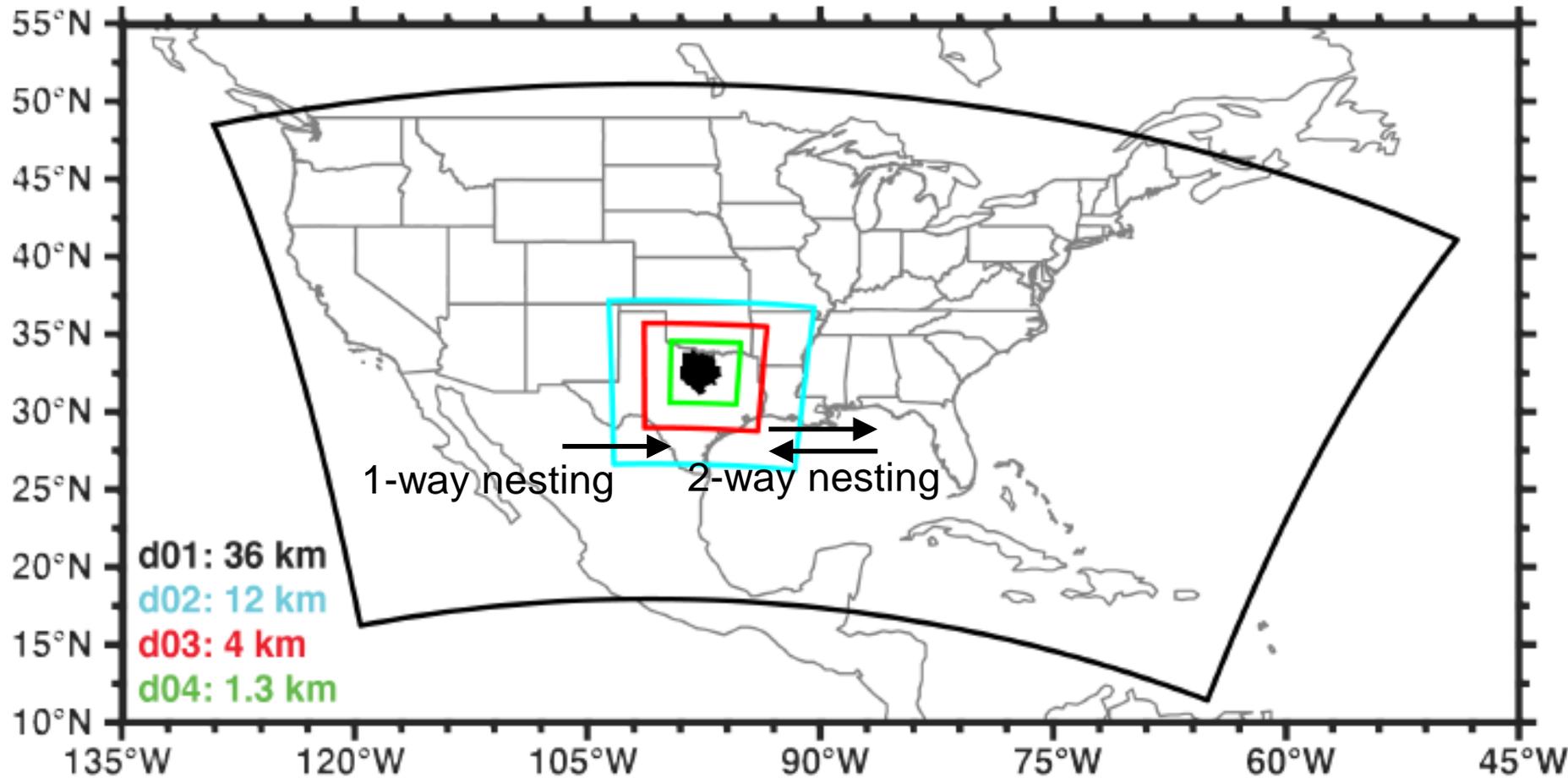


stretched

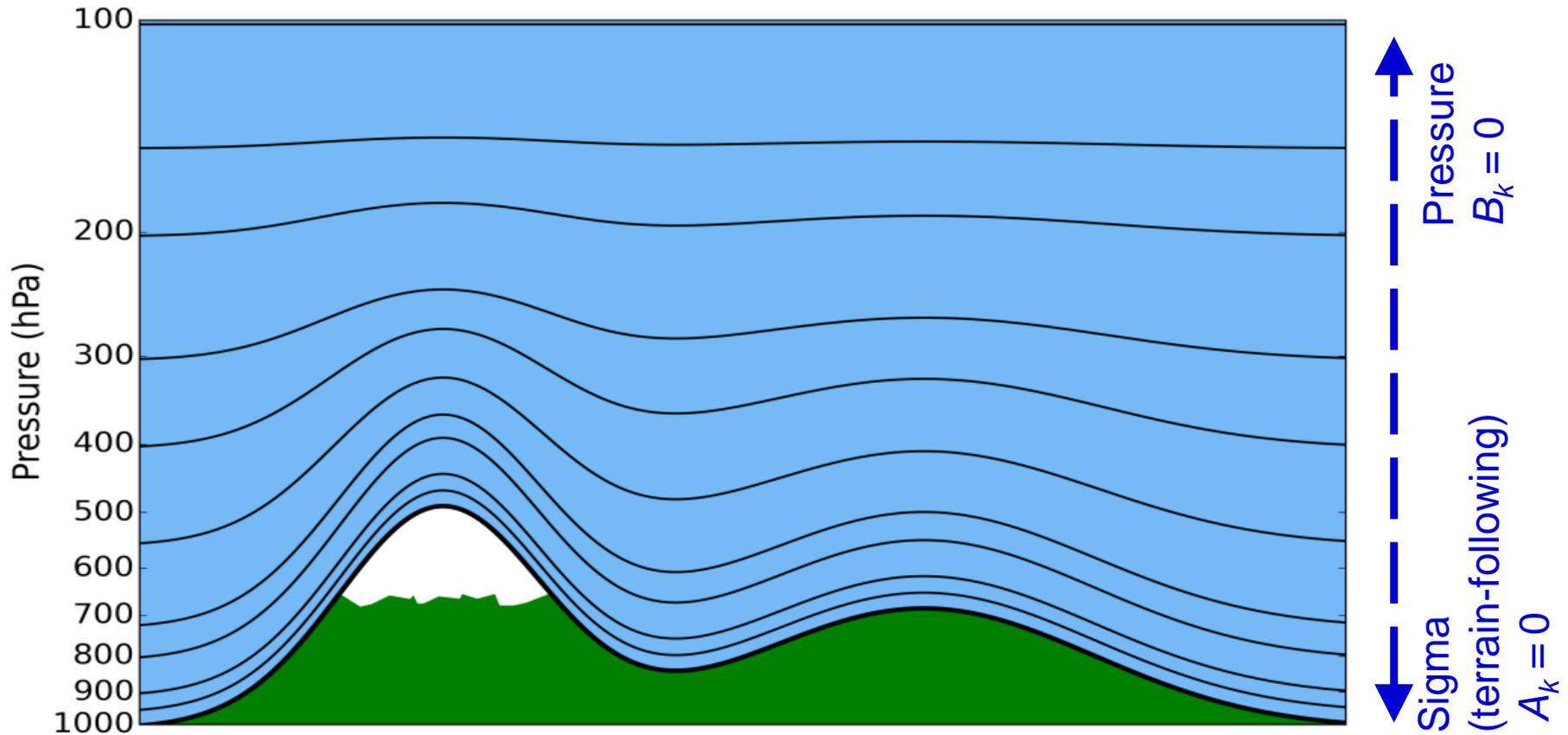


Regional models: limited domain, boundary conditions at edges

WRF domain with successive nests



Hybrid sigma-pressure vertical coordinate system



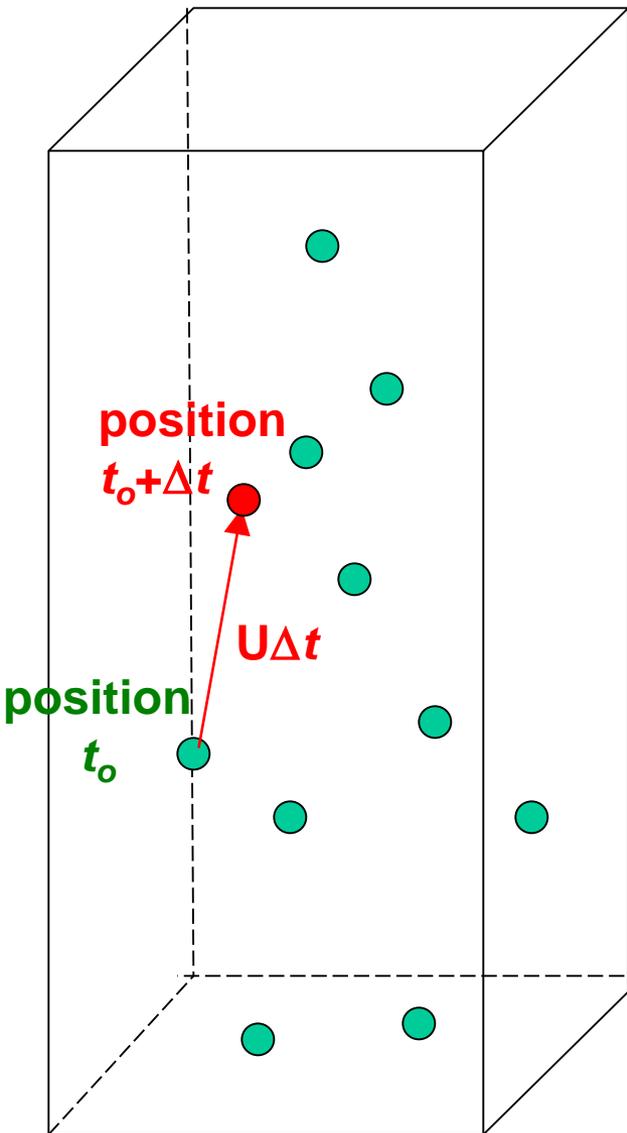
Midpoint pressure at level k :

$$p_k = A_k p_0 + B_k p_s$$

coefficients

↑ sea level ↑ surface

Lagrangian models track points in model domain (no grid)



- Transport large number of points with trajectories from input meteorological data base (\mathbf{U}) over time steps Δt
- Points have mixing ratio or mass but no volume
- Determine local concentrations in a given volume by the statistics of points within that volume or by interpolation

PROS over Eulerian models:

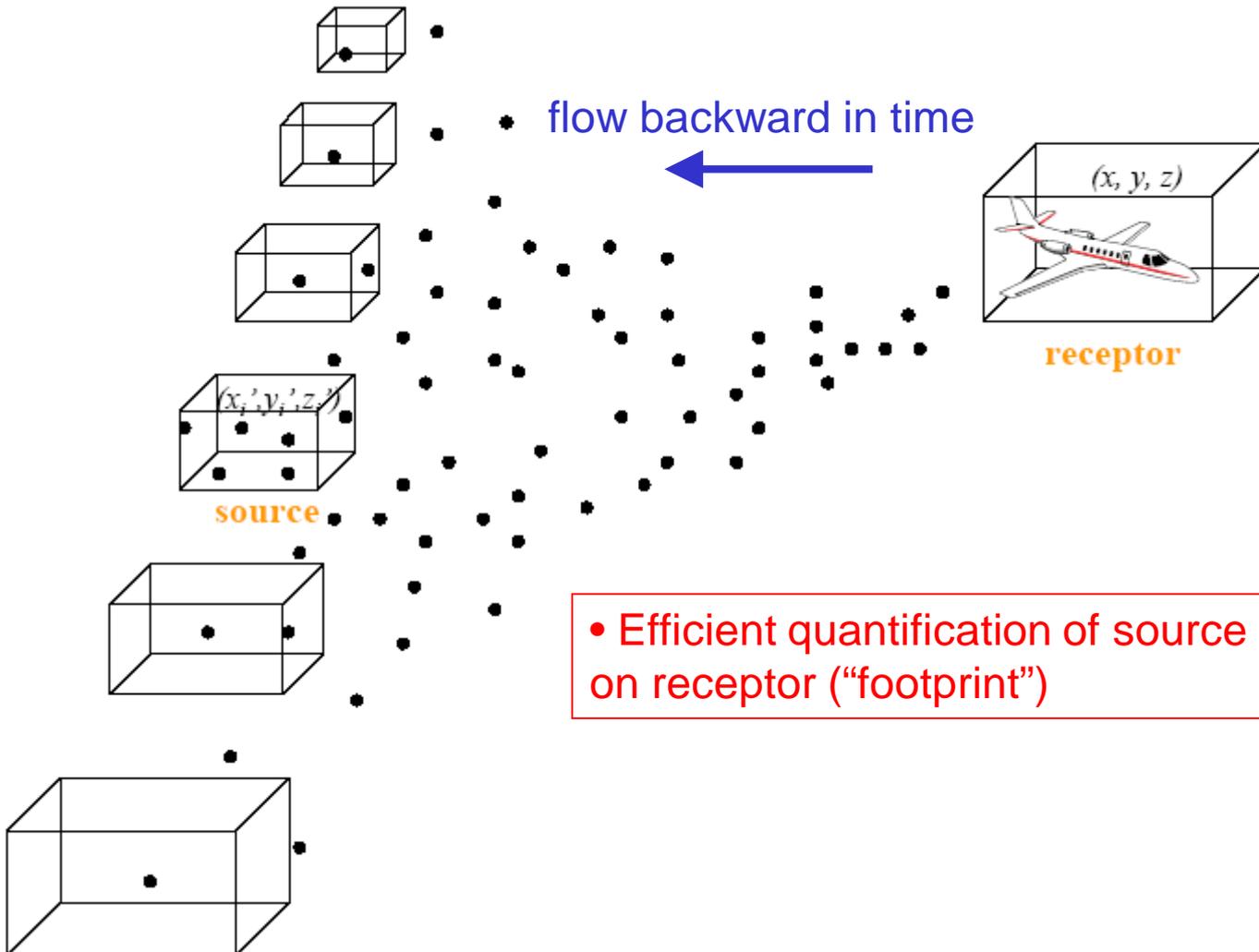
- stable for any wind speed
- no error from spatial averaging
- easy to parallelize
- easily track air parcel histories
- efficient for receptor-oriented problems

CONS:

- need very large # points for statistics
- inhomogeneous representation of domain
- individual trajectories do not mix
- cannot do nonlinear chemistry
- cannot be conducted on-line with meteorology

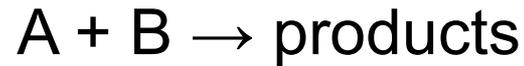
Lagrangian receptor-oriented modeling

Run Lagrangian model backward from receptor location,
with points released at receptor location only



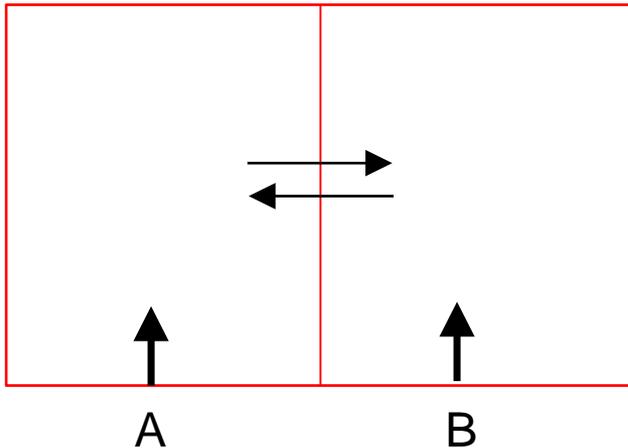
Representing non-linear chemistry

Consider two chemicals A and B emitted in different locations, and reacting by



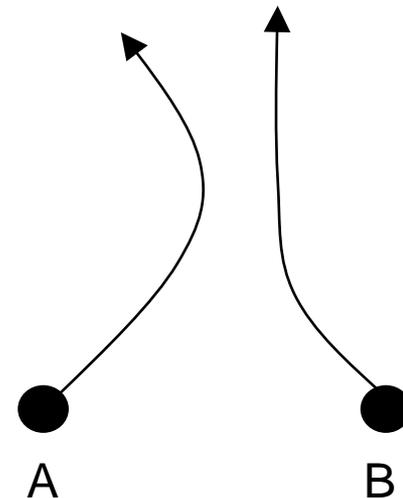
Eulerian model

gridboxes



A and B react following the mixing of gridboxes

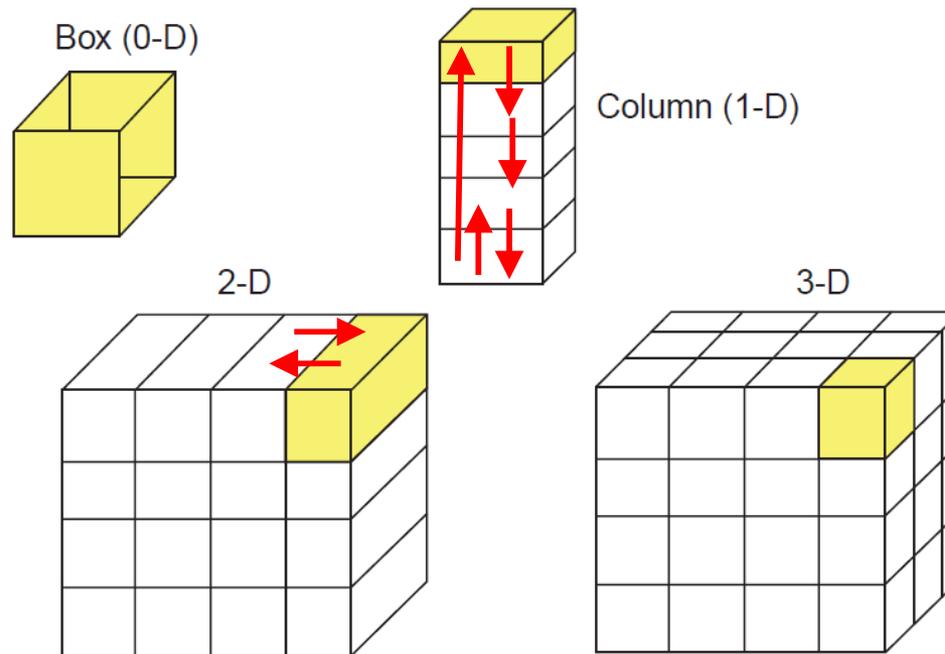
Lagrangian model



A and B never react

Reducing model dimension

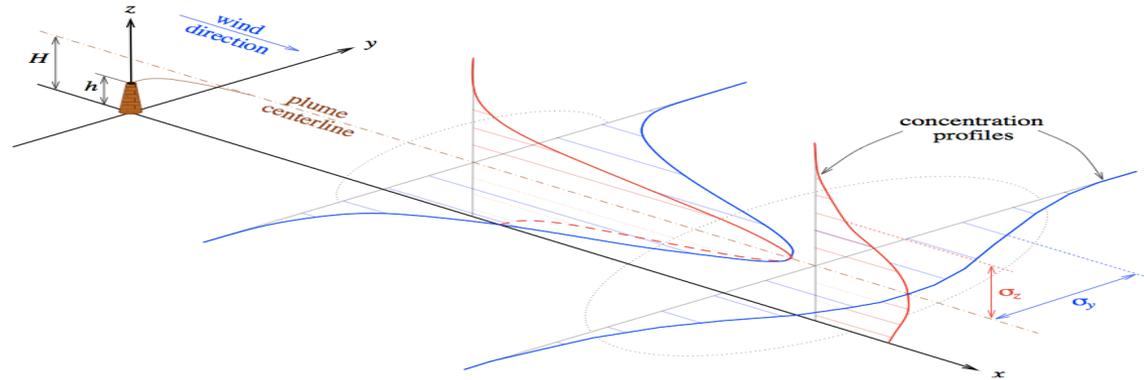
...because a model shouldn't be more complicated than it needs to be



- 2-D models used in stratosphere where zonal gradients are weak
 - Turbulent diffusion parameterization of eddy meridional transport
- 1-D models used for boundary layer mixing when horizontal gradients are weak
 - Turbulent diffusion and non-local (convective) transport parameterizations
- 0-D (box) models when flux divergence can be neglected in continuity equation
 - Chemical steady-state models for short-lived species
 - Global box models for long-lived species
 - Chemical mechanism diagnostic studies

Gaussian plume modeling of point source dispersion

Point source
with emission rate q



Transport in cross-wind direction is parameterized as diffusive process:

$$\frac{\partial C}{\partial t} = -u \frac{\partial C}{\partial x} + K_y \frac{\partial^2 C}{\partial y^2} + K_z \frac{\partial^2 C}{\partial z^2} \quad \text{for steady wind, inert plume}$$

Turbulent diffusion coefficients

Steady state solution with suitable boundary conditions:

$$C(x, y, z) = \frac{q}{4\pi(K_y K_z)^{1/2} x} \exp\left[-\frac{u}{4x} \left(\frac{y^2}{K_y} + \frac{z^2}{K_z}\right)\right]$$

On-line and off-line approaches to chemical modeling

On-line: coupled to dynamics

GCM conservation equations:
air mass: $\partial \rho_a / \partial t = \dots$
momentum: $\partial \mathbf{u} / \partial t = \dots$
heat: $\partial \theta / \partial t = \dots$
water: $\partial q / \partial t = \dots$
chemicals: $\partial C_i / \partial t = \dots$

PROs of off-line vs on-line approach:

- computational cost
- simplicity
- stability (no chaos)
- compute sensitivities back in time

CONs:

- no fast chemical-dynamics coupling
- need for meteorological archive
- transport errors

**Chemical data assimilation, forecasts
best done on-line**

Off-line: decoupled from dynamics

GCM conservation equations:
air mass: $\partial \rho_a / \partial t = \dots$
momentum: $\partial \mathbf{u} / \partial t = \dots$
heat: $\partial \theta / \partial t = \dots$
water: $\partial q / \partial t = \dots$

meteorological archive
(averaging time ~ hours)

Chemical transport model:
 $\partial C_i / \partial t = \dots$

**Chemical sensitivity studies
may best be done off-line**

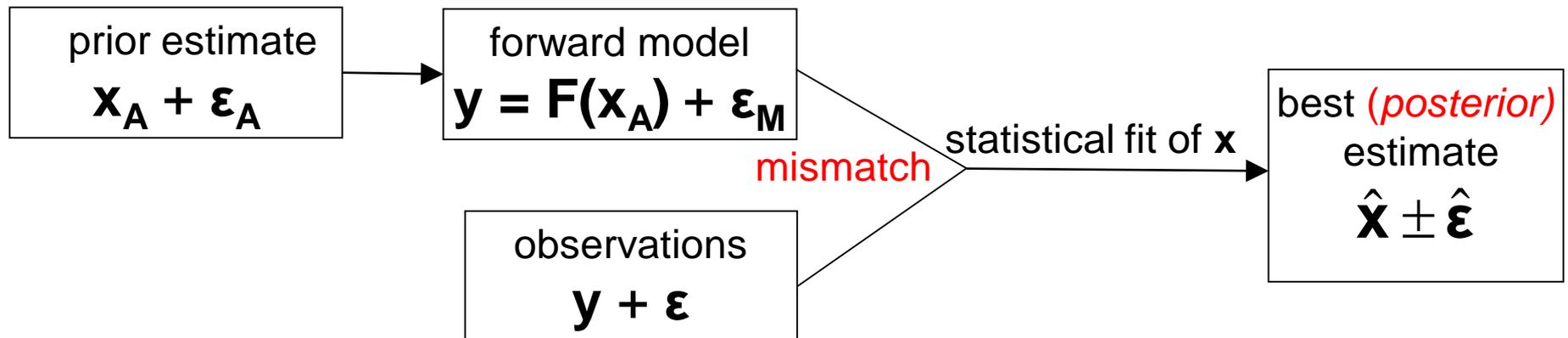


Inverse modeling and data assimilation

Quantify selected variables driving a physical system (*state vector \mathbf{x} , dim n*) by using:

- the observable manifestations of the system (*observation vector \mathbf{y} , dim m*)
- a physical model $\mathbf{y} = \mathbf{F}(\mathbf{x})$ (*forward model*)
- a *prior estimate \mathbf{x}_A* before the observations have been made
- Statistics for the errors $\boldsymbol{\varepsilon}$ in the different components of the problem

General approach:



Bayes' theorem: general basis for optimal estimation

$P(\mathbf{x})$ = probability density function (pdf) of \mathbf{x}

$P(\mathbf{x}, \mathbf{y})$ = pdf of (\mathbf{x}, \mathbf{y})

$P(\mathbf{y} | \mathbf{x})$ = pdf of \mathbf{y} given \mathbf{x}

$$P(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} \begin{cases} = P(\mathbf{x}) d\mathbf{x} P(\mathbf{y} | \mathbf{x}) d\mathbf{y} \\ = P(\mathbf{y}) d\mathbf{y} P(\mathbf{x} | \mathbf{y}) d\mathbf{x} \end{cases}$$

$$\Rightarrow \underbrace{P(\mathbf{x} | \mathbf{y})}_{\text{posterior pdf}} = \frac{\underbrace{P(\mathbf{y} | \mathbf{x})}_{\text{observation pdf}} \underbrace{P(\mathbf{x})}_{\text{prior pdf}}}{\underbrace{P(\mathbf{y})}_{\text{normalizing factor (unimportant)}}} \quad \text{Bayes' theorem}$$

Optimal estimate solution for \mathbf{x} given \mathbf{y} is $\max[P(\mathbf{x} | \mathbf{y})]$

$$\Rightarrow \text{solve for } \nabla_{\mathbf{x}} P(\mathbf{x} | \mathbf{y}) = \mathbf{0}$$

Solution of inverse problem minimizes cost function $J(\mathbf{x})$

corresponds to finding maximum of posterior pdf $P(\mathbf{x} | \mathbf{y})$

Error covariance matrices

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_A)^T \mathbf{S}_A^{-1} (\mathbf{x} - \mathbf{x}_A) + (\mathbf{y} - \mathbf{F}(\mathbf{x}))^T \mathbf{S}_o^{-1} (\mathbf{y} - \mathbf{F}(\mathbf{x}))$$

Prior terms

observation terms

$$-\ln P(\mathbf{x} | \mathbf{y}) = -\ln P(\mathbf{x})$$

$$-\ln P(\mathbf{y} | \mathbf{x})$$

Model adjoint

If a model can be linearized as a product of matrices,

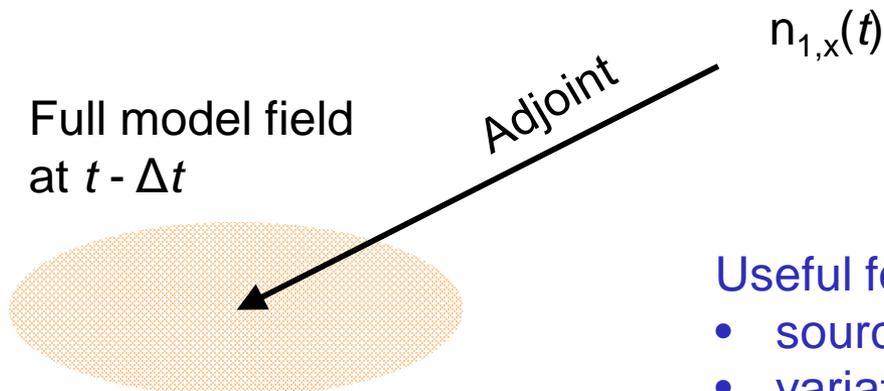
$$\mathbf{n}(t) = \mathbf{ACE}\mathbf{n}(t - \Delta t)$$

A = advection
C = chemistry
E = emissions
(for example)

Then the adjoint is the transpose $(\mathbf{ACE})^T = \mathbf{E}^T \mathbf{C}^T \mathbf{A}^T$

Application of the adjoint to a unit forcing $\mathbf{v} = (1, 0, \dots, 0)^T$ gives the sensitivity of $n_{1,x}(t)$ to the concentration field (and other model parameters) at the previous time step:

$$\mathbf{E}^T \mathbf{C}^T \mathbf{A}^T \mathbf{v} = \partial n_1(t) / \partial \mathbf{n}(t - \Delta t)$$

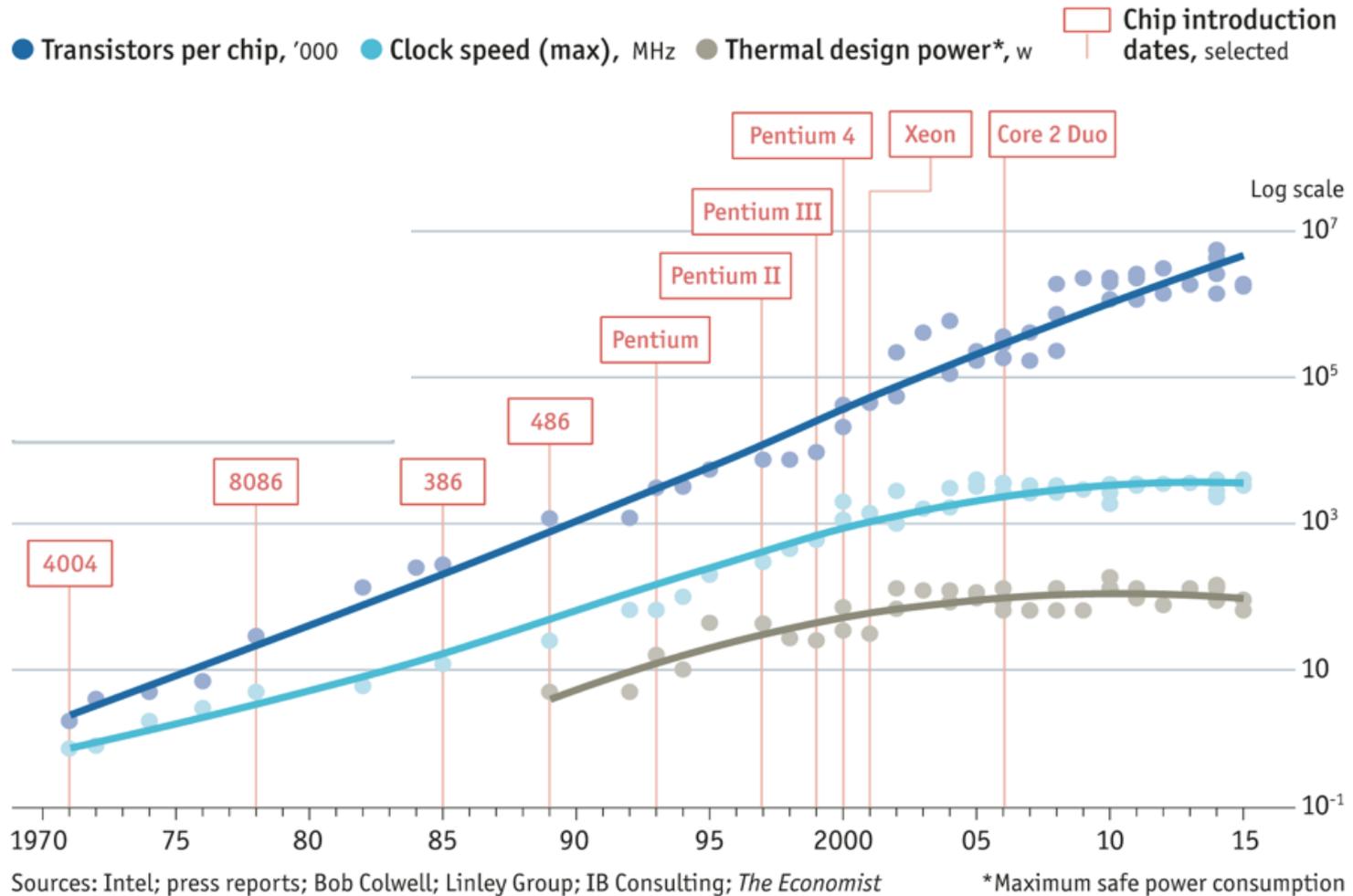


Useful for:

- source attribution in receptor-oriented problems
- variational minimization of cost function

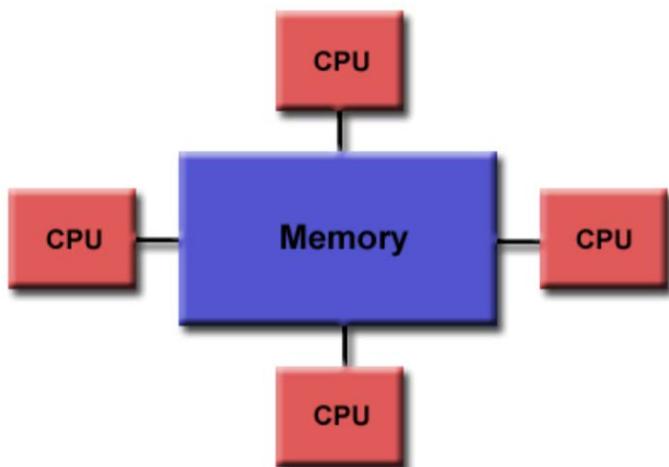
Increasing computational performance of models

Moore's law is slowing down, clock speeds are flat

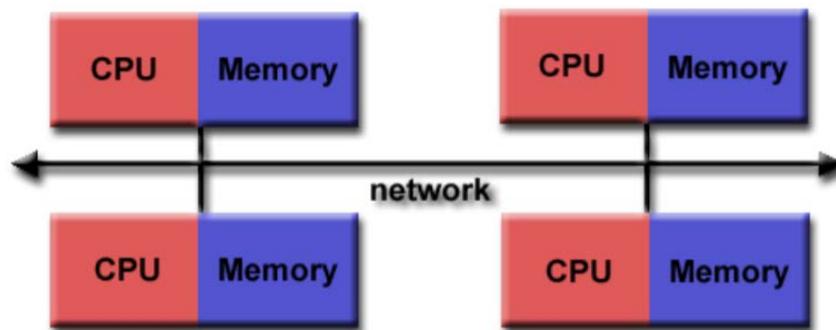


Since we cannot count on faster cores we need to increase the number of cores

Shared-memory vs distributed-memory model architectures



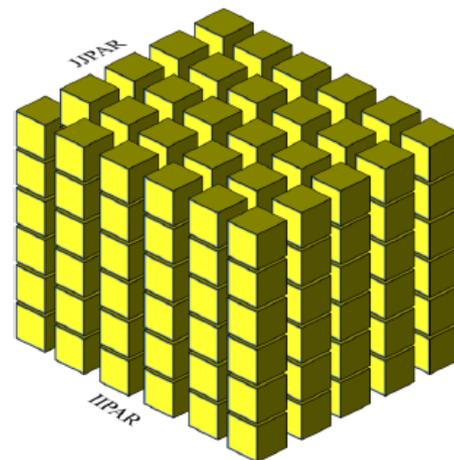
Shared memory architecture (OpenMP):
easy to code but does not scale well with
more than ~20 cores



Distributed-memory architecture (MPI):
harder to code but scales much better

Local operations $dC/dt = P_i(\mathbf{C}) - L_i(\mathbf{C})$ scale perfectly
column-by-column in massively parallel architectures;

Makes chemistry comparatively cheaper as #cores increases



The better a model becomes, the less accessible it is

Increase model resolution to capture finer-scale phenomena

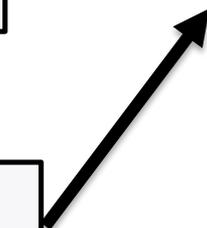
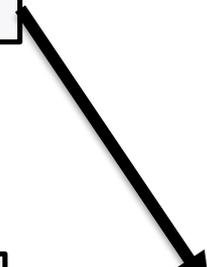
Need to download more input data

Use more advanced schemes to incorporate better scientific knowledge

Need much more computing power

Use advanced software (e.g. Earth System Modeling Framework) for better parallelization and model interoperability

Compiling and configuring models become much more difficult



A solution: cloud computing

- Swipe your credit card and rent a supercomputer when you need it



- GEOS-Chem code has been ported to the AWS cloud with loaded input data, libraries, etc. Just configure input file and run!
- Advantages for users:
 - No need to purchase an expensive system
 - No need to have systems experts to set up the model
 - No need to understand the model
 - Reproducibility of results from standard model is guaranteed
 - Download fully configured model to your own system





Atmospheric chemistry modeling has a bright future!

- Many problems to choose from: fundamental, applied, integrative
- Frontier of knowledge is near: many zeroth-order problems remain
- Next frontier in Earth system modeling
- Continual stream of new data sets to challenge us

To know more!

Guy P. Brasseur and Daniel J. Jacob

Modeling of Atmospheric Chemistry

Chapter 1: The concept of model

Chapter 2: Atmospheric structure and dynamics

Chapter 3: Chemical processes

Chapter 4: Model equations, numerical approaches

Chapter 5: Radiative, chemical, aerosol processes

Chapter 6: Numerical methods for chemical systems

Chapter 7: Numerical methods for advection

Chapter 8: Parameterization of small scales

Chapter 9: Surface fluxes

Chapter 10: Model evaluation

Chapter 11: Inverse modeling

Cambridge University Press, 2017