NONLINEAR DYNAMICS AND CHAOS

Patrick E McSharry
mcsharry@maths.ox.ac.uk
http://www.maths.ox.ac.uk/∼mcsharry
Tel: 01865 270521

DH24 (3rd Floor of Dartington House)

Trinity Term 2002, weeks 2 to 7

Mondays & Wednesdays 11:00am

Mathematical Institute and Department of Engineering Science

University of Oxford

LECTURE 10: WEATHER FORECASTING

- Numerical weather prediction
- Time and spatial scales
- Spectral methods
- Ensemble prediction
- Climate models
- Cross pollinate in time (CPT)
- Ensemble evaluation

Numerical weather prediction

- Horizontal momentum equation

$$\frac{\partial \mathbf{V}}{\partial t} = -(\mathbf{V}\cdot\nabla)\mathbf{V} - \omega\frac{\partial \mathbf{V}}{\partial p} - f\mathbf{k}\times\mathbf{V} - \nabla\phi + \mathbf{F}$$

- Hydrostatic equation

$$\frac{\partial p}{\partial \phi} = -\rho$$

- Thermodynamic equation

$$\frac{\partial T}{\partial t} = -\mathbf{V}\cdot\nabla T - \omega\left(\frac{\partial T}{\partial p} - \frac{RT}{c_p p}\right) + \frac{\dot{Q}}{c_p}$$

- Mass continuity

$$\nabla\cdot\mathbf{V} + \frac{\partial \omega}{\partial p} = 0$$

- Finite difference in time

$$\mathbf{x}_{t+1} = \mathbf{M}(\mathbf{x}_t)$$

Operation time and spatial scales

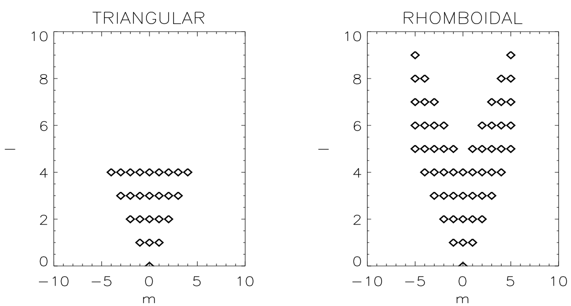
- Finite difference or spectral in space
- FFT required for nonlinear terms in spectral model
- Global model resolution ≈ 60 km
- $\approx 30 - 60$ vertical levels
- Number of variables $\approx 10^7$
- Timestep ≈ 20 minutes
- Computed out to ≈ 10 days
- Subgrid processes *parameterized*
- Observing network incomplete
- State estimation called *data assimilation*
- Predictability limit is on the order of a week

Spectral models

- Horizontal fields at each vertical level represented by *spherical harmonics*

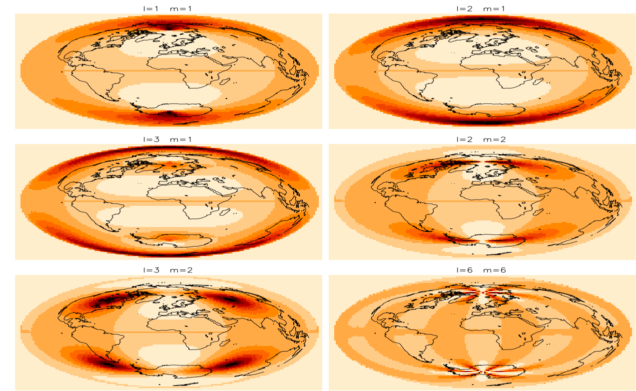
$$Y_{lm}(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(\cos\theta)e^{im\phi}$$

- where P_l^m are Legendre Polynomials
- SH form orthogonal basis set, they are solutions to Laplace's equation
- For global models—polar problem avoided
- "triangular" and "rhomboidal" truncations (ECMWF use T511/T255)



Resolution

- Spatial resolution $\approx 20000/m$ km
- Linear terms evaluated in spectral space
- For large models best to evaluate nonlinear terms in real space and transform back
- Requires fast transforms: use FFT along circles of latitude



4-D variational assimilation

- NWP model

$$\mathbf{x}_{t+1} = \mathbf{M}(\mathbf{p},\mathbf{x}_t)$$

- Constrained optimization

$$L = J + \sum_{t=0}^{T} \lambda_t'(\mathbf{x}_{t+1} - \mathbf{M}(\mathbf{p},\mathbf{x}_t))$$

- Differentiation gives

$$\frac{\partial J}{\partial \mathbf{x}_t} - \left(\frac{\partial \mathbf{M}(\mathbf{p},\mathbf{x}_t)}{\partial \mathbf{x}_t}\right)' \lambda_t + \lambda_{t-1} = 0$$

- Rearranging gives *adjoint model*

$$\lambda_{t-1} = \left(\frac{\partial \mathbf{M}(\mathbf{p},\mathbf{x}_t)}{\partial \mathbf{x}_t}\right)' \lambda_t + \mathbf{G}_t$$

where $\mathbf{G}_t = -\frac{\partial J}{\partial \mathbf{x}_t}$

- Need to calculate

$$\frac{\partial J}{\partial \mathbf{x}_0} = -\sum_{t=0}^{T}\left(\frac{\partial \mathbf{x}_t}{\partial \mathbf{x}_0}\right)' \mathbf{G}_t$$

4-D variational assimilation (continued)

- By induction

$$\left(\frac{\partial \mathbf{x}_t}{\partial \mathbf{x}_0}\right)' = \left(\frac{\partial \mathbf{x}_{t-1}}{\partial \mathbf{x}_0}\right)' \left(\frac{\partial \mathbf{M}(\mathbf{p},\mathbf{x}_{t-1})}{\partial \mathbf{x}_{t-1}}\right)'$$

$$= \left(\frac{\partial \mathbf{x}_{t-1}}{\partial \mathbf{x}_0}\right)' \mathcal{M}_{t-1}'$$

where $\mathcal{M}_t' = \left(\frac{\partial \mathbf{M}(\mathbf{p},\mathbf{x}_t)}{\partial \mathbf{x}_t}\right)'$

- This can be written as

$$\frac{\partial J}{\partial \mathbf{x}_0} = -\sum_{t=0}^{T} \mathcal{M}_1'\mathcal{M}_2'...\mathcal{M}_{t-1}'\mathbf{G}_t$$

- and

$$\lambda_{t-2} = \mathcal{M}_{t-1}'\lambda_{t-1} + \mathbf{G}_{t-1}$$

$$\lambda_{t-2} = \mathcal{M}_{t-1}'(\mathcal{M}_t'\lambda_t + \mathbf{G}_t) + \mathbf{G}_{t-1}$$

- And so on...

$$\lambda_0 = \sum_{t=0}^{T} \mathcal{M}_1'\mathcal{M}_2'...\mathcal{M}_{t-1}'\mathbf{G}_t = -\frac{\partial J}{\partial \mathbf{x}_0}$$

4-D variational assimilation (continued)

- The gradient vector of the cost function can be calculated with one integration of the model and one reversed integration of its adjoint!
- Operationally done with *tangent linear model*
- Conjugate gradient method used to minimize J
- J given by

$$J = \frac{1}{2}\sum_{t=0}^{T}(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t)'\mathbf{C}(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t)$$

- $\mathbf{y}_t$ are observations
- $\mathbf{H}$ is "observation function"
- $\mathbf{C}$ describes error covariances
- Method gives a true model trajectory—unlike "nudging" methods (e.g. Kalman filters)
- Any quantity predicted by the model can be assimilated in an attempt to remove observational uncertainty

Climate modelling

- Want statistics of attractor, not trajectories
- Ocean provides atmospheric boundary conditions
- General (Global) Circulation Models (GCMs)
- Atmosphere coupled to ocean
- Feedback processes
  - cloud (height dependent)
  - ice-albedo
  - land-surface
  - and many more. . .
- Solar variation vs. anthropogenic
- Multi-model results for Intergovernmental Panel on Climate Change (IPCC)
- Other ensemble approaches relevant

Hadley centre climate model

- Hadley Centre HADCM3
  - atmosphere: $2.5° \times 3.75°$ 19 levels (≈ T42)
  - ocean: $1.25° \times 1.25°$ 20 levels
- used for seasonal forecasting as well as CO_2 experiments

Budyko climate model

- radiation absorbed=radiation emitted

$$\pi R^2 F(1-A) = 4\pi R^2 \epsilon\sigma T^4$$

$$F = \frac{4\epsilon\sigma T^4}{(1-A)}$$

- ice-albedo feedback

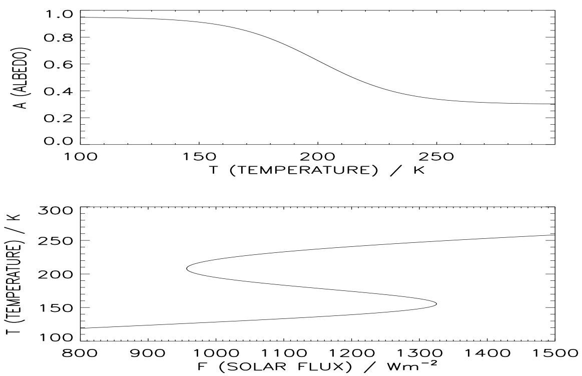
$$A = \frac{ae^{-\alpha(T-T_0)}}{1+e^{-\alpha(T-T_0)}} + b$$

- Milankovitch ⇒ variations in F
- stable "snowball" solution when $F < 1320\,\mathrm{Wm}^{-2}$
- geological evidence of equatorial glaciation *close to sea level* in the Neoproterozoic (600-7 MY ago)!!
- greenhouse effect changes emissivity ϵ

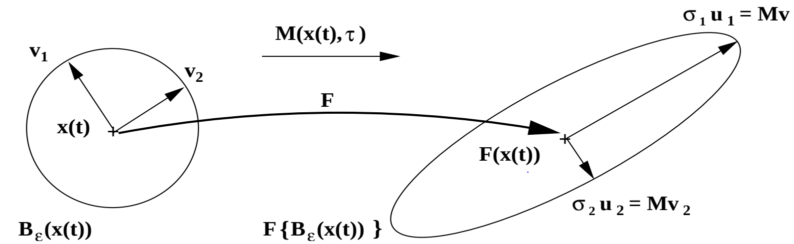
Budyko climate model (continued)

- Model parameter values

  $\sigma = 5.7 \times 10^{-8}\text{Jm}^{-2}\text{s}^{-1}\text{K}^{-4}$, $\epsilon = 1$, $\alpha = 1/18\text{K}^{-1}$, $T_0 = 200\text{K}$, $a = 0.65$, $b = 0.3$



Singular Value Decomposition (SVD)



- SVD provides a description of the linear propagation of a ball of initial conditions
- Linear growth implies that a sphere becomes an ellipsoid
- SVD of the linear propagator:

  $$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

  where both matrices $\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices,

  $$\mathbf{U}\mathbf{U}^T = \mathbf{I} = \mathbf{U}^T\mathbf{U}$$

  describing the rotational properties of $\mathbf{M}$
- The matrix $\mathbf{\Sigma}$ is diagonal and contains the singular values σ_i
- $\mathbf{\Sigma}$ quantifies the stretching/contraction characteristics of the matrix $\mathbf{M}$

Ensemble construction

- The initial condition $\mathbf{x}_0$ is in 10^7 dimensional space
- Computational constraint limits ensemble members to ≈ 50
- What are the best initial perturbations to make?
- Make perturbations in space defined by leading *singular vectors* (sampling directions that are going to grow the most)
- NWP model can be written as

  $$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x})$$

- $\delta\mathbf{x}$ is an infinitesimal perturbation
- Tangent-linear equation is

  $$\frac{d\delta\mathbf{x}}{dt} = D\mathbf{F}(\mathbf{x})\delta\mathbf{x}$$

- Propagator $\mathbf{M}(\mathbf{x}, \tau)$ is solution

  $$\delta\mathbf{x}(t+\tau) = \mathbf{M}(\mathbf{x}, \tau)\delta\mathbf{x}(t)$$

Ensemble construction

- Singular value decomposition

  $$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

- $\mathbf{\Sigma}$ is diagonal matrix of singular values
- $\mathbf{u}_i$ (columns of $\mathbf{U}$) are left singular vectors, $\mathbf{v}_i$ (columns of $\mathbf{V}$) are right singular vectors
- Special properties:

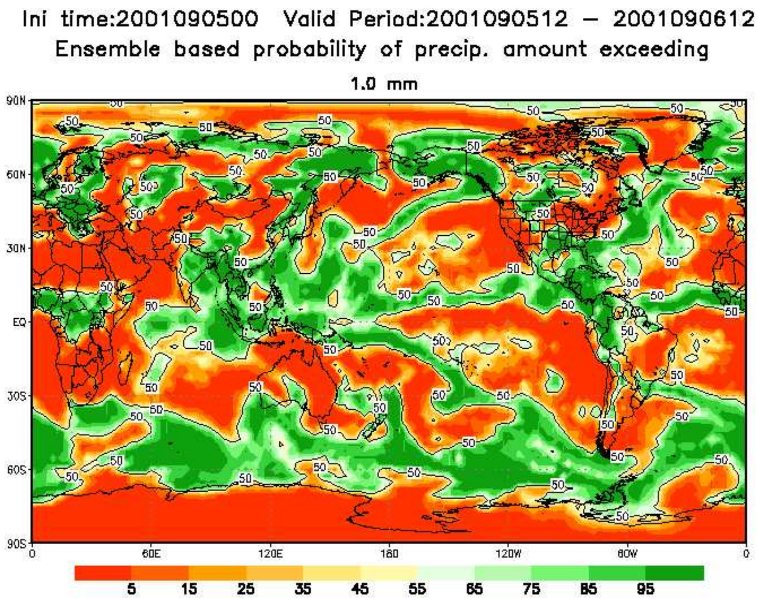
  $$\mathbf{M}\mathbf{v}_i = \sigma_i \mathbf{u}_i \quad \mathbf{M}^T\mathbf{u}_i = \sigma_i \mathbf{v}_i$$

  and

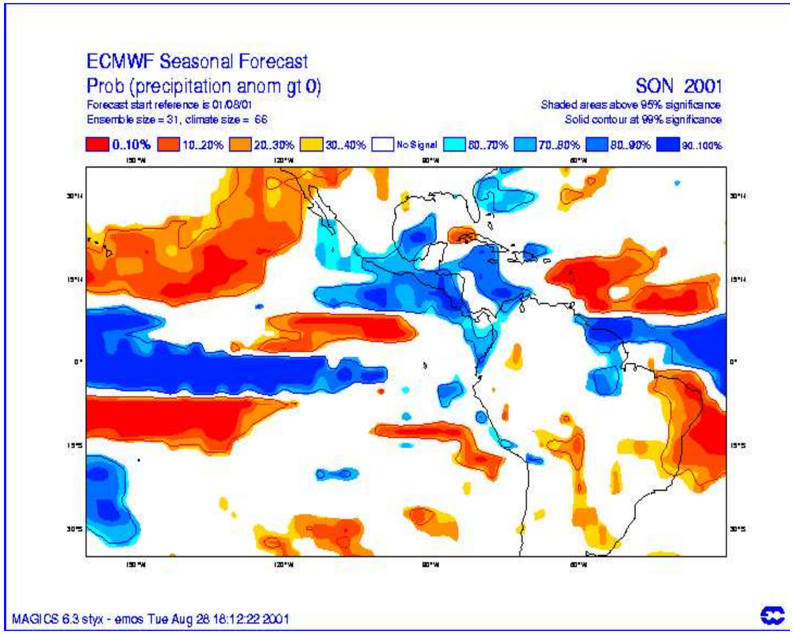
  $$\mathbf{M}^T\mathbf{M}\mathbf{v}_i = \sigma_i^2 \mathbf{v}_i \quad \mathbf{M}^T\mathbf{M}\mathbf{u}_i = \sigma_i^2 \mathbf{u}_i$$

- $\mathbf{u}_i$ are patterns (directions) that give maximum error growth in the linear model over tim
- ECMWF makes $\pm$ perturbations in space of leading 25 singular vectors
- NCEP uses "bred vectors" (the directions that have been growing the most)
- The great debate: higher resolution vs. bigger ensembles
- Additional information:
  - Palmer, T.N., Buizza, R. and Molteni, F., Singular vectors and the predictability of weather and climate, *Philos. T. Roy. Soc. A, 348*, 459-475, 1994
  - Palmer, T.N., Predicting uncertainty in forecasts of weather and climate, *Rep. Prog. P* *63,* 71-116, 2000

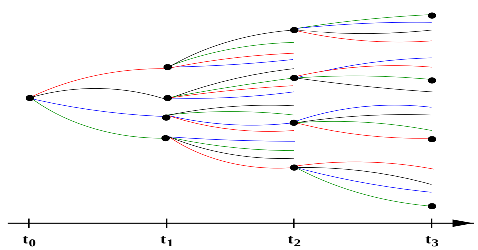
Probabilistic forecast for recipitation (NCEP)



Probabilistic forecast for recipitation (ECMWF)

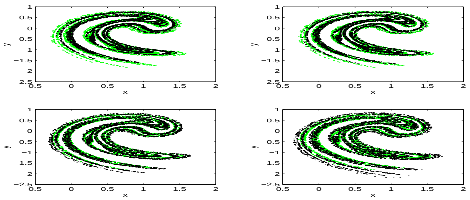


Cross Pollination in Time (CPT) Ensembles



- No perfect model exists for real systems
- Multiple models of varying quality
- How can one best combine these models?
- Cross pollination in time
- At each step, use all models to evolve initial conditions
- Prune set of initial conditions while maintaining spread
- CPT model is a hybrid model

Imperfect models in two-dimensions

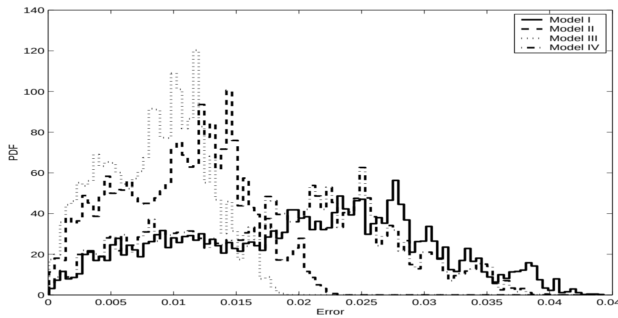


- Ikeda equations:

$$\begin{aligned} x_{i+1} &= 1 + \mu(x_i \cos(t) - y_i \sin(t)), \\ y_{i+1} &= \mu(x_i \sin(t) + y_i \cos(t)), \\ t &= 0.4 - 6.0/(x_i^2 + y_i^2 + 1), \end{aligned}$$

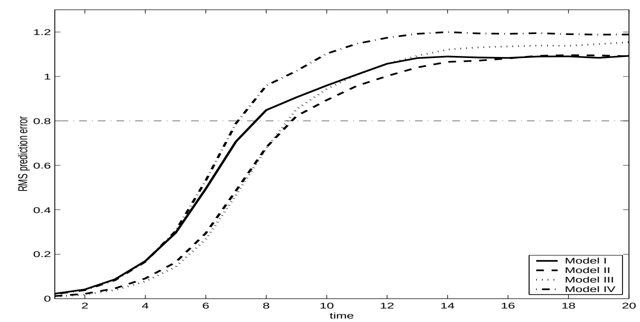
- System: μ = 0.881.
- Models: μ = 0.86, 0.87, 0.89, 0.9
- No structural error

One-step prediction errors



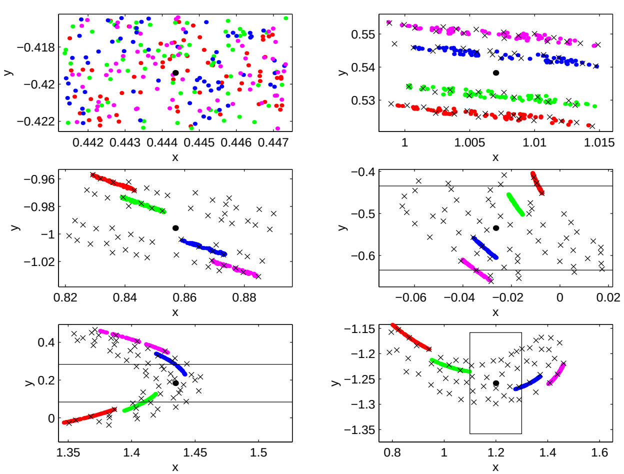
- Which model is best?
- Ranking by size of mean error: III, II, IV, I

RMS multi-step prediction error

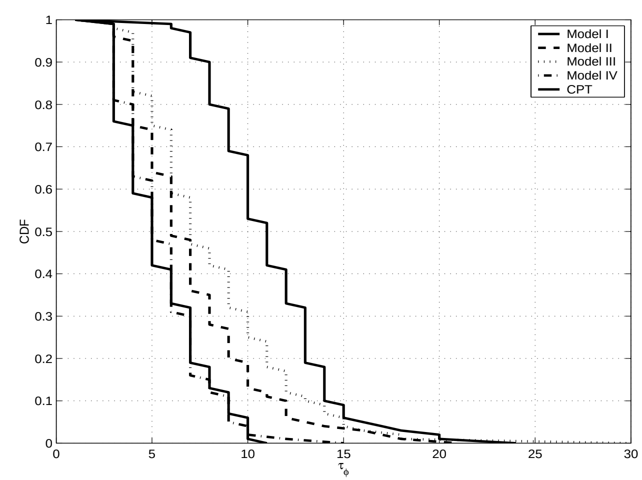


- Model III is best for $t \in [0, 7]$
- Model IV is better than model I for $t \in [0, 4]$
- Model I is better than model IV for $t \in [4, 10]$

Ensemble prediction zoom



Shadowing times

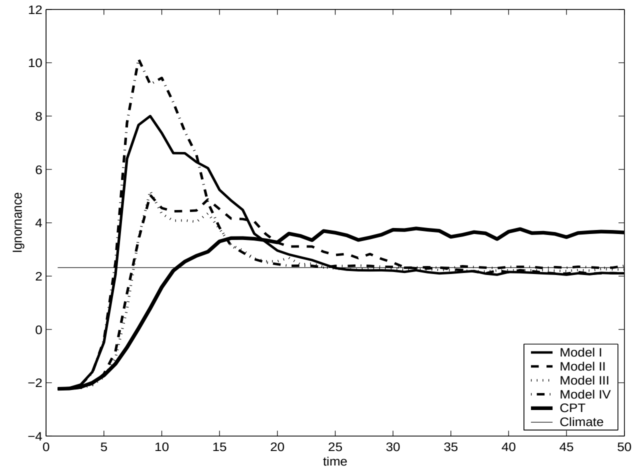


Ensemble evaluation: Ignorance

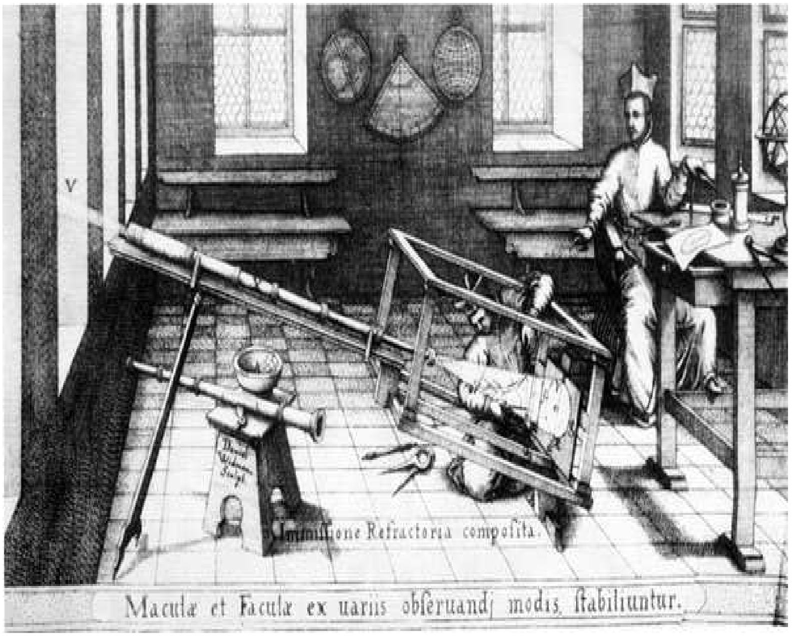
- Information theory can be used to measure the quality of a probabilistic forecast based on its ability to provide data compression
- Ignorance (Roulston & Smith, 2002) is a logarithmic scoring rule based on relative entropy
- Ignorance has the appealing property that it is equivalent to the expected return generated by placing bets which are proportional to the forecast probabilities
- Suppose that at each forecast time t, the probabilistic forecast is given by $\{p(t)_i\}_{i=1}^N$ where $p(t)_i$ is the probability of obtaining outcome i at time t
- If there are T forecast times in total and the actual outcome at time t is $i(t)$, then the ignorance of the forecasting system (averaged over all forecast-realisation pairs) is

$$I = -\frac{1}{T}\sum_{t=1}^{T} \log_2 p(t)_{i(t)}$$

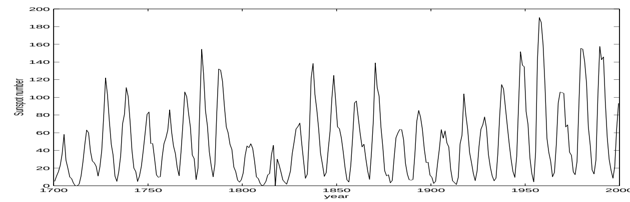
Ignorance



Early Sunspots



Sunspot index



- Five genetic algorithm models by George G. Szpiro, Phys. Rev. E **55**(3): 2557-2568, us data from 1700-1899
-

$$x_t = 7.7762 + x_{t-1}\left(0.8205 + \frac{x_{t-9} - x_{t-3}}{Q}\right)$$

$$\begin{aligned} Q^{(I)} &= 4x_{t-1} + 0.25x_{t-9} \quad Q^{(II)} = 9.4 + 2x_{t-2} + x_{t-9} + x_{t-13} \\ Q^{(II)} &= 4.8x_{t-1} + (1/9.5)x_{t-9} \quad Q^{(IV)} = 3.9x_{t-1} + x_{t-11} \\ Q^{(V)} &= 3.9x_{t-1} + x_{t-13} \end{aligned}$$

Sunspot index ensemble prediction: 1900-1910