

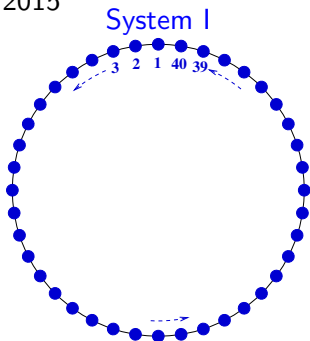
Predictability and Ensemble Forecasting with Lorenz-95 systems

Martin Leutbecher



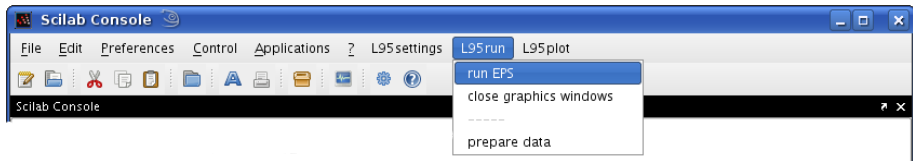
Training Course 2015

- 1 scilab
- 2 overview of the forecast model and the systems
- 3 **You experiment with ensemble forecasting yourself ...**
 - ▶ perfect model scenario
 - ▶ imperfect model scenario



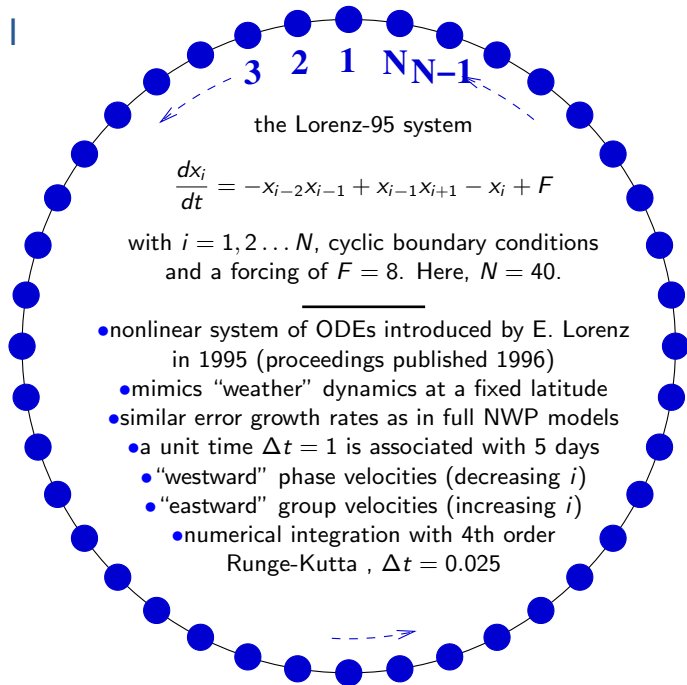
Using scilab

- scilab can be downloaded at www.scilab.org
- get a local terminal window (green icon bottom left)
- enter the command `/scratch/rd/nel/tc/install_tutorial`
This command obtains a copy of the tutorial scripts and input data.
- enter the command `tutorial`,
(use a new terminal or enter the full path `~/bin/tutorial` if tutorial is not found)



“prepare data” not required as input data is obtained when `install_tutorial` runs.

System I



the Lorenz-95 system

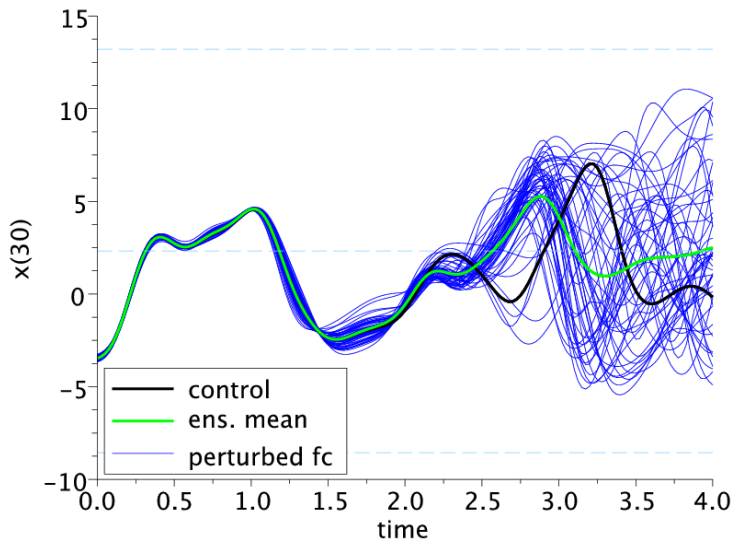
$$\frac{dx_i}{dt} = -x_{i-2}x_{i-1} + x_{i-1}x_{i+1} - x_i + F$$

with $i = 1, 2, \dots, N$, cyclic boundary conditions
and a forcing of $F = 8$. Here, $N = 40$.

- nonlinear system of ODEs introduced by E. Lorenz in 1995 (proceedings published 1996)
- mimics “weather” dynamics at a fixed latitude
- similar error growth rates as in full NWP models
 - a unit time $\Delta t = 1$ is associated with 5 days
 - “westward” phase velocities (decreasing i)
 - “eastward” group velocities (increasing i)
 - numerical integration with 4th order Runge-Kutta , $\Delta t = 0.025$

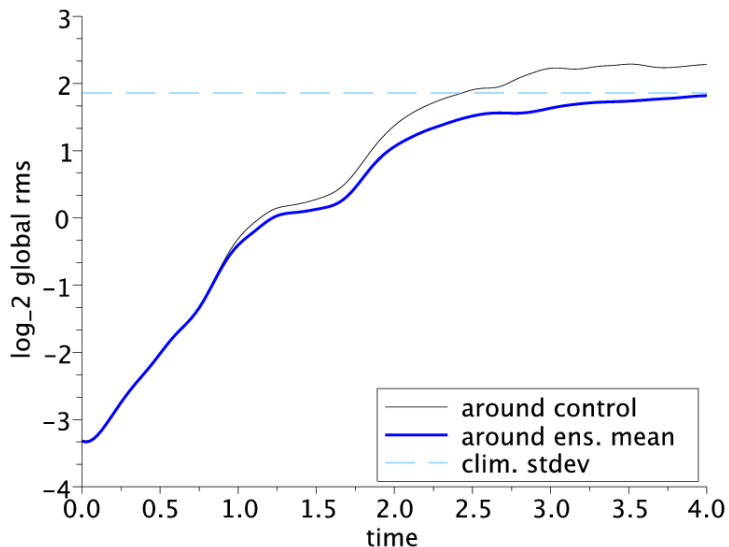
System I is chaotic

Sensitive dependence on initial conditions

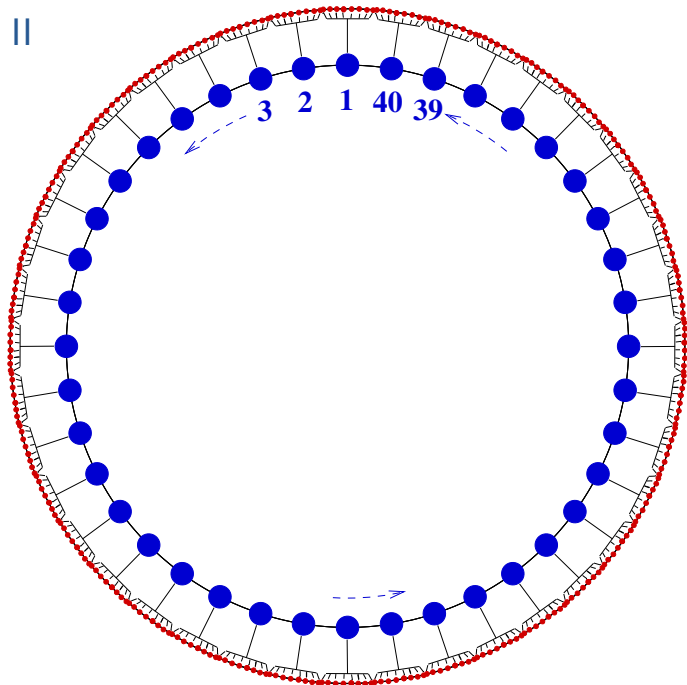


System I is chaotic

Perturbation growth



System II



System and model equations

The **system** comprises **slow variables** x_k and **fast variables** y_j

$$\frac{dx_k}{dt} = -x_{k-1}(x_{k-2} - x_{k+1}) - x_k + F - \frac{hc}{b} \sum_{j=J(k-1)+1}^{Jk} y_j \quad (1)$$

$$\frac{dy_j}{dt} = -cby_{j+1}(y_{j+2} - y_{j-1}) - cy_j + \frac{c}{b}F_y + \frac{hc}{b}x_{1+\lfloor \frac{j-1}{J} \rfloor} \quad (2)$$

with $k = 1, \dots, K$ and $j = 1, \dots, JK$. Here, $K = 40, J = 8$.

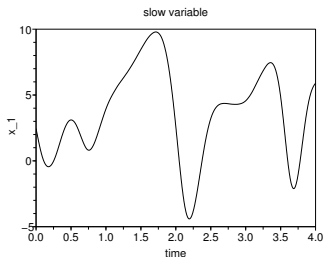
The **forecast model** is given by

$$\frac{dx_k}{dt} = -x_{k-1}(x_{k-2} - x_{k+1}) - x_k + F - g_U(x_k) + \eta_k(t). \quad (3)$$

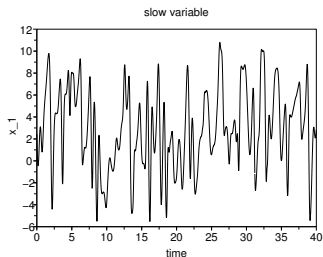
- g_U represents a **deterministic parameterization** of the net effect of the fast variables on the slow variables, see slide 11.
- η_k is a **stochastic forcing term** which represents the uncertainty due to the forcing of the fast variables, see slide 12.

Time series for system II

x_1

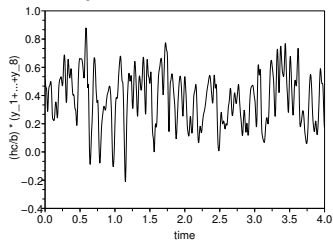


x_1

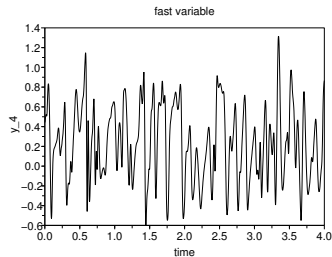


$$hc/b \times \sum_1^8 y_k$$

forcing of slow variable due to fast variables for x_1



y_4



(system II, version T2)

Truth integrations

System constants F and h and integration time step dt for the five different systems

name	T0	T1	T2	T5	T10
h	0	0.1	0.2	0.5	1.0
F	8.0	8.2	8.4	9.0	10.0
$10^3 dt$	25	2.5	2.5	2.5	2.5

The other variables are set to

$b = 10$ **amplitude ratio** between slow variables and fast variables

$c = 10$ **time-scale ratio** between slow and fast variables

$F_y = F$ **forcing amplitude**

For all systems, the climatological mean of the slow variables is about 2.4 and their climatological standard deviation is about 3.5.

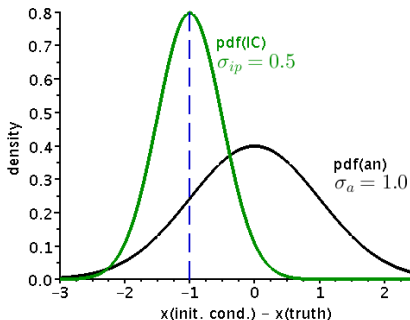
Initial conditions and their uncertainty

- Initial states (“analyses”) have been prepared for all truth systems (T0–T10). Values of the standard deviation of the analysis uncertainty of

$\sigma_a = 0.01, 0.02, 0.05, 0.1, 0.2, 0.5$ and 1.0

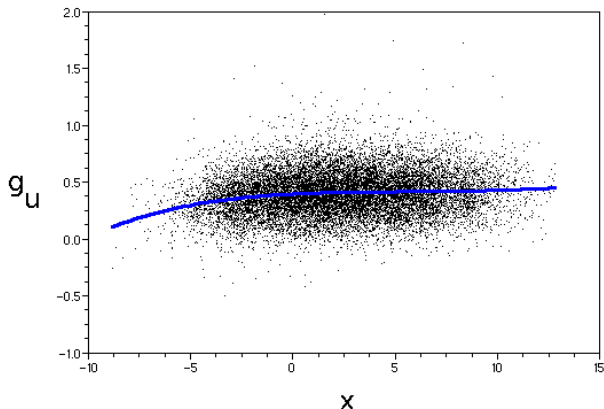
can be selected.

- The initial conditions for the ensemble forecasts are sampled from a Gaussian distribution centred on the analysis. The standard deviation of the initial condition perturbations σ_{IP} can be varied, the default value is σ_a .
- Ensemble forecasts are started every 0.4 time units (2 days).



Deterministic parameterisation of unresolved scales

The unresolved scales (**fast y -variables of system II**) have a net effect on the resolved scales (**slow x -variables**). A term $g_U(x)$ is subtracted from the RHS of model I to account for the net effect of the unresolved scales.

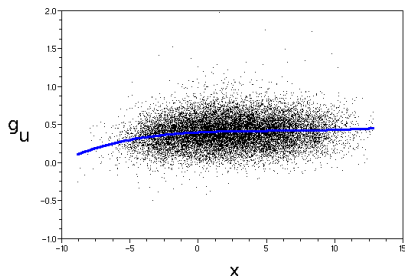


Deterministic parameterisation g_U (blue curve) of the net effect of the fast y -variables in system T2 on the slow x -variables.

Black dots represent the actual forcing due to y -variables

Representation of model uncertainty

In order to represent the scatter of the slow-variable tendencies due to the fast variables, i.e. the deviation of the black dots from the blue curve, a stochastic forcing term η_k can be activated in the ensemble forecasts.



- It is represented by independent AR(1)-processes for each slow variable

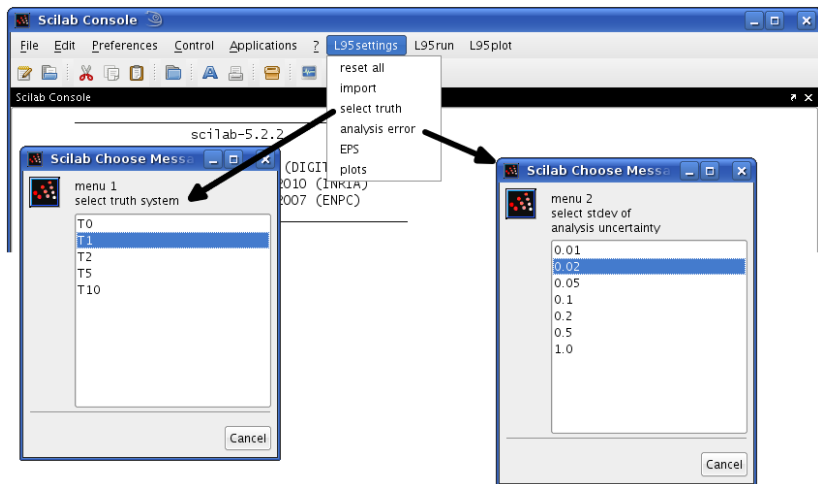
$$\eta_k(t + \Delta t) = \phi \eta_k(t) + \sigma_e (1 - \phi^2)^{1/2} z_k(t), \quad (4)$$

where the $z_k(t)$ are drawn from a Gaussian distribution $N(0, 1)$.

- the standard deviation σ_e and lag-1 autocorrelation ϕ over one integration time-step Δt can be set in menu 3 (EPS configuration: `sigma_e` and `phi`);
- see Wilks (2005) for further details.

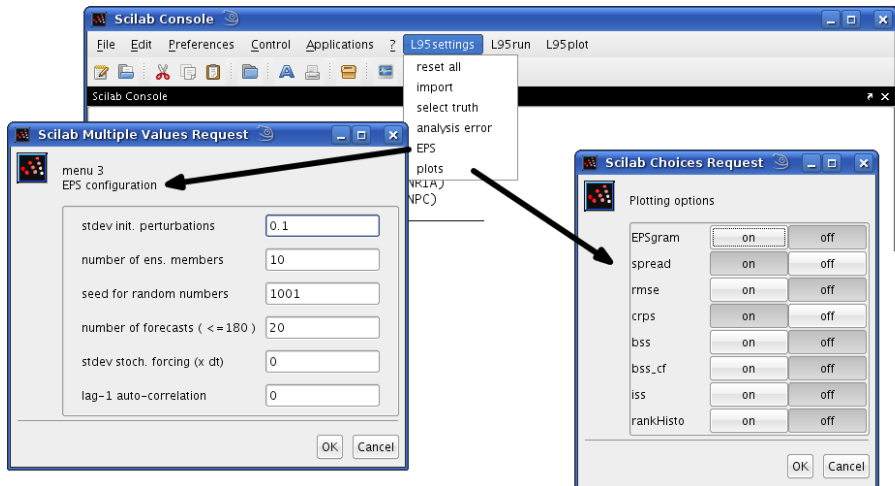
Numerical experimentation

Description of the menus



Numerical experimentation

Description of the menus



Numerical experimentation

Run Ensemble Prediction System



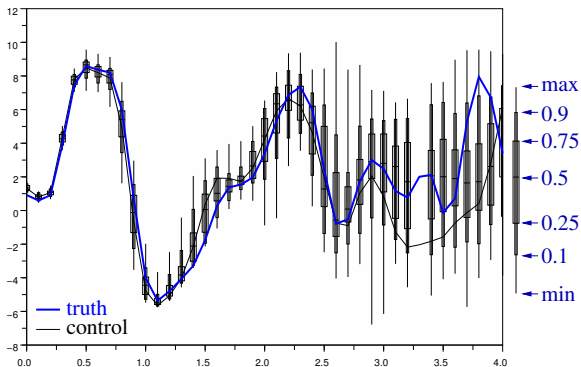
=====

This is experiment

```
system ID           = T0
stdev init. uncertainty = 0.10000
stdev init. perturbations = 0.10000
dt x stdev stoch. forcing = 0.00000
lag-1 auto-correlation = 0.00000
number of pert. forecasts = 10
number of cases       = 20
seed for random numbers = 1001
```

	f o r e c a s t		s t e p		
Score	0.0	0.5	1.0	2.0	4.0
Ens.Std	0.09	0.25	0.55	1.73	3.08
Ens.Mean RMSE	0.10	0.26	0.58	1.75	3.28
Ens.Std/RMSE	0.91	0.93	0.96	0.99	0.94

EPSgrams



The EPSgrams are saved in png-files if they have been activated as plot prior to running the ensemble. They can be viewed with

```
xv $SCRATCH/ML7data/ee/epsgram*.png
```

where ee is the respective experiment ID.

Numerical experimentation

Settings

- settings will be remembered from one run to the next
- the defaults can be restored with
→ `L95settings` → `reset all`
- settings from a previous experiment can be obtained with
→ `L95settings` → `import`
- The current settings can be printed with
→ `L95settings` → `print info`

Numerical experimentation

Comparison of different experiments

The screenshot displays the Scilab interface with a plot and several dialog boxes. The plot shows CRPS (Continuous Ranked Probability Score) on the y-axis (ranging from 0.0 to 2.5) against lead time on the x-axis (ranging from 0.0 to 4.0). Two lines are plotted: a black line for experiment 'xg' and a blue line for experiment 'xh'. Both lines show an increasing trend, with 'xh' generally having a higher CRPS than 'xg' after a lead time of 1.5. A legend in the bottom right of the plot identifies the lines.

The 'L95plot' menu is open, showing options: 'select experiments', 'select scores', 'make plots', 'compare experiments', and 'close graphics windows'. Arrows point from these options to the dialog boxes and the plot.

The 'Scilab Multiple Values Request' dialog box is open, with the text 'select Expts for plotting' and a list of expts 'a1,a2[,x5 ..]'. The input field contains 'xg,xh'. Buttons for 'OK' and 'Cancel' are visible.

The 'Scilab Choices Request' dialog box is open, showing 'Plotting options' with checkboxes for 'EPSgram', 'spread', 'rmse', and 'crps', each with 'on' and 'off' buttons. Below this, there is a table of settings:

	xg	xh
sigma_ip	0.100	0.050

Below the table, 'common settings' are listed:

```
sysID = T0
sigma_a = 0.100
npf = 10
```

Topics for the tutorial

- 1 Perfect model, reliable initial perturbations
 - ▶ characteristics of a “perfectly reliable ensemble”
 - ▶ sensitivity to initial uncertainty
 - ▶ sensitivity to ensemble size
- 2 Perfect model, *unreliable* initial perturbations
- 3 *Imperfect model*, reliable initial perturbations, no representation of model uncertainty in the ensemble
- 4 Imperfect model with *stochastic parameterization*, reliable initial perturbations
- 5 Optimization of an ensemble for an unknown system

It's your turn now ...

Perfect model, reliable initial perturbations

- start from the default settings (system T0, analysis uncertainty $\sigma_a = 0.1$, standard deviation of initial perturbations $\sigma_{ip} = \sigma_a$, 10 member ensemble)
- *look at the different verifications statistics*
 - ▶ What is the relationship between the RMS error of the ensemble mean and the spread?
 - ▶ Compare RMS error of unperturbed forecast (control), perturbed forecasts and ensemble mean. What are their values in the early and late forecast ranges?

Perfect model, reliable initial perturbations

Measures of Probabilistic Skill

- How does the Continuous Ranked Probability Score (CRPS) evolve with lead time?
- Compare the CRPS of the ensemble with the CRPS of the ensemble mean (=Mean absolute error of Ens. mean), with the CRPS of the control forecast, with the CRPS of the climate distribution (approximated by a Gaussian in this case).
- Logarithmic score (ls_g for Gaussian pdf)
- Logarithmic Score (LS) and Brier Score (BS) for particular events. Why are results noisier than for the scores of the continuous variable? Why do asymptotic values of the score depend on the event thresholds?
The binary events are:
 - ▶ values larger than the climatological mean ($>m$)
 - ▶ values larger than the clim. mean + one standard deviation ($>m+s$)
 - ▶ values larger than the clim. mean - one standard deviation ($>m-s$)
- Are the rank histograms (Talagrand diagrams) flat?

Perfect model, reliable initial perturbations

Sensitivity to initial uncertainty

- Change the analysis error standard deviation (\rightarrow L95settings \rightarrow analysis error) to a small value, e.g. 0.01.
- Change the standard deviation of the initial perturbations to the same value $\sigma_{ip} = \sigma_a$ (\rightarrow L95settings \rightarrow EPS)
- Run the EPS and compare the results with the default run.
- You may want to repeat the exercise for a situation with large initial uncertainty, e.g. $\sigma_{ip} = \sigma_a = 1.0$
- What are the gains/losses in lead time in terms of probabilistic skill as the initial uncertainty is decreased and increased?

Perfect model, reliable initial perturbations

Sensitivity to ensemble size

- Go back to the default settings
→ L95settings → reset all
- Vary the number of members, say between 2 and 64
→ L95settings → EPS
- Understand how the RMS error of the ensemble mean changes with increasing ensemble size.
- What happens to the ensemble standard deviation (=RMS of deviations from ensemble mean)?
- Can you work out correction factors that have to be applied to finite ensemble size estimates of ensemble standard deviation and ensemble mean RMS error in order to obtain the limit values for an infinite ensemble size?
- Study the sensitivity of the various measures of probabilistic skill to ensemble size.

Perfect model but unreliable initial perturbations

- Return again to the default settings
→ L95settings → reset all
- Now, configure the ensemble to use either too small or too large initial perturbation amplitude.
→ L95settings → EPS
stdev init. perturbations < 0.1 or > 0.1
- Study the effects on ensemble standard deviation (spread) and ensemble mean RMS error.
- How do measures of probabilistic skill change due to the modified initial perturbations?
- How do the RMS errors of ensemble mean and perturbed forecasts change relative to the fixed RMS error of the control forecast?
- Is the skill indeed maximized for the reliable initial perturbations, i.e. $\sigma_{ip} = \sigma_a = 1.0$?
- Cave: not all scores are proper!

Imperfect model with reliable initial perturbations

- Start with default settings but then set the system representing truth to T5.
- Run the ensemble and compare ensemble spread and CRPS of the default (truth T0) with this run.
Is model error dominant?
- Now repeat the runs with T0 and T5 using more accurate initial conditions, e.g. $\sigma_a = \sigma_{ip} = 0.02$.
Is model error dominant now?

Imperfect model with stochastic parameterisation of model uncertainty

and with reliable initial perturbations

- use system T5 and $\sigma_a = \sigma_{ip} = 0.02$
- a stochastic representation of the effect of the unresolved fast scales that is consistent with the actual statistics of the tendency errors is an AR(1) with
 - ▶ a lag-1 correlation of $\text{phi}=0.46$
 - ▶ a stochastic forcing of the AR(1) of
`dt x stdev stoch. forcing = 0.010`
- Compare the characteristics of ensemble runs with and without the stochastic parameterisation of model uncertainty
- Now run with stochastic parameterisation but without accounting for temporal correlation of tendency errors, i.e. with
 - ▶ a lag-1 correlation of $\text{phi}=0.0$
 - ▶ a stochastic forcing of the AR(1) of
`dt x stdev stoc. forcing = 0.010`

Unknown initial uncertainty and unknown model uncertainty

- select system Un
- try to find “optimal” values for
 - ▶ the standard deviation of initial perturbations,
 - ▶ the standard deviation of the stochastic forcing and
 - ▶ the lag-1 autocorrelation of the AR(1) process representing model uncertainty
- What are the lowest values for CRPS and LS_G for a 10 member ensemble you can achieve for a sample of 45 cases?
- Hint, you will not be able to get much better than this:

	f o r e c a s t s t e p				
	0.0	0.5	1.0	2.0	4.0
CRPS	0.09	0.31	0.79	1.84	2.23
LS_G	-0.43	0.76	1.68	2.68	2.87

Fisher, M., M. Leutbecher, and G. A. Kelly, 2005: On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation. *Quart. J. Roy. Meteor. Soc.*, **131**, 3235–3246.

Leutbecher, M., R. Buizza, and L. Isaksen, 2007: Ensemble forecasting and flow-dependent estimates of initial uncertainty. In *Flow-dependent aspects of data assimilation*, Workshop Proceedings, ECMWF, Shinfield Park, Reading, UK, 185–201. Available at www.ecmwf.int/publications/library/ecpublications/_pdf/workshop/2007/Data_assimilation.pdf

Lorenz, E. N., 1995: Predictability: A problem partly solved. In *Seminar on Predictability*, volume Vol. I, ECMWF, Reading, UK, 1–18. Available at http://www.ecmwf.int/publications/library/ecpublications/_pdf/seminar/1995.

Lorenz, E. N., 2005: Designing chaotic models. *J. Atmos. Sci.*, **62**, 1574–1587.

Lorenz, E. N. and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, **55**, 399–414.

Wilks, D. S., 2005: Effects of stochastic parametrizations in the Lorenz '96 system. *Quart. J. Roy. Meteor. Soc.*, **131**, 389–407.

Leutbecher, M.: [Tutorial on data assimilation based on the Lorenz-95 system](#)

www.ecmwf.int/newsevents/training/meteorological_presentations/MET_DA.html

How to obtain the scilab scripts?

- 1 download from ECMWF website:
`http://www.ecmwf.int/newsevents/training/meteorological_presentations/MET_PR.html`
- 2 during the course:
`cd /home/unix/presentations/TC_NWP2015/PR/Leutbecher`

PR-TC.tar.gz

is a compressed tar-file containing all scilab scripts needed.
You can email it to yourself as an attachment.