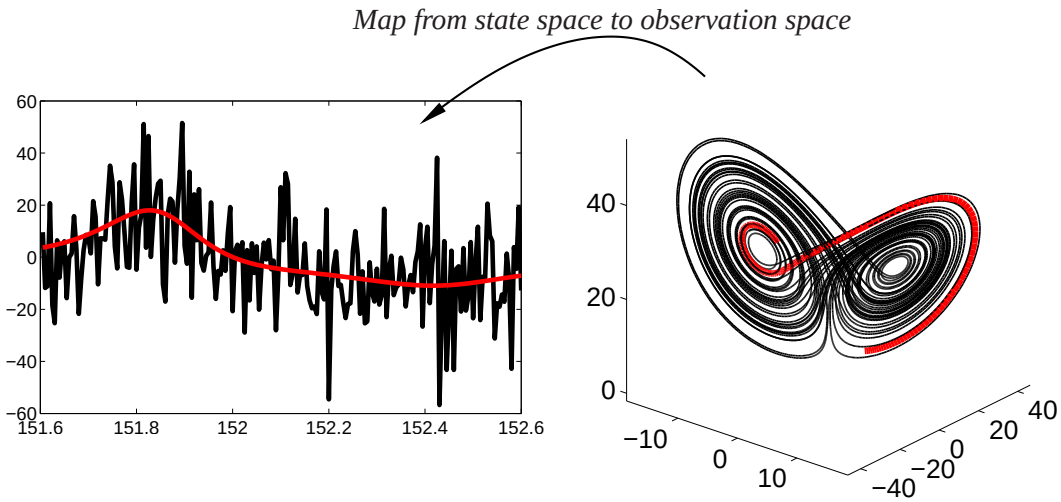
Validation and statistical assessment of data assimilation algorithms

Supervisors: Jochen Broecker, Peter Jan van Leeuwen

Data Assimilation (or DA for short), for the purpose of this project, refers to the following problem: given a dynamical system (e.g. a weather model) and given a series of observations (e.g. wind measurements from the real weather), find a trajectory of the system that matches the observed data. For more information on DA, see www.met.reading.ac.uk/~darc. The picture below is supposed to give a rough idea of the concept.



Given some observations (black line, left panel) and a dynamical model (see "butterfly" phase portrait, right panel), find a trajectory of the model (red line, right panel), which, when mapped into observation space (symbolised by the arrow) follows the observations (red line, left panel).

The aim of is project is not to develop new DA algorithms, but rather to develop a formalism to assess existing ones. The problem here is this: Once the data has been used to estimate the trajectory, it should not be used again (without precautions) to evaluate that trajectory. The trajectory need not be a 'good' one just because it fits the observations well. The DA algorithm might essentially just reproduce the observations without capturing any genuine features of the underlying dynamics. This phenomenon is known in statistics as overfitting; like students who are given problem sheets with solutions during the module, and are then asked the very same problems again during the exams. Clearly, even if they are doing well it does not mean they have understood much.

The easiest solution would be to use independent weather observations from the same period and region as the original data. Such data however is hardly ever available. But this problem of incestuous evaluation appears everywhere in statistics, and people have come up with various ways to deal with it (e.g. BIC, AIC, Mallows' Cp). A good way to get started with this PhD project would be to check which of these methods could be adapted or modified for data assimilation purposes.

Apart from theoretical studies and numerical experiments, this project might involve analysing DA experiments with real world (or laboratory) observations (there are cases where independent observations *do* exist which could be used to check any theories developed during this project). The student will have some freedom in choosing her or his focus.

Student profile:
Necessary background: Suitable for students with a degree in mathematics, statistics and probability theory, or environmental physics. Students with a strong background in probability theory and stochastic processes are especially encouraged to apply.

Funding particulars:
N/A