General introduction to Statistical Downscaling

VALUE 1st Training school:
6-15 November in Santander (Spain)
Rasmus E. Benestad

Empirical Statistical Downscaling – ESD - modules

"Introduction to Dynamical and Statistical Downscaling"

Definitions

Philosophy

Diagnostics
&
Predictions

Linear models

Clim.pact

limitations

Test yourself...

Empirical Statistical Downscaling

Definitions

Definitions - terms

Models

Equations, predictors & predictands

Empirical-statistical downscaling & statistical downscaling

Calibration

Dependency

Noise & stochastic variables

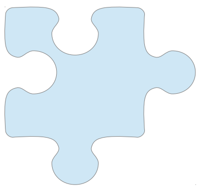
Weather generator & Monte-Carlo simulations

Information

Statistics

Ensembles & determinism

Empirical Orthogonal Functions (EOFs)



Models

The term 'model' is rather generic.

http://www.thefreedictionary.com

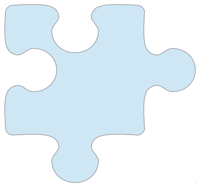
"A schematic description of a system, theory, or phenomenon that accounts for its known or inferred properties and may be used for further study of its characteristics"

Climate models - general circulation models (GCMs).

Global climate models (GCMs)

Regional climate models (RCMs)

Statistical models (e.g. regression) or mathematical equations.



Equations, predictors & predictands

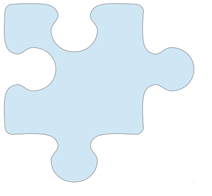
Statistical models usually consist of an equation for which a set of parameters are estimated so that it best represents patterns found in the data.

small-scale = f(geography + large-scale) + noise:

$y = f(g,X) + n.$

predictand = quantification of small-scale (y)

predictor = quantification of large-scale (X)



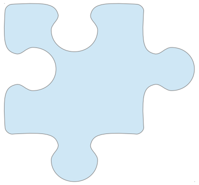
Empirical-statistical downscaling & statistical downscaling

Synonyms.

ESD & SDS: draw information from both empirical data (1) as well as statistical theory (2).

(1) Measurements

(2) Mathematical framework (regression, distributions, optimisation, probabilities)

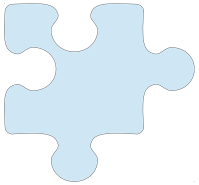


Calibration

The act of estimating the parameters of the statistical models which provide best description of desired aspects found in data.

'Training' a model on data.

Various strategies for optimisation.

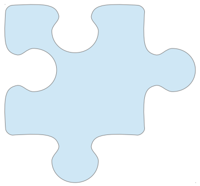


Dependency

An inherent and systematic (physics-based) connection between parts. Can be often be quantified:

$y = f(x)$

May be exposed through correlation and provides the basis for statistical models.

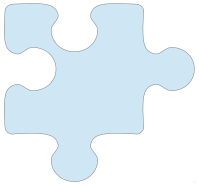


Noise & stochastic variables

Wikipedia:

Stochastic (from the Greek *στόχος* for *aim* or *guess*) is an adjective that refers to systems whose behavior is intrinsically non-deterministic, sporadic, and categorically *not* intermittent (i.e. random).

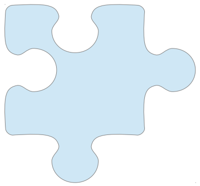
- Unknown dependencies and unpredictable in terms of external factors.
- If stationary, some statistics (distributions, likelihood, persistence) may very well be predictable.
- Also referred to as 'noise' & useful for describing uncertainties



Weather generator & Monte-Carlo simulations

A model that generates a stochastic variable with similar statistical characteristics as chosen weather elements: distribution & time structure.

Can be combined with other statistical models describing dependencies. WGs add little information other than observed distribution and persistence. A form of 'Monte-Carlo simulations', and can be used for hypothesis testing.



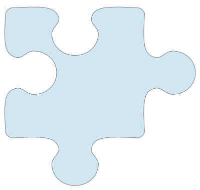
Information

Wikipedia:

"...in its most restricted technical sense, is a sequence of symbols that can be interpreted as a message".

Related to the mathematical concepts of combinations and permutations. In statistics, the information content is related to the degrees of freedom (DOF):

"Estimates of statistical parameters can be based upon different amounts of information or data. The number of independent pieces of information that go into the estimate of a parameter is called the degrees of freedom (df)".

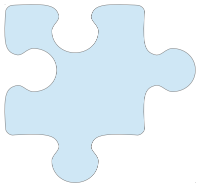


Statistics

Means different things to different people. Common understanding are tables and quantification. Also abstract mathematics and mathematical modelling (least-squares fit versus maximum likelihood).

wikipedia: ***"Statistics** is the study of the collection, organization, analysis, interpretation, and presentation of data"*

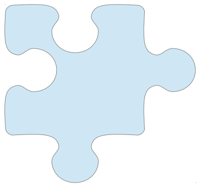
Can combine uncertain data with absolute certain mathematical framework.



Ensembles & determinism

A collection of equal model simulations, representing a set of possible realisations. Represent non-deterministic systems, where there is no information about which model is best or worst.

Since ESD is cheap computation-wise, it is well suited for downscaling large GCM ensembles.

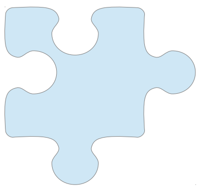


Empirical Orthogonal Functions (EOFs)

Different 'flavours' - all a mathematically form of **eigenfunctions on the data space**.

Provide maps of spatially coherent anomalies and time evolution of these.

Also known as principal component analysis (PCA), and belong to the field of linear algebra.

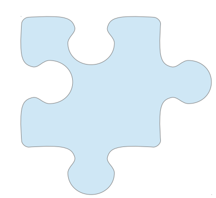


Non-stationarity

Relationships and dependencies change.

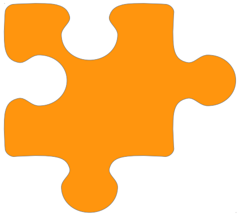
$y \neq f(X)$

Possibly because f(.) is not well-defined, not identified, not all factors are accounted for, or because the equation is ill-defined and there may not be a real dependency.



Empirical Statistical Downscaling

The philosophy



Philosophy

Why?

What is meant by 'scales'?

Inter-scalar dependencies

The problem of downscaling

The analysis aspect

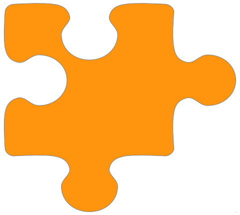
The requirement of predictors

Global climate models - 'GCMs'

Statistics

ESD advantages & drawbacks

Probabilistic & large ensembles

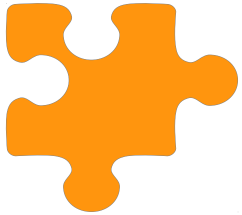


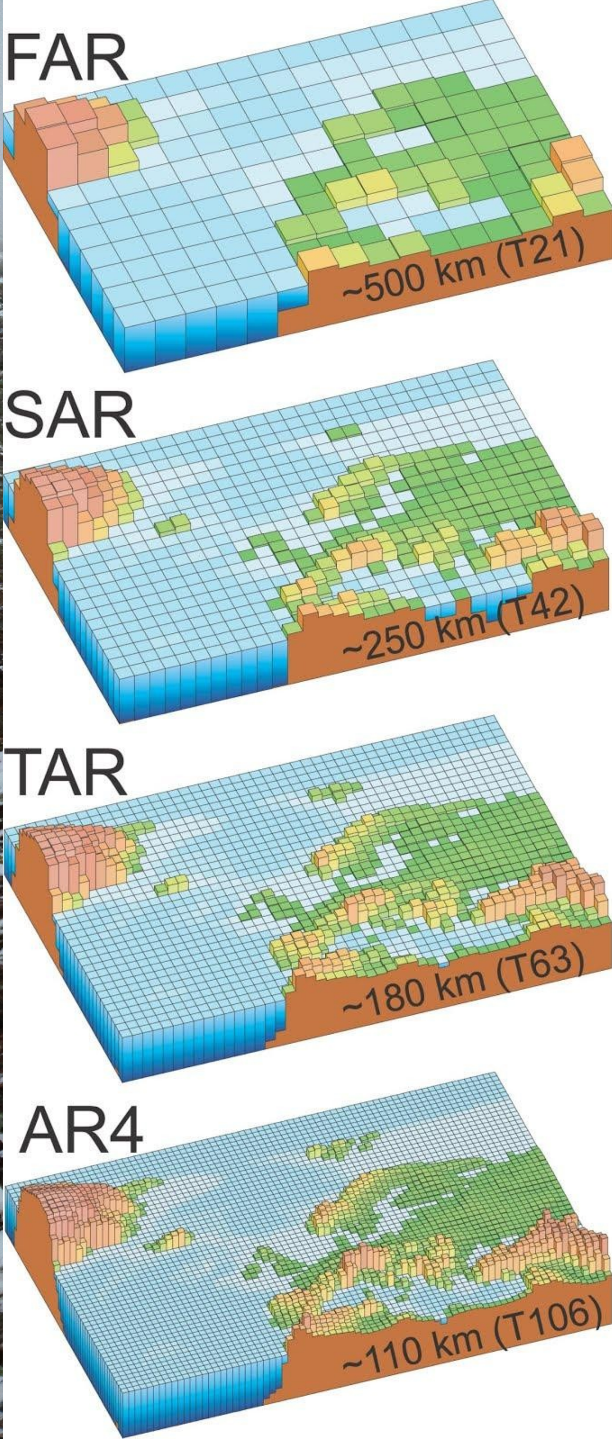
Why downscaling?

Models provide the 'large picture' and describe the essential aspects & processes. They do not represent all the details.

If there are dependencies between large-scale structure and small-scale details, then there is redundancy with respect to information content.

This information can be utilised to infer some of the characteristics about small scales.




FAR
~500 km (T21)
SAR
~250 km (T42)
TAR
~180 km (T63)
AR4
~110 km (T106)

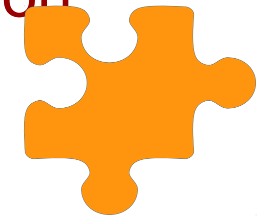
What is meant by 'scale'?

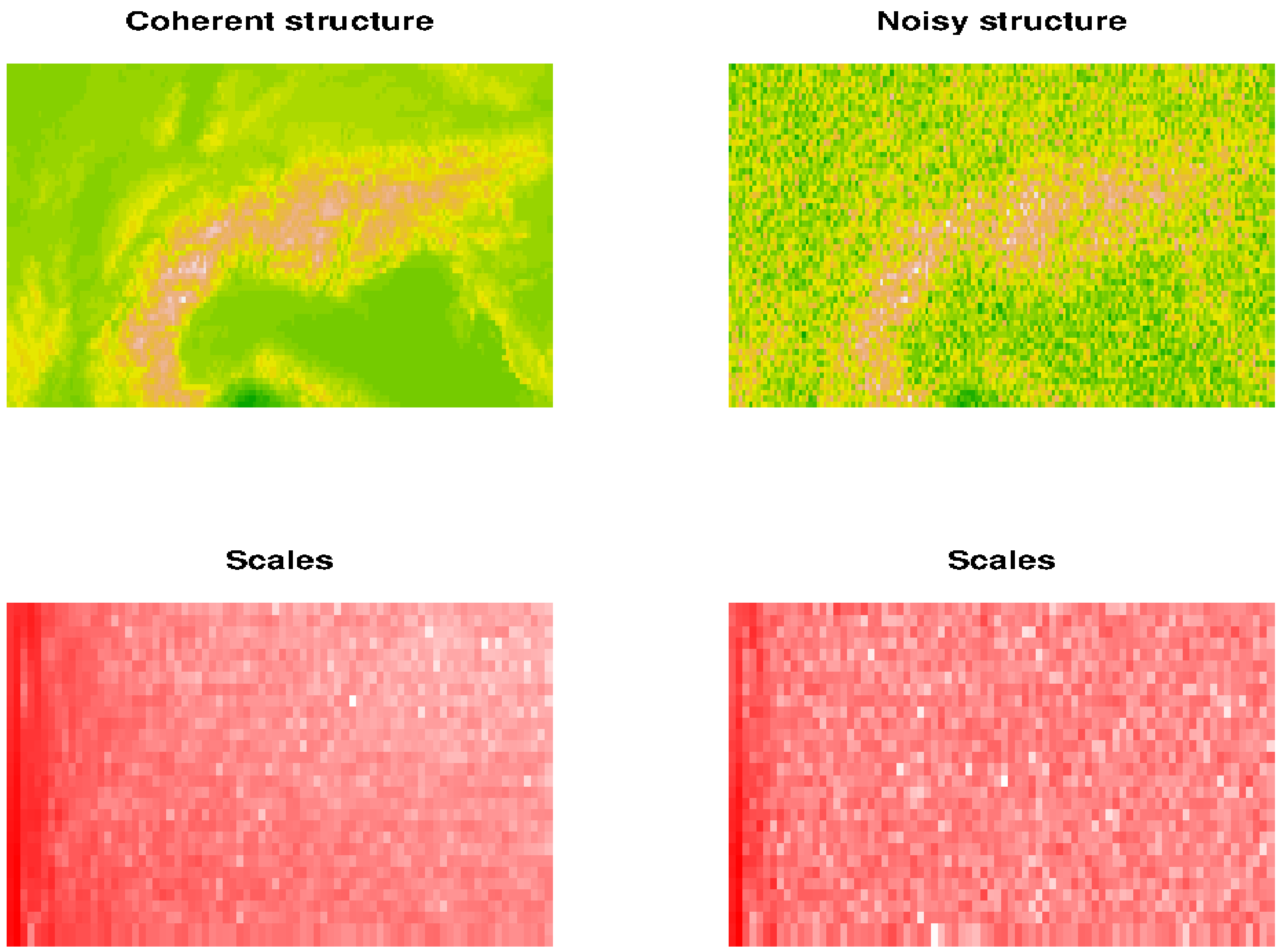
Description of the physical size or extent of an object, event, or phenomenon.

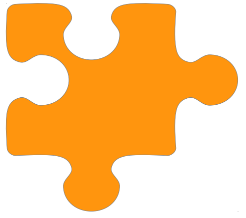
Statistics: scale parameter describes the degree of the spread in a distribution.

Physical size is related to model resolution and their minimum skillful scale (MSS).

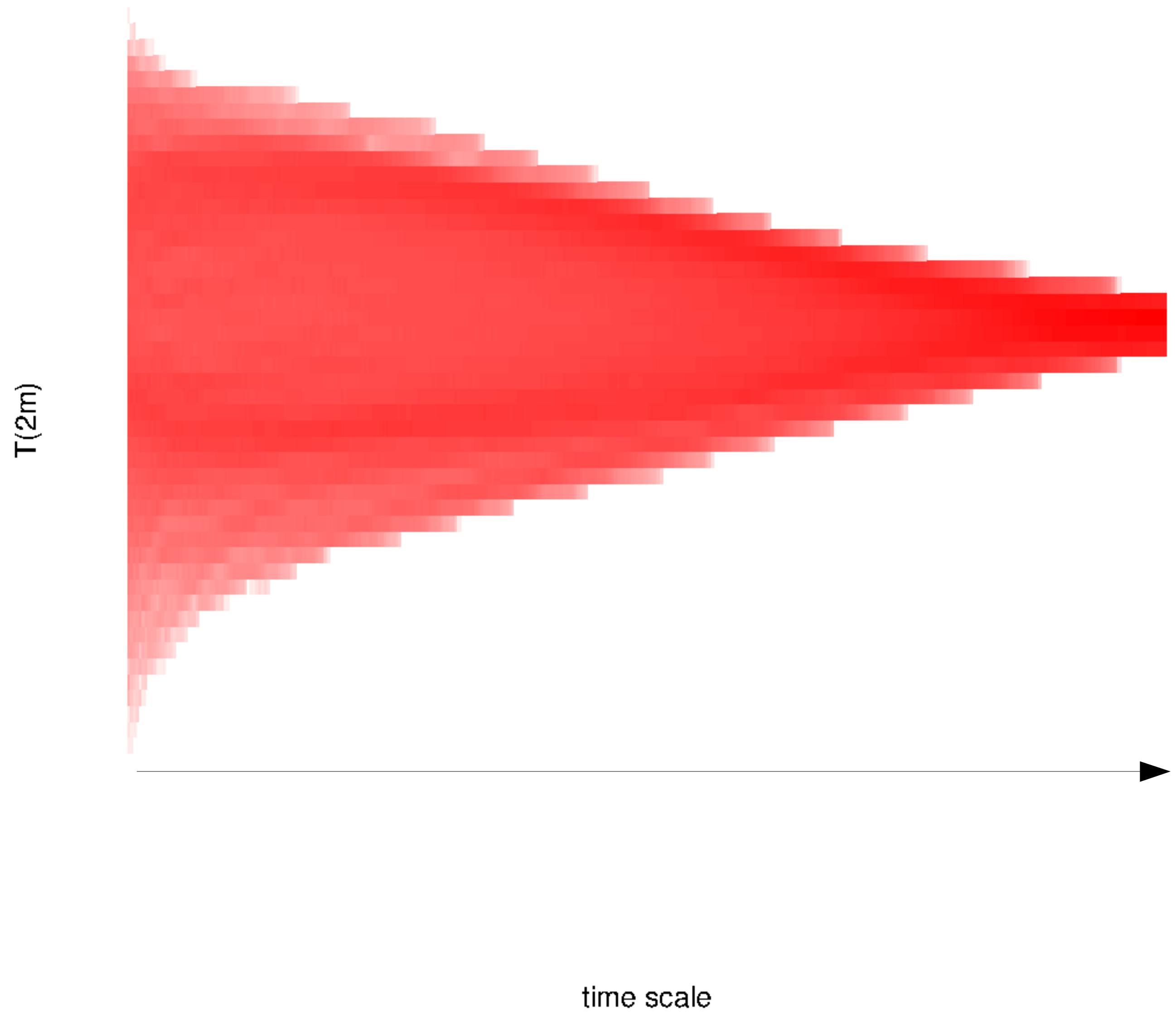
MSS = f(discrete numerical algorithms, parameterisation spatial/temporal resolution, representation of physical processes)

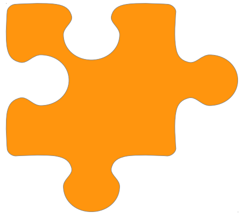



Coherent structure
Noisy structure
Scales
Scales

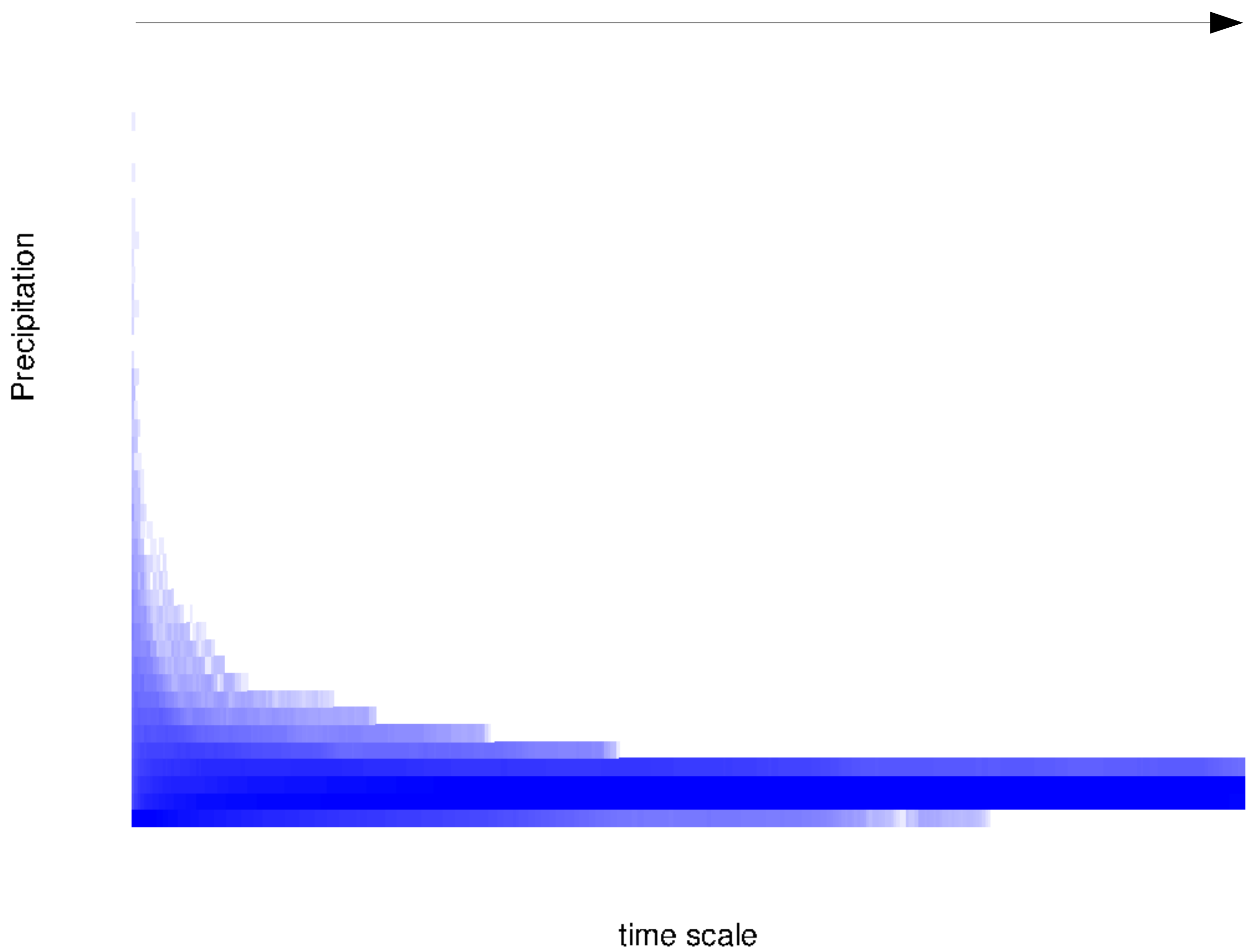


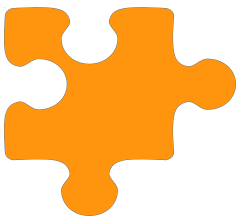
Probability distribution function





Probability distribution function



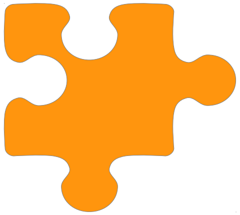


Inter-scalar dependencies

Smooth variation in space/time rather than abrupt: non-zero autocorrelation.

Persistence and spatially smooth functions.
The climate system has some redundant information, which can be described by models.

Deterministic models and Bayesian statistics/conditional probabilities.



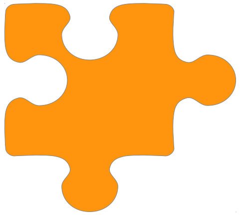
The problem of downscaling

A case of inter-scale dependency, where one tries to identify the link between large and small spatial scales.

Not necessarily a one-to-one correlation.

Local climate is influenced by local geography, large-scale situation, and local small-scale processes (noise):

$y = f(g,X) + n.$

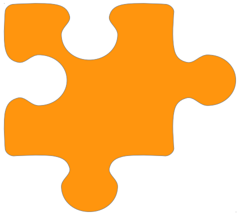


The analysis aspect

Downscaling will provide a description - a number - of some local quantity.

For ESD, this is a result from a more comprehensive analysis, which can shed light on interactions and enhance our physical understanding.

There is also a validation aspect.



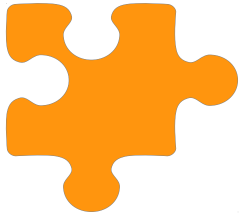
The requirement of predictors

Must capture the 'signal' - contain information about the essential variations.

Must be well-predicted.

Should have a strong connection with the predictand.

Should be part of a stationary link.



Global climate models - 'GCMs'

Coupled general circulation models of at least atmosphere and oceans.

Describe the dynamics as well as thermodynamics, in addition to simple ways for representing the effect of unresolved small-scale processes (parameterisation).

'Tuned' where parameterisation schemes are tweaked to give realistic results - within confidence intervals.

Radiative forcing and feedback processes most essential for climate change.

Issue of non-stationarity

Affects all model work - parameterisation schemes in GCMs/RCMs and ESD.

Physics-based models for which equations are known are less prone – the left and right hand sides of the equation:

$y = f(X)$