Understanding models and modelling experiments

Gab Abramowitz
Climate Change Research Centre, UNSW

Outline

- What is a model?
- Why are models so hard to use?
- Understanding modelling uncertainty
- Measuring model performance
- What makes a good model?
- Using models to investigate scientific questions

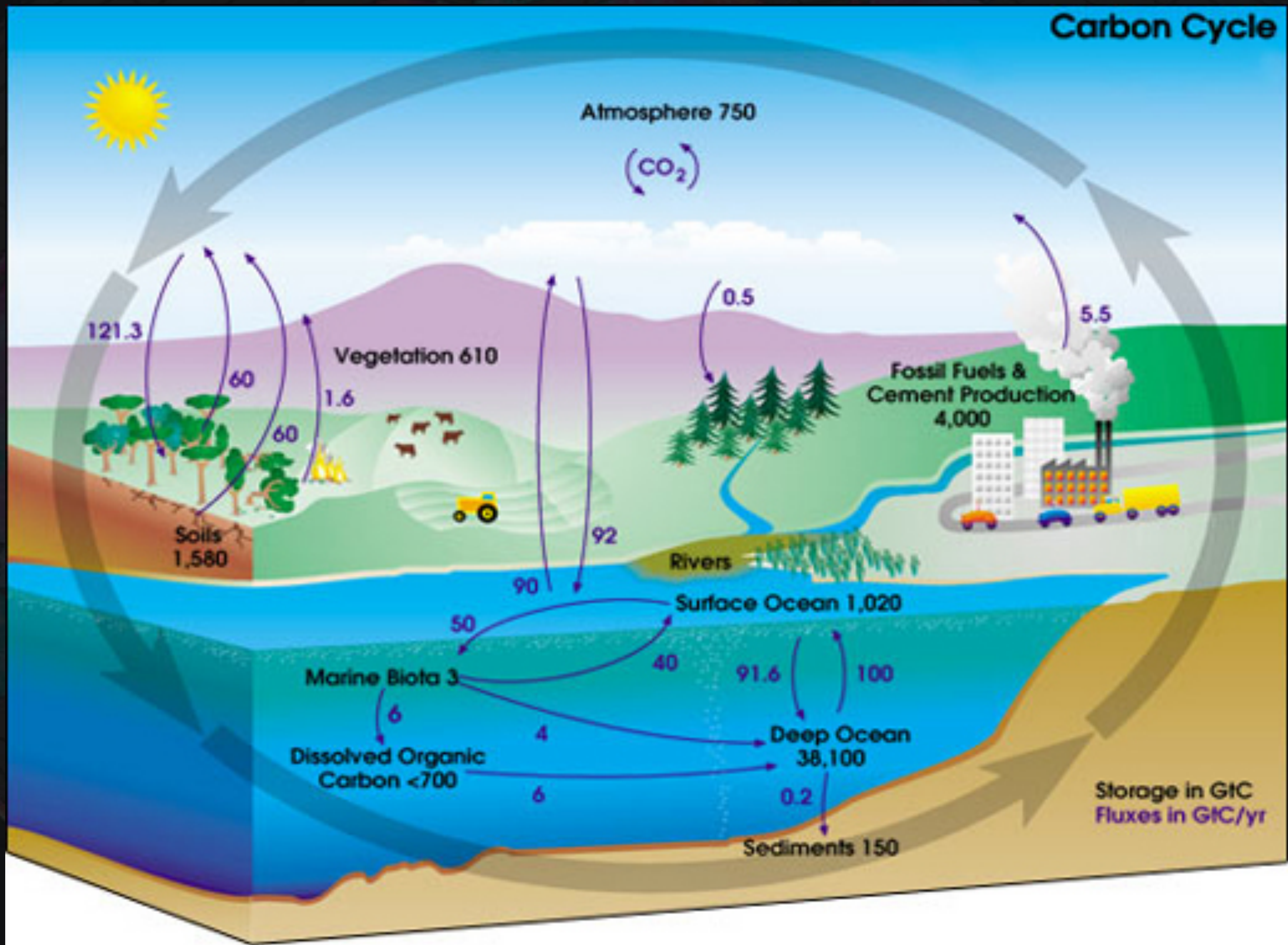
A model is not reality

- It is a scientist's job to understand how they differ.
- Models are **never** 100% correct.
- They can be very useful – how do we know when they're useful?
- Why are they not correct?
  1. A model is a closed system but simulates an open system (see Oreskes et al, Science 1994)
  2. Many relationships are based on empirical approximations rather than physical laws
     - These relationships are approximations based on a certain period in time, spatial scale, set of circumstances
  3. Their spatial and temporal aggregations mean that even physical laws need to be parametrised as 'net effects'.

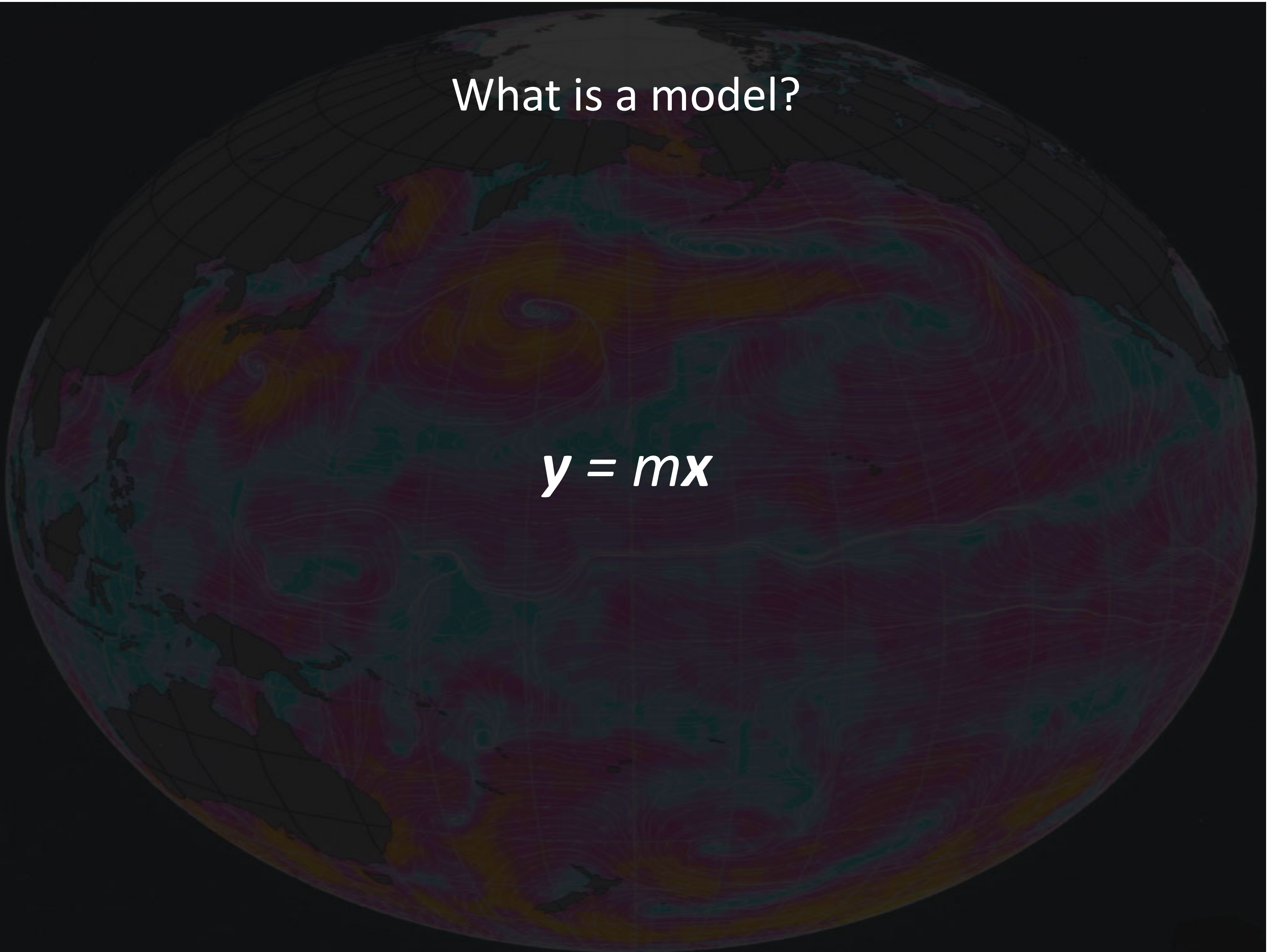
What is a model?

- A model is a simplified representation of a natural system.

What is a model?



What is a model?

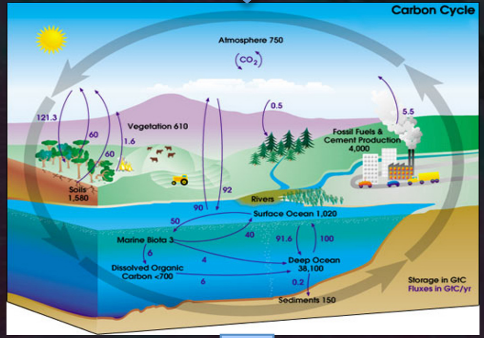


$$\boldsymbol{y} = m\boldsymbol{x}$$

Differences between models – steps in model development

- PERCEPTUAL MODEL – identify features of the system

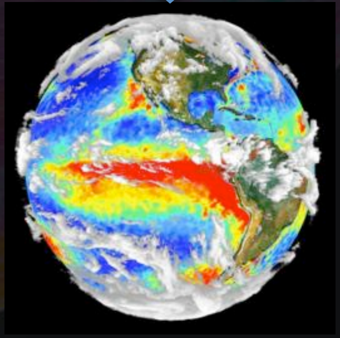
- CONCEPTUAL MODEL – identify relationships between features/processes in the perceptual model



- MATHEMATICAL/SYMBOLIC MODEL – identify equations that describe the conceptual model

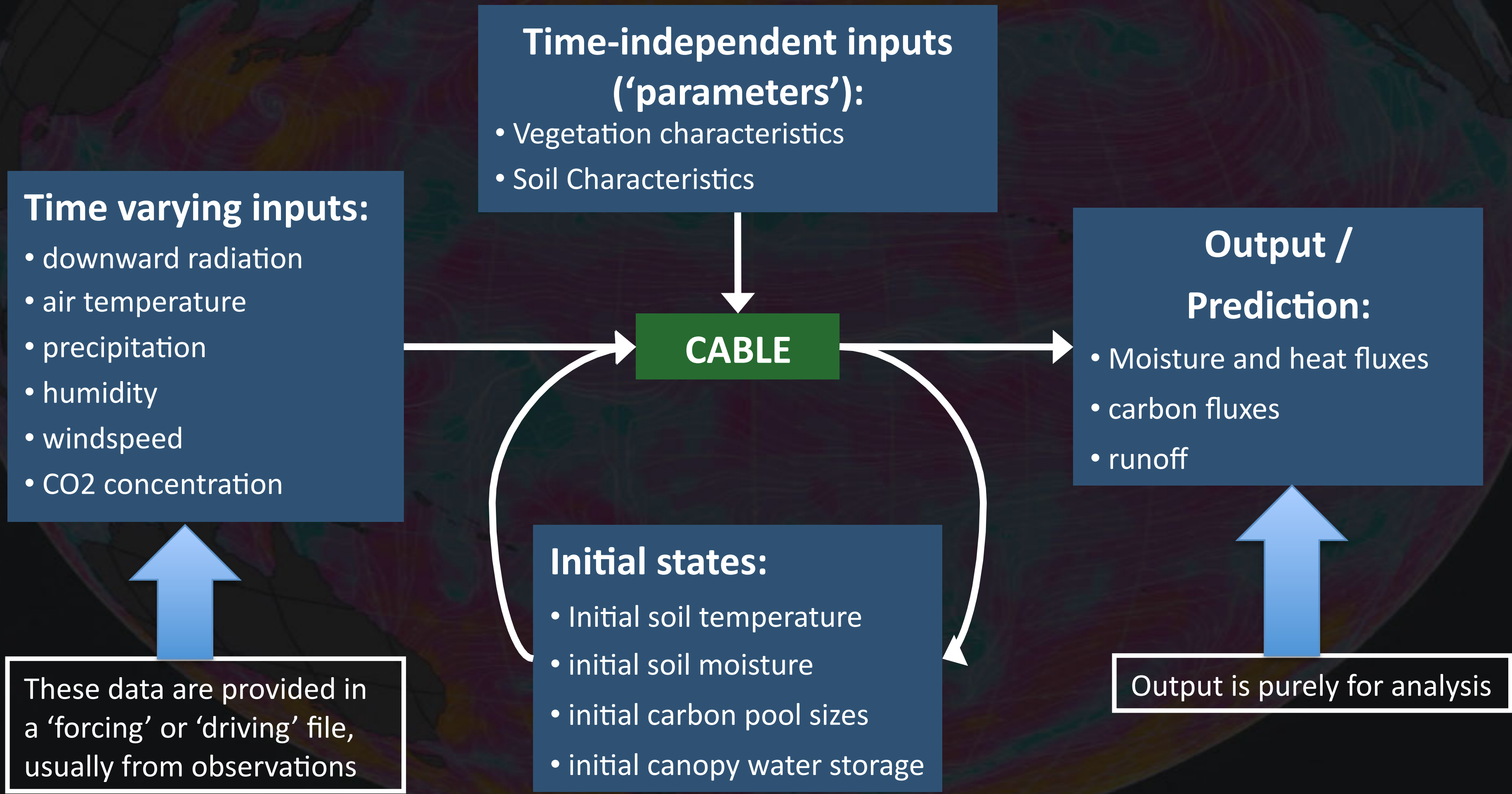
$$\frac{\partial(\eta_{Asat}\,\eta_{lf})}{\partial t} = \frac{\partial}{\partial z}\left(K_s \psi_s b \eta_{lf}^{b+2} \frac{\partial \eta_{lf}}{\partial z} - K_s \eta_{lf}^{2b+3} \right) + r(z).$$

- NUMERICAL MODEL – codification of equation solutions, spatial and temporal aggregation choices; implementation on a computer system.



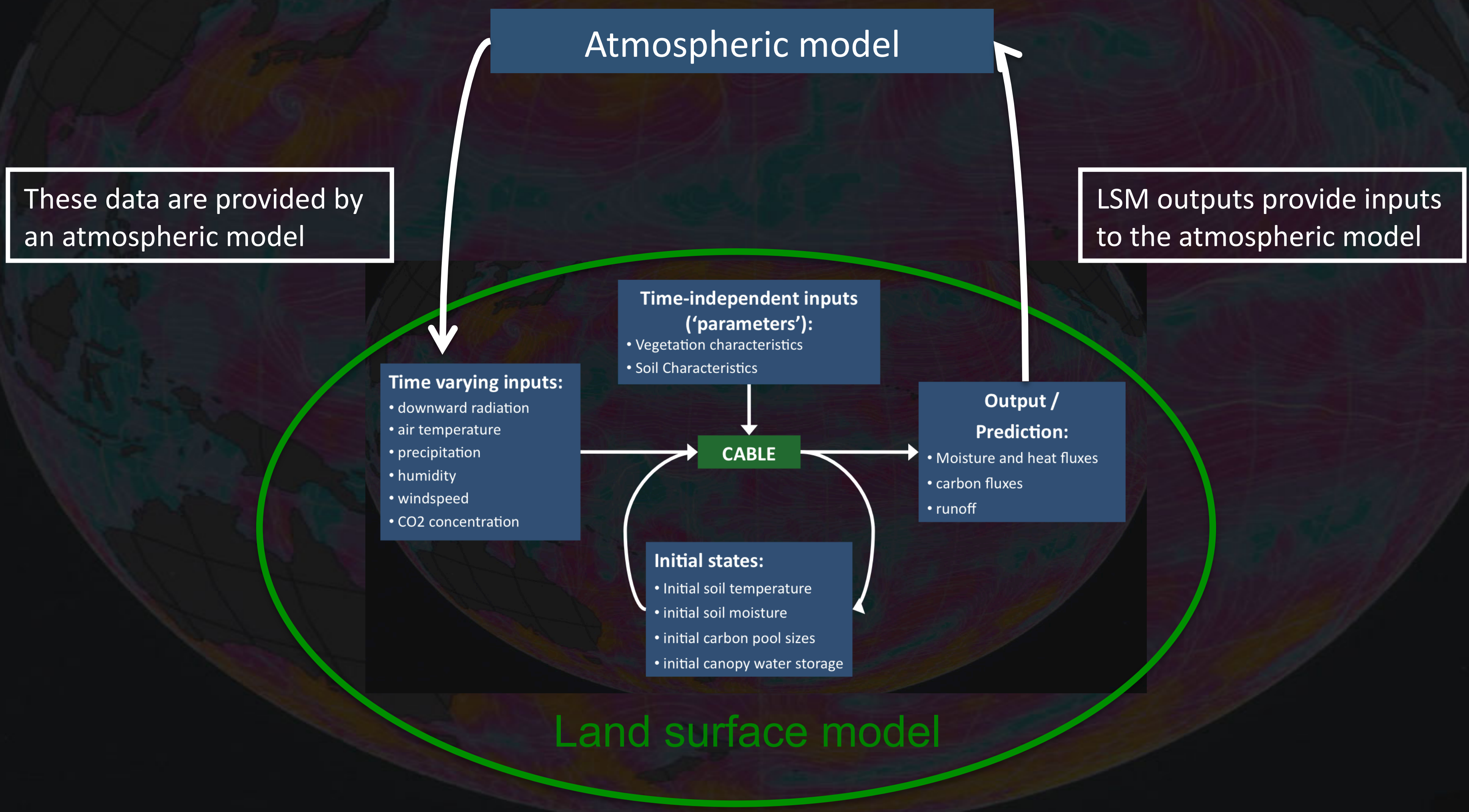
Coupled models and component (uncoupled) models

1. Uncoupled / offline / component models
e.g. **land surface model**

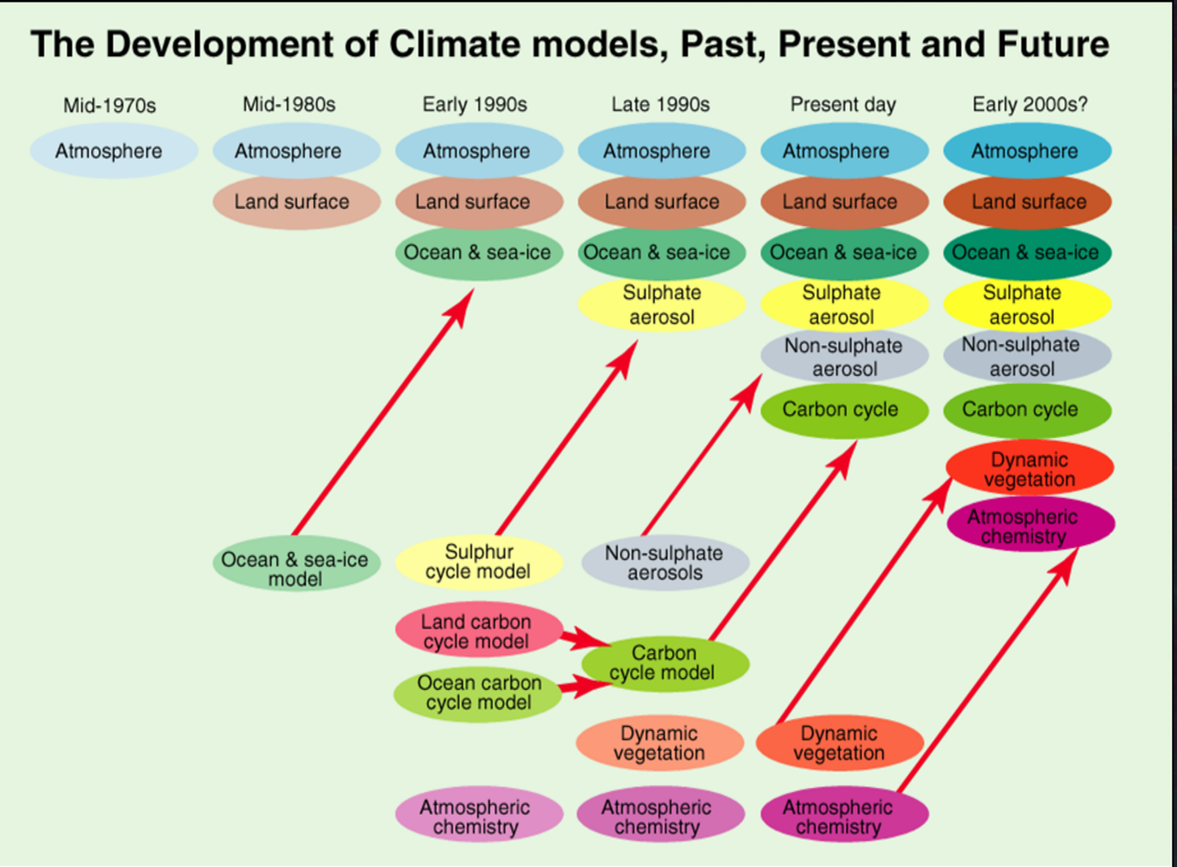


Coupled models and component (uncoupled) models

2. Coupled model system including a **land surface model**



Coupled models and component (uncoupled) models



The development of climate models over the last 25 years showing how the different components are first deveolped seperately and later coupled into comprehensive climate models (IPCC WG1, 2001)

Coupled models, e.g.

- Global Climate Model (GCM)
- Earth System Model (ESM)
- Earth system Model of Intermediate Complexity (EMIC)
- Single column model

Examples: ACCESS model, Mk3L, CSIRO Mk3.6, CCAM, WRF

Behave very differently to component models

- May have chaotic response to smooth variations in initial conditions or parameter values

Empirical vs physically based models

- The basic idea: processes that are well understood are modelled using "physical laws". Those not so well understood are modelled using empirical approximations. Most models have a combination of both.
- All treatments – including "physical laws", are in some sense empirical
- "First principles" to "heavily parameterised" to "fixed" to "ignored"
- Known physical mechanisms (e.g. gas law) are relatively scale-independent – model will usually improve with increasing resolution.
- An empirical parametrisation must often make assumptions about functional dependence, spatial scale of importance, time scale of importance
  - This is one reason why changing coupled model resolution is so difficult – "tuning".
- Distinction between physical and empirical is based on 'free' unmeasurable parameters e.g. Ginzburg and Jensen – will talk about this towards the end.

Outline

- What is a model?
- Why are models so hard to use?
- Understanding modelling uncertainty
- Measuring model performance
- What makes a good model?
- Using models to investigate scientific questions

Why are models so hard to use and understand?

- Most models are written for their creator(s) and not anyone else.
- In coupled models, each component is usually written by a different person/group
- Most funding agencies do not view model development or model refinement as core science – they are just tools
- Most research organisations do not view model documentation / model support as core science
- Unfortunately, this community loves Annoyingly Cryptic References Or Names that Yield Meaninglessness, and use them whenever they can
- The blue-red model spectrum:

Why are model so hard to use and understand?

Model has technical documentation	Model has no technical documentation
Technical documentation matches what is in the model code	Technical documentation related to what was in the code 5 years ago
Model is open source, community oriented and has hundreds of users	Model is only used by a few people in one organisation
All development of the model is contained in a version control system	Individuals maintain and manage multiple versions in home directories/desktop
Model has a clear user interface and user guide	Model has no user guide and no specific interface
Code is clearly commented, and logically structured	Code is not commented at all and structure is ad hoc
Variable names are consistent throughout the code and relate to their function	Variable names change in each subroutine call and are meaningless
Model changes meet prescribed performance/realism/functionality checks	Changes are accepted purely on the basis of personal preference.

Outline

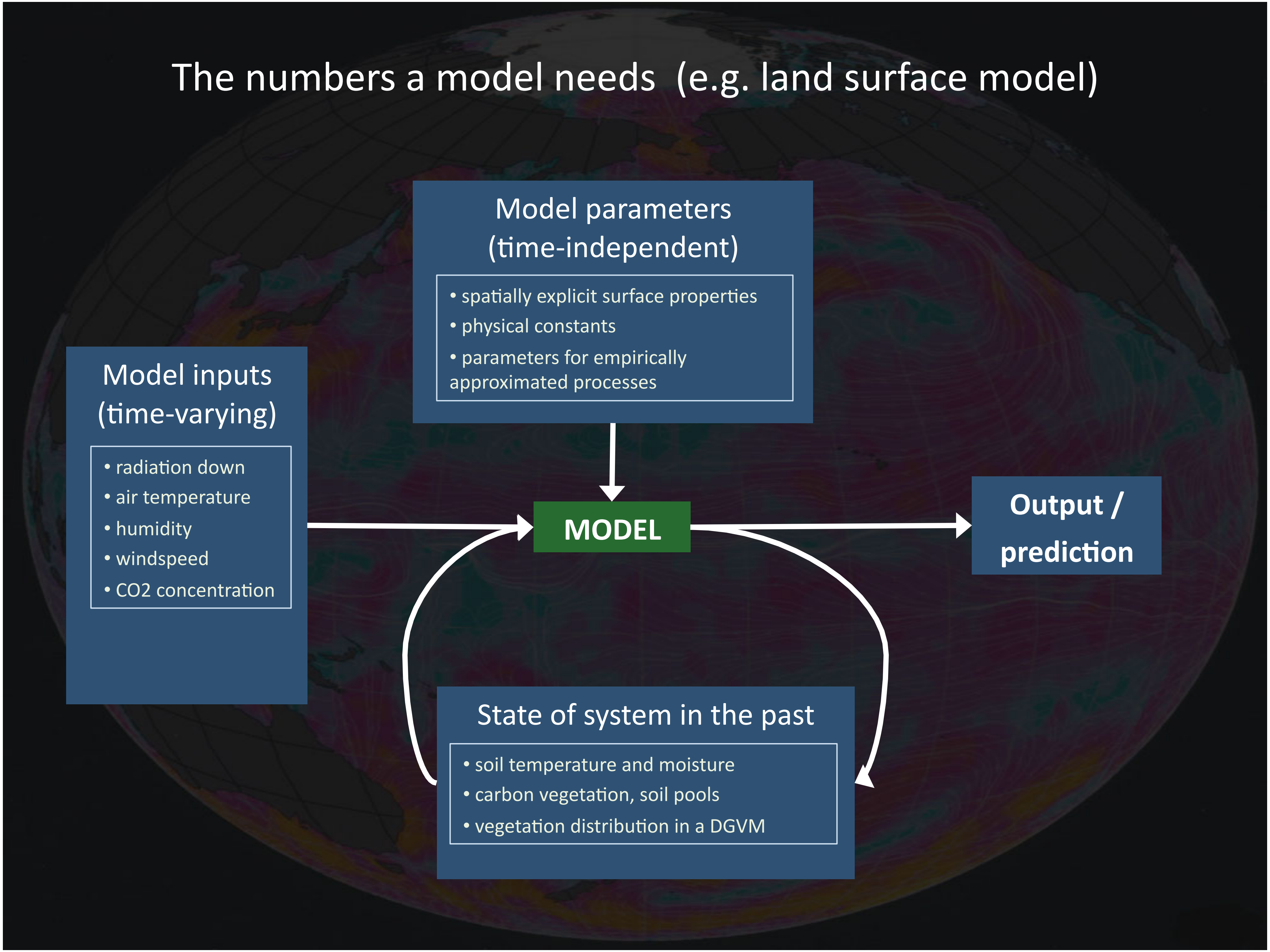
- What is a model?
- Why are models so hard to use?
- Understanding modelling uncertainty
- Measuring model performance
- What makes a good model?
- Using models to investigate scientific questions

Uncertainty

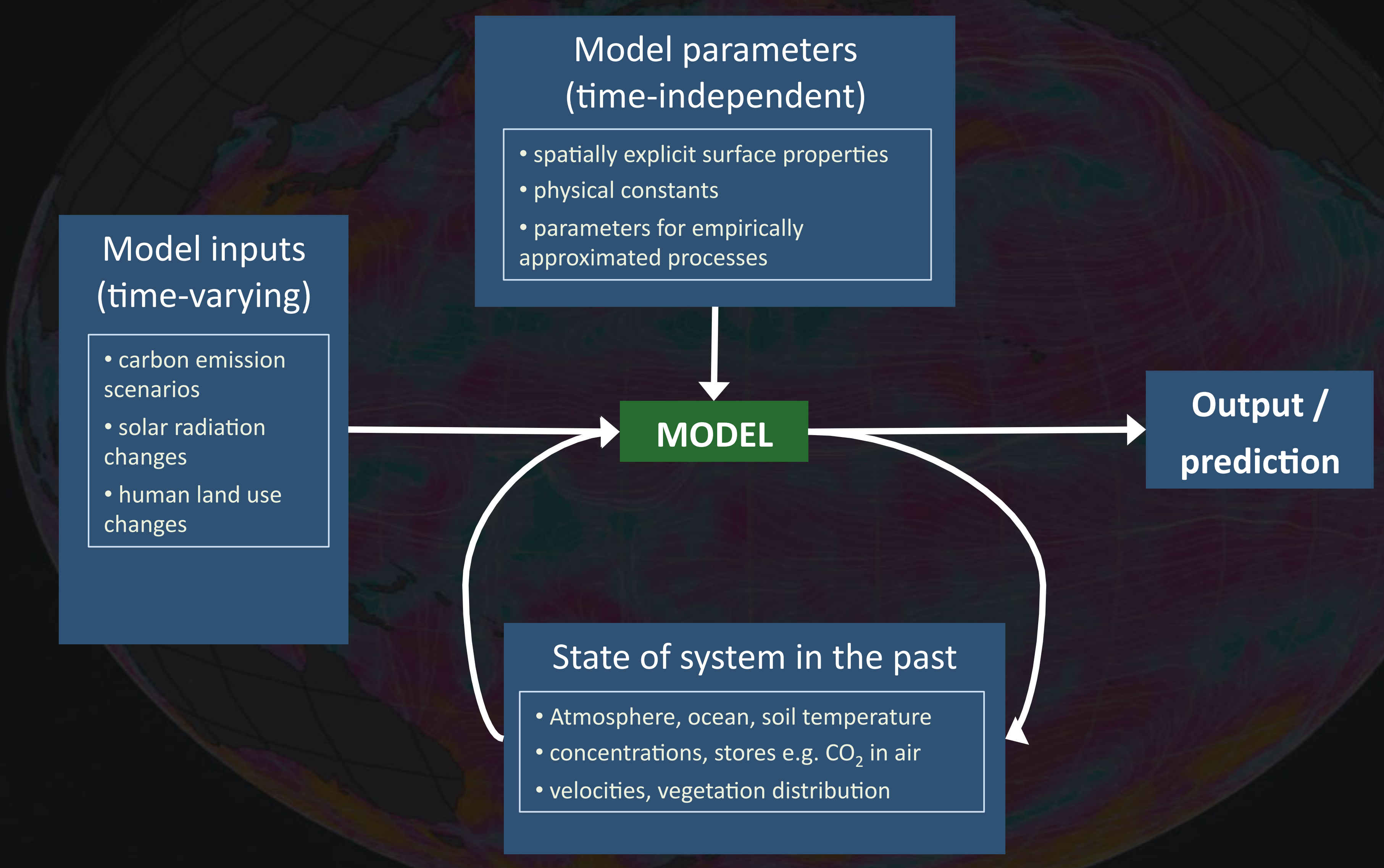
- Sources of uncertainty
  – Input uncertainty
  – Initial state uncertainty
  – Parameter uncertainty

- Dealing with uncertainty: ensemble simulation and stochastic variables

- Model space uncertainty
  – Model independence


The numbers a model needs (e.g. land surface model)
Model parameters (time-independent)
• spatially explicit surface properties
• physical constants
• parameters for empirically approximated processes
Model inputs (time-varying)
• radiation down
• air temperature
• humidity
• windspeed
• CO2 concentration
MODEL
Output / prediction
State of system in the past
• soil temperature and moisture
• carbon vegetation, soil pools
• vegetation distribution in a DGVM

The numbers a model needs (e.g. coupled climate model)



Why distinguish between parameters and inputs?

- I assert that some real world relationship is well approximated by a linear model: ***y***=*m**x*** (*m*: parameter, ***x***: input)

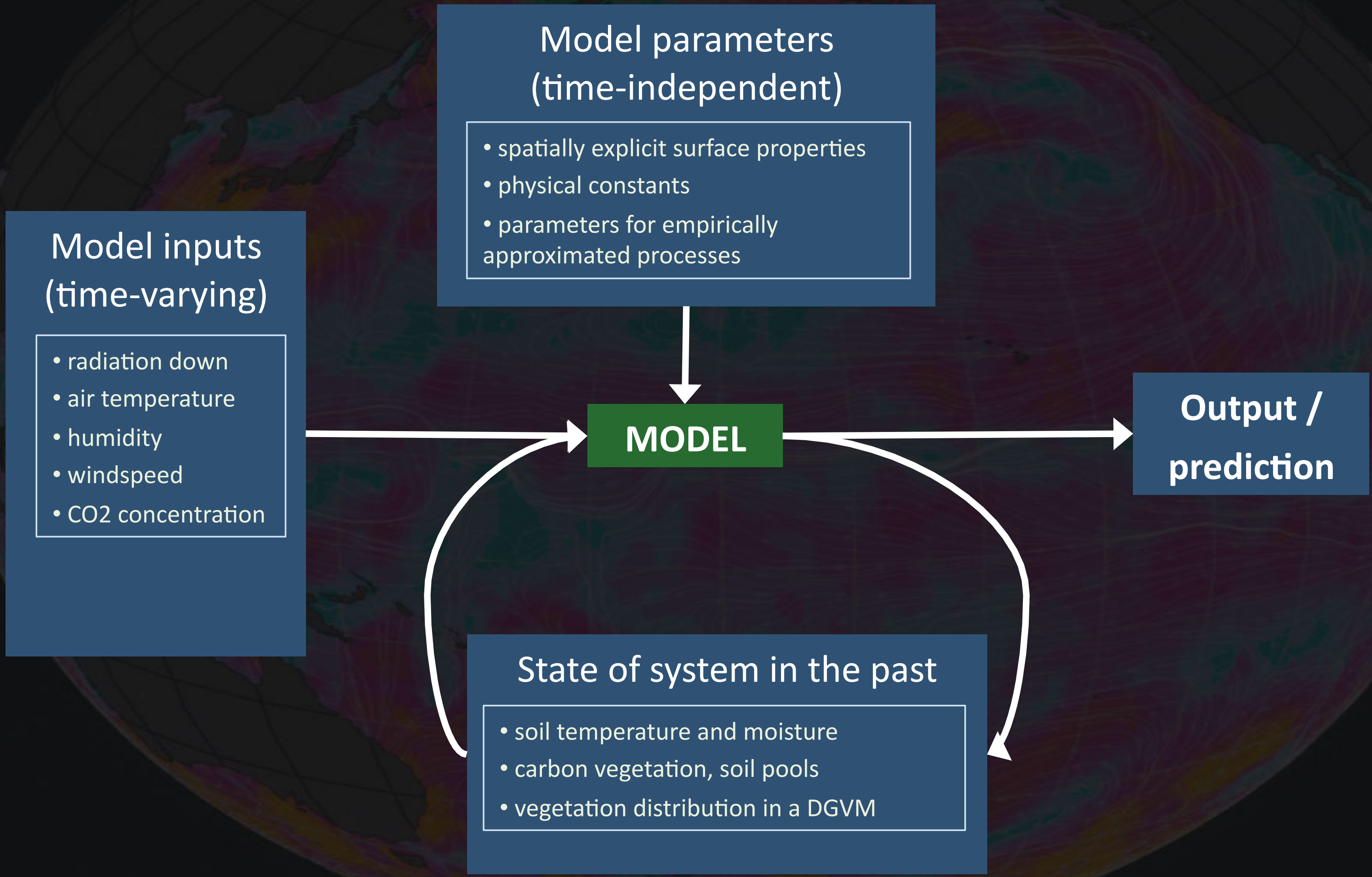
- After comparing with observed data I then suggest that *m* needs to vary with time (i.e. that *m* is an input, rather than a parameter)

- My suggestion that I now need another model to model *m* in time is an implicit admission that my linear model has failed to give adequate insight into the relationship between ***y*** and ***x***.

- By holding on to my original model I am just re-defining a non-linear problem.

- The distinction between inputs and parameters is a fundamental part of a model's defintion - a "good" model separates parameters and inputs appropriately

Sources of uncertainty



Uncertainty in inputs

Model inputs (time-varying)

- radiation down
- air temperature
- humidity
- windspeed
- CO2 concentration

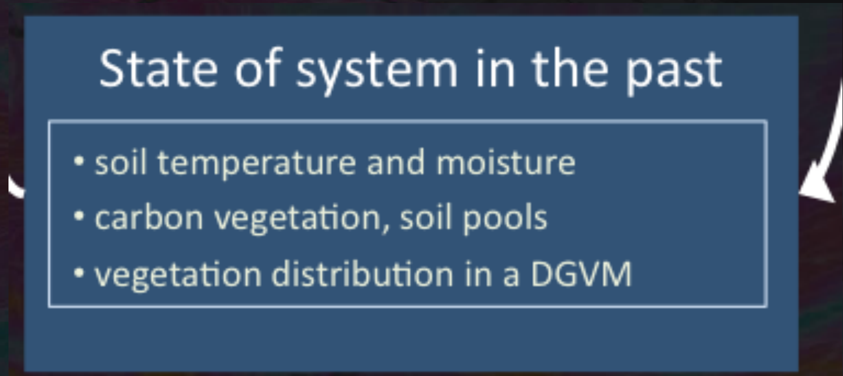
- Measurement uncertainty of inputs (offline):
  - Precision & accuracy
  - Instrument failure and gap-filling
  - Time aggregation

- If coupled to a climate model:
  - Errors/uncertainty from other model components
  - Cumulative effect of errors with coupled feedbacks

- How would you quantify this uncertainty in your final prediction variables?

Uncertainty in initial conditions

- Soil moisture and soil temperature
- Carbon pools in soil, vegetation
- Canopy water storage



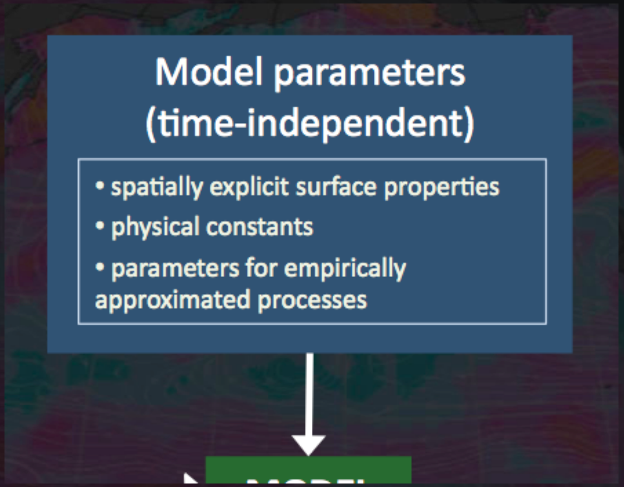
- Most commonly dealt with by 'spin-up' of models.
- Component models commonly have separate spin up first.
  - Convergence to reality?
  - Reflect model biases? State values become model-specific
  - What if we used "true" values? Would the model be stable?

- NWP use of soil moisture nudging in data assimilation

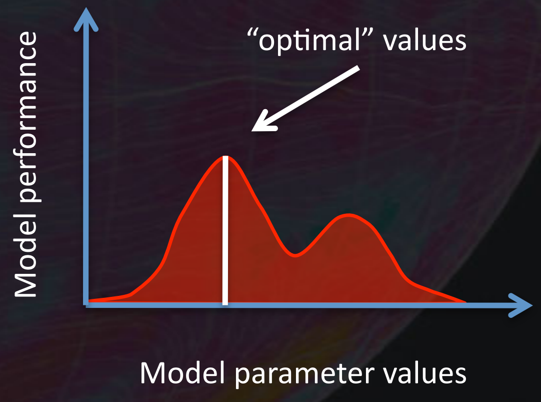
- How would you quantify this uncertainty in your final prediction variables?

Uncertainty in parameter values

- Spatially explicit vegetation and soil characteristics that are time-invariant for the simulation
- Physical constants
- Parameters for empirically approximated processes



- Most commonly dealt with by ‘calibrating’ (parameter estimation; ‘tuning’ etc):
  - Can be automated parameter estimation (usually with a component model, e.g. land surface model)
  - Or manual “expert guess” calibration



Automated calibration – more detail

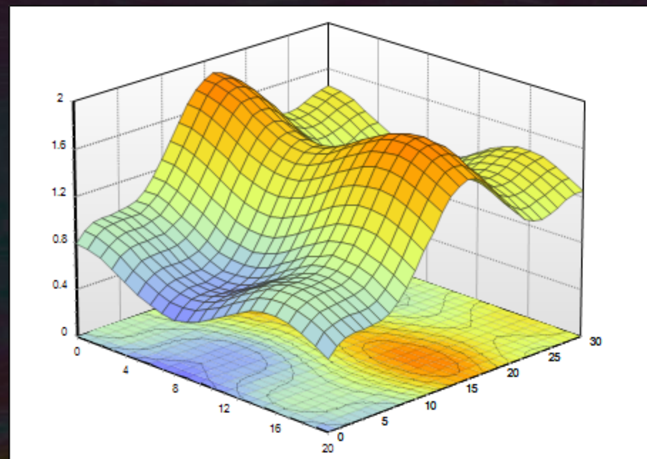
- Find an observed data set of a model output that is likely to give information about parameter values

- Select realistic ranges for parameter values

- Decide on a cost / error function

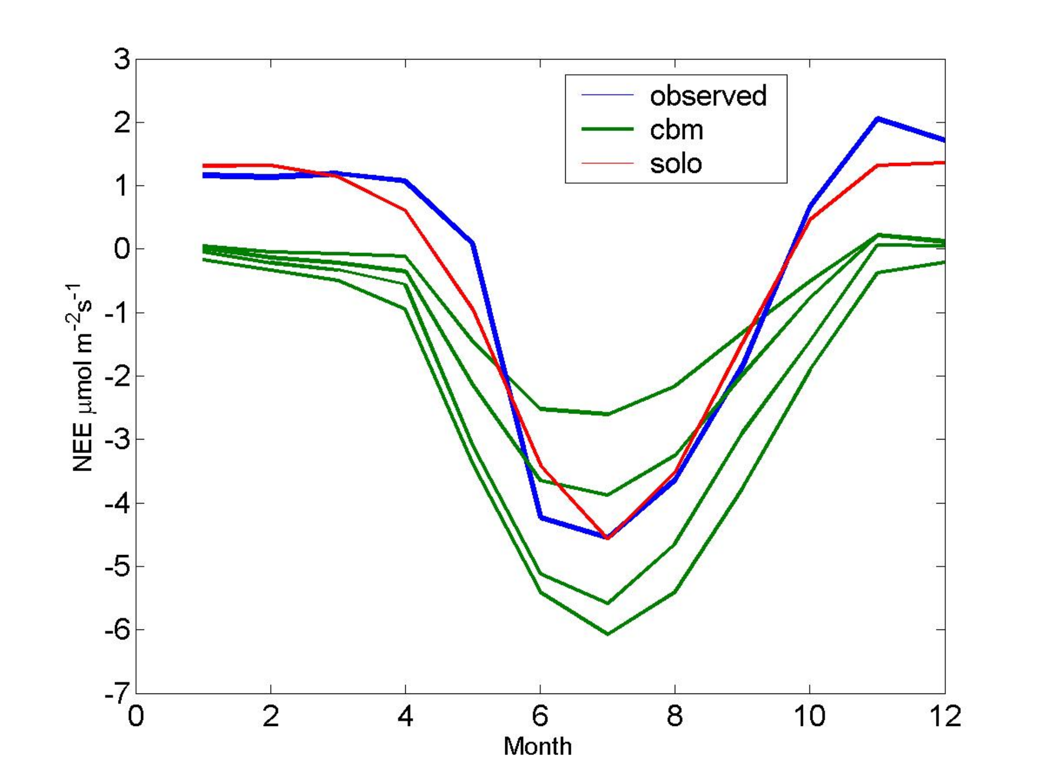
$$E = w_1 \sqrt{\sum_k \frac{(\mathrm{mod}_1 - obs_1)^2}{k}} + w_2 \sqrt{\sum_k \frac{(\mathrm{mod}_2 - obs_2)^2}{k}}$$

- Find the parameter values that minimise the cost function in this acceptable range



Automated calibration – potential problems

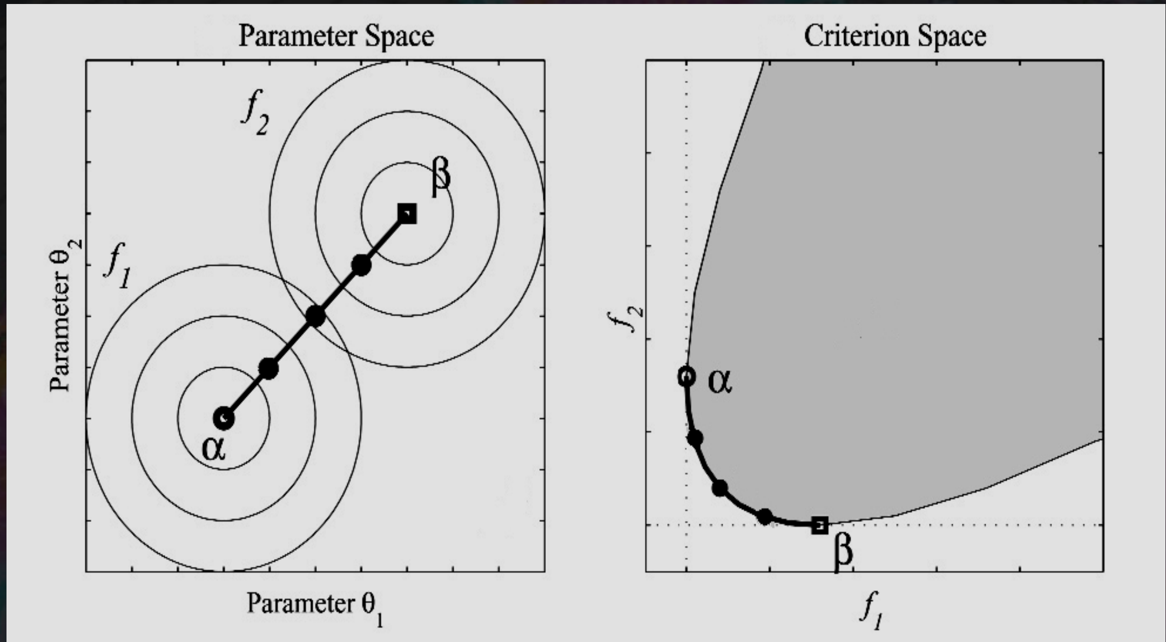
- Is there any guarantee that parameter values obtained in this way are meaningful?
  - Assumes the model is perfect
  - Values may be those that best compensate for model biases/errors



Automated calibration – potential problems

- Is there any guarantee that parameter values obtained in this way are meaningful?
  - Assumes the model is perfect
  - Values may be those that best compensate for model biases/errors
- Calibration often 'tunes' a model to particular data set, time period or cost function:
  - There is a danger of moving more toward an empirical model
- Is there any better way of selecting parameters? Probably not.
- Can give us insight into model deficiencies though…

Multiple criteria calibration

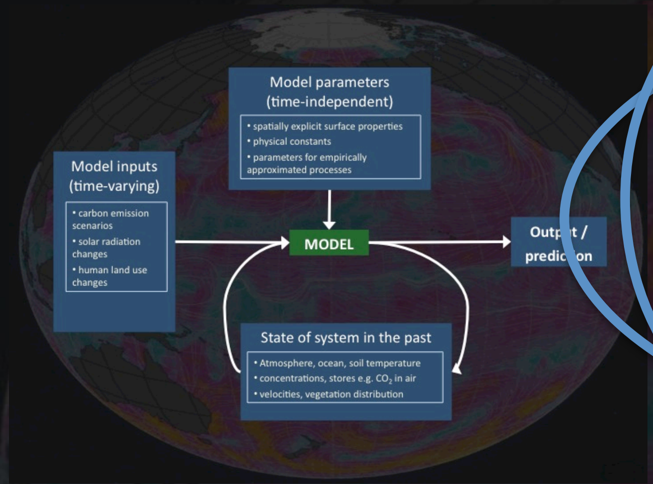


- Illustrates a common principle in modelling a complex system with a simple model
  - "conservation of crap"

- The amount of separation of the two minima is an indication of the parameter-independent error in the model.

- How can we account for parameter uncertainty?

Deterministic, ensemble, and stochastic simulation

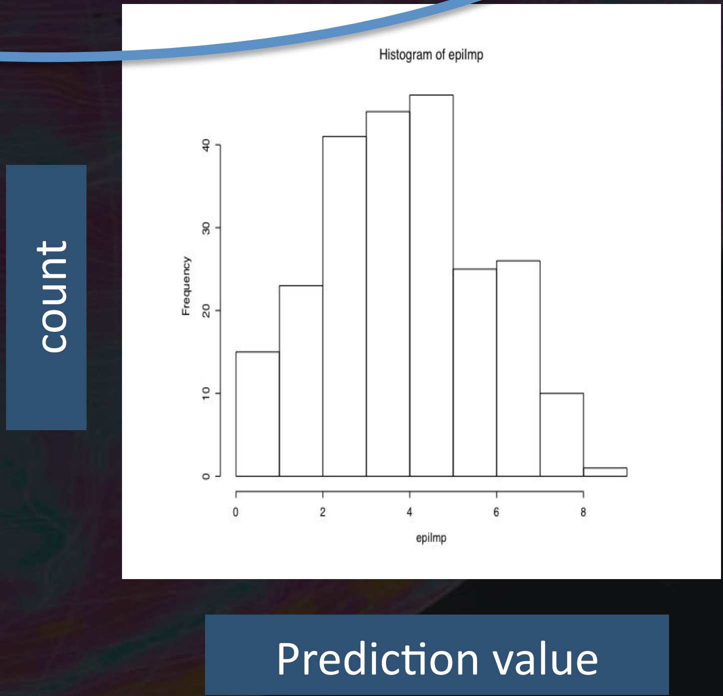


Model parameters 1

Model inputs 1

Initial states 1

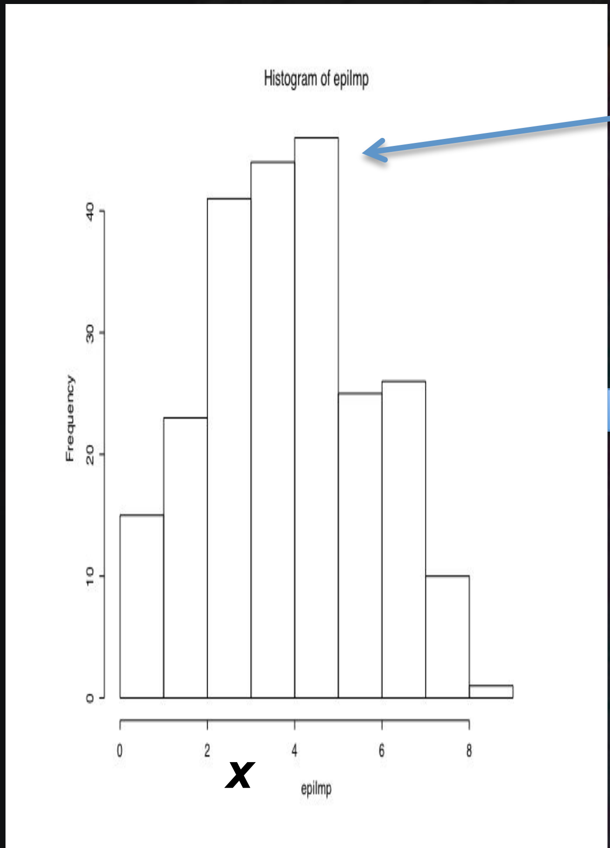
Prediction 1



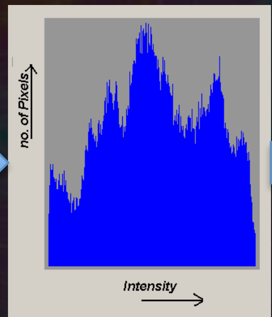
- Estimate different but equally plausible parameter values, input values and initial states

- Run model for possible combinations to get a statistical characterisation of the prediction: "**ensemble simulation**"

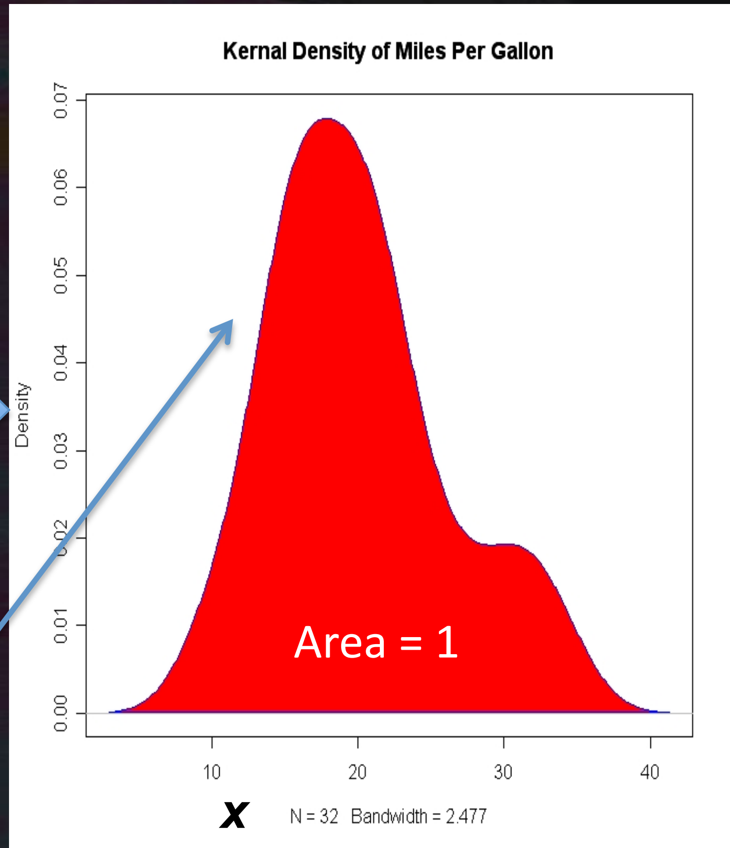
What is a probability density function?



Relative frequency of certain values of ***x***



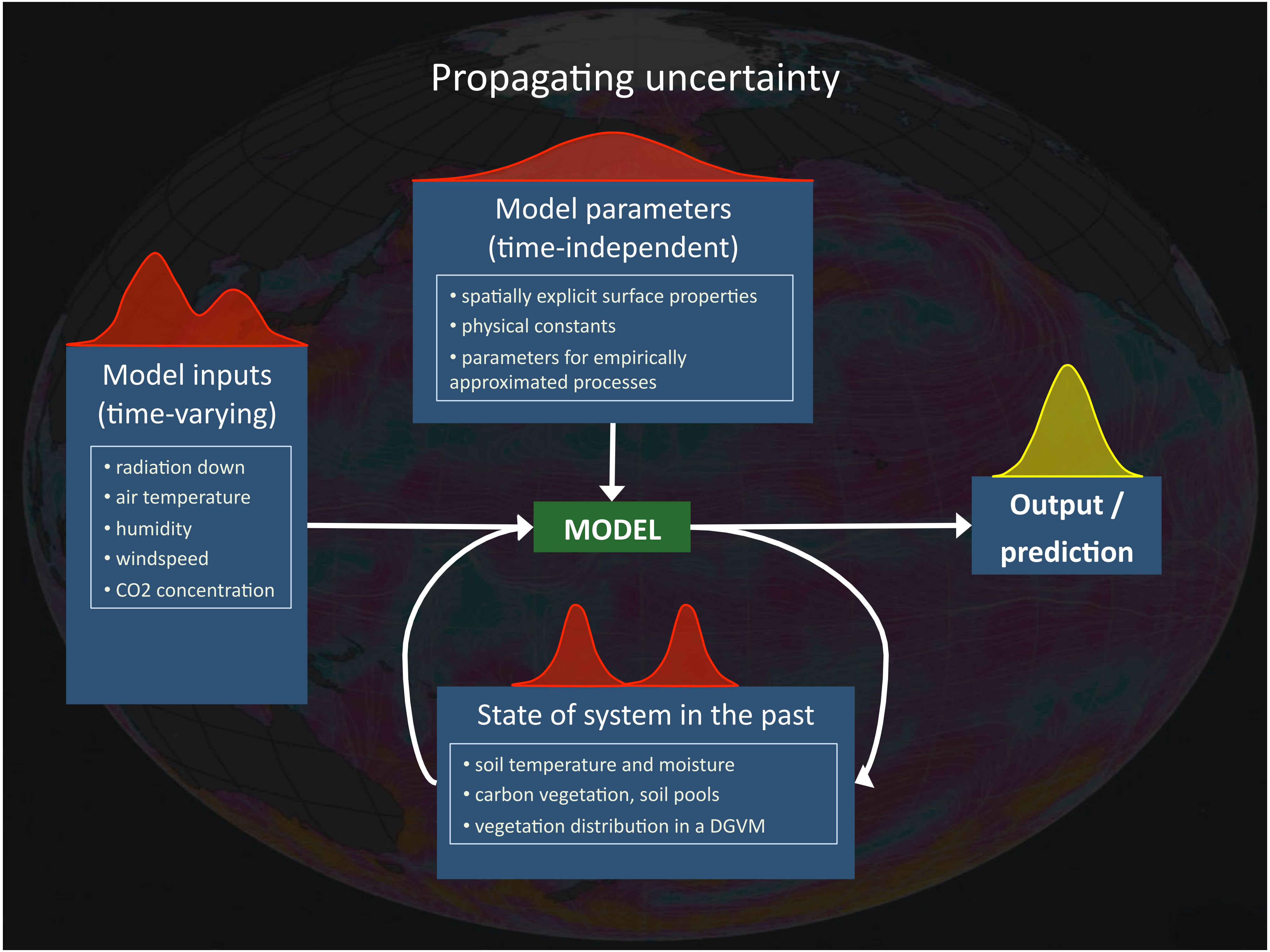
Relative probability of certain values of ***x***



- With many estimates of ***x*** we can estimate the probability density function (PDF) of ***x***

Deterministic vs. stochastic simulation

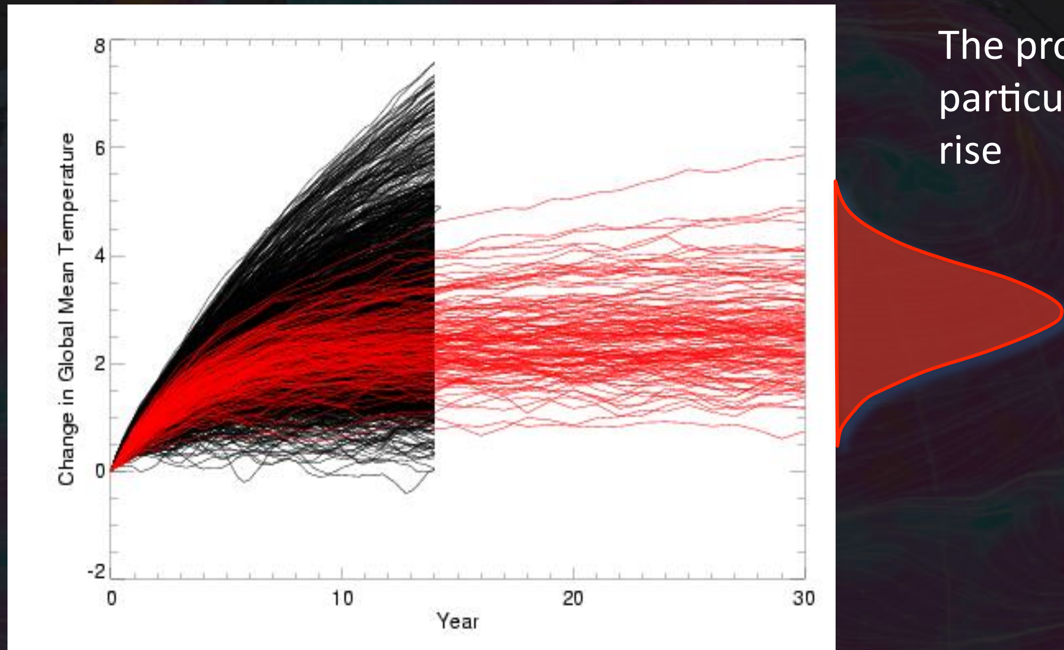
- A variable described by a PDF (rather than a single value) is called a ‘stochastic variable’ or ‘random variable’
- Particularly important for describing the probability of extreme events
- By better sampling uncertainties in a stochastic simulation we can be more confident about the predictions.
- Running a land surface model 10,000 times to generate a pdf is possible
  - e.g. I have a parameter perturbation driver for CABLE (just ask).


Propagating uncertainty
Model parameters
(time-independent)
• spatially explicit surface properties
• physical constants
• parameters for empirically approximated processes
Model inputs
(time-varying)
• radiation down
• air temperature
• humidity
• windspeed
• CO2 concentration
MODEL
Output /
prediction
State of system in the past
• soil temperature and moisture
• carbon vegetation, soil pools
• vegetation distribution in a DGVM

Quantifying uncertainty – ensemble simulations

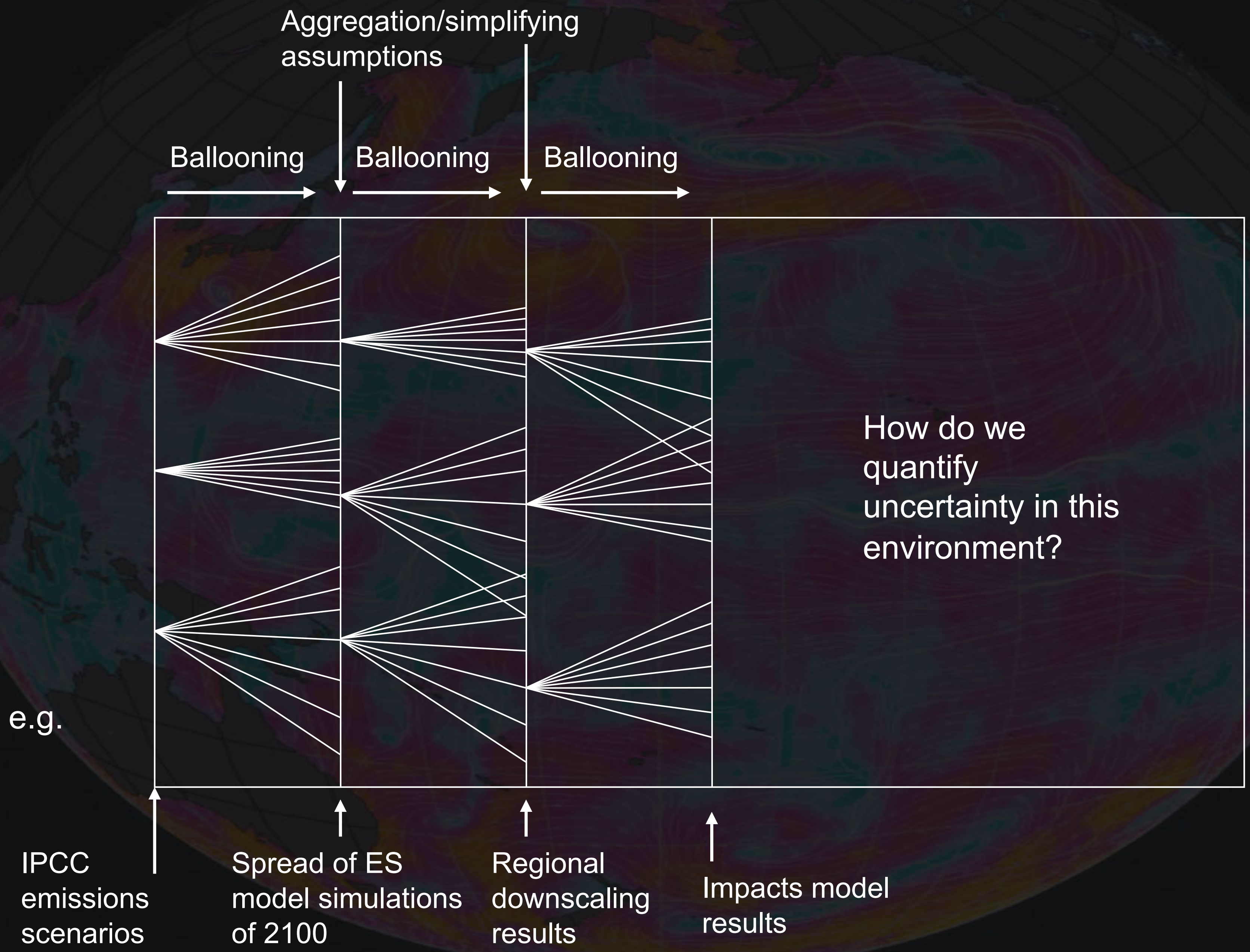
climateprediction.net

HadCM3, perturb parameter values, initial conditions



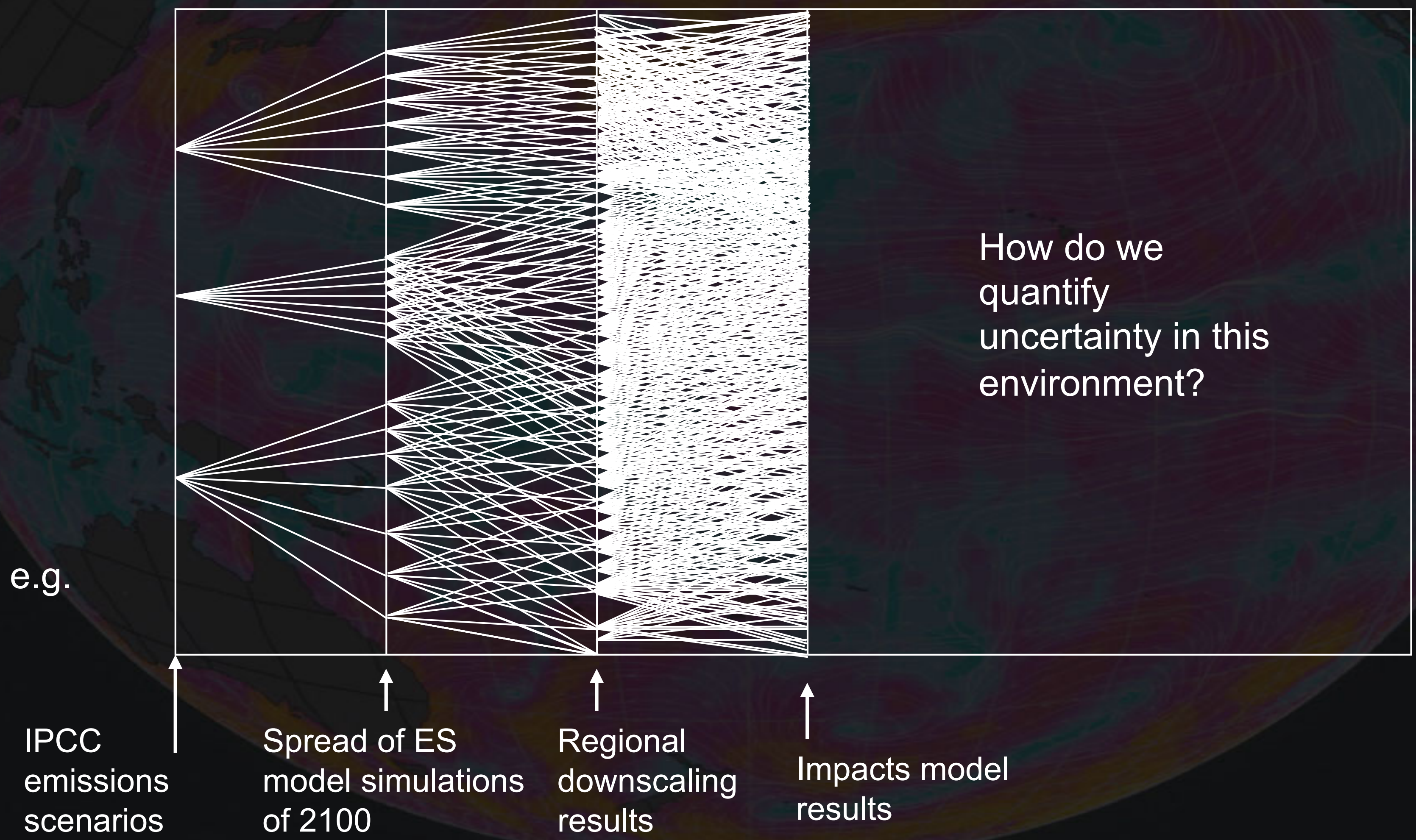
The probability of a particular temperature rise

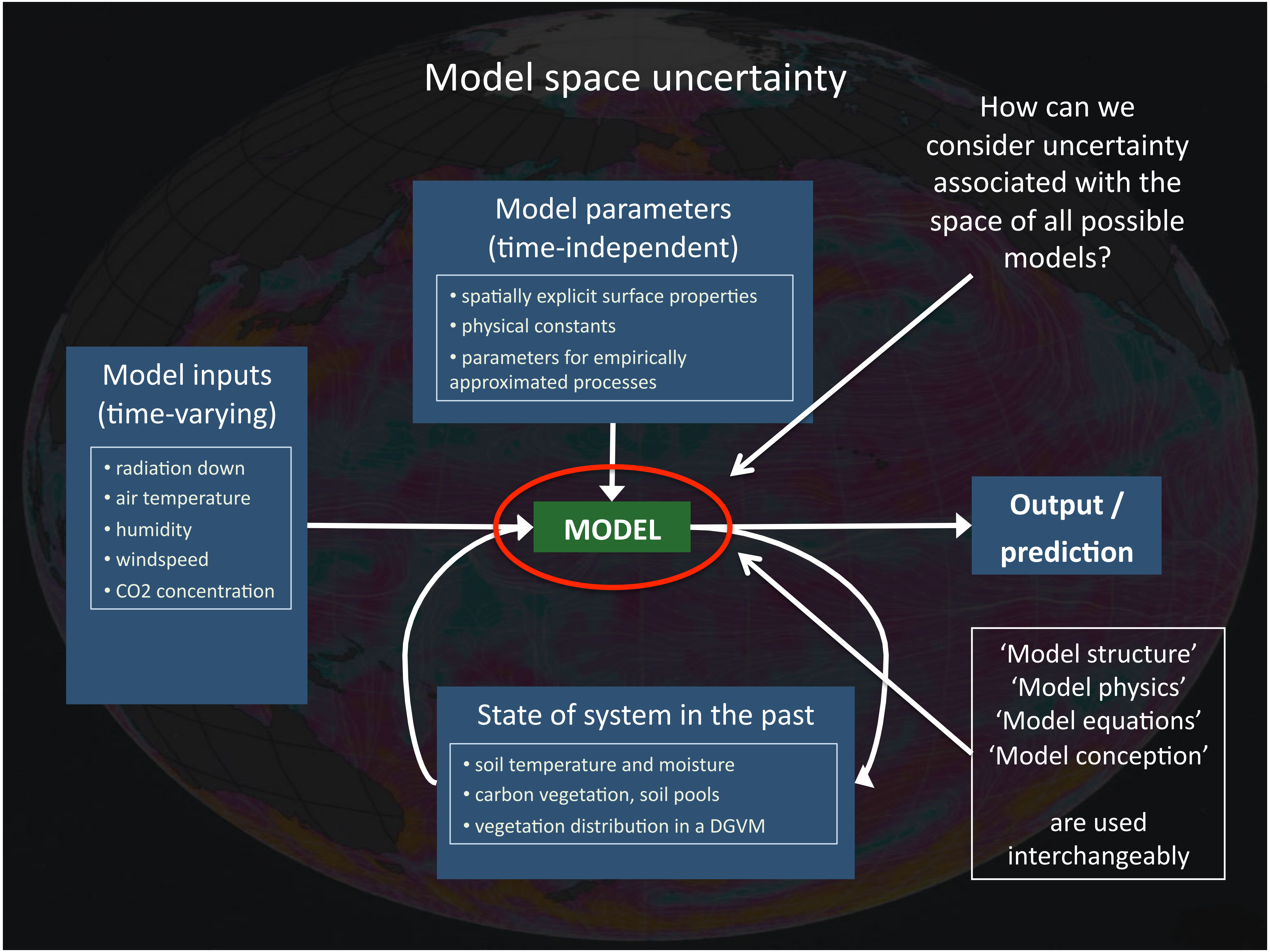
Cascading uncertainty in prediction problems

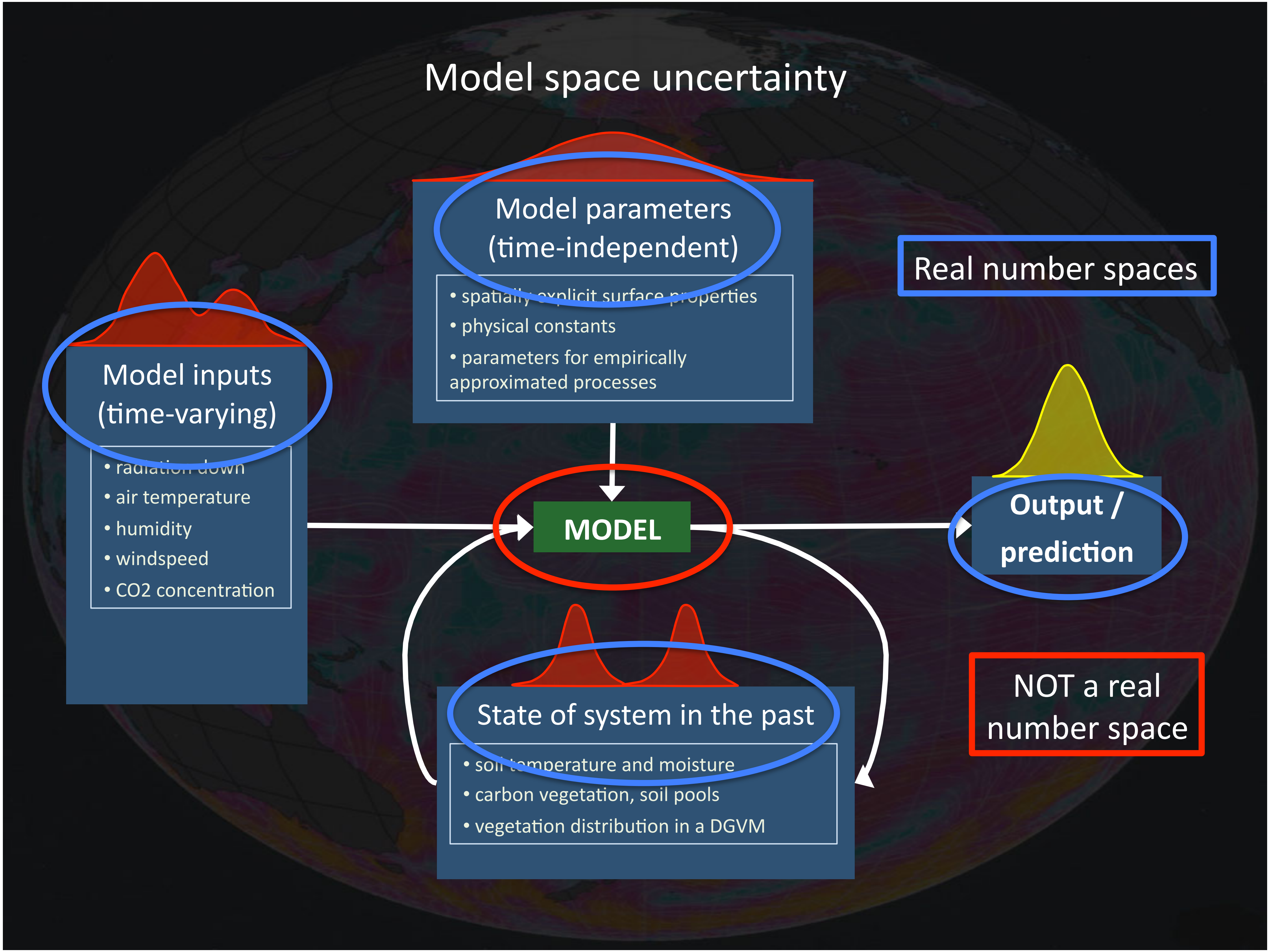


Cascading uncertainty in prediction problems

- Dealing with uncertainty comprehensively is very difficult
- Potentially very computationally expensive
- We're not very good at it yet




Model space uncertainty
How can we consider uncertainty associated with the space of all possible models?
Model parameters (time-independent)
• spatially explicit surface properties
• physical constants
• parameters for empirically approximated processes
Model inputs (time-varying)
• radiation down
• air temperature
• humidity
• windspeed
• CO2 concentration
MODEL
Output / prediction
State of system in the past
• soil temperature and moisture
• carbon vegetation, soil pools
• vegetation distribution in a DGVM
'Model structure'
'Model physics'
'Model equations'
'Model conception'
are used interchangeably


Model space uncertainty
Model parameters (time-independent)
• spatially explicit surface properties
• physical constants
• parameters for empirically approximated processes
Real number spaces
Model inputs (time-varying)
• radiation down
• air temperature
• humidity
• windspeed
• CO2 concentration
MODEL
Output / prediction
State of system in the past
• soil temperature and moisture
• carbon vegetation, soil pools
• vegetation distribution in a DGVM
NOT a real number space

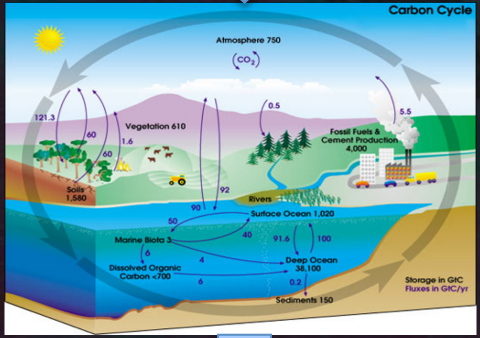
What’s in the model space?

- PERCEPTUAL MODEL – identify features of the system
  - Defines the variables in the input, parameter and state spaces

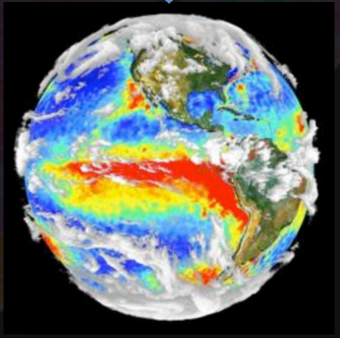
- CONCEPTUAL MODEL – identify relationships between features/processes in the perceptual model

- MATHEMATICAL/SYMBOLIC MODEL – identify equations that describe the conceptual model

- NUMERICAL MODEL – codification of equation solutions, spatial and temporal aggregation choices; implementation on a computer system.



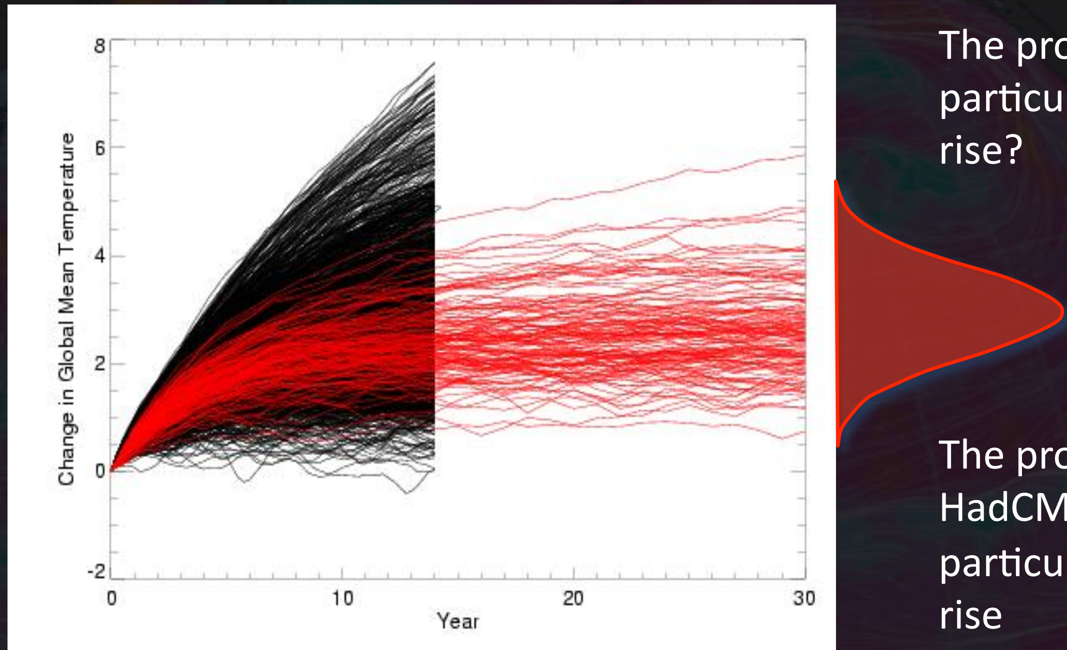
$$\frac{\partial(\eta_{Asat}\,\eta_{lf})}{\partial t} = \frac{\partial}{\partial z}\left(K_s \psi_s b\, \eta_{lf}^{b+2} \frac{\partial \eta_{lf}}{\partial z} - K_s \eta_{lf}^{2b+3} \right) + r(z).$$



Quantifying uncertainty – ensemble simulations

climateprediction.net

HadCM3, perturb parameter values, initial conditions



The probability of a particular temperature rise?

NO.
The probability that HadCM3 will simulate a particular temperature rise

- To get an unbiased estimate, we need **multi-model ensembles** with:
  - many independent estimates of parameter values
  - many independent estimates of initial conditions
  - **many independent estimates of the MODEL itself**

Model independence

Modelling groups share literature, data sets, parametrisations, even model code:

- How independent are models built by different research groups (think IPCC - impacts)?
- How should we define independence?

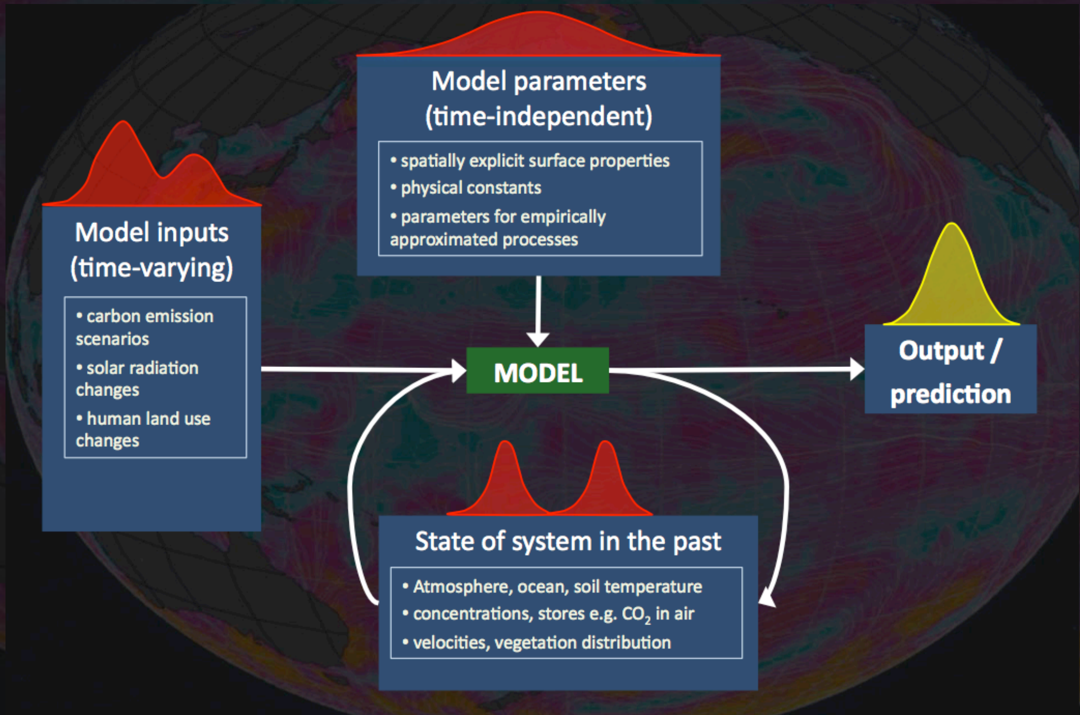
At least two different ways to think about model independence:

1. Classify models based on their simulation values – analogy with Linnaean taxonomy
   - the amount by which models differ in some output/s in the same conditions reflect the level of their independence

2. Classify models based on the independence of their structure – analogy with evolutionary taxonomy
   - what proportion of the treatment of particular processes do models share?

Model independence – an example

- Abramowitz & Gupta (2008) tried to develop a measure of distance between models as a proxy for independence:
  - Based on differences in models' output in similar circumstances
  - Distance measure (metric) could allow statistical characterisation of model space



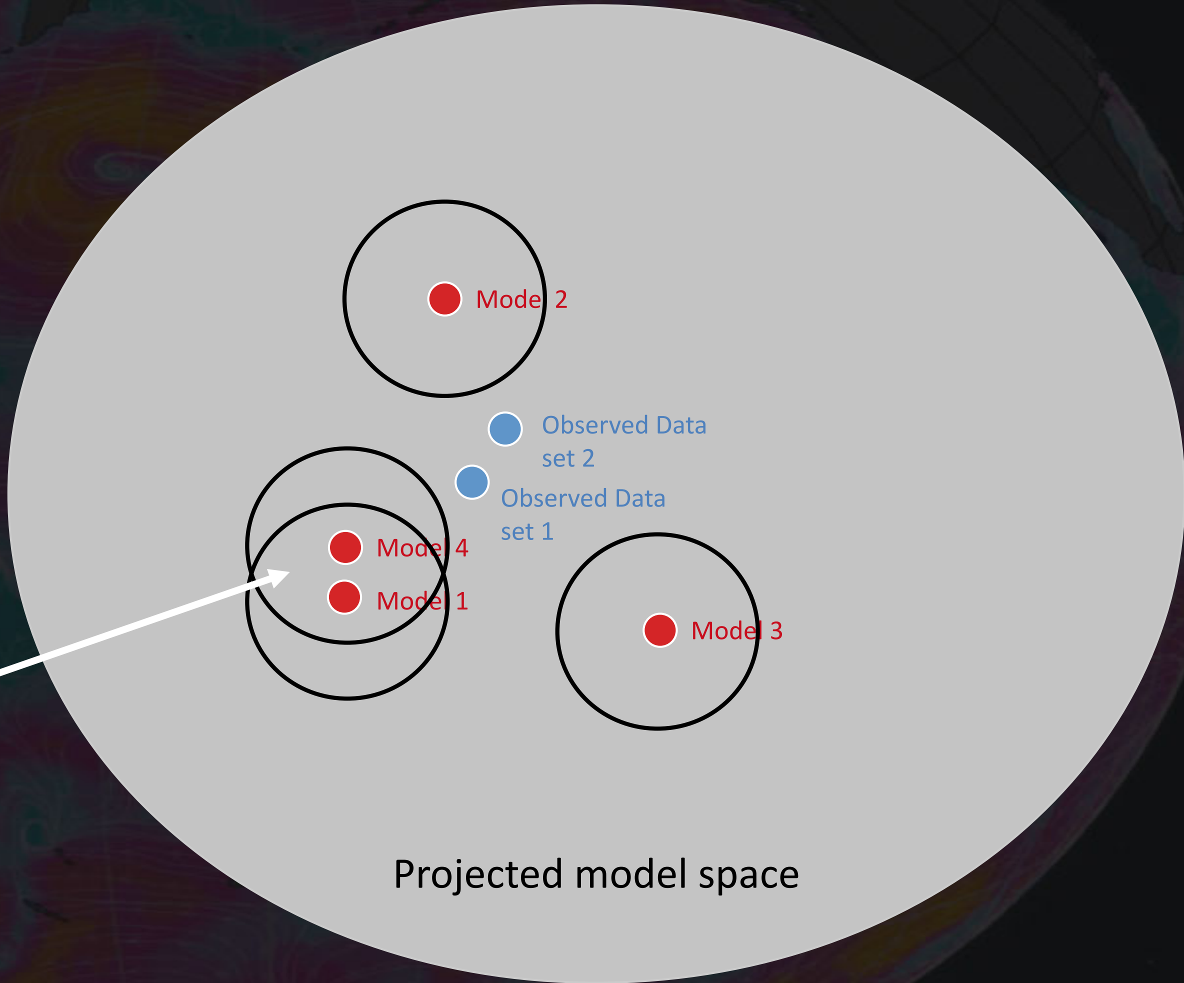
Model independence – an example

- Abramowitz & Gupta (2008) tried to develop a measure of distance between models as a proxy for independence:
  - Based on differences in models' output in similar circumstances
  - Distance measure (metric) could allow statistical characterisation of model space

- Even if we do have a distance metric for the model space as a proxy for independence, using this information is not easy:
  - How do we weight model independence vs. weighting model performance?
  - [Background: impacts applications usually use multi-model ensemble average]

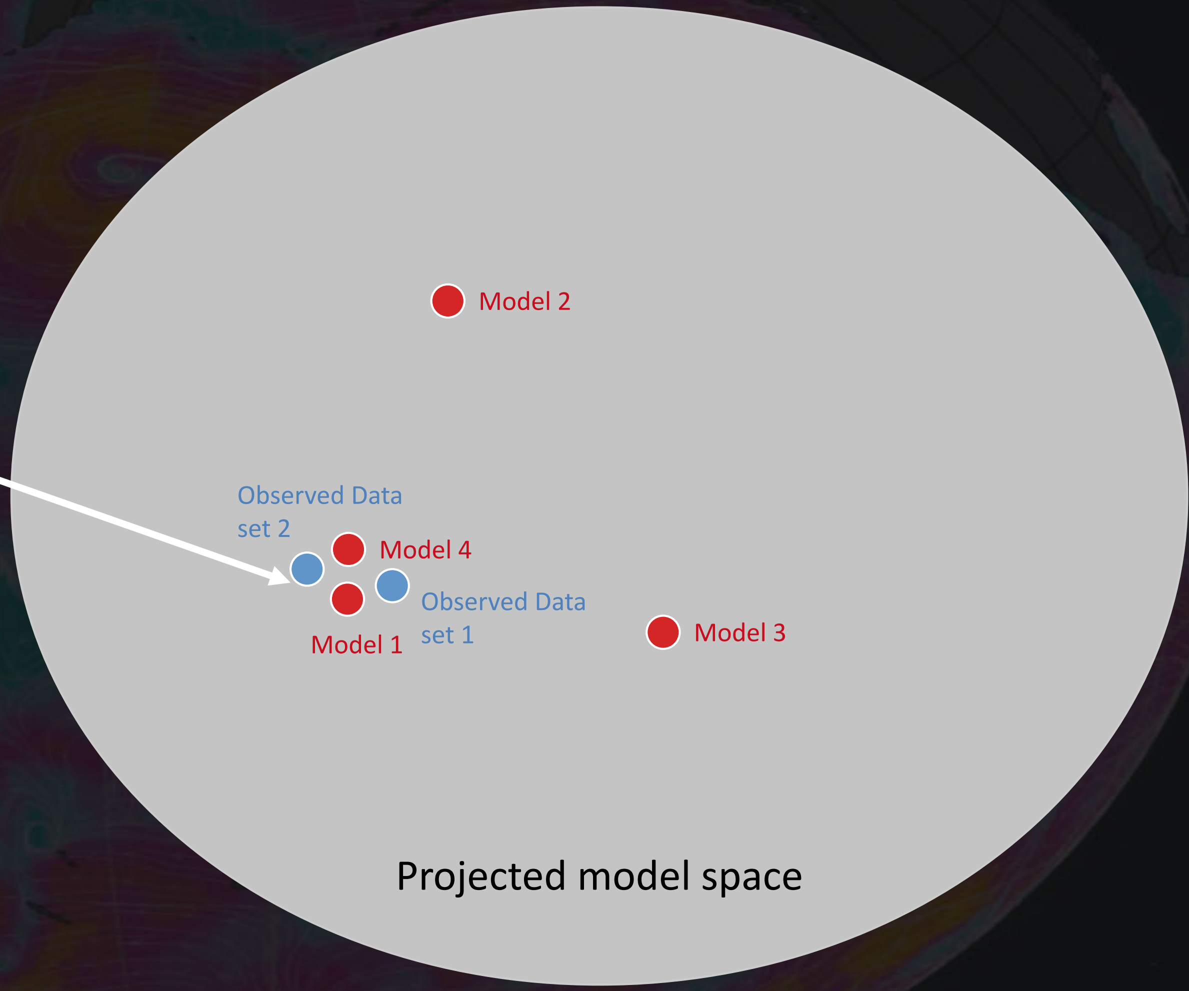
How to weight independence vs. performance in ensembles?

- Assume we have a metric on a projection of the model space:
  - d(model, obs) = performance
  - d(model1,model2) = dependence
- Assume we want to simply include/exclude models from an ensemble based on dependence
- Use a “dependence radius” to decide
- Model 1 and 4 appear quite dependent
- Only if they perform poorly…



How do we weight independence vs. performance?

Model 1 and 4 appear quite independent, but are the same distance apart!



Similar predictions might just mean that both models are correct – especially difficult if observations are of uncertain quality

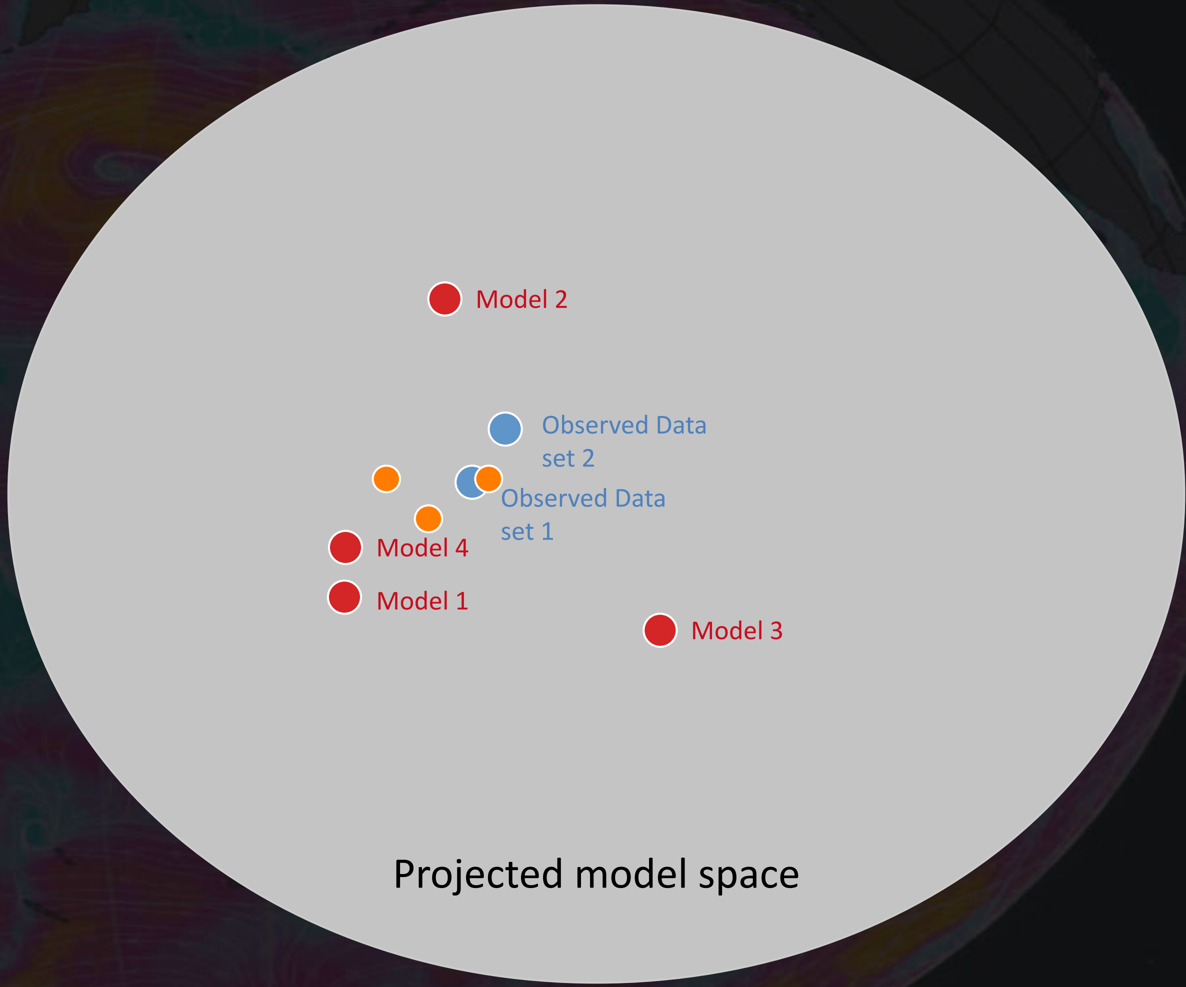
“Do we want independence? Surely we want models to converge to the truth.”

- Assume we have a metric on a projection of the model space:
  - d(model, obs) = performance
  - d(model1,model2) = dependence

- Assume we want to simply include/exclude models from an ensemble

1. All ensemble average
2. Remove worst performing
3. Remove the most dependent

- Analogy with hilltop estimation by walkers



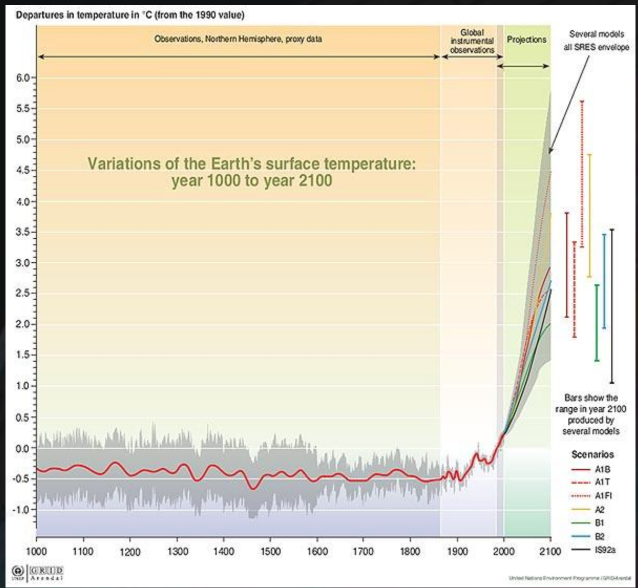
Outline

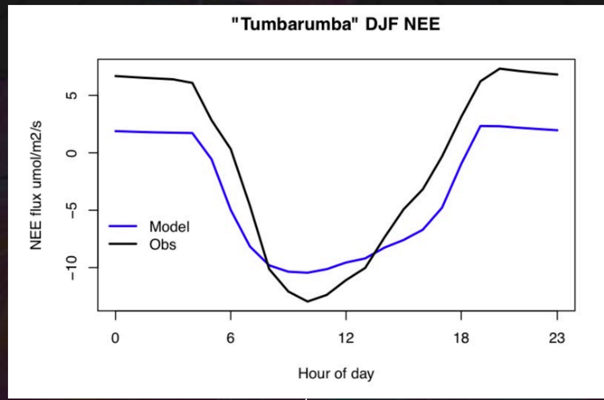
- What is a model?
- Why are models so hard to use?
- Understanding modelling uncertainty
- Measuring model performance
- What makes a good model?
- Using models to investigate scientific questions

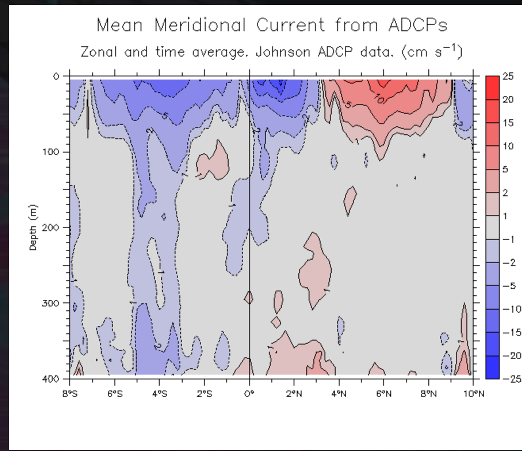
Measuring performance

- Spatial representation choices
- Temporal representation choices
- Cost function choices
- Multiple variables
- “Verification”, “validation” and “evaluation”
- How good should a model be? How can we decide on model benchmarks?

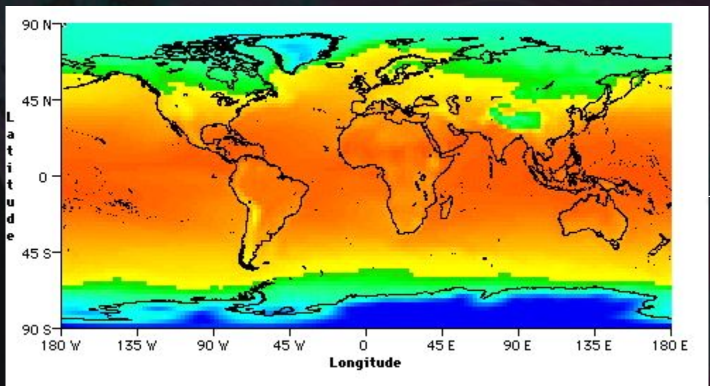
Measuring performance – spatial representation:

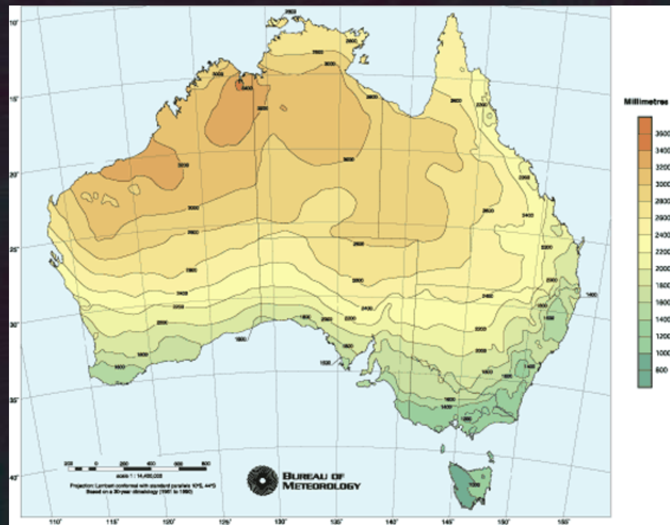






Relationship between these?

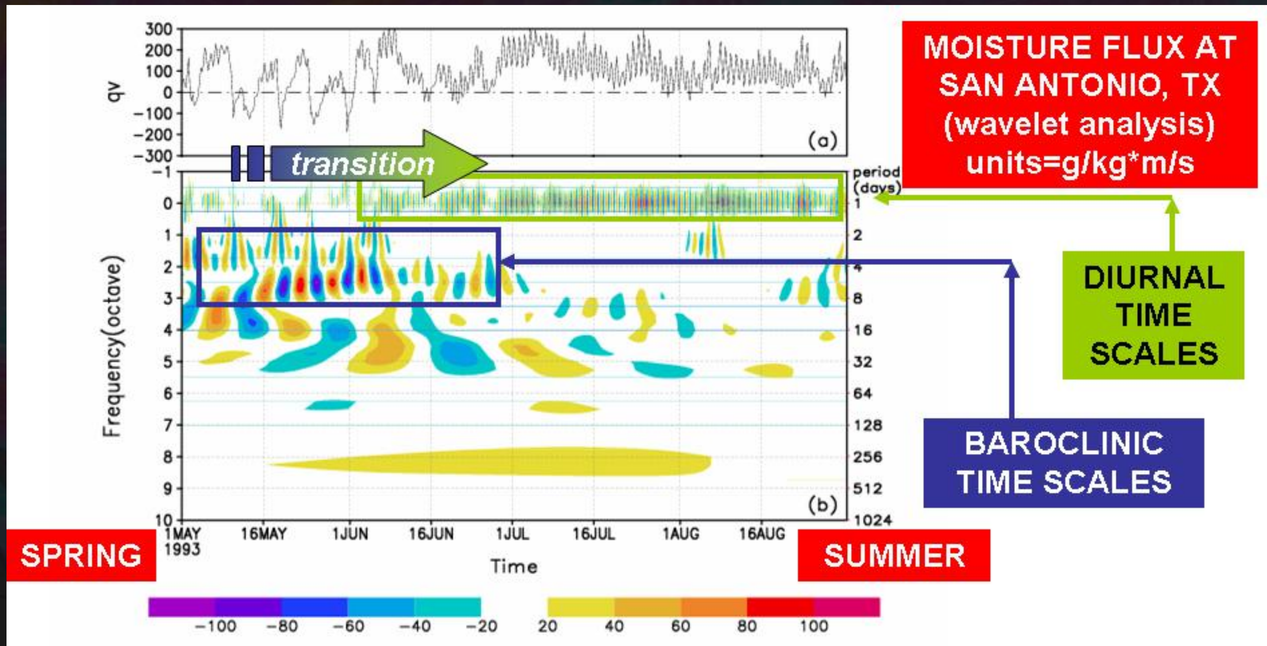




- Different spatial spatial representations will give very different results
  - Good performance in one does not guarantee good performance in another
  - May mask issues

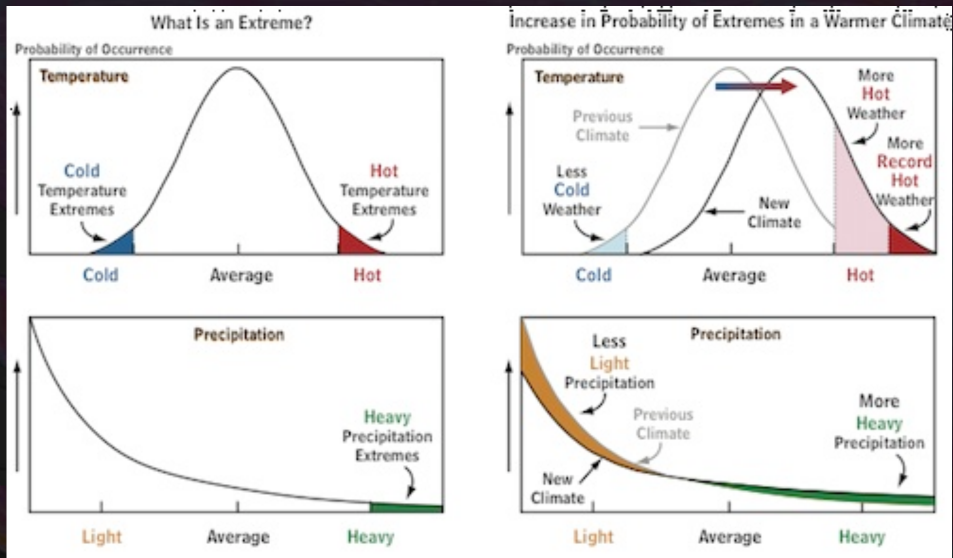
Measuring performance – temporal representation:

- Hourly, daily, monthly, annual averages – increasing loss of information
- Frequency domain measures, e.g. wavelet transforms



Measuring performance – temporal representation:

- Hourly, daily, monthly, annual averages – increasing loss of information
- Frequency domain measures, e.g. wavelet transforms
- Statistical characterisation (e.g. using pdfs)

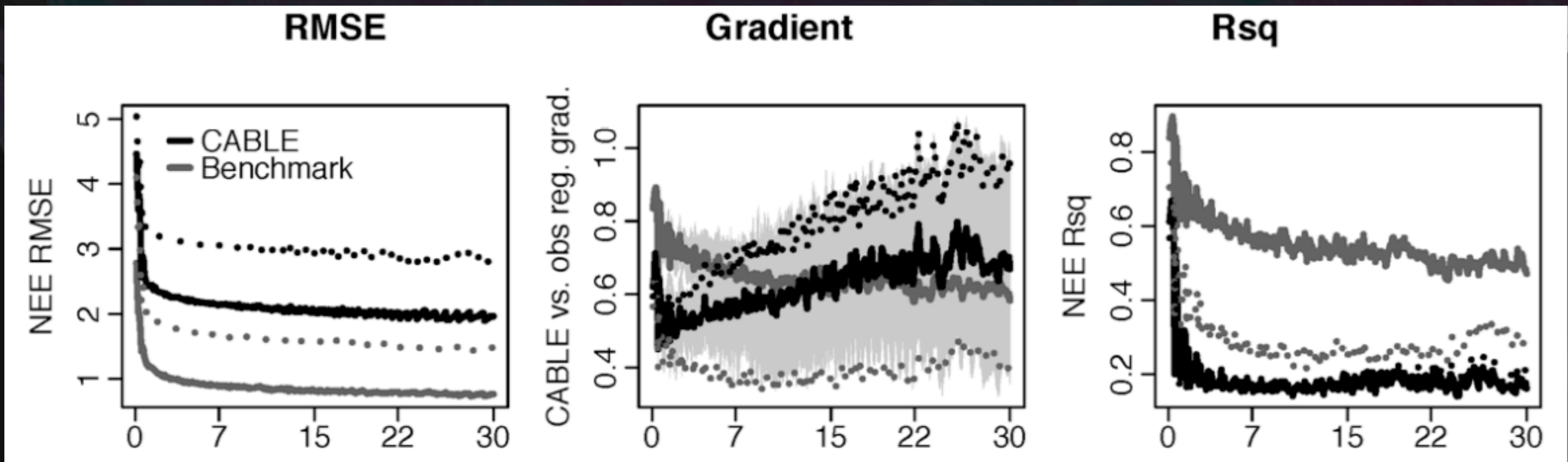


Measuring performance – cost functions

- Cost functions – RMSE; mean; maximum; minimum; variance/std; correlation; model-observation regression gradient, intercept, r^2; categorical histograms; PDF overlap; likelihood

- All give different information about performance

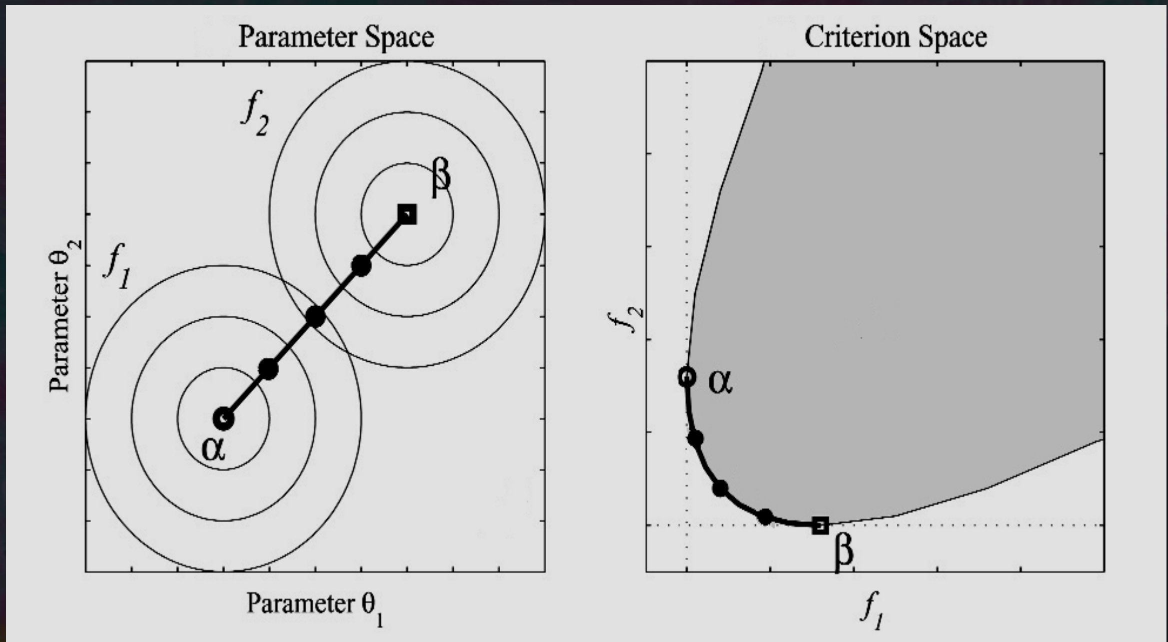
- Good performance in one doesn't guarantee good performance in another



Measuring performance – multiple outputs

- For example, a land surface model might predict:
  - Latent heat flux
  - Sensible heat flux
  - Runoff
  - Drainage to water table
  - Net Ecosystem Exchange of CO_2:
    - Is uptake / GPP right?
    - Is respiration term right?

- How can / should they be treated simultaneously?



Measuring performance – evaluation pedantry

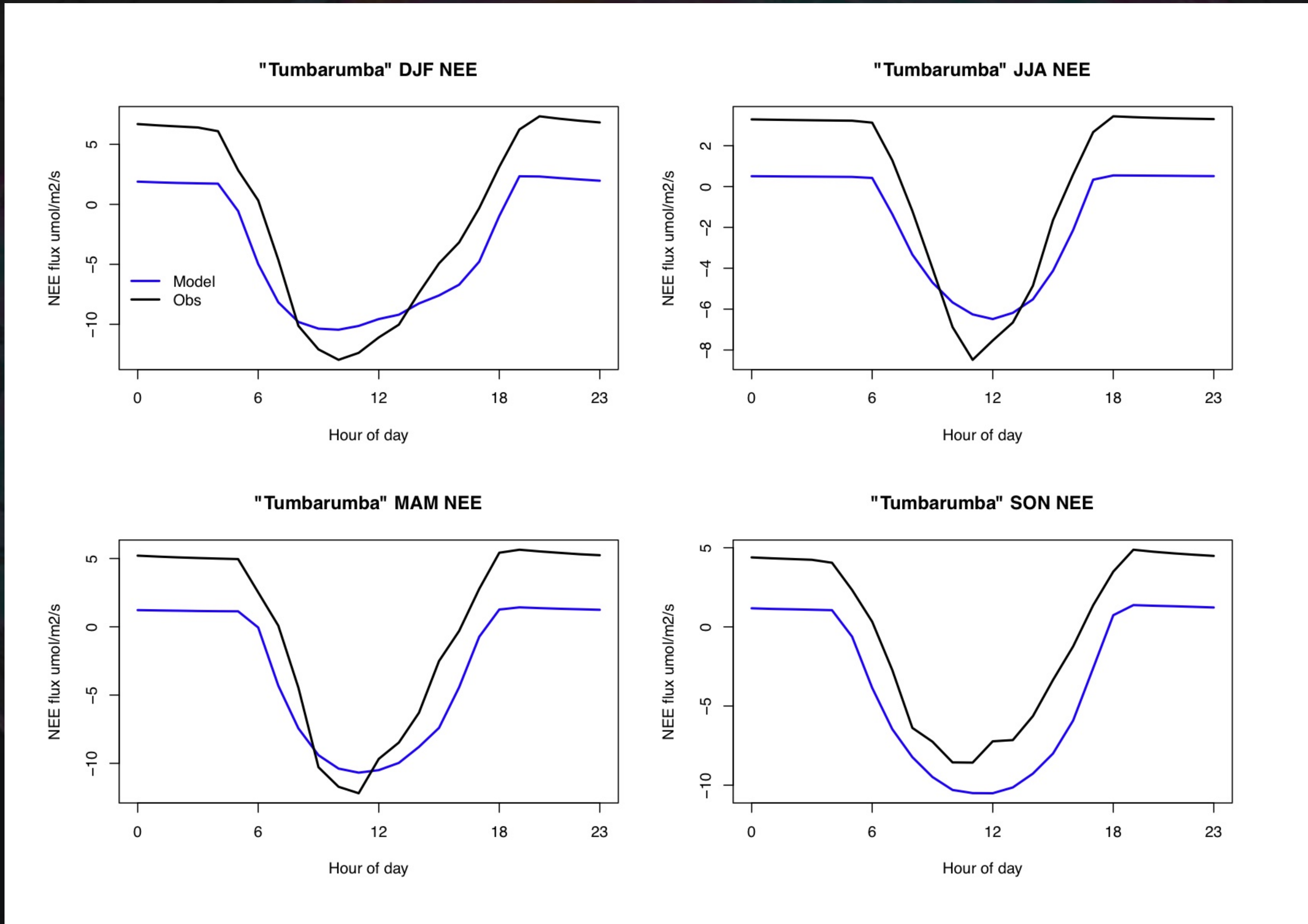
- “Verification” – literally means testing for truth

- “Validation” – valid for a particular purpose
  - Specific spatial representation
  - Specific temporal representation
  - Specific cost function

- “Evaluation” – a general term for looking at performance

Measuring performance – benchmarks

- How good should a model be?



Measuring performance – benchmarks

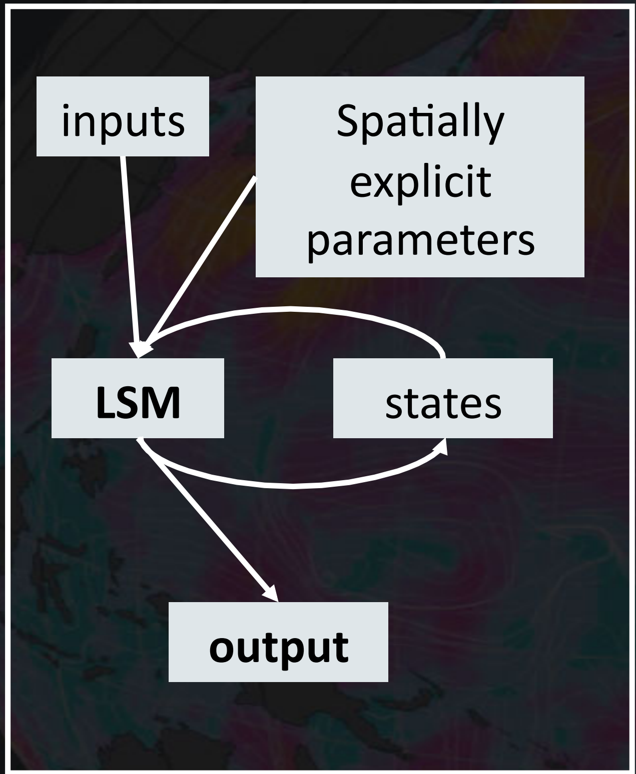
- How good should a model be?
  - What level of performance should we expect in a given performance measure?

- Hierarchy of benchmarks:
  - Physical consistency within closed system – energy and mass conservation [weak]
  - …….
  - Within observational uncertainty [strong]

- One example: *the level of performance you should expect depends on the amount of information provided to the model*
  - Expect a simple model with few inputs/parameters to be outperformed by a complex model
  - This information content of inputs can be quantified using an empirical model

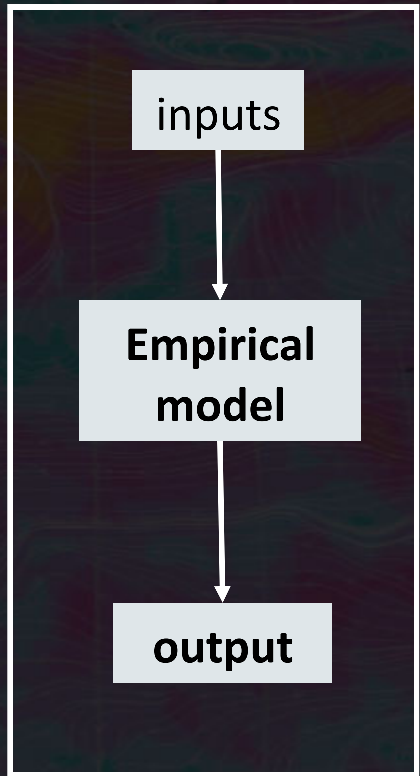
Empirical benchmarking

Normal model:



Physically based

Empirical model:



Statistically based

COMPARE

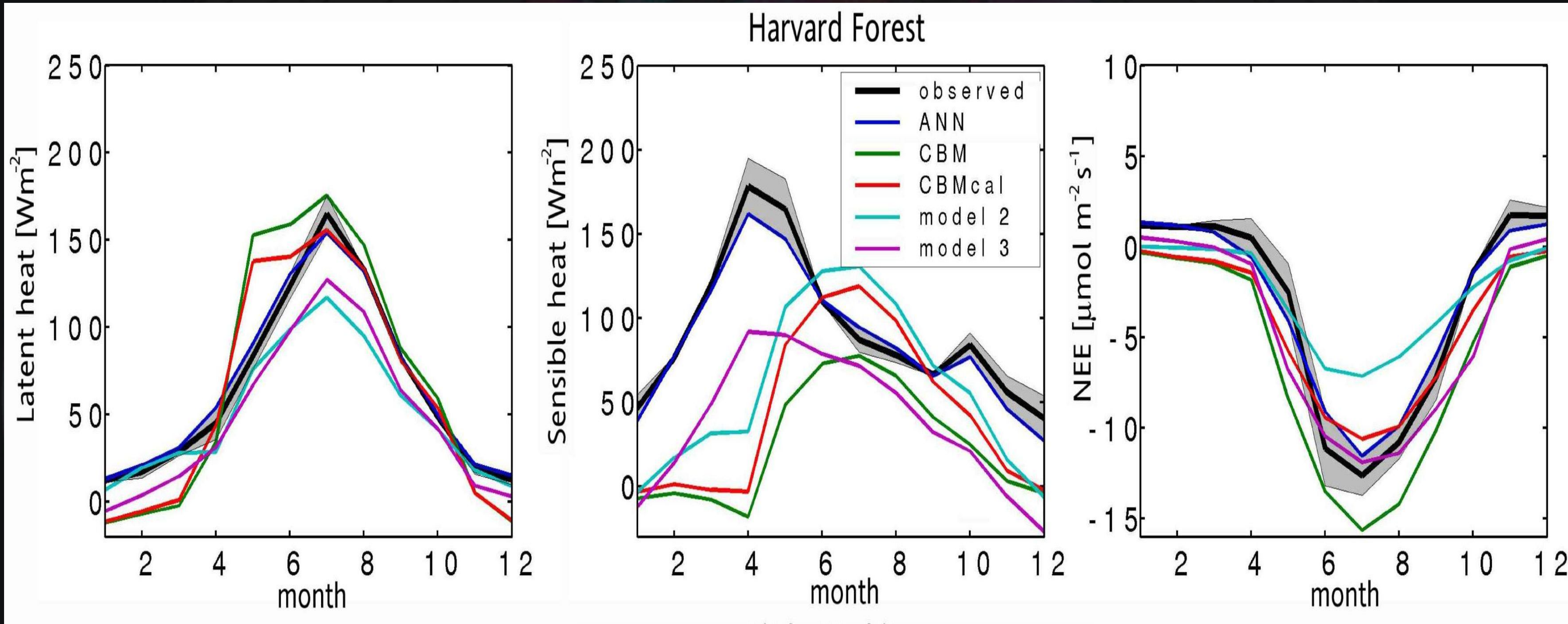
e.g. NEE CO_2, latent heat, sensible heat

- Multiple linear regression
- Neural Network
- SOLO
- other machine learning…

Requires observed data to train and test empirical model (flux tower data here)

By manipulating the relationship between training and testing data sets we can test how well a LSM utilises the information available to it…

Empirical benchmarking



- To make this a fair comparison, we can manipulate:
  - The quality/ ability of the empirical model (linear regression, ANNs, others)
  - The relationship between the training and testing sets for the empirical model
  - Which inputs the empirical model can use (i.e. more or less information about outputs)

Outline

- What is a model?
- Why are models so hard to use?
- Understanding modelling uncertainty
- Measuring model performance
- What makes a good model?
- Using models to investigate scientific questions

What makes a good model?

- The best performing model?
- No – one that encompasses understanding of the primary mechanisms and causal relationships affecting the variability of the system.
  - A good fit to observations does not guarantee this
  - Favour a simple explanation/mechanism over a complicated one [Occam's razor]
- Example of motion of planets (Ginzburg and Jensen, TREE, 2004):

1. Ptolemy's epicycles:

2. Newton-Copernicus

	Total number of parameters[a]	Number of unsupported parameters[b]	Number of parameters describing the theory goal	'Degree of overfitting'[c]
Theories explaining the motions of other planets in our solar system				
Newton (1687)	5[d]	0	5[j]	0
Ptolemy (*ca.* 150)	~10[e]	~10	5[j,k]	+5

Using models to investigate scientific questions

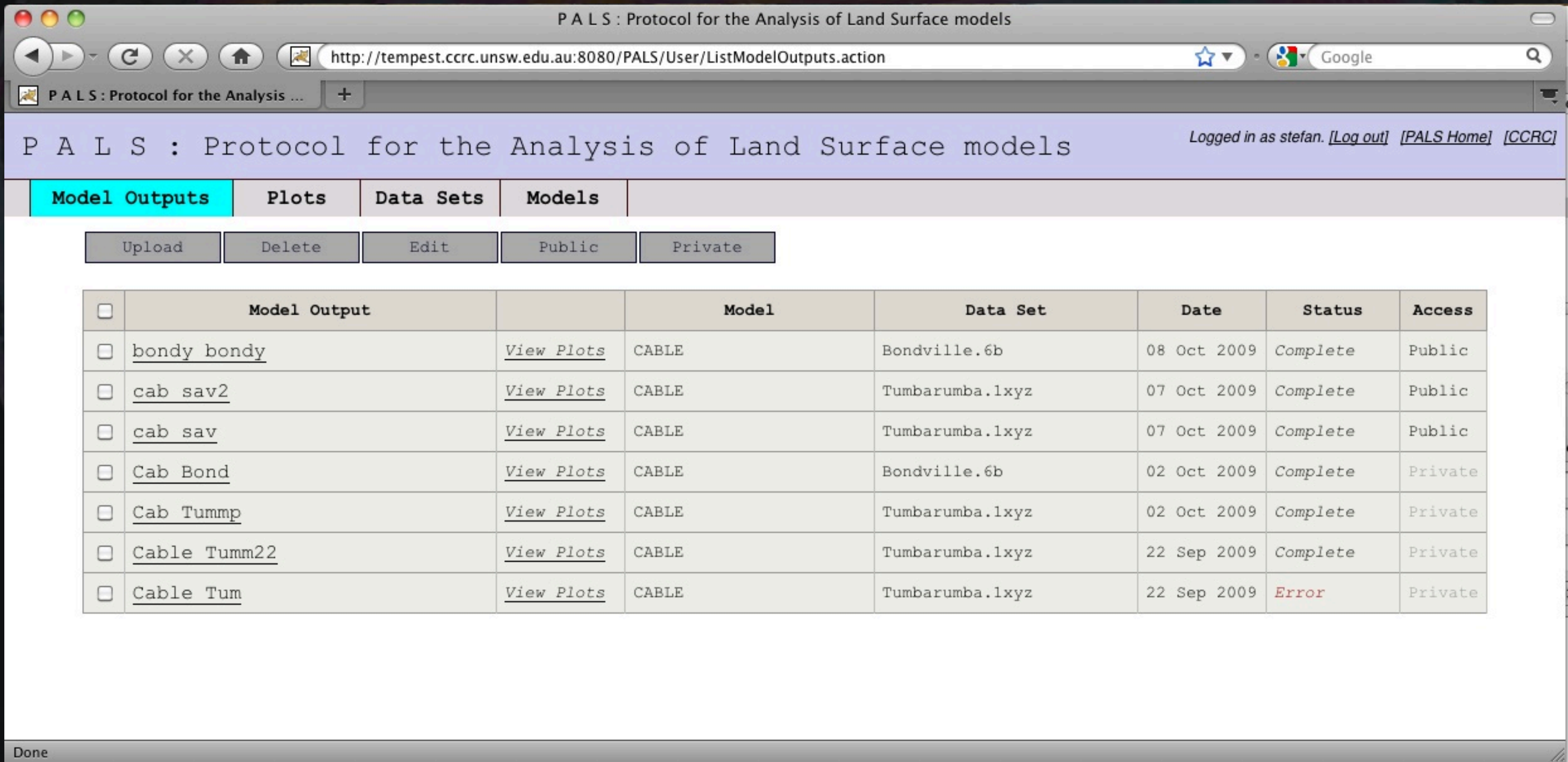
- We often assume that the scale of the phenomena is the scale of its causes
- To what extent is the natural system modellable? Which processes are functionally predictable and which are chaotic ("butterfly effect")? How do they interact?
- What are the major sources of uncertainty in the experiment?
- How does the uncertainty in the simulation affect the conclusions?
- Is the measure of performance appropriate?

Conclusions

- Complex systems modelling is a relatively new and complicated scientific tool
- Clear criteria for establishing whether simulations are meaningful or not are not firmly established yet – it's your job to be convincing
- Estimating and interpreting uncertainty is very difficult
- Evaluating performance involves interpretation of appropriate spatial and temporal representation, cost functions and variable combinations.
- You are not going to be able to deal with all of these issues, but being aware of them is important.

Protocol for the Analysis of Land Surface models (PALS)

- Screenshots below…
- For use by data collectors and land surface modellers
- Contact me if you're interested – alpha version to be released in April 2010 (gabsun@gmail.com)



P A L S : Protocol for the Analysis of Land Surface models

http://tempest.ccrc.unsw.edu.au:8080/PALS/User/ListModelOutputs.action

P A L S : Protocol for the Analysis of Land Surface models

Logged in as stefan. [Log out] [PALS Home] [CCRC]

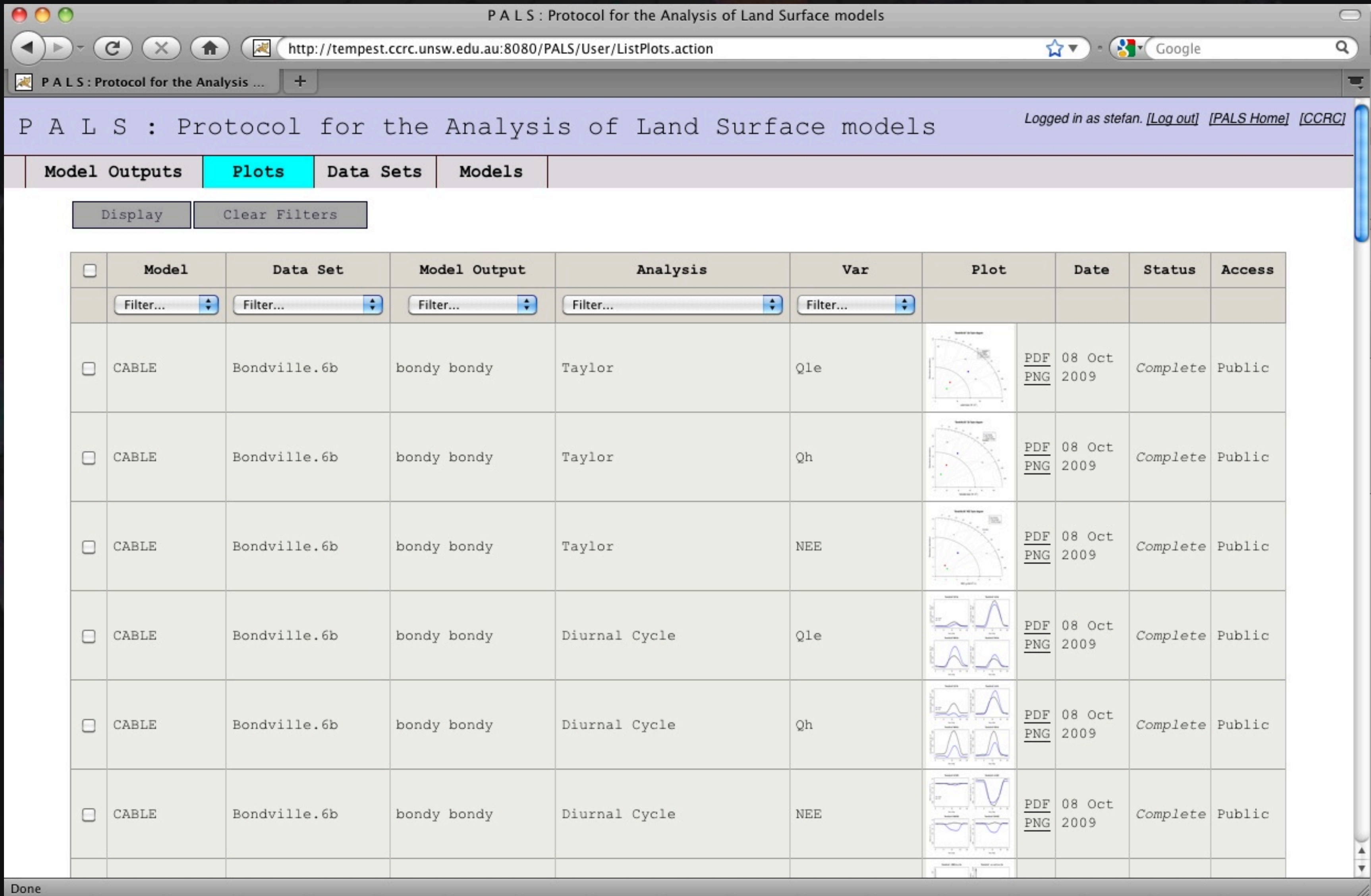
Model Outputs | Plots | Data Sets | Models

Upload | Delete | Edit | Public | Private

Model Output		Model	Data Set	Date	Status	Access
bondy bondy	View Plots	CABLE	Bondville.6b	08 Oct 2009	Complete	Public
cab sav2	View Plots	CABLE	Tumbarumba.1xyz	07 Oct 2009	Complete	Public
cab sav	View Plots	CABLE	Tumbarumba.1xyz	07 Oct 2009	Complete	Public
Cab Bond	View Plots	CABLE	Bondville.6b	02 Oct 2009	Complete	Private
Cab Tummp	View Plots	CABLE	Tumbarumba.1xyz	02 Oct 2009	Complete	Private
Cable Tumm22	View Plots	CABLE	Tumbarumba.1xyz	22 Sep 2009	Complete	Private
Cable Tum	View Plots	CABLE	Tumbarumba.1xyz	22 Sep 2009	Error	Private

Done

Protocol for the Analysis of Land Surface models (PALS)


P A L S : Protocol for the Analysis of Land Surface models

http://tempest.ccrc.unsw.edu.au:8080/PALS/User/ListPlots.action

P A L S : Protocol for the Analysis of Land Surface models

Logged in as stefan. [Log out] [PALS Home] [CCRC]

Model Outputs | Plots | Data Sets | Models

Display | Clear Filters

	Model	Data Set	Model Output	Analysis	Var	Plot		Date	Status	Access
	Filter...	Filter...	Filter...	Filter...	Filter...					
	CABLE	Bondville.6b	bondy bondy	Taylor	Qle		PDF PNG	08 Oct 2009	*Complete*	Public
	CABLE	Bondville.6b	bondy bondy	Taylor	Qh		PDF PNG	08 Oct 2009	*Complete*	Public
	CABLE	Bondville.6b	bondy bondy	Taylor	NEE		PDF PNG	08 Oct 2009	*Complete*	Public
	CABLE	Bondville.6b	bondy bondy	Diurnal Cycle	Qle		PDF PNG	08 Oct 2009	*Complete*	Public
	CABLE	Bondville.6b	bondy bondy	Diurnal Cycle	Qh		PDF PNG	08 Oct 2009	*Complete*	Public
	CABLE	Bondville.6b	bondy bondy	Diurnal Cycle	NEE		PDF PNG	08 Oct 2009	*Complete*	Public

Done

Protocol for the Analysis of Land Surface models (PALS)