GODAE OceanView International School

New Frontiers in Operational Oceanography

2-13 October 2017, Club Pollentia Resort, Mallorca, Spain.

New frontiers in ocean circulation modelling

Julien Le Sommer

Institut des Géosciences de l'Environnement, CNRS, Grenoble, France

julien.lesommer@univ-grenoble-alpes.fr

IGE CNRS Université Grenoble Alpes

About the lectures

About the lectures

Lecture #1 : Ocean circulations models : scope, usage and fundamentals.

- Thursday: 11:45-12:30

Lecture #2 : Representation of physical processes ocean circulation models

- Saturday: 9:45-10:30

Lecture #3 : Towards data-driven, probabilistic ocean circulation models

- Saturday: 11:45-12:30

About the lectures

Lecture #1 : Ocean circulations models : scope, usage and fundamentals.

- Thursday: 11:45-12:30

Lecture #2 : Representation of physical processes ocean circulation models

- Saturday: 9:45-10:30

Lecture #3 : Towards data-driven, probabilistic ocean circulation models

- Saturday: 11:45-12:30

Slides of the lectures :

https://github.com/lesommer/2017-lectures-godae-ocean-view

Book chapter :

https://github.com/lesommer/2017-book-chapter-godae-ocean-view

Lecture #1

Ocean circulations models : scope, usage and fundamentals.

Ocean circulation models

- are based on **first principle physics**;
- are used for a **wide range of applications**;
- are a key **building block of operational oceanography**;
- require **group effort and collaborations** over long time scales.

We have also learned that :

- **subgrid closures are rate controling** components of ocean models;
- **numerical discretization is a key** aspect of ocean circulation modelling;
- besides the numerical kernel, **interfaces are also important** (forcing function, coupling interfaces, etc…)
- ocean models are now (arguably) **able to describe mesoscale motions at global scale**;
- new frontier : **submesoscale-permitting modelling at global scale**.

Lecture #2

Representation of physical processes ocean circulation models

- ocean circulation models are **continuously increasing the range of physical processes** they account for;
- but defining objectively **what fraction of ocean variability is adequately represented** in ocean model solution is difficult;
- the **notion of effective resolution** is a simple (simplistic ?) concept that serves as a rationale for separating resolved and unresolved scales in ocean models;
- Interaction with **computer scientist and applied mathematicians will be essential** to further increasing the range of scale resolved in ocean models (schemes, HPC, …)
- the design of **SGS closures is a key activity** of ocean model design but (1) this is not an easy task (2) the problem is not that well posed (average ?, LES, RANS ?)
- the **coupling with third party models** is likely to become a natural way for ocean circulation model to account for certain classes of physical processes

Lecture #3 :

Towards data-driven, probabilistics ocean circulation models

Topics covered

- the **ill-posed nature of deterministic** geoscienctific modelling
- **stochastic closures** in ocean circulation models
- **stochastic modelling** to deal with various sources of uncertainties
- **ensemble simulations** in ocean circulation modelling
- the Bayesian framework for **dealing with uncertainty in models**
- **data-driven development** of ocean circulation models
- ocean modelling at the age of **data science**

Outline of Lecture #3

Part 1. Stochastic modelling in ocean circulation models

Outline of Lecture #3

Part 1. Stochastic modelling in ocean circulation models

Part 2. Uncertainty in ocean circulation modelling

Outline of Lecture #3

Part 1. Stochastic modelling in ocean circulation models

Part 2. Uncertainty in ocean circulation modelling

Part 3. Probabilistic ocean modeling with ensemble simulations

Outline of Lecture #3

Part 1. Stochastic modelling in ocean circulation models

Part 2. Uncertainty in ocean circulation modelling

Part 3. Probabilistic ocean modeling with ensemble simulations

Part 4. Towards data-driven ocean modelling

Part 1.

Stochastic modelling in ocean circulation models

Beyond deterministic subgrid scales closures

Beyond deterministic subgrid scales closures

The **SGS closure problem** : estimating a good model $\mathcal{M}(\bar{\mathbf{s}}) = \mathcal{N}(\bar{\mathbf{s}}) - \overline{\mathcal{N}(\mathbf{s})}$.

Beyond deterministic subgrid scales closures

The **SGS closure problem** : estimating a good model $\mathcal{M}(\bar{\mathbf{s}}) = \mathcal{N}(\bar{\mathbf{s}}) - \overline{\mathcal{N}(\mathbf{s})}$.

- as discuused, assuming that you can **draw a local deterministic model** that is fully asserved to resolved scales of motions is very strong assumption.
- physically your are assuming that the **net result of micro-states at macro-scale** is not dependent on the micro-scales.

Beyond deterministic subgrid scales closures

The **SGS closure problem** : estimating a good model $\mathcal{M}(\bar{\mathbf{s}}) = \mathcal{N}(\bar{\mathbf{s}}) - \overline{\mathcal{N}(\mathbf{s})}$.

- as discuused, assuming that you can **draw a local deterministic model** that is fully asserved to resolved scales of motions is very strong assumption.

- physically your are assuming that the **net result of micro-states at macro-scale** is not dependent on the micro-scales.

but a number of recent works have started to reconsider this assumption.

Beyond deterministic subgrid scales closures

The **SGS closure problem** : estimating a good model $\mathcal{M}(\bar{\mathbf{s}}) = \mathcal{N}(\bar{\mathbf{s}}) - \overline{\mathcal{N}(\mathbf{s})}$.

- as discuused, assuming that you can **draw a local deterministic model** that is fully asserved to resolved scales of motions is very strong assumption.
- physically your are assuming that the **net result of micro-states at macro-scale** is not dependent on the micro-scales.

but a number of recent works have started to reconsider this assumption.

- these approaches allow the closure to be only weakly constrained by the resolved scales.

Beyond deterministic subgrid scales closures

The **SGS closure problem** : estimating a good model $\mathcal{M}(\bar{\mathbf{s}}) = \mathcal{N}(\bar{\mathbf{s}}) - \overline{\mathcal{N}(\mathbf{s})}$.

- as discuused, assuming that you can **draw a local deterministic model** that is fully asserved to resolved scales of motions is very strong assumption.
- physically your are assuming that the **net result of micro-states at macro-scale** is not dependent on the micro-scales.

but a number of recent works have started to reconsider this assumption.

- these approaches allow the closure to be only weakly constrained by the resolved scales.
- a certain amont of **randomness is introduced with a stochastic procedure** to account for the possible varibility of fluxes at macroscales

Beyond deterministic subgrid scales closures

The **SGS closure problem** : estimating a good model $\mathcal{M}(\bar{\mathbf{s}}) = \mathcal{N}(\bar{\mathbf{s}}) - \overline{\mathcal{N}(\mathbf{s})}$.

- as discuused, assuming that you can **draw a local deterministic model** that is fully asserved to resolved scales of motions is very strong assumption.
- physically your are assuming that the **net result of micro-states at macro-scale** is not dependent on the micro-scales.

but a number of recent works have started to reconsider this assumption.

- these approaches allow the closure to be only weakly constrained by the resolved scales.
- a certain amont of **randomness is introduced with a stochastic procedure** to account for the possible varibility of fluxes at macroscales
- Stochastic parameterizations seem in particular well suited to the representation of **the cross scale exchanges of energy and momentum** for SGS balanced turbulence (in particular energy backscatter)

Beyond deterministic subgrid scales closures

The **SGS closure problem** : estimating a good model $\mathcal{M}(\bar{\mathbf{s}}) = \mathcal{N}(\bar{\mathbf{s}}) - \overline{\mathcal{N}(\mathbf{s})}$.

- as discuused, assuming that you can **draw a local deterministic model** that is fully asserved to resolved scales of motions is very strong assumption.
- physically your are assuming that the **net result of micro-states at macro-scale** is not dependent on the micro-scales.

but a number of recent works have started to reconsider this assumption.

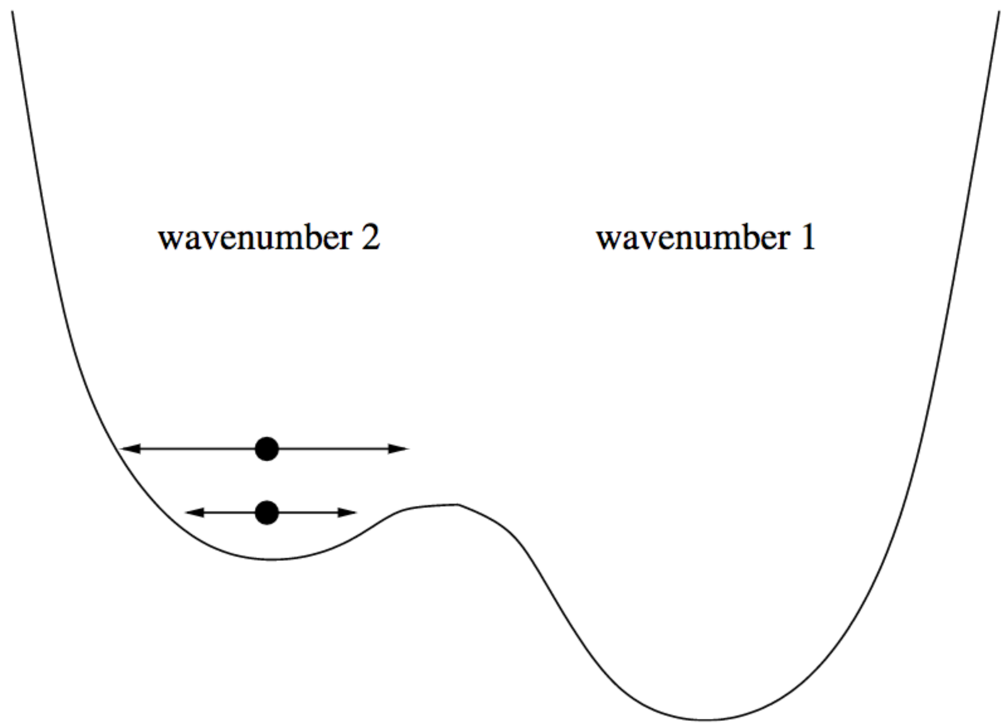
- these approaches allow the closure to be only weakly constrained by the resolved scales.
- a certain amont of **randomness is introduced with a stochastic procedure** to account for the possible varibility of fluxes at macroscales
- Stochastic parameterizations seem in particular well suited to the representation of **the cross scale exchanges of energy and momentum** for SGS balanced turbulence (in particular energy backscatter)
- notable improvment of Gulf stream dynamics have also been obtained through the stochastic representation of upscaling due to the **nonlinear** nature of the **equation of state** of sea-water.

Potential wells and marginal stability

There is actually **another rationale that supports this idea of introducing random perturbations** in ocean models.

Potential wells and marginal stability

There is actually **another rationale that supports this idea of introducing random perturbations** in ocean models.

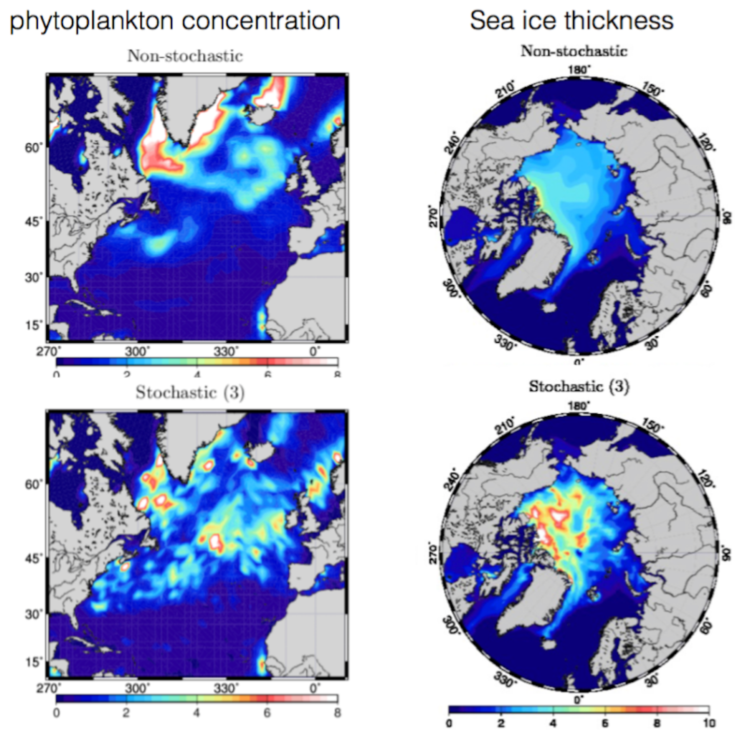


The idea is that **perturbations could allow your system to escape from local potential wells** and therefore explore more reliably the range of possible states.

Range of applications of stochastic ocean modelling

Stochastic approaches are therefore not limited to the representation of subgrid processes, they have been suceesfully applied to account for :

- the **uncertainty** in sea ice model parameters
- the **uncertainty** in biogeochemical models parameters



Brankart et al. (2015)

Part 2.

Uncertainty in ocean circulation modelling

Multiple sources of uncertainty in ocean models

- ocean model solutions are affected by **several sources of errors** that combine and interact.

Multiple sources of uncertainty in ocean models

- ocean model solutions are affected by **several sources of errors** that combine and interact.
- ocean model solutions are affected by **errors in their initial condition** : this is particularly true at depth;

Multiple sources of uncertainty in ocean models

- ocean model solutions are affected by **several sources of errors** that combine and interact.
- ocean model solutions are affected by **errors in their initial condition** : this is particularly true at depth;
- ocean model solutions are affected by **errors in the forcing function** : uncertain atmospheric fields, uncertain parameter in bulk formula...

Multiple sources of uncertainty in ocean models

- ocean model solutions are affected by **several sources of errors** that combine and interact.
- ocean model solutions are affected by **errors in their initial condition** : this is particularly true at depth;
- ocean model solutions are affected by **errors in the forcing function** : uncertain atmospheric fields, uncertain parameter in bulk formula...
- ocean model solutions are affected by **errors associated with subgrid closures** : closures are imperfect models, with usually weakly constrained parameters

Multiple sources of uncertainty in ocean models

- ocean model solutions are affected by **several sources of errors** that combine and interact.
- ocean model solutions are affected by **errors in their initial condition** : this is particularly true at depth;
- ocean model solutions are affected by **errors in the forcing function** : uncertain atmospheric fields, uncertain parameter in bulk formula...
- ocean model solutions are affected by **errors associated with subgrid closures** : closures are imperfect models, with usually weakly constrained parameters
- ocean model solutions are affected by **the physical approximation** use in for their continuous formulation : eg. non-traditional Coriolis terms

Multiple sources of uncertainty in ocean models

- ocean model solutions are affected by **several sources of errors** that combine and interact.
- ocean model solutions are affected by **errors in their initial condition** : this is particularly true at depth;
- ocean model solutions are affected by **errors in the forcing function** : uncertain atmospheric fields, uncertain parameter in bulk formula…
- ocean model solutions are affected by **errors associated with subgrid closures** : closures are imperfect models, with usually weakly constrained parameters
- ocean model solutions are affected by **the physical approximation** use in for their continuous formulation : eg. non-traditional Coriolis terms
- ocean model solutions are affected by **numerical discretization errors** : even high order schemes have non-zero errors, these errors cumulate in time.

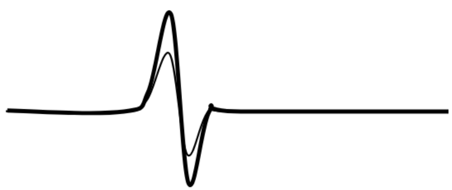
Multiple sources of uncertainty in ocean models

- ocean model solutions are affected by **several sources of errors** that combine and interact.
- ocean model solutions are affected by **errors in their initial condition** : this is particularly true at depth;
- ocean model solutions are affected by **errors in the forcing function** : uncertain atmospheric fields, uncertain parameter in bulk formula...
- ocean model solutions are affected by **errors associated with subgrid closures** : closures are imperfect models, with usually weakly constrained parameters
- ocean model solutions are affected by **the physical approximation** use in for their continuous formulation : eg. non-traditional Coriolis terms
- ocean model solutions are affected by **numerical discretization errors** : even high order schemes have non-zero errors, these errors cumulate in time.

Overall, **uncertainty is a major property of ocean circulation models** that physicists tend to neglect to often.

The complex nature of numerical discretization errors

- numerical discretization errors can be classified in two categories :



dissipative errors affect the amplitude

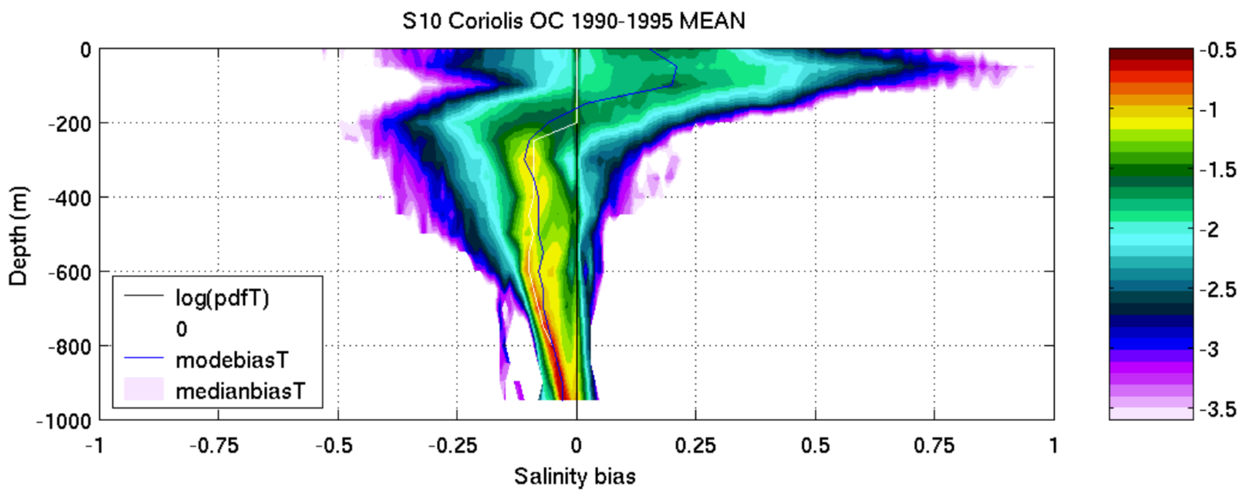
dispersion errors affect the phase of the solution

- numerical discretization errors may **propagate spatially** in the solution
- numerical discretization errors may **accumulate over time**

after integration over long times, **numerical discretization errors can show complex patterns**.

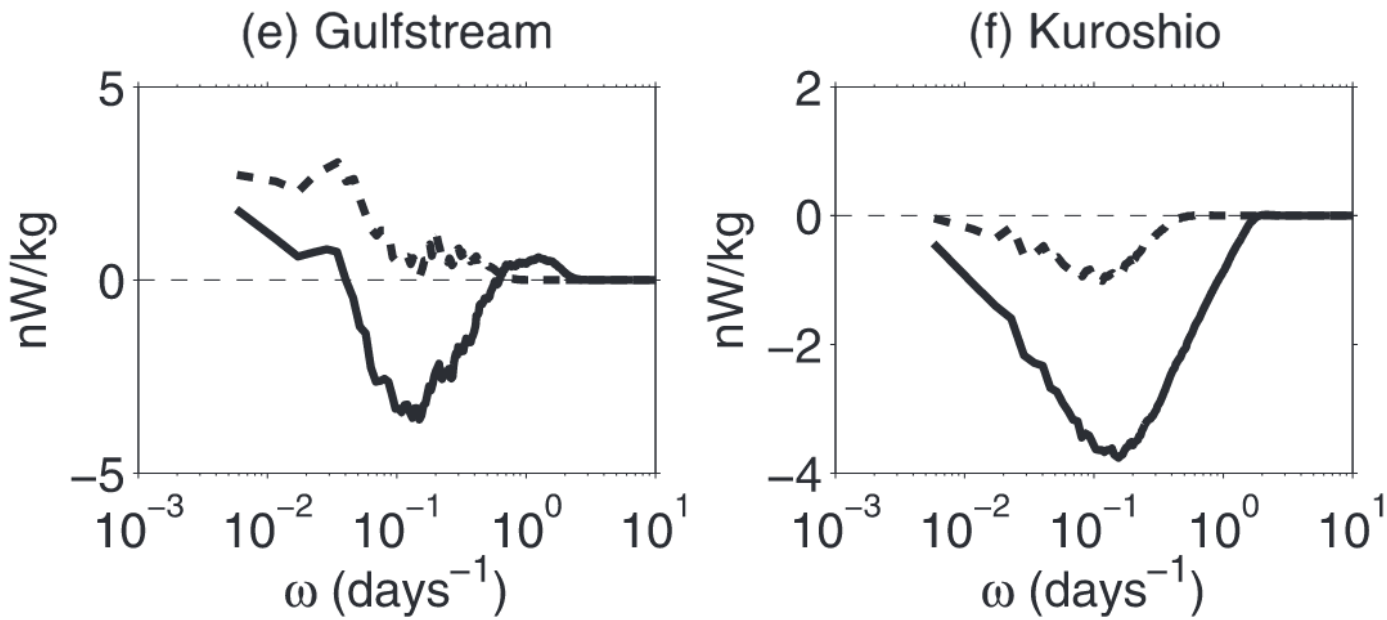
Practical expression of errors in ocean model solution

(Melet et al.)



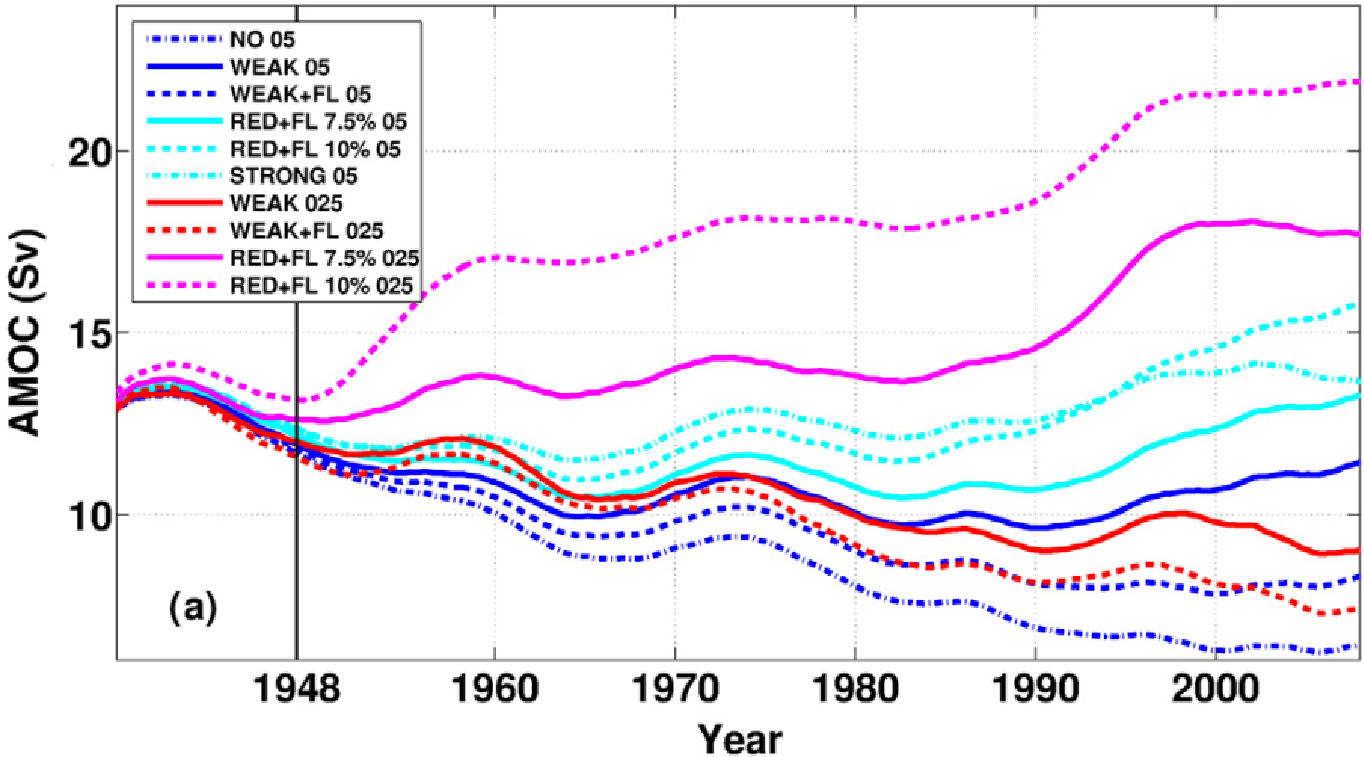
- In practice, ocean model error manifest itself through the **time-mean state** of the model solution, its **trend** at its **variability** at all time scale.
- Model errors are also usually **correlated both in space and in time** (see above).
- Model error is usually due to a combination of different factors

Example 1 : the phase of mesoscale eddy variability



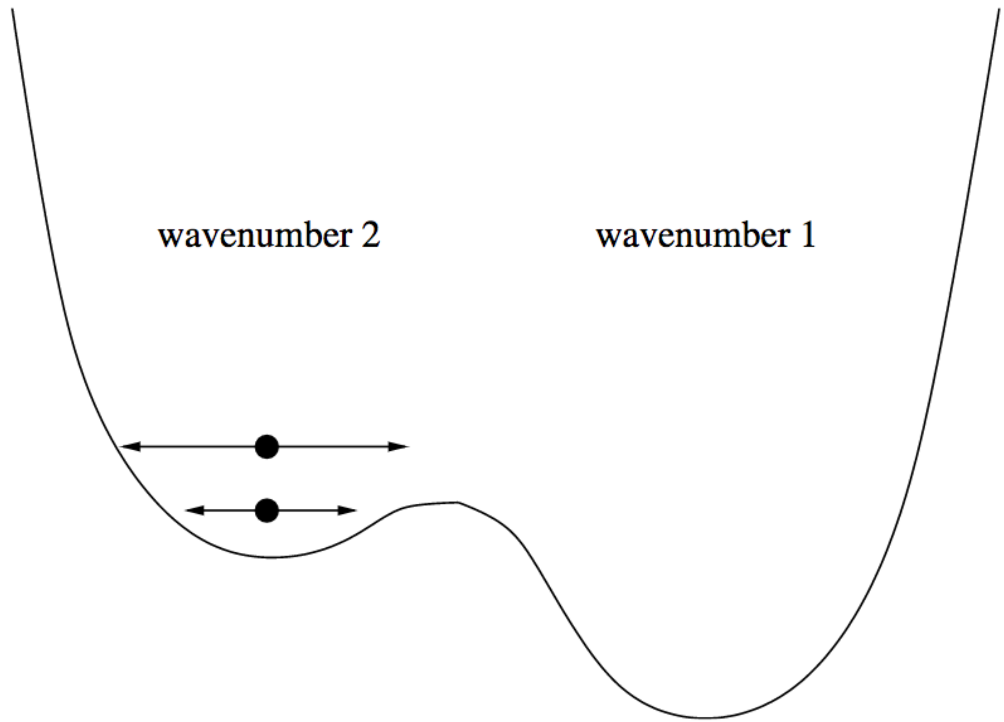
- the **phase of mesoscale** variability is obvioulsy **not predicted** by forced, non-assimilated ocean models.
- but implicitely, it is tempting to think that the phase of mesoscale flow features is not affecting too much the low frequency variability.
- Arbic et al. (2012) have shown that energy cascades in time to long time scales so that the phase of **mesoscale may actually affect long time scales**…

Example 2: AMOC in forced ocean models



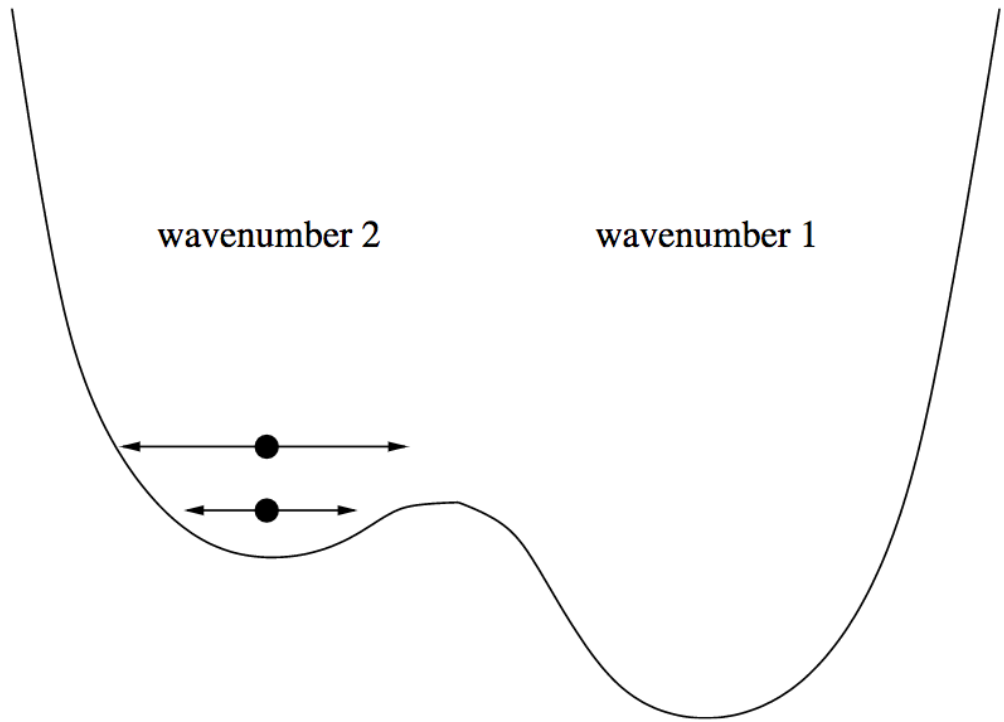
- A key prediction that is expected from ocean models is **how the atlantic meridional overturning circulation (AMOC) varies** in time.
- Behrens et al. (2013) have shown that although the interannual variability of the AMOC is rather predictable, **surface salinity restoring is a major control** of the decadal trend of AMOC.

Multimodality of ocean states



- The **topology of the phase space of non-linear system** can be rather complicated, possible showing muliple possible local energy minima.

Multimodality of ocean states

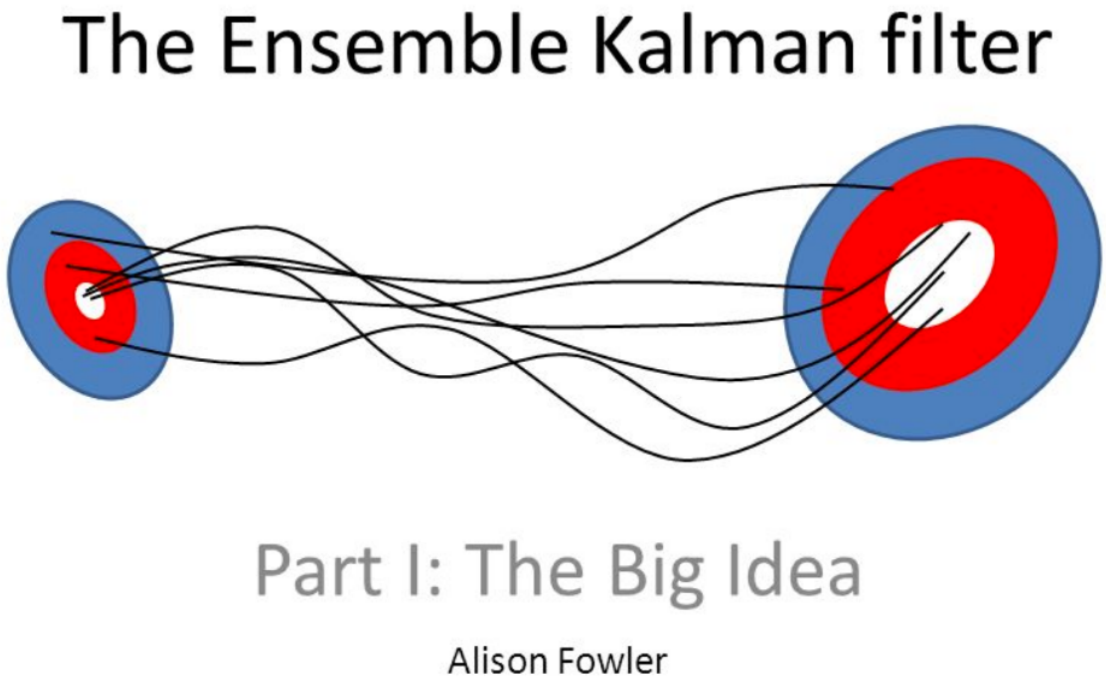


- The **topology of the phase space of non-linear system** can be rather complicated, possible showing muliple possible local energy minima.
- Deshayes et al. (2013) found in a global model that the freshwater transport in the southern atlantic was such that the **AMOC could be unstable to forcing perturbations**. *but can we trust this model prediction ?*

Part 3.

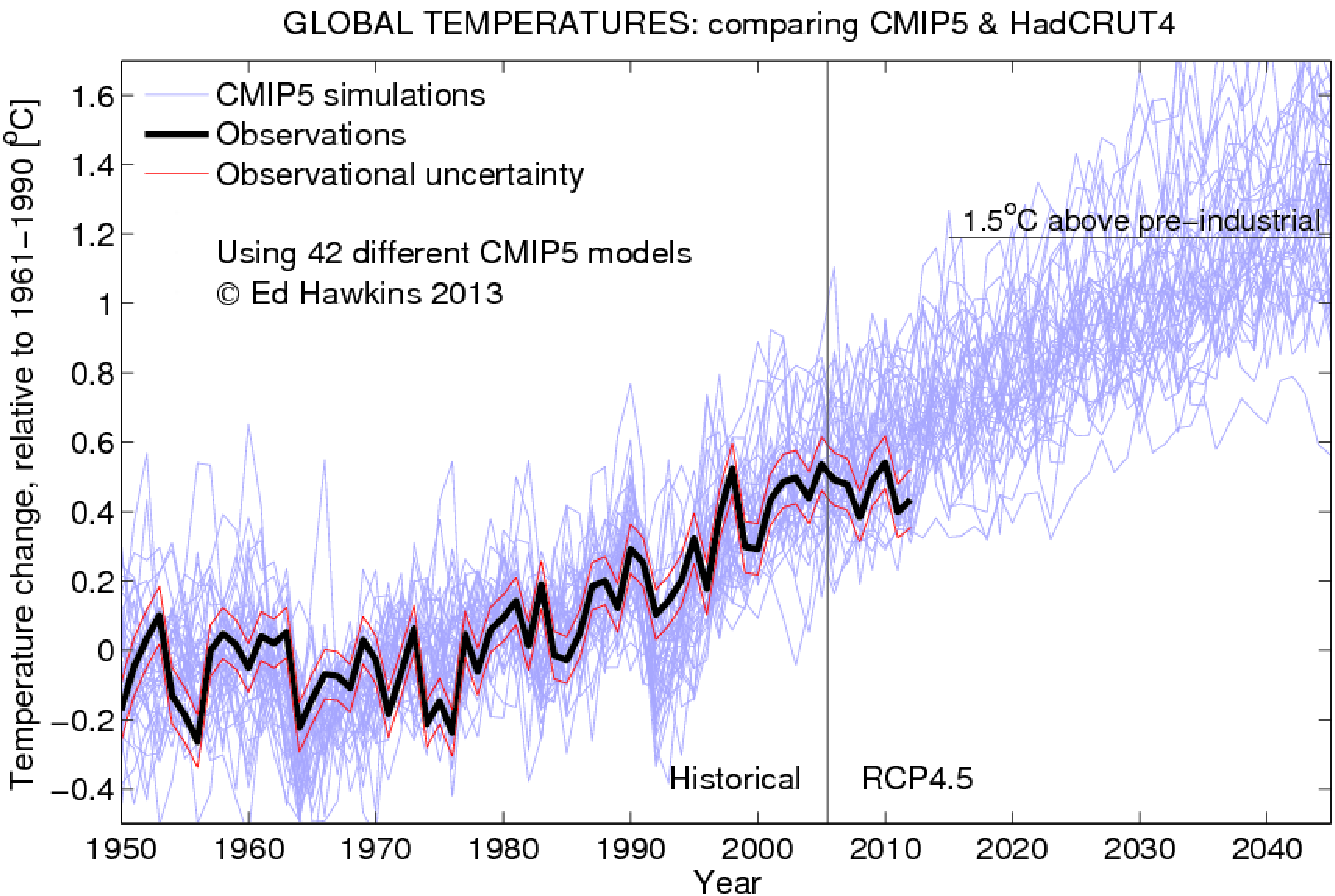
Probabilistic ocean modelling with ensemble simulations

Ensemble simulation in data assimilation



Ensemble modelling is used in data assimilation for sampling the possible likely states of the system.

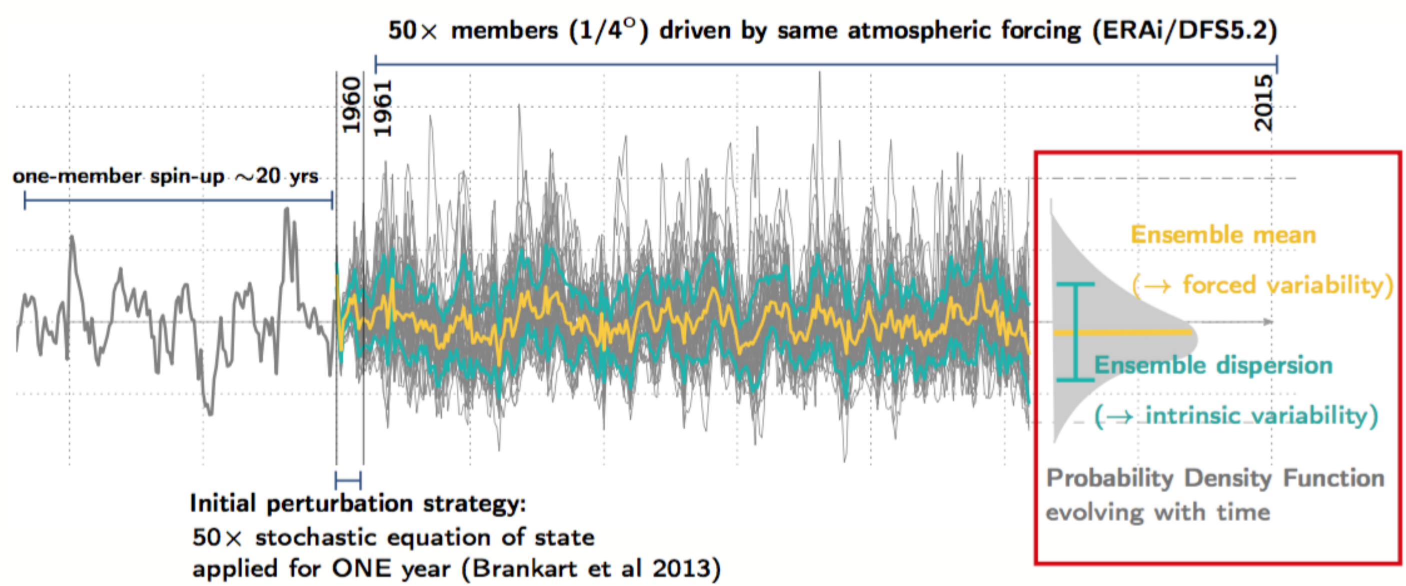
Ensemble simulation in climate modelling



Ensemble modelling is climate projections in order to reduce the uncertainty associated with spurious individual model bias/sensitivity.

Intrinsic ocean chaos revealed by ensemble simulation

The OCCIPUT ensemble of global eddying ocean model solutions



this set of simulation allows to investigate wjat fraction of oceanic variability is **governed by the atmosphere** versus the spontaneous **variability due to oceanic chaos**

see the posters by **Stephanie Leroux** and **Ixetl Garcıa Gomez**.

Benefits of ensemble modelling

Ensemble modelling not only allows to tackles specific questions regarding oceanic variability but also :

Benefits of ensemble modelling

Ensemble modelling not only allows to tackles specific questions regarding oceanic variability but also :

- is a very good strategy in terms of HPC : ensemble runs avec 100% scallable !

Benefits of ensemble modelling

Ensemble modelling not only allows to tackles specific questions regarding oceanic variability but also :

- is a very good strategy in terms of HPC : ensemble runs avec 100% scallable !
- more generally probabilistic ocean modelling (ensemble + stochastic) allow to combine nicely data driven and physics driven approaches (see later)

Benefits of ensemble modelling

Ensemble modelling not only allows to tackles specific questions regarding oceanic variability but also :

- is a very good strategy in terms of HPC : ensemble runs avec 100% scallable !
- more generally probabilistic ocean modelling (ensemble + stochastic) allow to combine nicely data driven and physics driven approaches (see later)
- therefore allowing ocean modellers to discuss more happily with data assimilation experts.

Benefits of ensemble modelling

Ensemble modelling not only allows to tackles specific questions regarding oceanic variability but also :

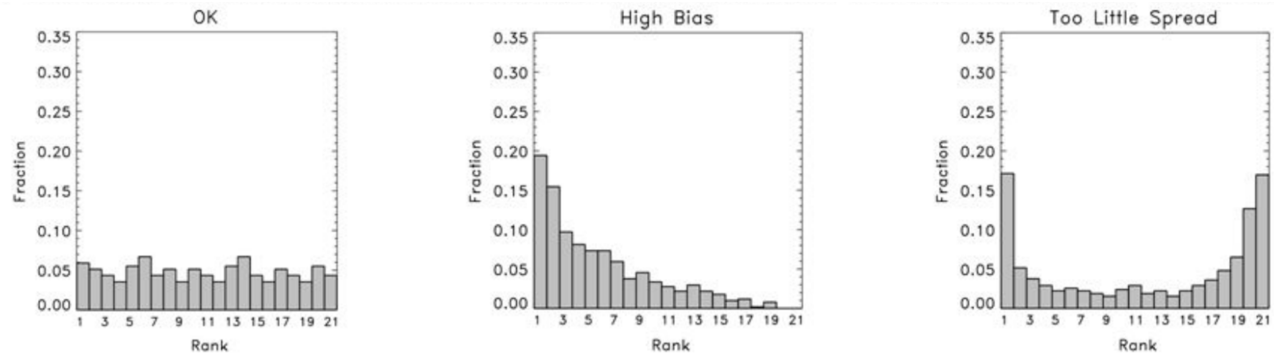
- is a very good strategy in terms of HPC : ensemble runs avec 100% scallable !
- more generally probabilistic ocean modelling (ensemble + stochastic) allow to combine nicely data driven and physics driven approaches (see later)
- therefore allowing ocean modellers to discuss more happily with data assimilation experts.
- most importantly, ensemble modelling allows allows for a more objective comparison of model solutions with observations.

Benefits of ensemble modelling

Ensemble modelling not only allows to tackles specific questions regarding oceanic variability but also :

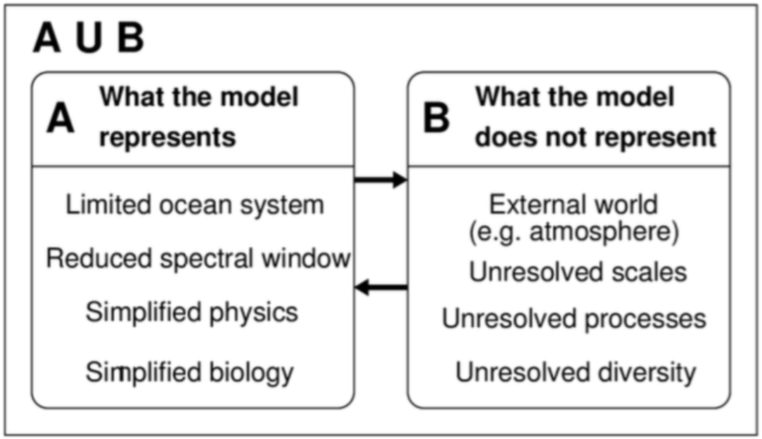
- is a very good strategy in terms of HPC : ensemble runs avec 100% scallable !
- more generally probabilistic ocean modelling (ensemble + stochastic) allow to combine nicely data driven and physics driven approaches (see later)
- therefore allowing ocean modellers to discuss more happily with data assimilation experts.
- most importantly, ensemble modelling allows allows for a more objective comparison of model solutions with observations.

example : Rank histograms and model reliability

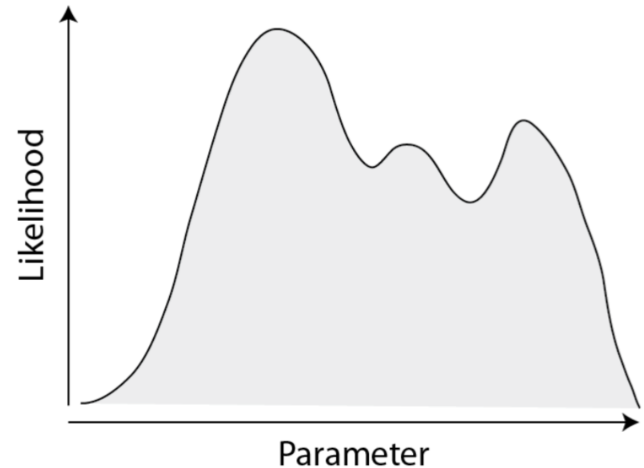


A system approach to ocean modelling

A system approach to ocean modelling

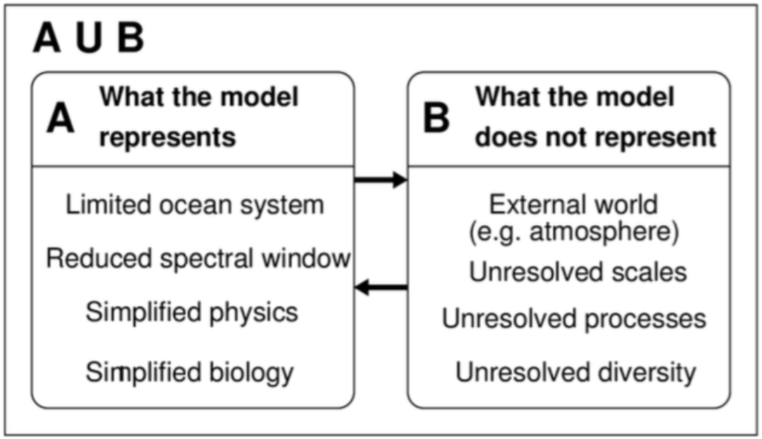


System approach to ocean modelling

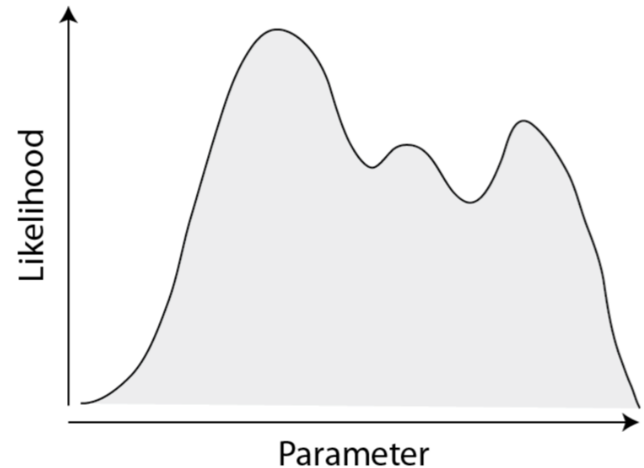


multi-modality of ocean distributions

A system approach to ocean modelling



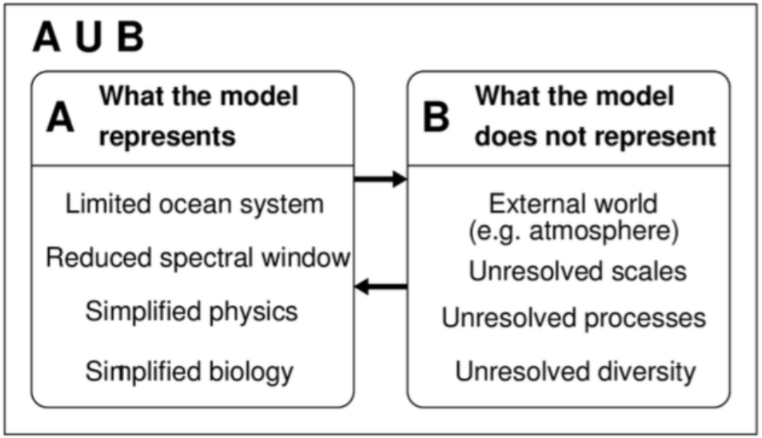
System approach to ocean modelling



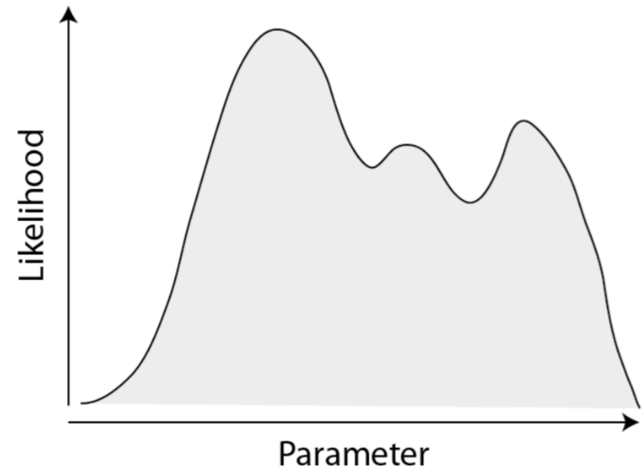
multi-modality of ocean distributions

- dealing more objectively with **model uncertainty** through a combination of **ensemble and stochastic methods**

A system approach to ocean modelling



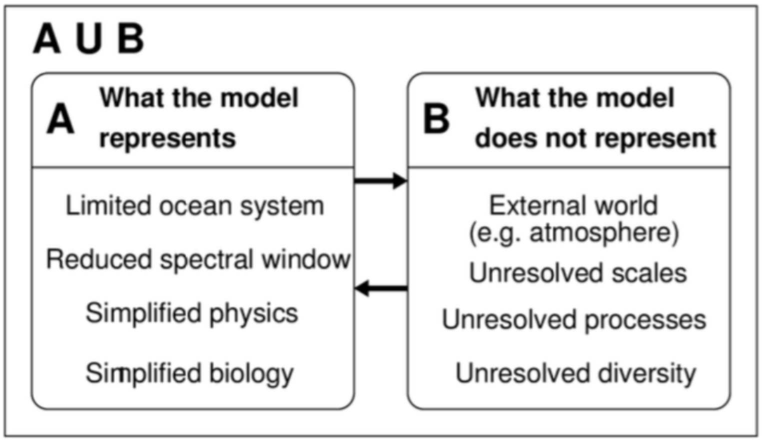
System approach to ocean modelling



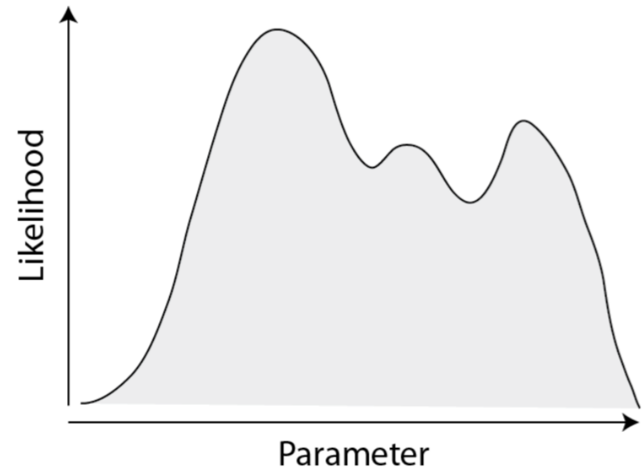
multi-modality of ocean distributions

- dealing more objectively with **model uncertainty** through a combination of **ensemble and stochastic methods**
- allowing to **explicitely sample the distribution of possible states** given explicitely formulated uncertainties (Bayesian approach)

A system approach to ocean modelling



System approach to ocean modelling

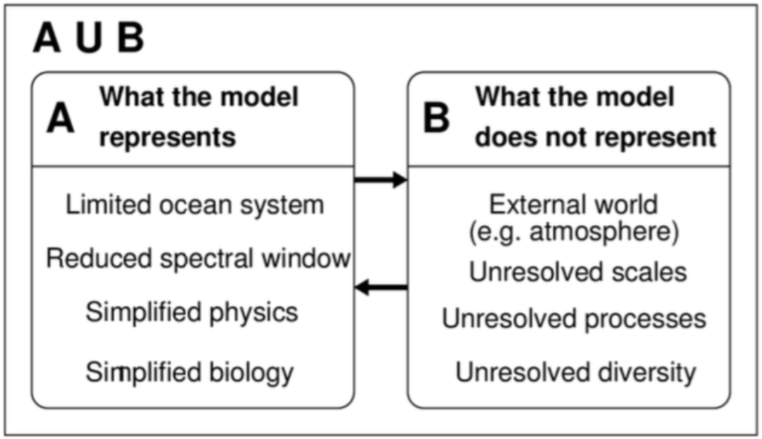


multi-modality of ocean distributions

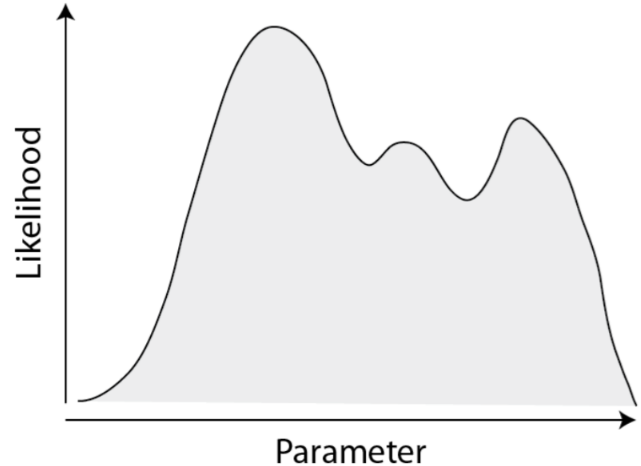
- dealing more objectively with **model uncertainty** through a combination of **ensemble and stochastic methods**
- allowing to **explicitely sample the distribution of possible states** given explicitely formulated uncertainties (Bayesian approach)

arguably a paradigm shift in ocean circulation modelling ?

A system approach to ocean modelling



System approach to ocean modelling



multi-modality of ocean distributions

- dealing more objectively with **model uncertainty** through a combination of **ensemble and stochastic methods**

- allowing to **explicitely sample the distribution of possible states** given explicitely formulated uncertainties (Bayesian approach)

arguably a paradigm shift in ocean circulation modelling ?

- but still unclear how to deal with the **daemon of dimensionality** (sampling a 10^n degrees of freedom pdf with 50 members...)

Part 4.

Towards data-driven ocean modelling

The changing usage of geoscientific data

- as we all know the **flow of geoscientific data** is increasing at an accelerating rate

The changing usage of geoscientific data

- as we all know the **flow of geoscientific data** is increasing at an accelerating rate
- this is true for **observational data** due to new satellite missions with increased downlink capacity, and evolution of in-situ networks. SWOT mission : 360 Mb/s versus 30kb/s for conventional altimeters

The changing usage of geoscientific data

- as we all know the **flow of geoscientific data** is increasing at an accelerating rate
- this is true for **observational data** due to new satellite missions with increased downlink capacity, and evolution of in-situ networks. SWOT mission : 360 Mb/s versus 30kb/s for conventional altimeters
- this is also true for **model data** with increase model resolution at higher temporal sampling. Basin to global scale submesoscale permitting model produce 100s Tb of data.

The changing usage of geoscientific data

- as we all know the **flow of geoscientific data** is increasing at an accelerating rate
- this is true for **observational data** due to new satellite missions with increased downlink capacity, and evolution of in-situ networks. SWOT mission : 360 Mb/s versus 30kb/s for conventional altimeters
- this is also true for **model data** with increase model resolution at higher temporal sampling. Basin to global scale submesoscale permitting model produce 100s Tb of data.
- sure, this implies that **in our scientific studies**, we use larger and larger dataset;

The changing usage of geoscientific data

- as we all know the **flow of geoscientific data** is increasing at an accelerating rate
- this is true for **observational data** due to new satellite missions with increased downlink capacity, and evolution of in-situ networks. SWOT mission : 360 Mb/s versus 30kb/s for conventional altimeters
- this is also true for **model data** with increase model resolution at higher temporal sampling. Basin to global scale submesoscale permitting model produce 100s Tb of data.
- sure, this implies that **in our scientific studies**, we use larger and larger dataset;
- this also motivates the design innovative data-mining strategies and **data analysis pipelines**

The changing usage of geoscientific data

- as we all know the **flow of geoscientific data** is increasing at an accelerating rate
- this is true for **observational data** due to new satellite missions with increased downlink capacity, and evolution of in-situ networks. SWOT mission : 360 Mb/s versus 30kb/s for conventional altimeters
- this is also true for **model data** with increase model resolution at higher temporal sampling. Basin to global scale submesoscale permitting model produce 100s Tb of data.
- sure, this implies that **in our scientific studies**, we use larger and larger dataset;
- this also motivates the design innovative data-mining strategies and **data analysis pipelines**
- several scientific fields have experienced **strong paradigm shifts** due to similar transitions over the last 10 years : genomics, neurosciences, and arguably social sciences.

The changing usage of geoscientific data

- as we all know the **flow of geoscientific data** is increasing at an accelerating rate
- this is true for **observational data** due to new satellite missions with increased downlink capacity, and evolution of in-situ networks. SWOT mission : 360 Mb/s versus 30kb/s for conventional altimeters
- this is also true for **model data** with increase model resolution at higher temporal sampling. Basin to global scale submesoscale permitting model produce 100s Tb of data.
- sure, this implies that **in our scientific studies**, we use larger and larger dataset;
- this also motivates the design innovative data-mining strategies and **data analysis pipelines**
- several scientific fields have experienced **strong paradigm shifts** due to similar transitions over the last 10 years : genomics, neurosciences, and arguably social sciences.

Will the change in the flow of data we experience in geoscience change the way how we produce knowledge in the future ?

How do ocean modellers use observationnal data ? (1/2)

How do ocean modellers use observationnal data ? (1/2)

- Observational data are used for constraining ocean model **initial states**. in practice, it is becoming more common to initialize models with ocean reanalysis (ex: NATL60)

How do ocean modellers use observationnal data ? (1/2)

- Observational data are used for constraining ocean model **initial states**. in practice, it is becoming more common to initialize models with ocean reanalysis (ex: NATL60)
- Observational data are used for constraining ocean model **forcing functions at the surface**; this is based on satellite data and atmospheric reanalysis data (not technically speaking observations)

How do ocean modellers use observationnal data ? (1/2)

- Observational data are used for constraining ocean model **initial states**. in practice, it is becoming more common to initialize models with ocean reanalysis (ex: NATL60)
- Observational data are used for constraining ocean model **forcing functions at the surface**; this is based on satellite data and atmospheric reanalysis data (not technically speaking observations)
- after running their simulations, ocean modellers **assess their model solutions** with observational data; (how seriously they do it often depend on reviewers…)

How do ocean modellers use observationnal data ? (1/2)

- Observational data are used for constraining ocean model **initial states**. in practice, it is becoming more common to initialize models with ocean reanalysis (ex: NATL60)
- Observational data are used for constraining ocean model **forcing functions at the surface**; this is based on satellite data and atmospheric reanalysis data (not technically speaking observations)
- after running their simulations, ocean modellers **assess their model solutions** with observational data; (how seriously they do it often depend on reviewers…)
- Observational data may also be used **for designing subgrid closures** and/or **assessing the representation of physical processes** in ocean models.

How do ocean modellers use observationnal data ? (1/2)

- Observational data are used for constraining ocean model **initial states**. in practice, it is becoming more common to initialize models with ocean reanalysis (ex: NATL60)
- Observational data are used for constraining ocean model **forcing functions at the surface**; this is based on satellite data and atmospheric reanalysis data (not technically speaking observations)
- after running their simulations, ocean modellers **assess their model solutions** with observational data; (how seriously they do it often depend on reviewers…)
- Observational data may also be used **for designing subgrid closures** and/or **assessing the representation of physical processes** in ocean models.

But overall, usages of observational data are pretty straighforward.

How do ocean modellers use observational data ? (2/2)

we can also observe the emergence of new usages which bridge from ocean modelling to data-assimilation :

How do ocean modellers use observational data ? (2/2)

we can also observe the emergence of new usages which bridge from ocean modelling to data-assimilation :

- **parameters of subgrid closures** can be adjusted through with approaches from data assimilation (Wirth and verron)

How do ocean modellers use observational data ? (2/2)

we can also observe the emergence of new usages which bridge from ocean modelling to data-assimilation :

- **parameters of subgrid closures** can be adjusted through with approaches from data assimilation (Wirth and verron)
- the **parameters of bulk formulae** can be adjusted using observational data with data assimilation (Meinvielle et al.)

How do ocean modellers use observational data ? (2/2)

we can also observe the emergence of new usages which bridge from ocean modelling to data-assimilation :

- **parameters of subgrid closures** can be adjusted through with approaches from data assimilation (Wirth and verron)
- the **parameters of bulk formulae** can be adjusted using observational data with data assimilation (Meinvielle et al.)
- **increments from data assimilation schemes** used in ocean reanalysis can be used for better understanding the origin and nature of ocean model bias.

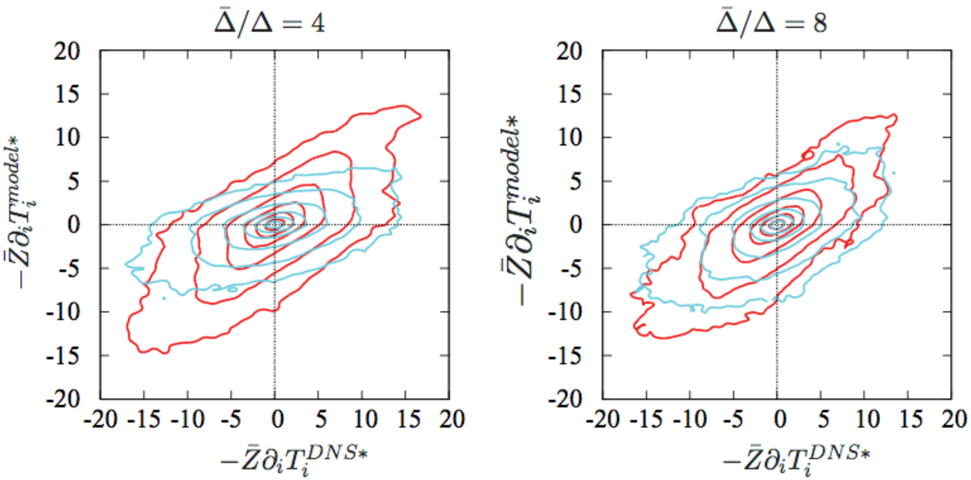
How do ocean modellers use observational data ? (2/2)

we can also observe the emergence of new usages which bridge from ocean modelling to data-assimilation :

- **parameters of subgrid closures** can be adjusted through with approaches from data assimilation (Wirth and verron)
- the **parameters of bulk formulae** can be adjusted using observational data with data assimilation (Meinvielle et al.)
- **increments from data assimilation schemes** used in ocean reanalysis can be used for better understanding the origin and nature of ocean model bias.

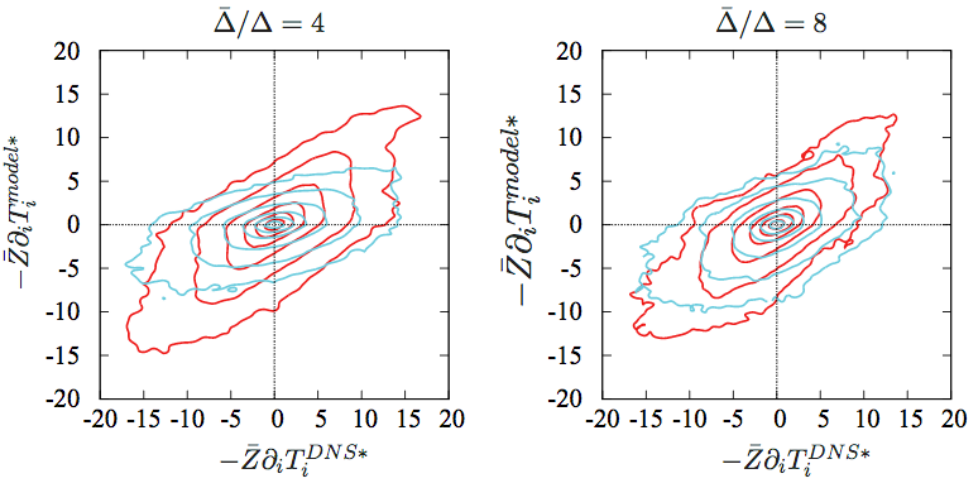
Arguably, ocean model design and usage is gradually using more methods coming from data science.

Data-driven design of subgrid scale closures



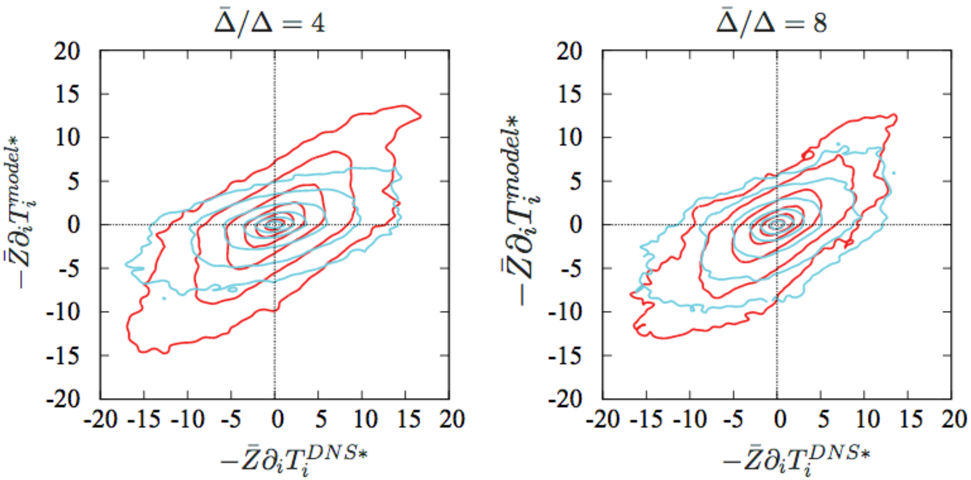
- Several groups in **fluid dynamics** have sucessfully designed SGS closures with data driven techniques, namely methods adapted to **find regularities and correlation in large datasets** :
    - Vollant et al (2017) : LES closures through machine-learning
    - Ling et al. (2016) : RANS modelling using deep neural networks

Data-driven design of subgrid scale closures



- Several groups in **fluid dynamics** have sucessfully designed SGS closures with data driven techniques, namely methods adapted to **find regularities and correlation in large datasets** :
  - Vollant et al (2017) : LES closures through machine-learning
  - Ling et al. (2016) : RANS modelling using deep neural networks
- some cloud micro-physics closures have already been designed from LES simulations using data-driven approaches.

Data-driven design of subgrid scale closures



- Several groups in **fluid dynamics** have sucessfully designed SGS closures with data driven techniques, namely methods adapted to **find regularities and correlation in large datasets** :
  - Vollant et al (2017) : LES closures through machine-learning
  - Ling et al. (2016) : RANS modelling using deep neural networks
- some cloud micro-physics closures have already been designed from LES simulations using data-driven approaches.
- There has been several attempts for designing SGS closures for ocean models with data-driven approaches, so far unsuccesfully.

Data-driven design of subgrid scale closures

$$\frac{\partial c_i}{\partial t} + \bar{u}_j \frac{\partial \langle c \rangle}{\partial x_j} = \frac{\partial}{\partial x_j}\left(K_{jj} \frac{\partial \langle c \rangle}{\partial x_j} \right)$$

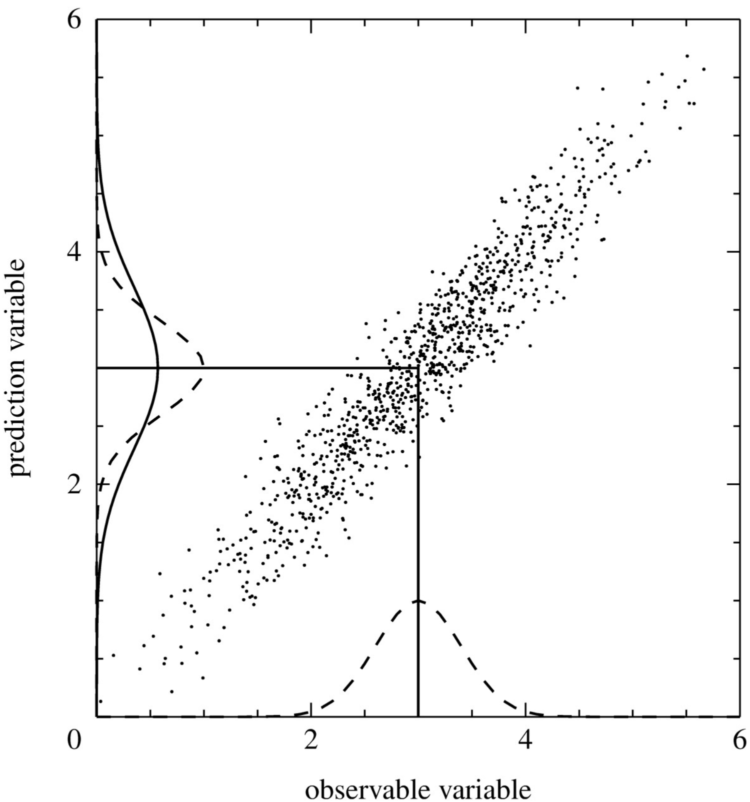
$$\langle u'_j, c' \rangle = -K_{jj} \frac{\partial \langle c \rangle}{\partial x_j}$$

Consider the **eddy diffusivity** closure

- central to several of our closures
- actually a statistical fit
- with a weak first principle basis…

Data-driven design of subgrid scale closures

$$\frac{\partial c_i}{\partial t} + \overline{u}_j \frac{\partial (c)}{\partial x_j} = \frac{\partial}{\partial x_j}\left(K_{jj} \frac{\partial (c)}{\partial x_j} \right)$$



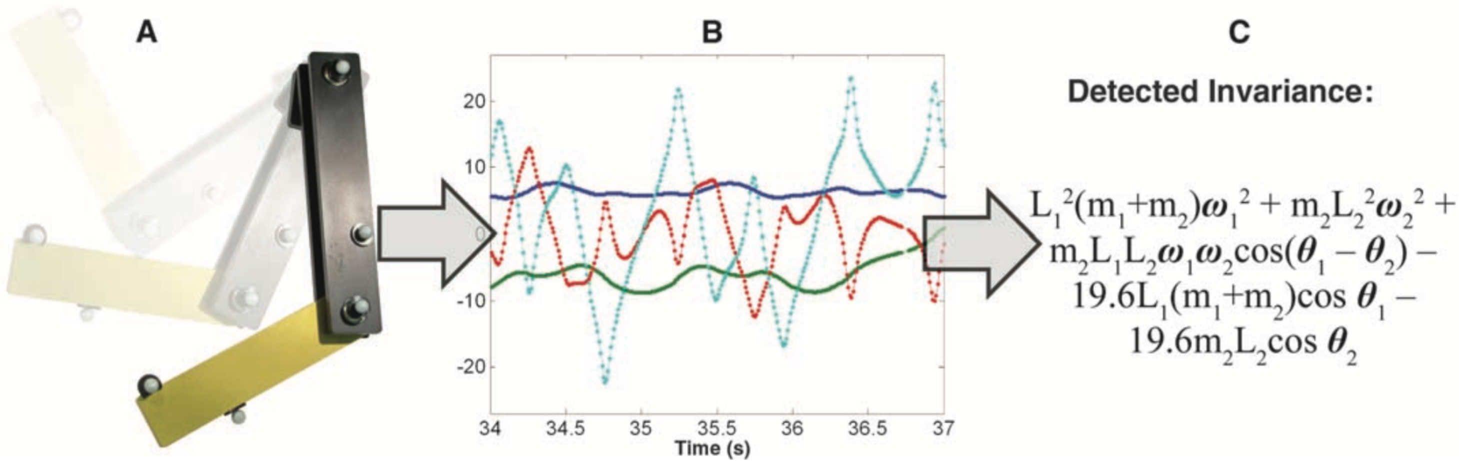
$$\langle u'_j, c' \rangle = -K_{jj} \frac{\partial (c)}{\partial x_j}$$

Consider the **eddy diffusivity** closure

- central to several of our closures
- actually a statistical fit
- with a weak first principle basis...

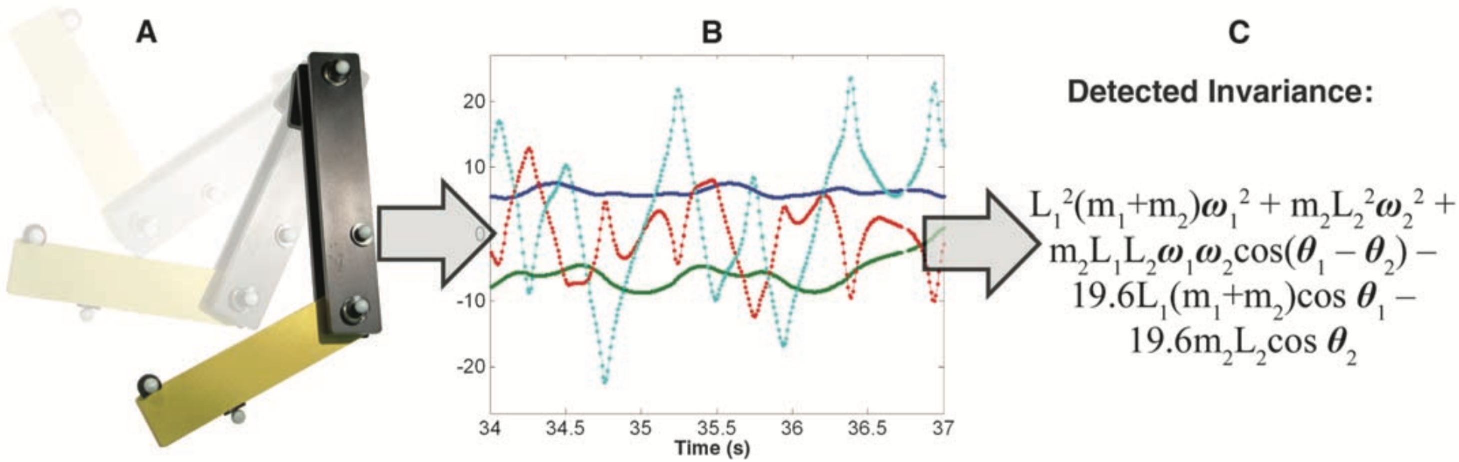
Physics-driven versus data driven ocean model design ?

Physics-driven versus data driven ocean model design ?



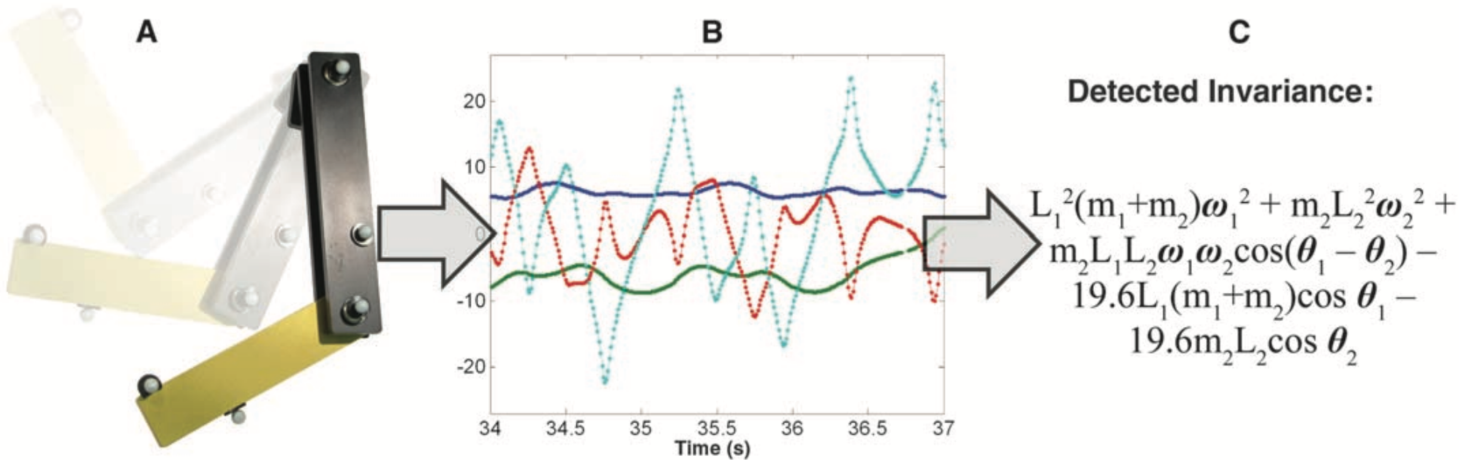
- data-driven techniques have be used for recovering the evolution equation of simple dynamical systems

Physics-driven versus data driven ocean model design ?



- data-driven techniques have be used for recovering the evolution equation of simple dynamical systems
- applied successsfully so far to low dimensional systems with perfect observations (Lipson 2009)

Physics-driven versus data driven ocean model design ?



- data-driven techniques have be used for recovering the evolution equation of simple dynamical systems
- applied successsfully so far to low dimensional systems with perfect observations (Lipson 2009)
- how / when would those techniques be applicable to high dimensional / imperfectly observed dyamical systems ?

Wrapping-up Lecture #3

Lecture #3

Towards data-driven, probabilistics ocean circulation models

Lecture #3

Towards data-driven, probabilistics ocean circulation models

Here we have

Lecture #3

Towards data-driven, probabilistics ocean circulation models

Here we have

- illustrated several aspect of the ill-posed nature of deterministic geoscienctific modelling.

Lecture #3

Towards data-driven, probabilistics ocean circulation models

Here we have

- illustrated several aspect of the ill-posed nature of deterministic geoscienctific modelling.
- shown (*said*) that **stochastic closures are a promising approach for reducing ocean circulation models bias**

Lecture #3

Towards data-driven, probabilistics ocean circulation models

Here we have

- illustrated several aspect of the ill-posed nature of deterministic geoscienctific modelling.
- shown (*said*) that **stochastic closures are a promising approach for reducing ocean circulation models bias**
- shown that **stochastic modelling actually allows to generically deal with various sources of uncertainties** inherent to ocean models

Lecture #3

Towards data-driven, probabilistics ocean circulation models

Here we have

- illustrated several aspect of the ill-posed nature of deterministic geoscienctific modelling.
- shown (*said*) that **stochastic closures are a promising approach for reducing ocean circulation models bias**
- shown that **stochastic modelling actually allows to generically deal with various sources of uncertainties** inherent to ocean models
- discussed how ensemble simulations can help **understanding the drivers of low frequency oceanic variability**

Lecture #3

Towards data-driven, probabilistics ocean circulation models

Here we have

- illustrated several aspect of the ill-posed nature of deterministic geoscienctific modelling.
- shown (*said*) that **stochastic closures are a promising approach for reducing ocean circulation models bias**
- shown that **stochastic modelling actually allows to generically deal with various sources of uncertainties** inherent to ocean models
- discussed how ensemble simulations can help **understanding the drivers of low frequency oceanic variability**
- advocated that the **Bayesian framework** allows to consider ocean circulation models for what they actually are, namely **probabilistic models of possible states of the ocean**.

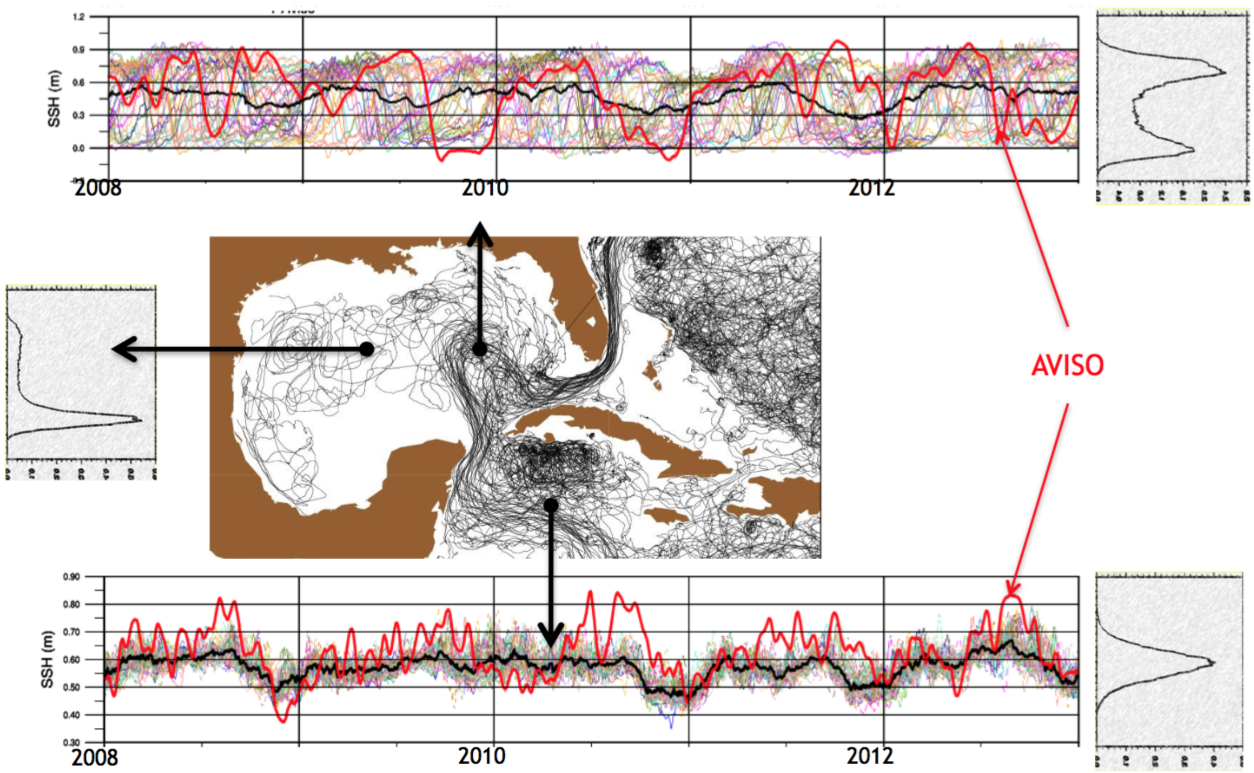
Lecture #3

Towards data-driven, probabilistics ocean circulation models

Here we have

- illustrated several aspect of the ill-posed nature of deterministic geoscienctific modelling.
- shown (*said*) that **stochastic closures are a promising approach for reducing ocean circulation models bias**
- shown that **stochastic modelling actually allows to generically deal with various sources of uncertainties** inherent to ocean models
- discussed how ensemble simulations can help **understanding the drivers of low frequency oceanic variability**
- advocated that the **Bayesian framework** allows to consider ocean circulation models for what they actually are, namely **probabilistic models of possible states of the ocean**.
- asked whether **ocean circulation model development** will be more **data-driven** in the future.

Additional material


SSH (m)
2008
2010
2012
AVISO
SSH (m)
2008
2010
2012