Modele de classification de profils (PCM) océaniques: application aux données Argo

Ifremer/LOPS:
G. Maze, H. Mercier, C. Feucher (PhD), F. Paul, C. Kermabon

Telecom Bretagne:
P. Tandeo,R. Fablet, P. Lenca, M. Radcenco, C. Le Goff

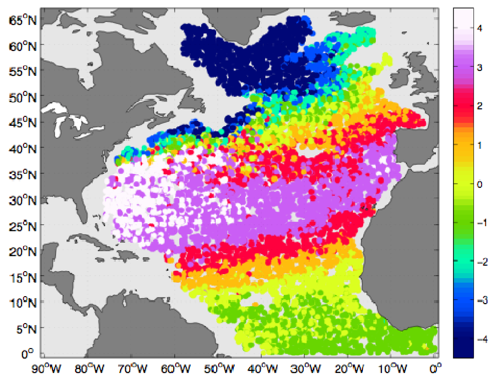
5eme reunion annuelle Projet NAOS
21/09/2016 à Villefranche sur Mer

Problematic 1/2

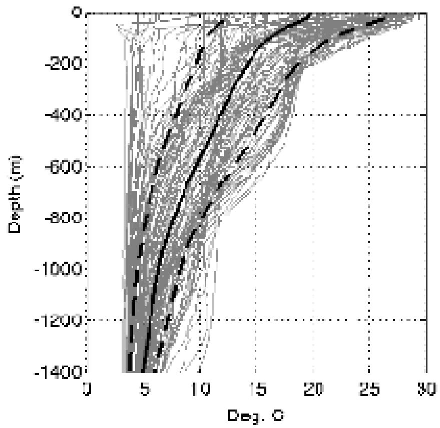
Heat is not evenly distributed in ocean

Heat is stored in reservoirs organized around fronts

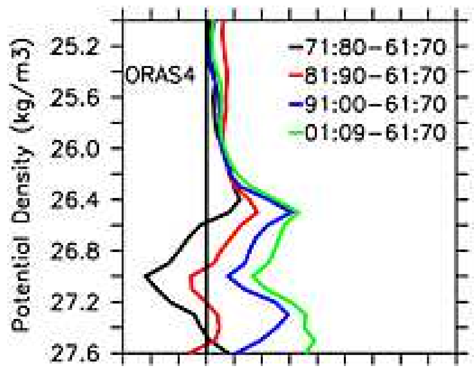
Heat changes are nonlinear, not uniform



Vertical mean temperature



Random sample of Argo temperature profiles



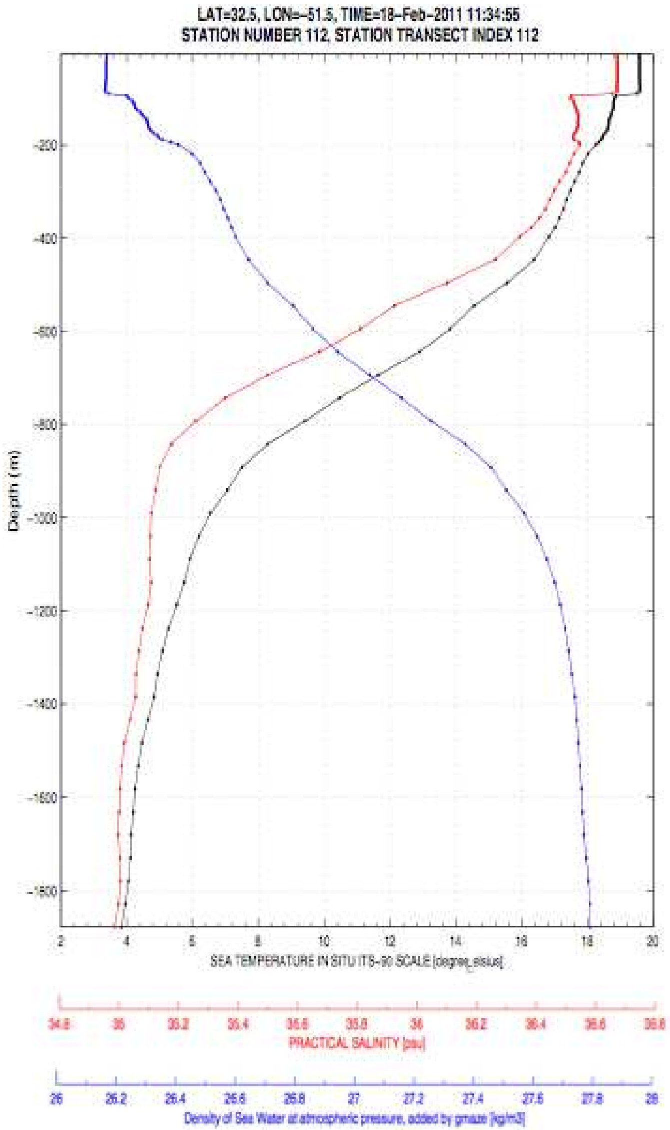
Hakkinen et al, GRL 2015

How to identify interior heat structure and its variability ?

Problematic 2/2

Diagnostics are not objective

Argo profile sample



Structure verticale complexe

Ce que l'opérateur identifie d'un clin d'oeil
est très difficilement descriptible
de manière objective:
- Profondeur de mélange ?
- Stratification principale ?
- Outliers ?

Diagnostiques complexes et peu performants
Arbres de décision construit “à la main”
Paramètres inadéquates
Large barre d'erreur
Nécessite temps opérateur

Identifying coherent structures or “patterns” in profiles

Profile Classification Model

with a Gaussian Mixture Modeling of the PDF

pattern = *typical* profile

typical profile = *recurrent* profile

recurrent profile = peak in the PDF

Classification methods have great properties:
- they can handle asymmetries
- they do not promote “homogeneous” patterns
- they can detect single/isolated patterns

It is a descriptive method to identify class of items so that items from a class are similar while being different from the items of the other classes.

It can be supervised (we know the classes)
or un-supervised (we have to discover classes)

Profile Classification Model
with a Gaussian Mixture Modeling of the PDF

Decomposition of a PDF into a weighted sum of Gaussians:

$$p(\mathbf{x}|\mu, \Sigma) = \sum_{k=1}^{K} \lambda_k \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k)$$

K sets the number of class of profiles (K≥ # of peaks)
Each class is modeled with a multi-dimensional Gaussian density

We will classify profiles according to there probability to “belong” to a class in the PDF

In practice, given:

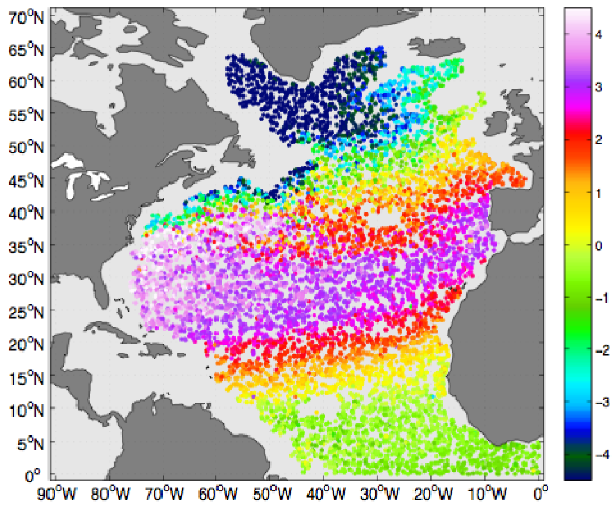
- a collection of data **x** (N vectors of D-dimensions, N profiles of D levels)
- a number of class **K**

the analysis determines weights λ_k and Gaussian parameters μ_k, Σ_k to maximize the (log) likelihood of the mixture.

The solving method is the “Expectation-Maximization” algorithm (not detailed here, see Bilmes, 1998)

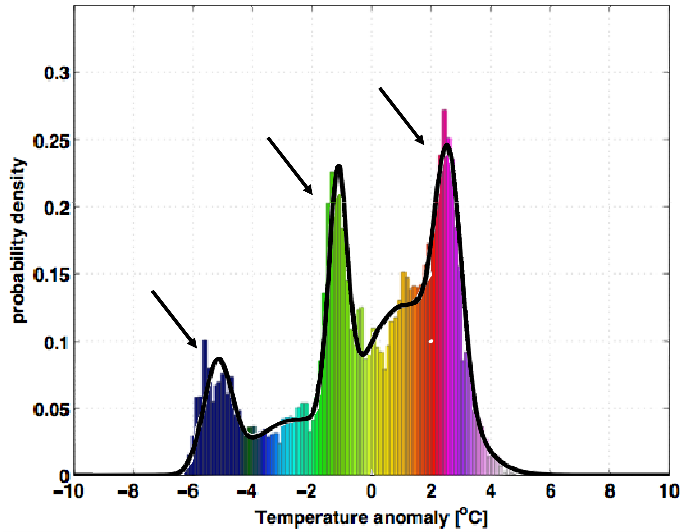
Classification of profiles according to their vertical mean temperature, i.e. integral heat content

Argo profiles - 0/1400m

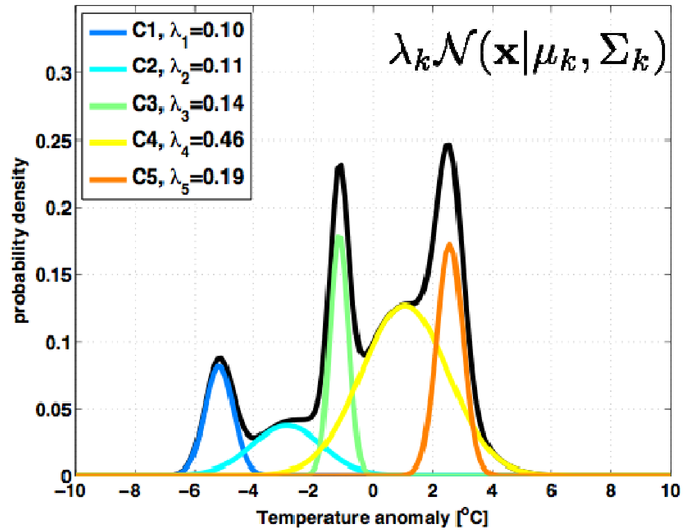


Vertical mean temperature / anom. vs domain average (~9ºC)

Observed PDF vs Model PDF (GMM, K=5)

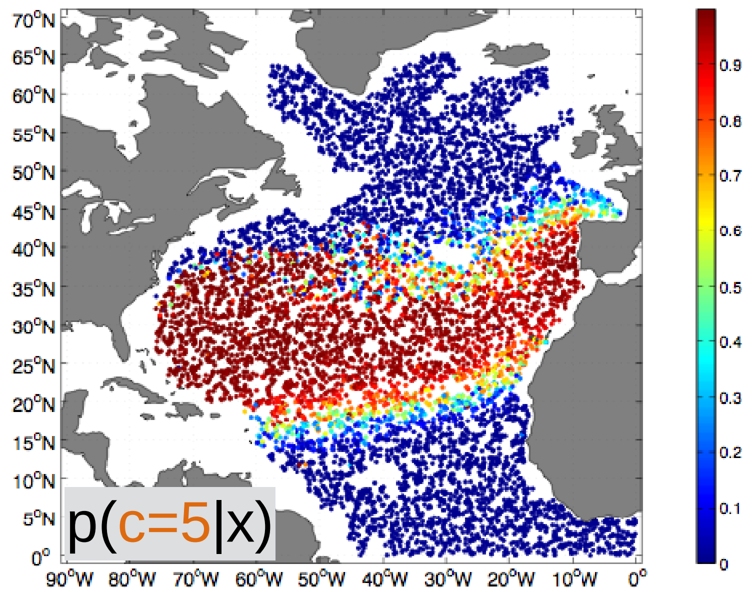


Details of K=5 GMM classes



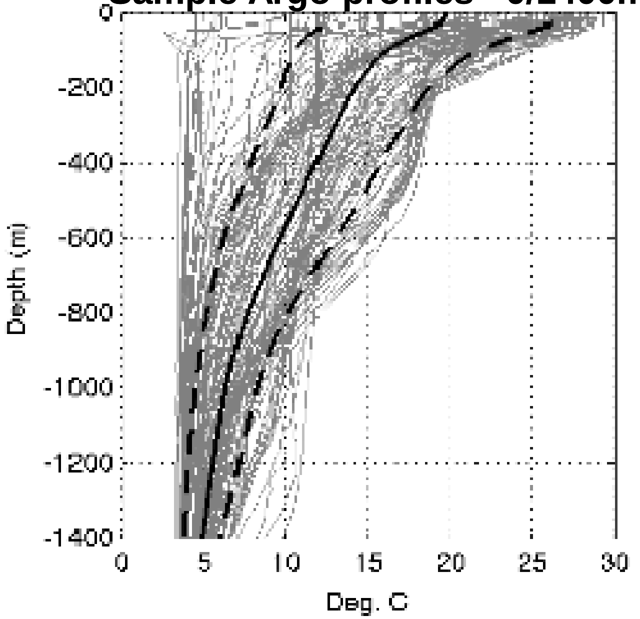
5 classes capture the essential PDF structure. In this trivial case, a class is defined by 1-D pair of μ_k, Σ_k

We can compute the probability for a profile to belong to the class k. This is the posterior probability p(c=k|x)

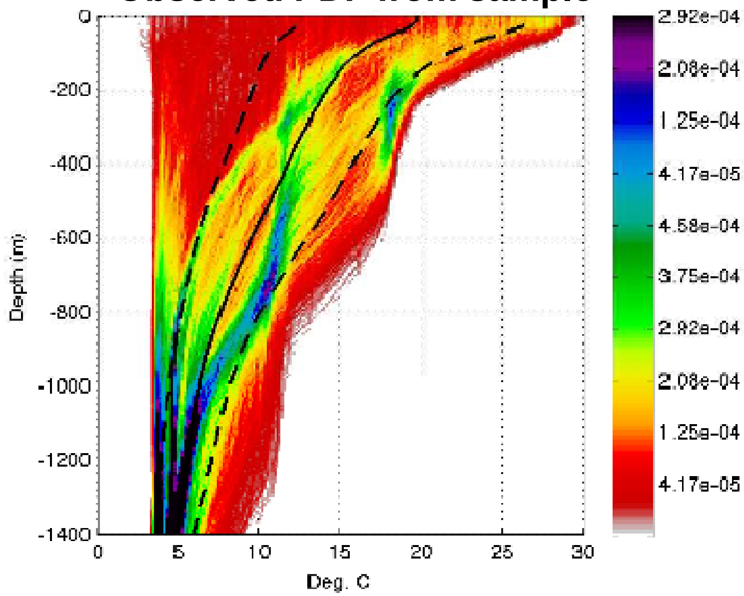


Classification of profiles according to their vertical temperature structures

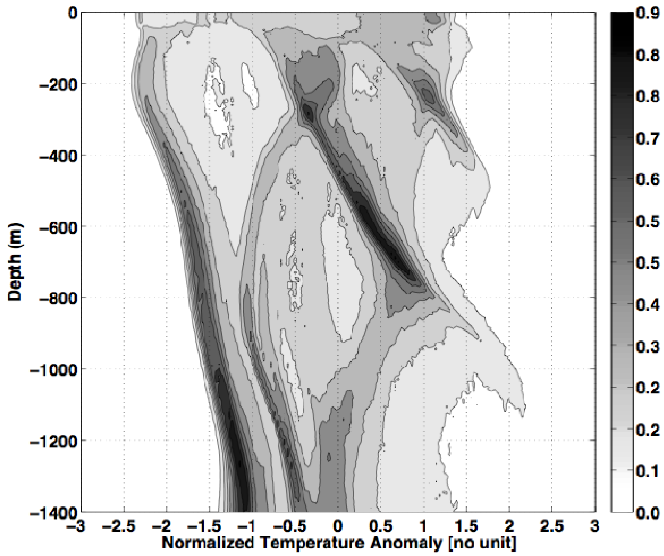
Sample Argo profiles - 0/1400m



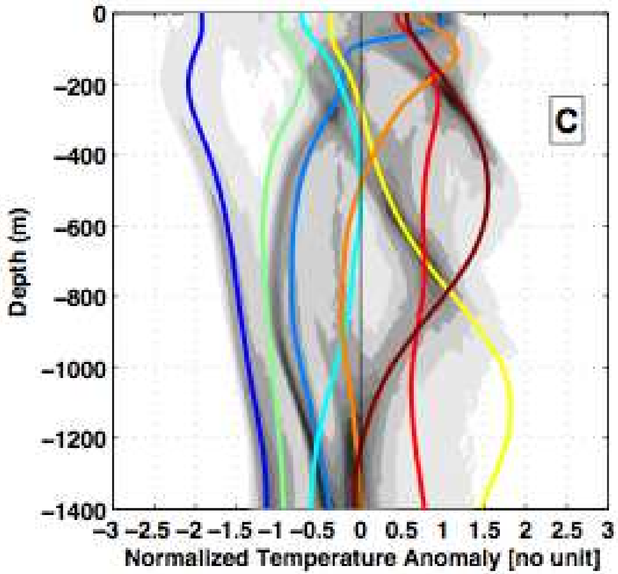
Observed PDF from sample



Complete Observed PDF (D-normalized)



K=8 GMM classes



We fit several PCM to the PDF of profiles

K=8 classes capture the essential PDF structure

Class #1
Class #2
Class #3
Class #4
Class #5
Class #6
Class #7
Class #8

Class #1
Homogeneously cold

Class #4
Cold & surface spread

Class #3
Near-neutral, large spread

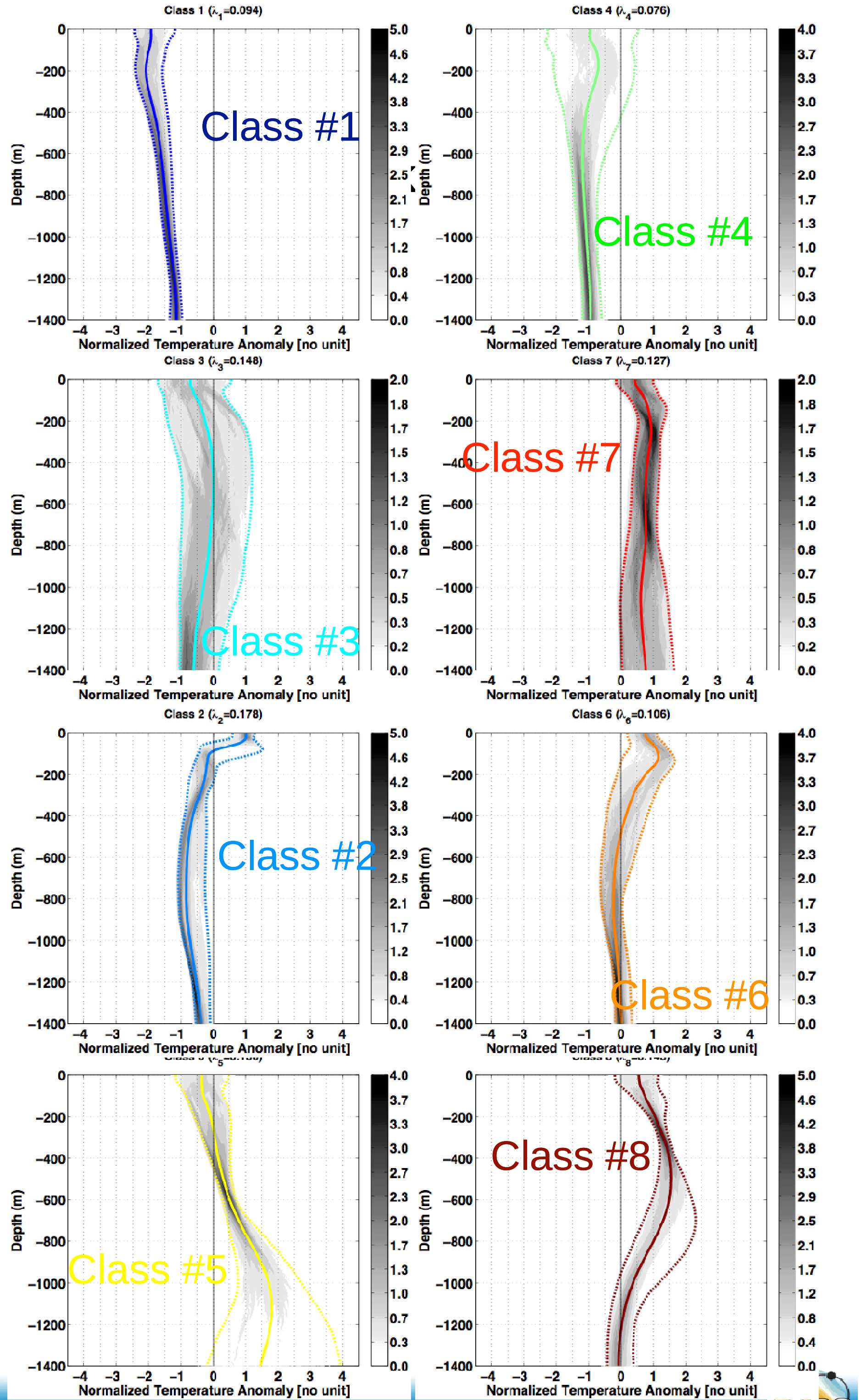
Class #7
Homogeneously warm

Class #2
Mid-depth cold & surface-trapped heat

Class #6
Neutral at depth & warm sub-surface

Class #5
Near neutral at the surface & warm at depth in the

Class #8
Near neutral at surface and depth & warm at mid-depth
in western s



Class #1
Homogeneously cold

Class #4
Cold & surface spread

Class #3
Near-neutral, large spr

Class #7
Homogeneously warm

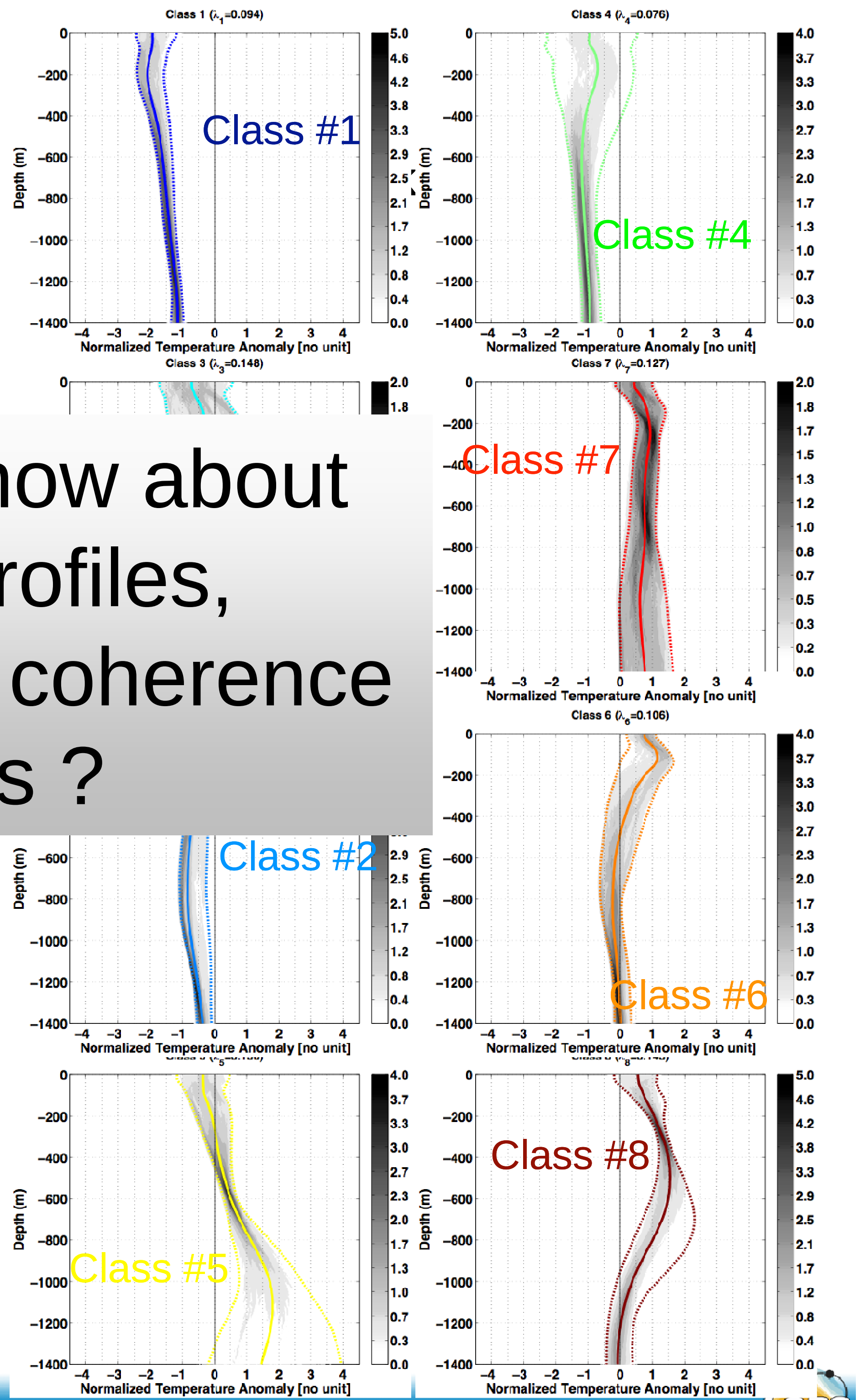
Class #2
Mid-depth cold & surfa

Class #6
Neutral at depth & warm sub-surface

Class #5
Near neutral at the surface & warm at depth in the

Class #8
Near neutral at surface and depth & warm at mid-depth
in western su

PCM doesn't know about location of profiles, is there a spatial coherence in classes ?



Classification of profiles according to their vertical temperature structures

Class #1
Homogeneously cold
in the subpolar gyre

Class #4
Cold & surface spread
north of the GS / NAC

Class #3
Near-neutral, large spread
in the inter-gyre / GS / NAC

Class #7
Homogeneously warm
in south-eastern subtropical gyre

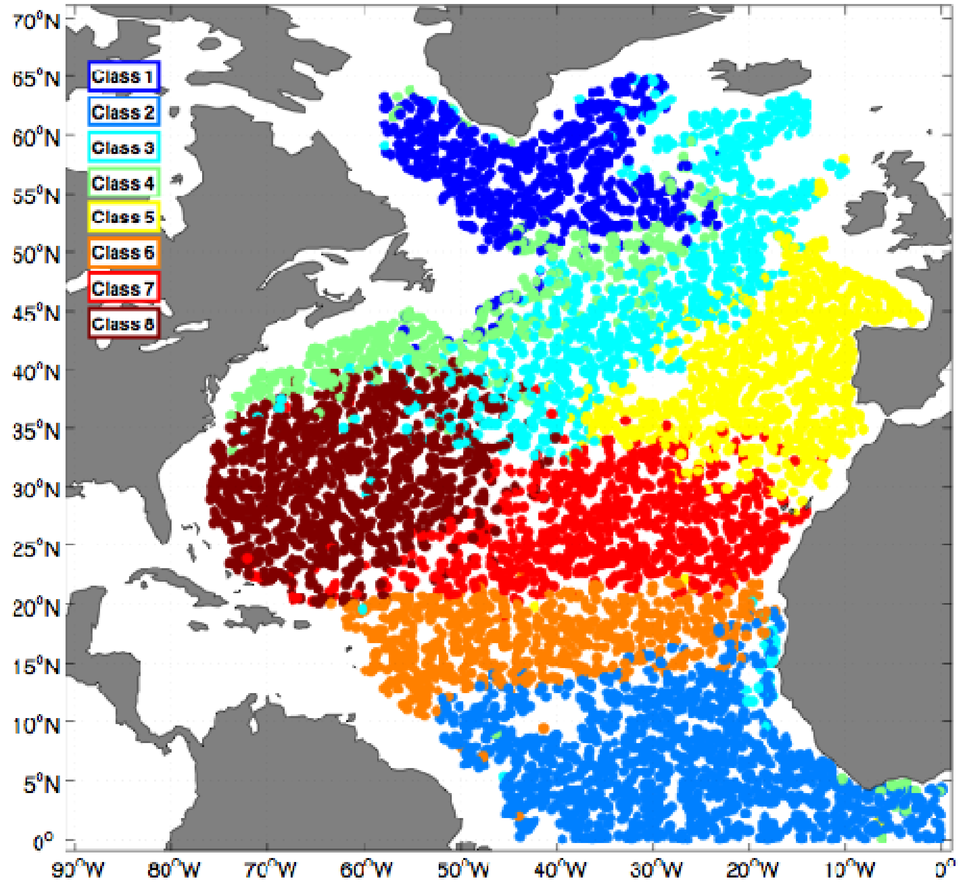
Class #2
Mid-depth cold & surface-trapped heat
in the equatorial band

Class #6
Neutral at depth & warm sub-surface
in the western tropical band

Class #5
Near neutral at the surface & warm at depth in the
Mediterranean outflow influence

Class #8
Near neutral at surface and depth & warm at mid-depth
in western subtropical gyre region

Each profile is *labelled* with the class # with maximum of probability p(c|x)



Patterns are spatially coherent: model of stack of water masses are unique

Recap

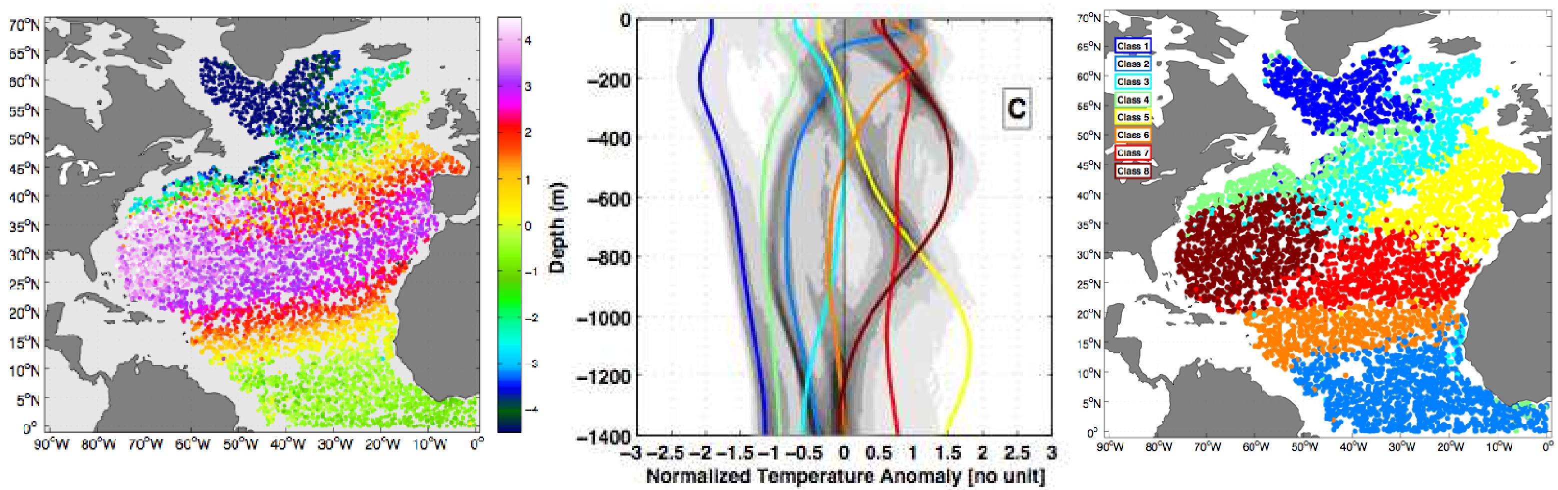
Maze et al, 2016, Prog. Oc.

We used un-supervised classification method to identify coherent structures or “patterns”

in a large collection of temperature profiles

Patterns are “data-driven models” of internal heat reservoirs

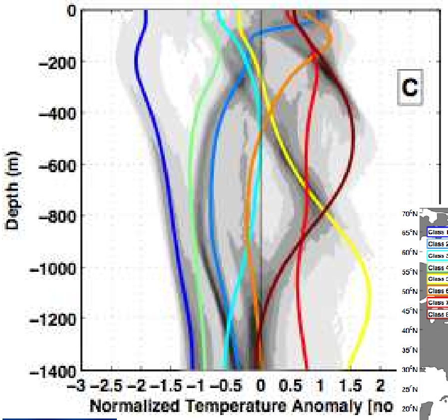
Patterns are spatially coherent: stack of water masses are unique

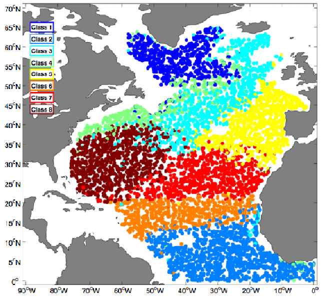


We call this a **Profile Classification Model (PCM)**

Model can be used to classify new data

Posterior probabilities p(c|x) provide a useful OHC decomposition

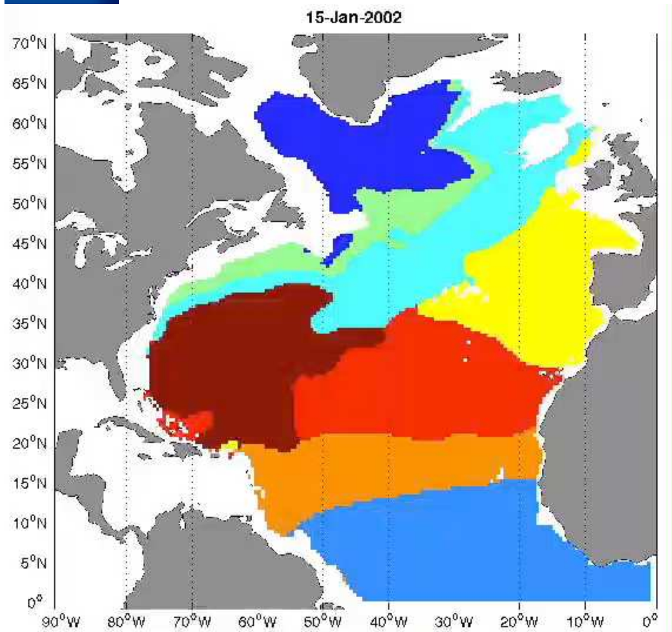


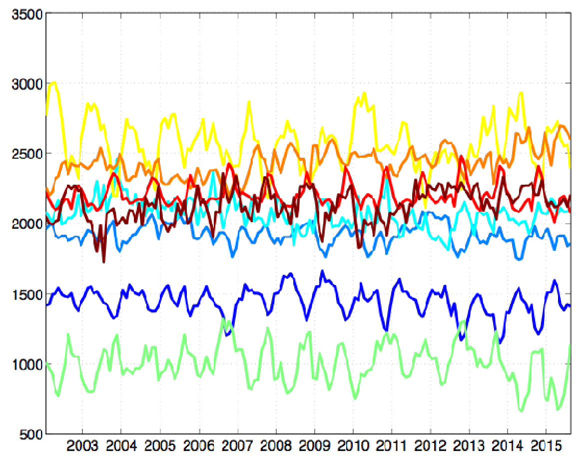


So, we derived a model for the heat content structure in the North Atlantic

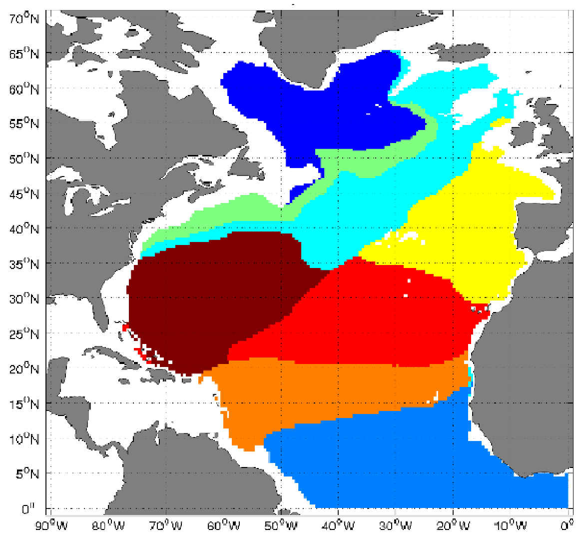
How can we use this model to characterize heat content variability ?

Oceanic re-analysis, ISAS-13 (Gaillard et al, J. Clim. 2016)
2002-2015,monthly, 1/2x1/2, 0-2000m fields from Argo OI
Take the Argo PCM and classify ISAS-3D gridded time series



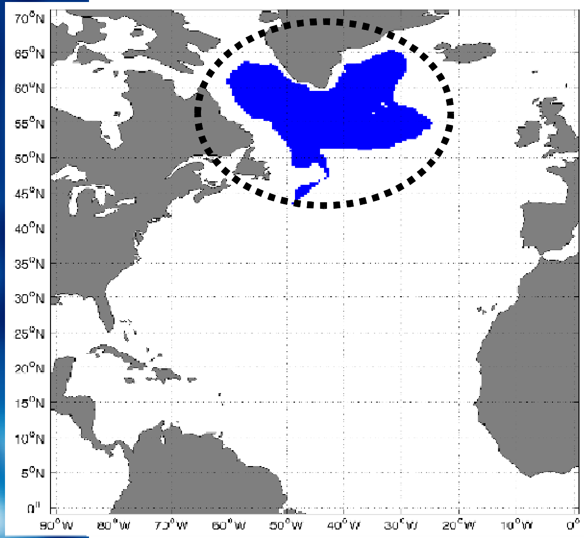


\# grid points / class



locally most frequent class

Class contours as “natural” regional boxes



Focus on the subpolar gyre Ocean Heat Content:

full OHC:

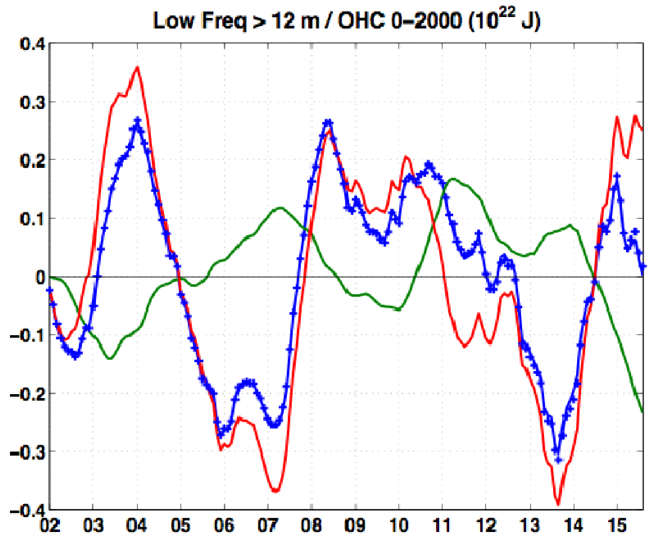
$$\iint_{x,y} \left(p(\mathbf{c}_k|\mathbf{x}) \int_{z=0}^{2000} \rho_0 c_p \theta(x,y,z) dz \right) dxdy$$

OHC with mean temperature / variable class contour:

$$\iint_{x,y} \left(p(\mathbf{c}_k|\mathbf{x}) \int_{z=0}^{2000} \rho_0 c_p \overline{\theta(x,y,z)} dz \right) dxdy$$

OHC with mean class contour / variable temperature:

$$\iint_{x,y} \left(\overline{p(\mathbf{c}_k|\mathbf{x})} \int_{z=0}^{2000} \rho_0 c_p \theta(x,y,z) dz \right) dxdy$$



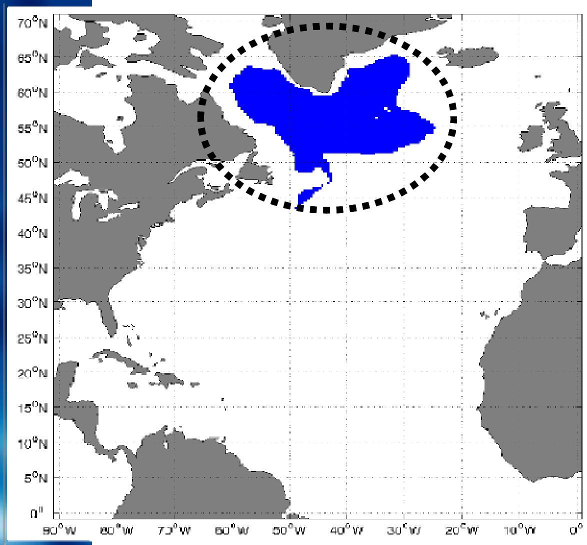
PCM property: $\sum_{k=1}^{K} p(\mathbf{c}_k|\mathbf{x}) = 1$

for each grid point (profile), a fraction $0<p(c_k|x)<1$ of OHC can be attributed to class k without loosing heat

“subpolar gyre” OHC variability
driven by its extent,
modulated by local temperature variability

What’s in these signals ?

Class contours as “natural” regional boxes



Focus on the subpolar gyre Ocean Heat Content:

full OHC:

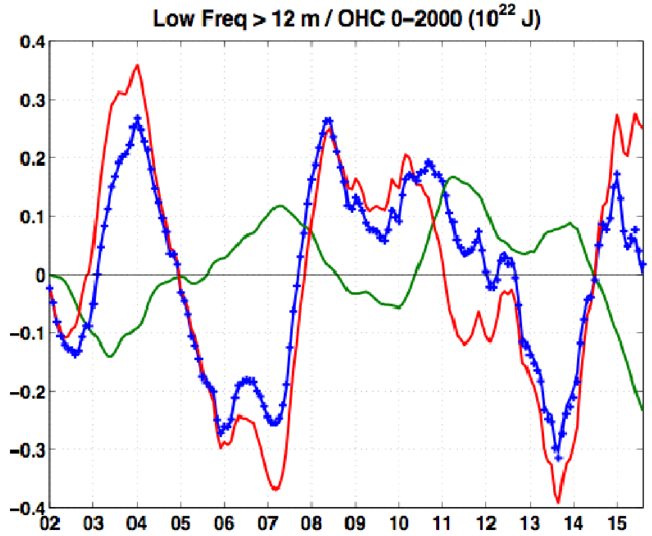
$$\iint_{x,y} \left(p(\mathbf{c}_k|\mathbf{x}) \int_{z=0}^{2000} \rho_0 c_p \theta(x,y,z) dz \right) dxdy$$

OHC with mean temperature / variable class contour:

$$\iint_{x,y} \left(p(\mathbf{c}_k|\mathbf{x}) \int_{z=0}^{2000} \rho_0 c_p \overline{\theta(x,y,z)} dz \right) dxdy$$

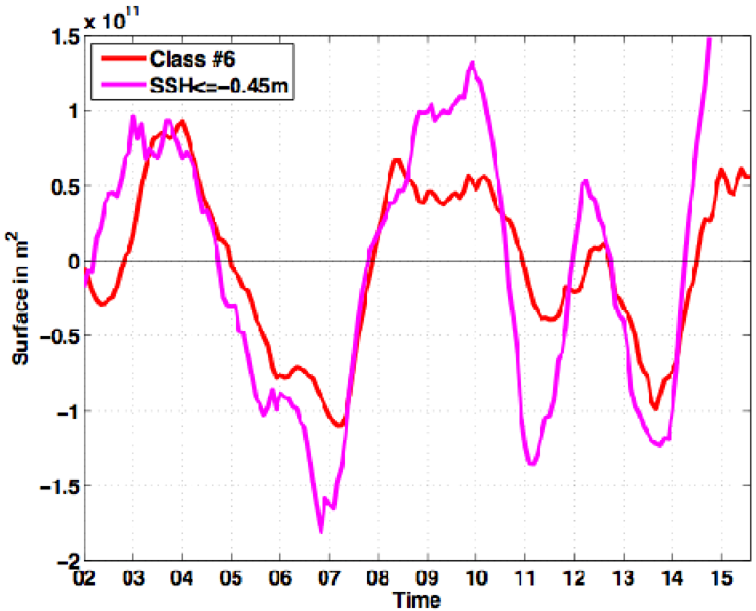
OHC with mean class contour / variable temperature:

$$\iint_{x,y} \left(\overline{p(\mathbf{c}_k|\mathbf{x})} \int_{z=0}^{2000} \rho_0 c_p \theta(x,y,z) dz \right) dxdy$$



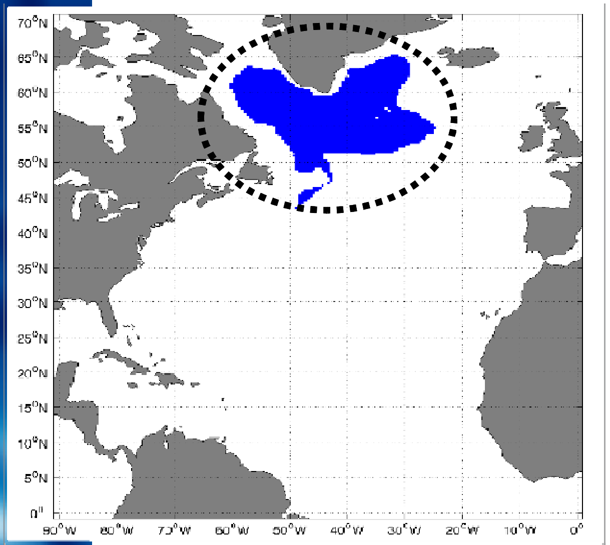
14

Class surface vs SSH contour extent



“Gyre” extent is dynamically constrained

Class contours as “natural” regional boxes



Focus on the subpolar gyre Ocean Heat Content:

full OHC:

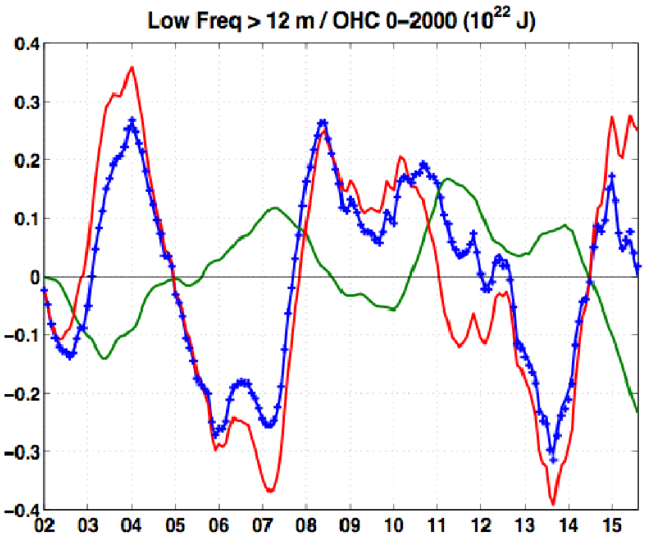
$$\iint_{x,y}\left(p(\mathbf{c}_k|\mathbf{x})\int_{z=0}^{2000}\rho_0 c_p \theta(x,y,z)dz\right)dxdy$$

OHC with mean temperature / variable class contour:

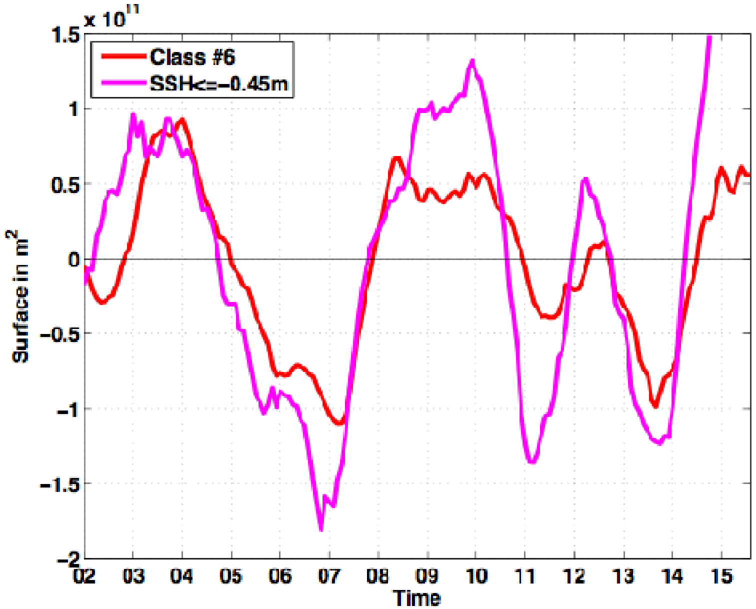
$$\iint_{x,y}\left(p(\mathbf{c}_k|\mathbf{x})\int_{z=0}^{2000}\rho_0 c_p \overline{\theta(x,y,z)}dz\right)dxdy$$

OHC with mean class contour / variable temperature:

$$\iint_{x,y}\left(\overline{p(\mathbf{c}_k|\mathbf{x})}\int_{z=0}^{2000}\rho_0 c_p \theta(x,y,z)dz\right)dxdy$$

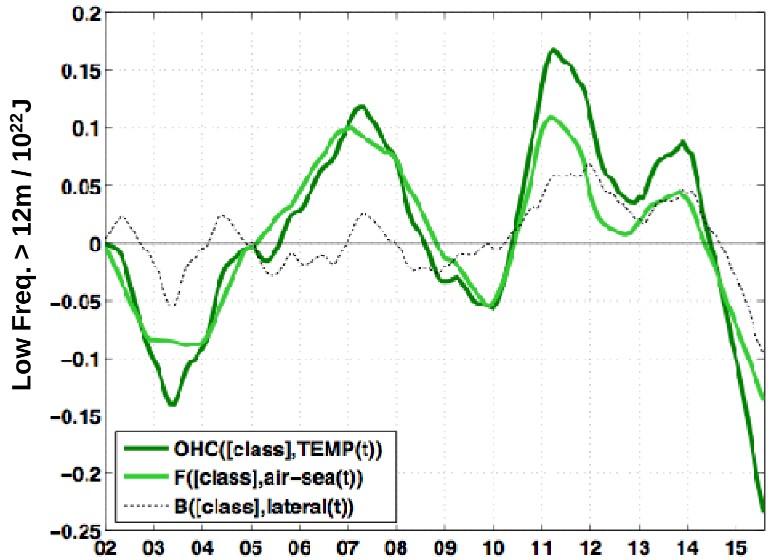


Class surface vs SSH contour extent

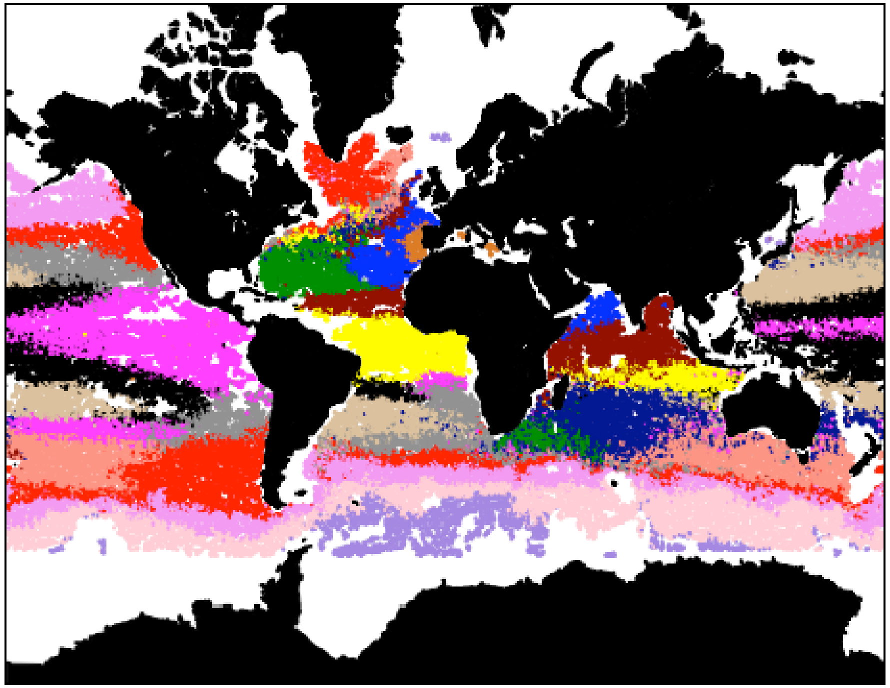


“Gyre” extent is dynamically constrained

Class temp. vs Air-sea Flux



“Gyre” temperature driven by Air-sea heat fluxes



Mining larger datasets

Spark

fast and general engine for large-scale data processing at LOPS-Cersat

Développement en cours: portage de la méthode PCM sur cluster big data

ANR SONIFIC à soumettre

x10

here: 15 years, N.Atl.: 0.1×10^6 profiles

x10

All Argo: 15 years, global: 1.5×10^6

ORA-S4: 50 years, monthly, global 1/1 gridded: 26×10^6

ISAS13+nrt: 13 years, monthly, global 1/2 gridded: 43×10^6

x10

HadGEM: 140 years, monthly, global 1/1 gridded: 92×10^6

x10

ORCA025: 40 years, weekly, global 1/4 gridded: $1\,400 \times 10^6$

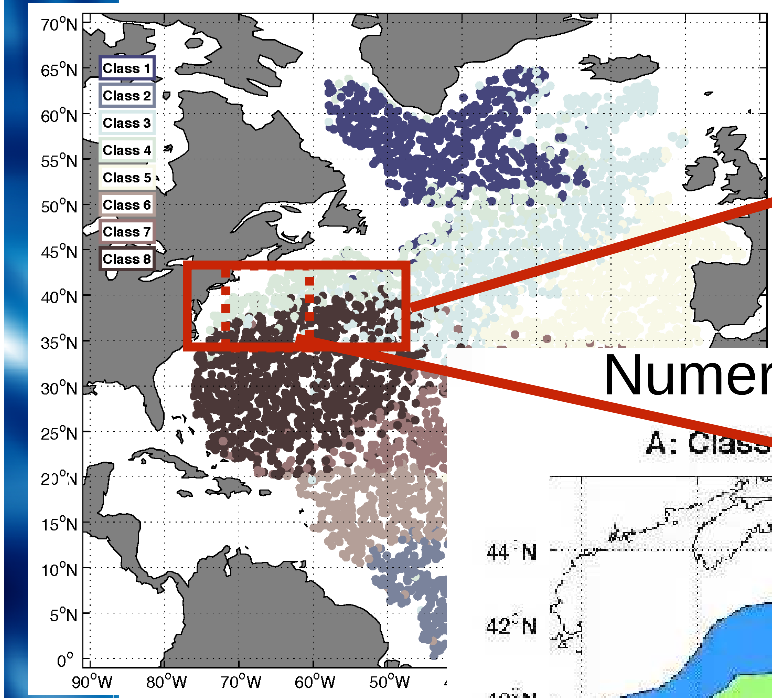
CMIP5: 50 years, monthly, global 1/1 gridded, 50 runs: $1\,500 \times 10^6$

DRAKKAR12: 20 years, weekly, global 1/12 gridded: $6\,400 \times 10^6$

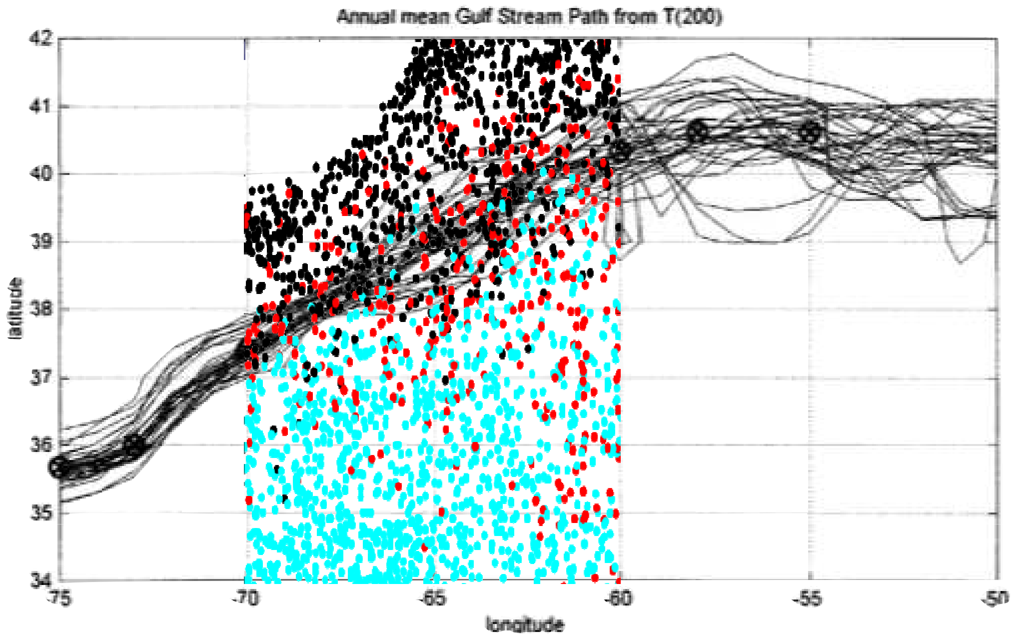
x10

OCCIPUT: 50 years, weekly, global 1/4 gridded, 50 runs: $8\,900 \times 10^6$

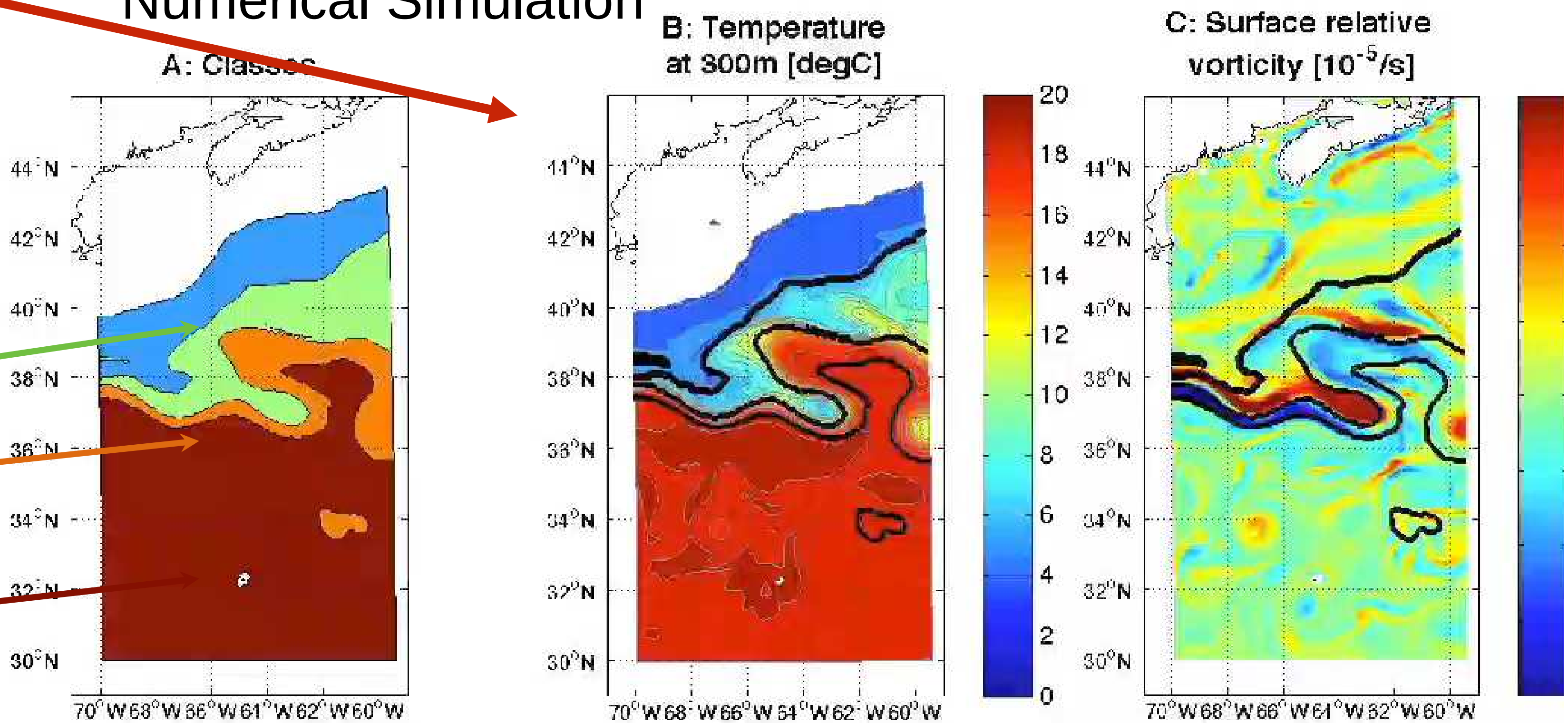
In frontal regions, a PCM is very efficient in identifying profiles from flanks and the core:



Argo



Numerical Simulation



Cold flank

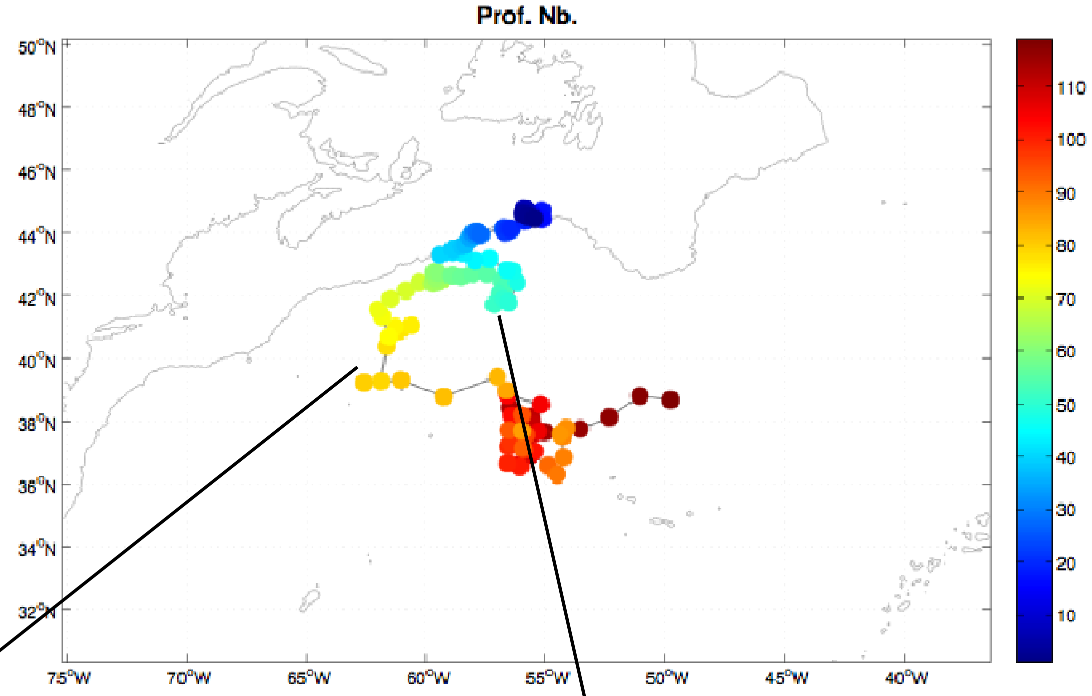
GS core

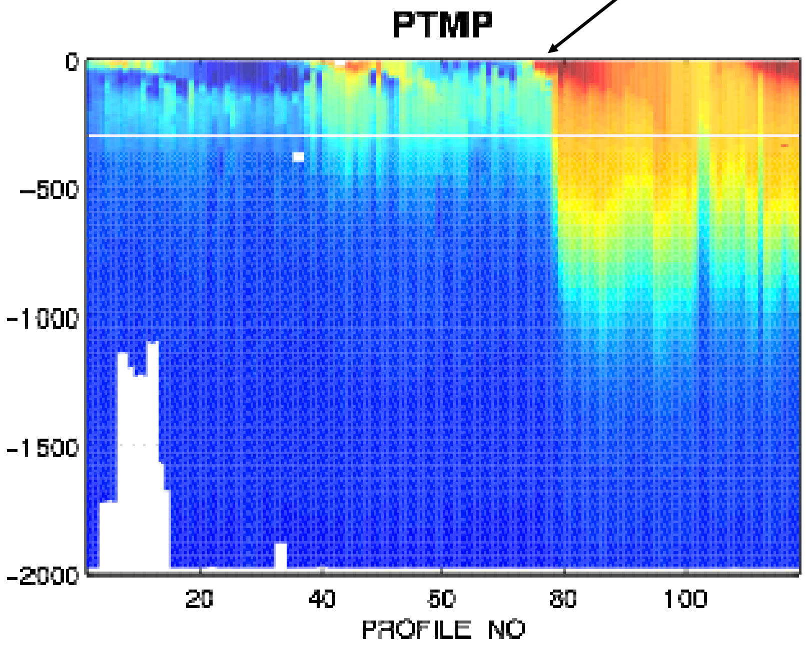
Warm flank

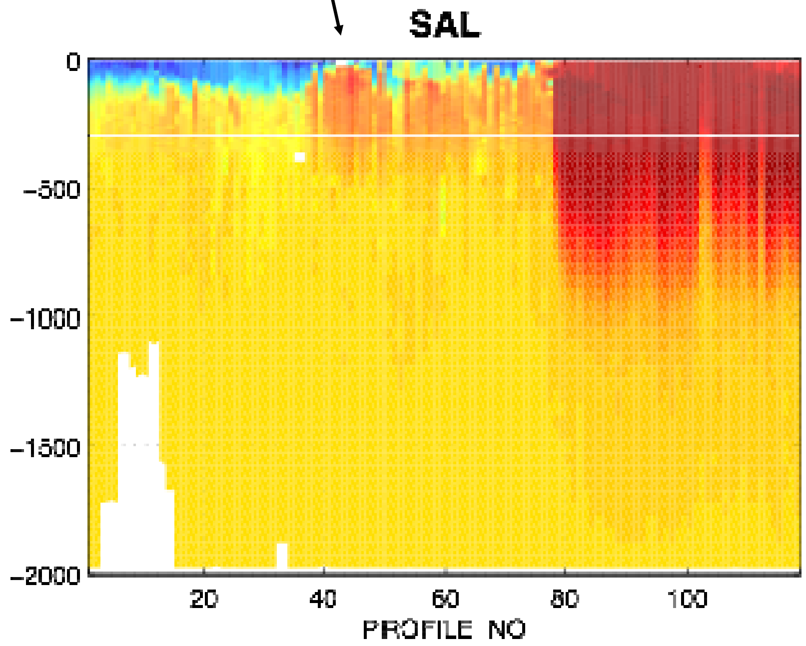
Improved referencing for in-situ data validation

eg: Argo Float 4900136

goes near the Gulf Stream at cyc. #50 and through it at cyc. #80

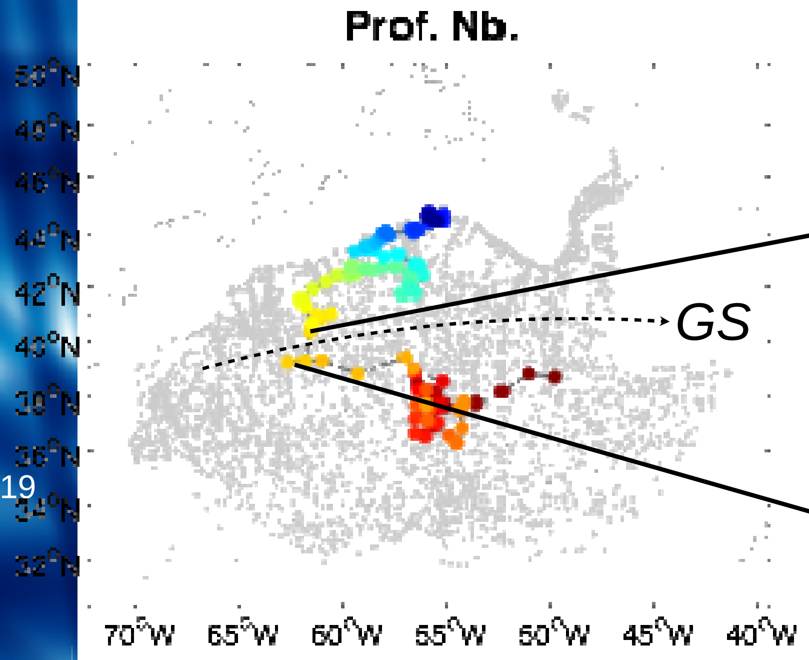


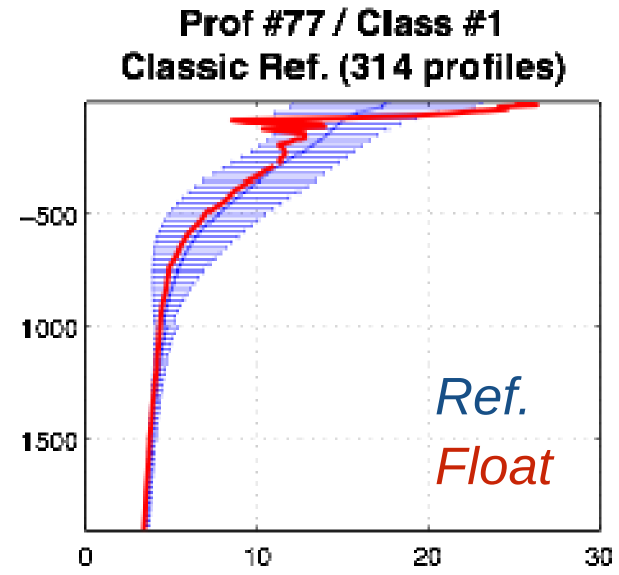


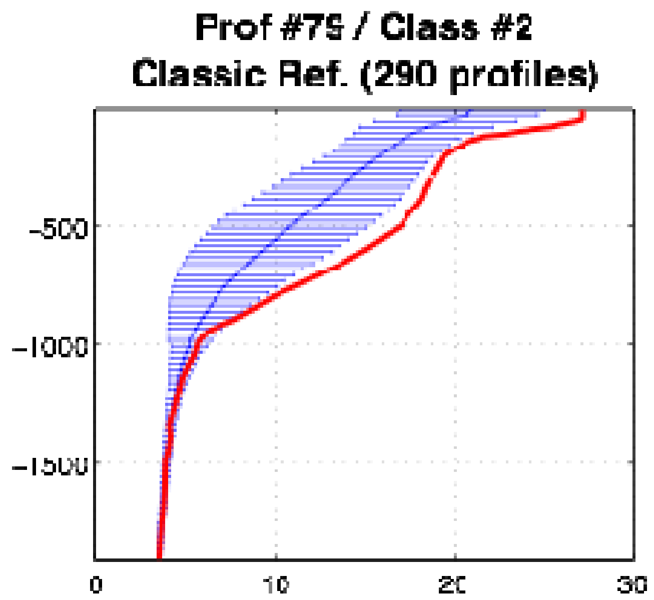


Improved referencing for in-situ data validation

Classic Dist. weighted ref. mean/std   Lx=300km;Ly=Lx/2



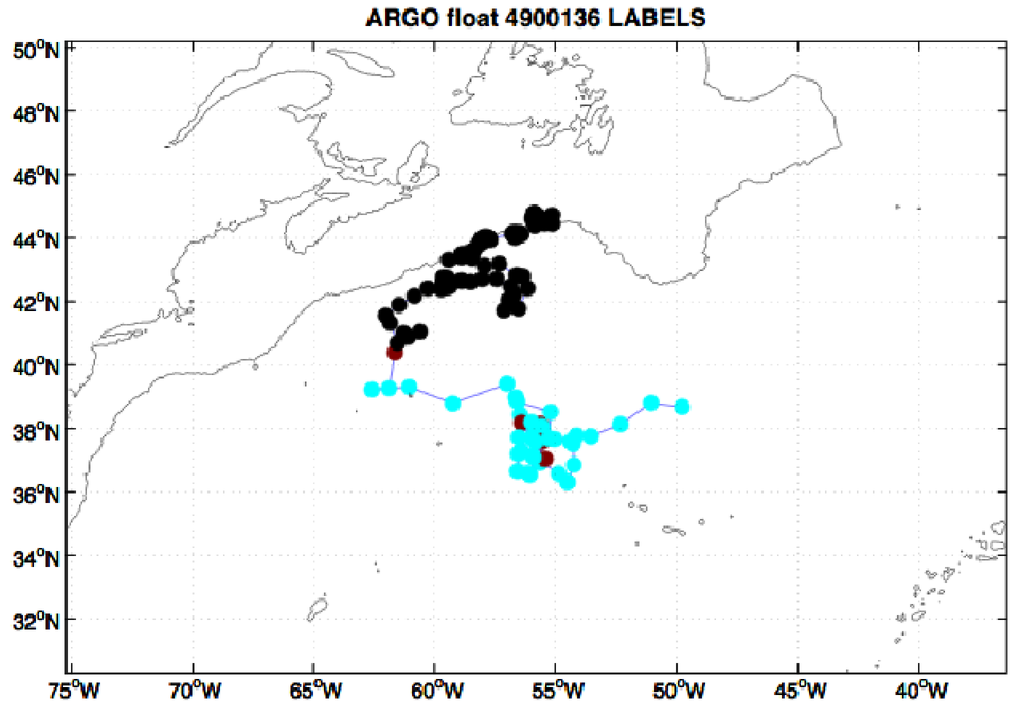




Argo Ref. and CTD database

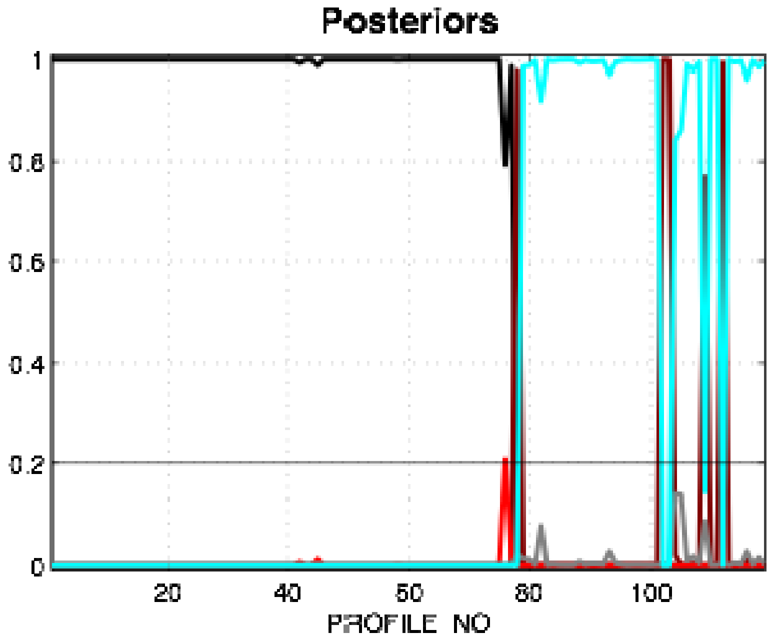
Profiles selected in the pre-processing of the modified OW method (Cabanes et al, 2016)

Improved referencing for in-situ data validation

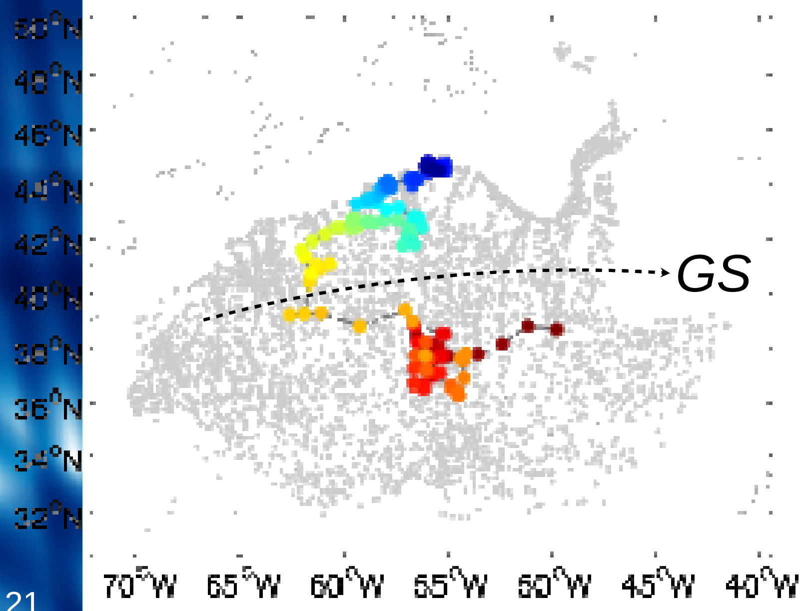


We train a PCM on the reference database and classify the Argo float profiles to validate

The classification is able to determine where, with regard to the front core, the profile is located



Improved referencing for in-situ data validation

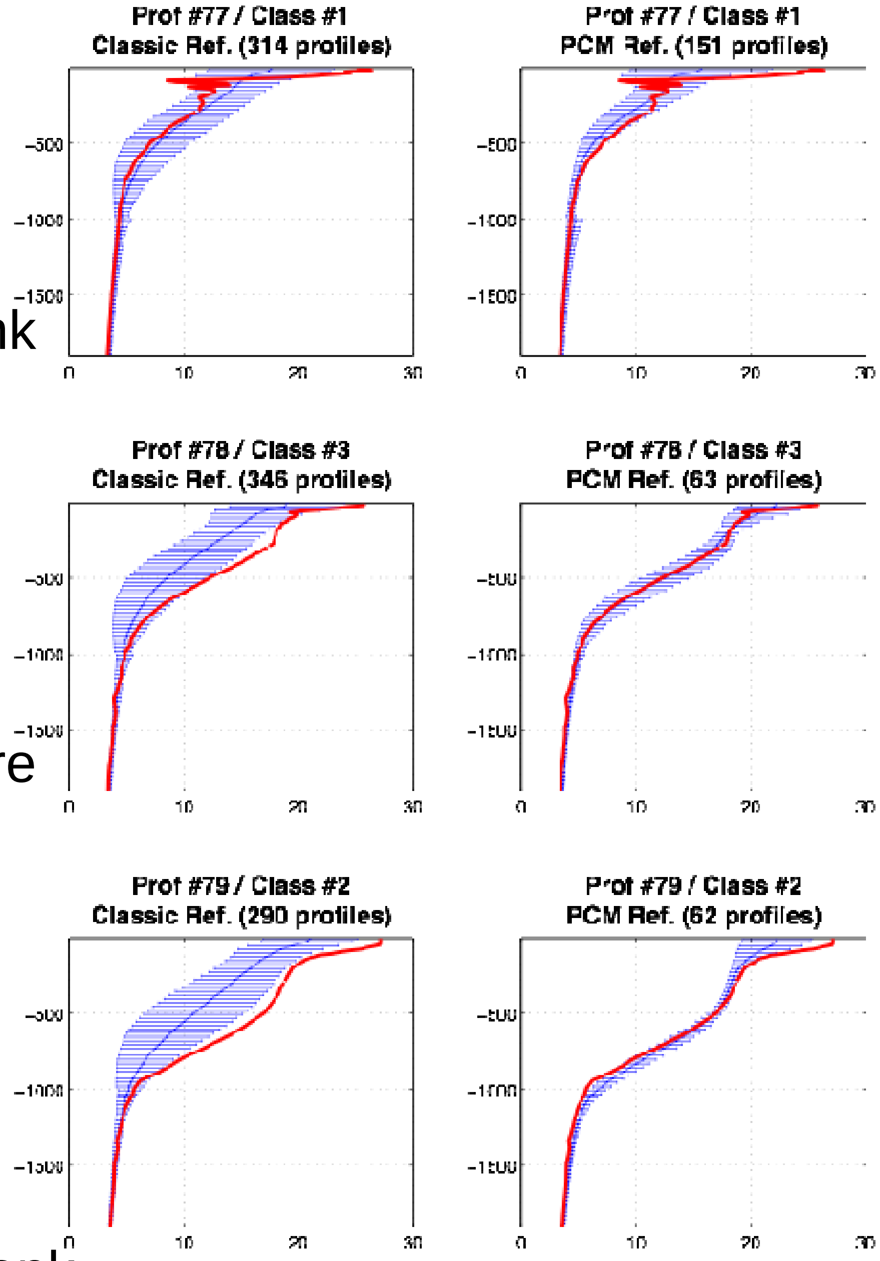


STANDARD | PCM

Cold flank

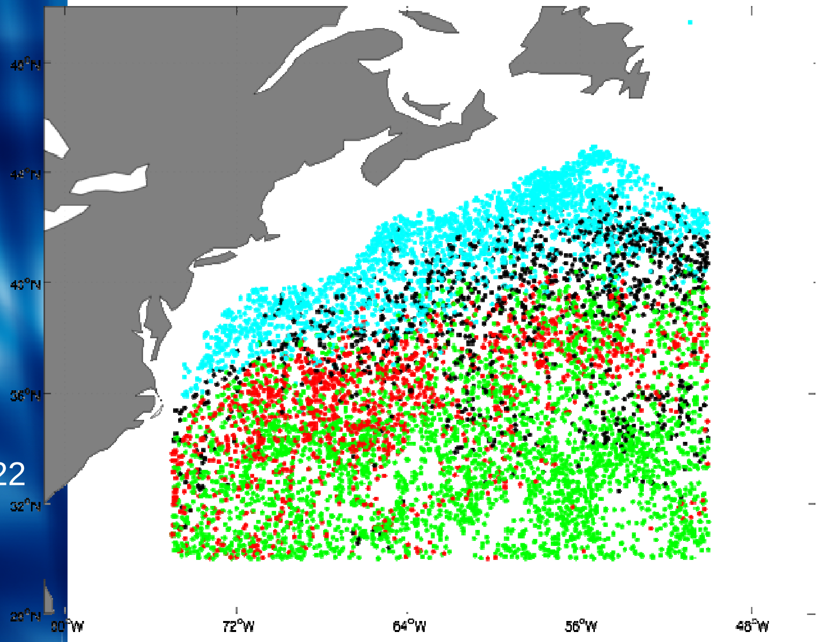
GS core

Warm flank

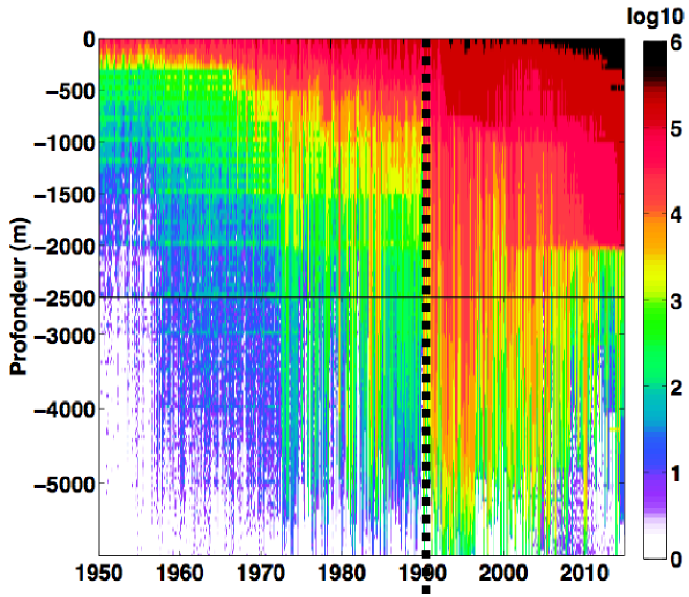


Reference stats are more appropriate:
quality control is improved

A new perspective on WBC variability



Evolution depuis 1950 du nombre de mesures océaniques acquises à différentes profondeurs, par tranche de 50m (source: base de données in situ CORA4.1)



start of altimetry for WBC monitoring

?

The PCM method could improve the WBC long term diagnostic because (i) it does not rely on altimetry and (ii) does not “blur” signals with mapping

Application to the CORA datasets in progress …

Suite:

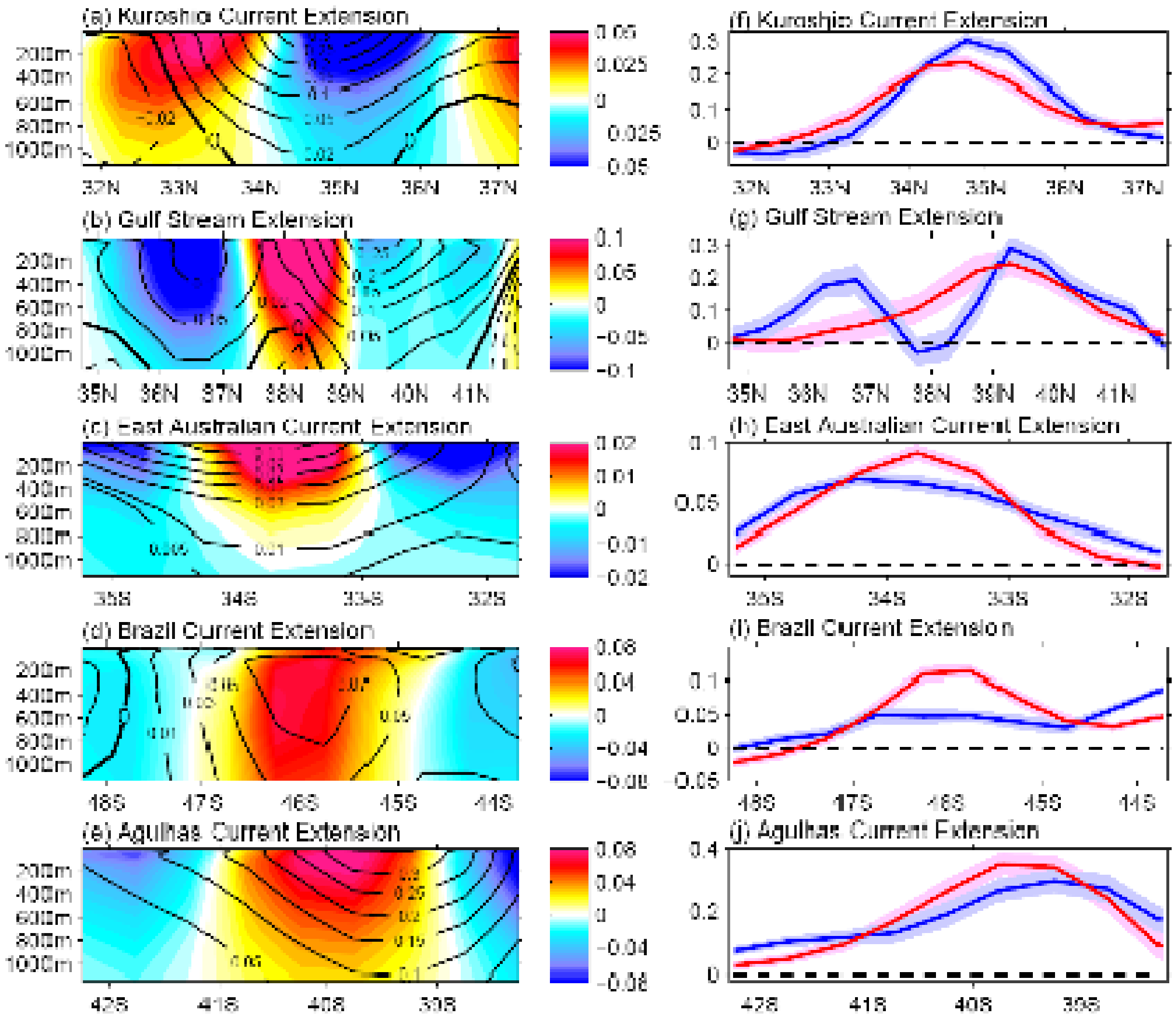
- global, variabilité, WBC, QC
- Projets ANR SONIFIC & INSU/LEFE SOMOVAR avec TB
- 2017 “2nd International workshop on Data Science & Environment: Focus on Oceanography”

A new perspective on WBC variability

Indications for poleward shift and/or intensification of WBC

WBC warming faster than the rest of the ocean

"However, ***uncertainties in detection and attribution of these warming trends remain****, pointing to a need for a long-term monitoring network of the global western boundary currents and their extensions."*



[U] m/s
contours: mean
color: trend

1950-2008

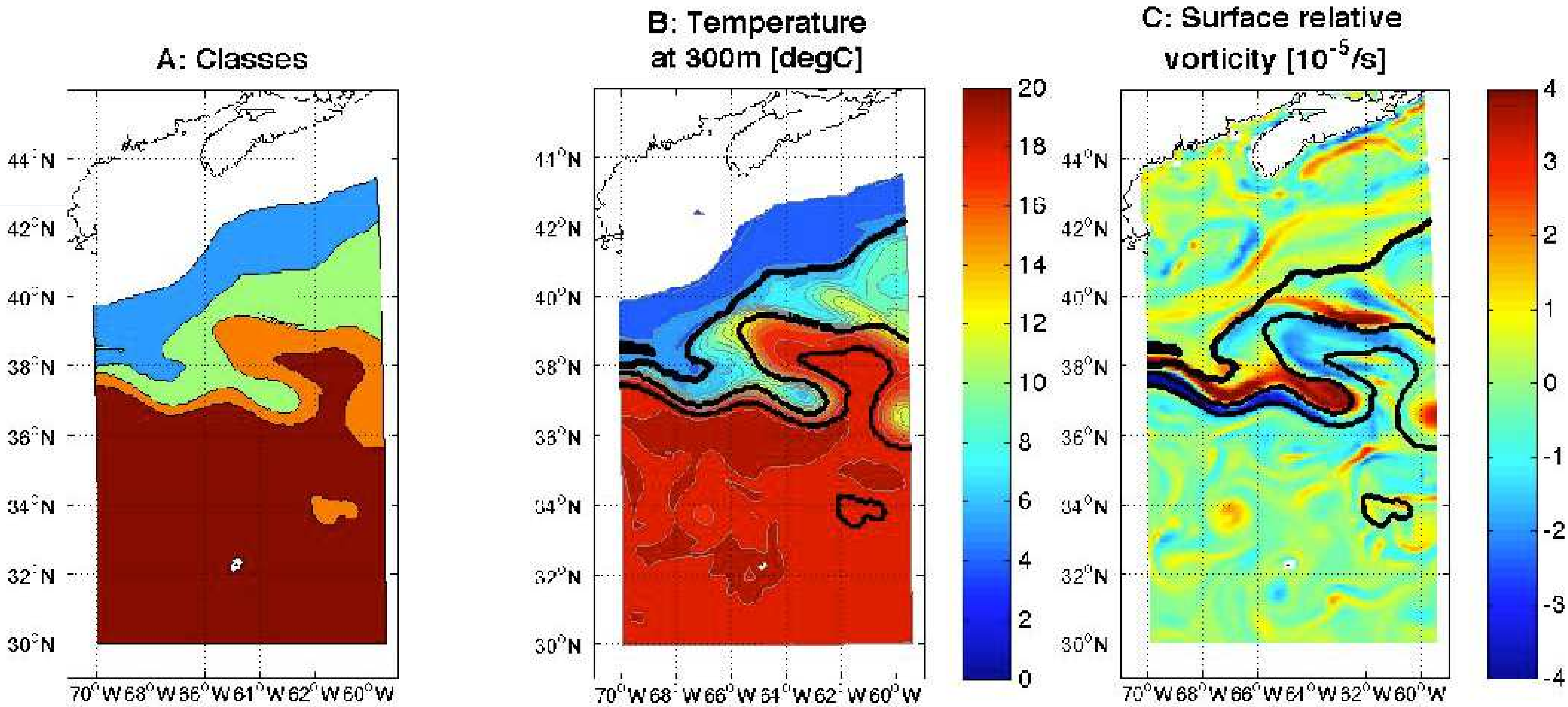
SODA
Wu et al, Nat. 2012

[U-0/500] m/s
red: mean
blue: trend

Can a PCM help ?

A new perspective on WBC variability

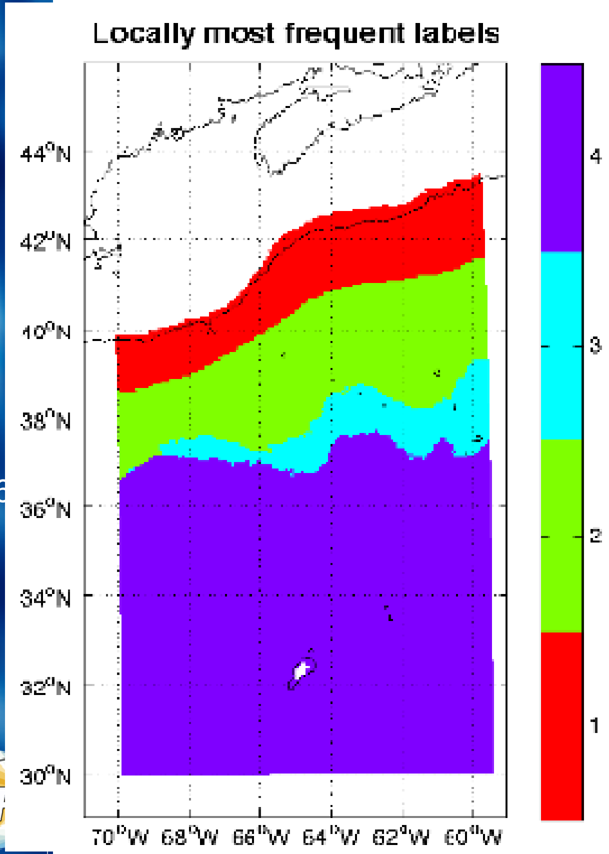
1/12 DRAKKAR model simulation, 3 years of 5-days mean

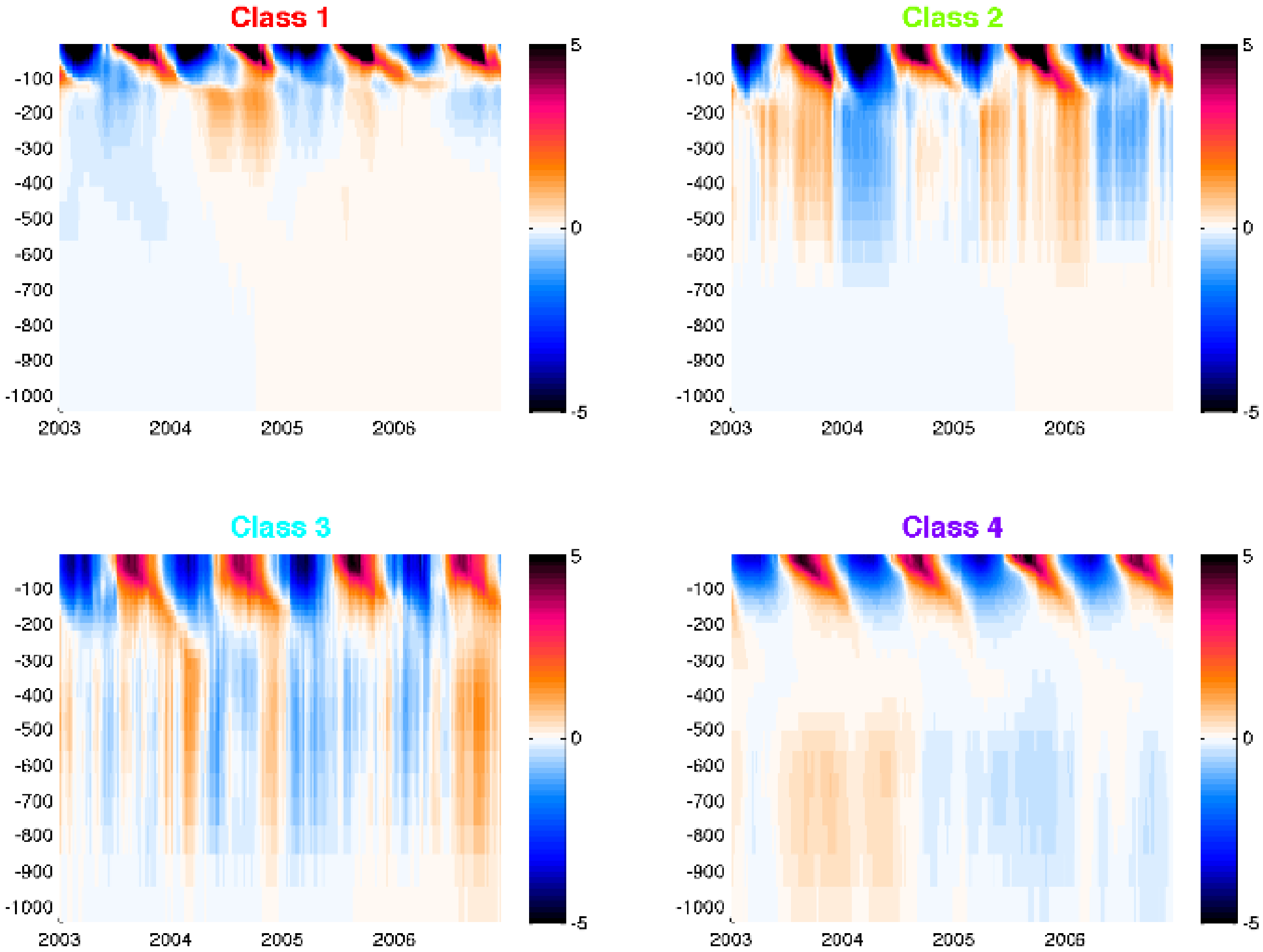


We can derive time series of profiles for each class

A new perspective on WBC variability

1/12 DRAKKAR model simulation, 3 years of 5-days mean





We can derive time series of profiles for each class