

Rapport de stage de M1
DAMTP, Université de Cambridge

Vadim LIMOUSIN – ENS PARIS

Dirigé par : Pr. Colm-cille P. CAULFIELD

Supervisé par : Dr. Miles COUCHMAN

Tuteur à l'ENS : Emmanuel DORMY

Février - Juin 2022

Table des matières

1	Organisation et déroulement du stage à l’université de Cambridge	2
1.1	Organisation du stage	2
1.1.1	Venir à Cambridge en période de Brexit	2
1.1.2	Se loger à Cambridge	2
1.2	Déroulement du stage	2
1.2.1	Ma place au sein de l’équipe	2
1.2.2	Activité scientifique du laboratoire	3
1.2.3	Animation scientifique au laboratoire	3
1.2.4	Vie sociale à Cambridge	4
2	Contenu scientifique du stage	5
2.1	L’importance du <i>turbulent mixing</i> ou mélange turbulent	5
2.2	Γ , l’efficacité de mélange	6
2.3	L’analyse de données en turbulence stratifiée	9
	Références	12

1 Organisation et déroulement du stage à l'université de Cambridge

1.1 Organisation du stage

1.1.1 Venir à Cambridge en période de Brexit

Mon stage s'est déroulé du 14 Février au 17 Juin 2022 à Cambridge. Il fallait, dans un premier temps, constituer un dossier pour obtenir l'autorisation de séjourner au Royaume-Uni durant cette période. Il n'est pas nécessaire de disposer d'un visa pour réaliser un projet de recherche de moins de 6 mois qui n'est pas rémunéré, s'il est spécifiquement appelé "projet de recherche" et non "stage" dans les documents. L'Université de Cambridge ne peut pas signer de "convention de stage", une "convention de projet de recherche" est nécessaire. Il a donc suffi dans mon cas de réunir les documents suivants (vérifier ici au cas où) :

1. Un passeport en cours de validité.
2. Une lettre avec en-tête écrite par l'université de Cambridge attestant que l'université m'invite à venir sur le territoire britannique.
3. Une convention de projet de recherche.

1.1.2 Se loger à Cambridge

La ville de Cambridge est organisée autour de ses 31 colleges, des établissements privés qui offrent aux élèves logements et bibliothèques (moyennant une inscription bien sûr). Essayer d'être logé dans un *colleges*, ce qui peut se faire à l'aide des contacts du tuteur, est pour moi la meilleure option. D'une part, la vie étudiante est grandement facilitée une fois qu'on trouve des contacts dans un *colleges*. D'autre part, c'est l'occasion de potentiellement être logé dans des bâtiments historiques magnifiques, le plus vieux *colleges* datant de 1284.

N'ayant, pour ma part, pas trouvé de logement dans un *colleges*, j'habitais dans une collocation pour un loyer de 700£/mois. D'après mes connaissances, le loyer dans un *colleges* est similaire.

1.2 Déroulement du stage

1.2.1 Ma place au sein de l'équipe

Durant mon stage, j'ai été accueilli au *Department of Applied Mathematics and Theoretical Physics* (DAMTP) du *Center for Mathematical Sciences* (CMS). J'ai

fait partie de l'équipe du Pr. Colm-cille P. Caulfield. Ce dernier étant le directeur du DAMTP, je ne l'ai rencontré qu'à cinq occasions pour discuter de mes résultats. Ces rencontres m'ont permis de progresser rapidement dans mon projet. Son post-doc le Dr. Miles Couchman m'a encadré.

Bien que je n'étais pas "stagiaire", toute l'équipe me considérait comme tel. J'étais très autonome et j'avais le champ libre, tant sur l'organisation de mon temps de travail que sur son contenu, ce qui correspond parfaitement à mon mode de travail.

1.2.2 Activité scientifique du laboratoire

Le CMS est organisé en huit pavillons. J'ai travaillé avec l'équipe du pavillon H. Les domaines de recherche étudiés gravitent autour de la mécanique des fluides, une partie travaille dans la limite des milieux visqueux (leurs travaux s'appliquant par exemple à la mécanique des fluides biologiques et à l'étude de la dynamique des organismes nageant dans des milieux visqueux), une autre étudie les fluides visco-élastiques et une dernière, dont j'ai fait partie, travaille sur la turbulence stratifiée (dont les applications sont souvent géophysiques). Le pavillon est divisé en deux grands groupes de chercheurs : l'un réfléchit principalement à travers des simulations numériques et l'autre mène un travail centré autour de l'expérimentation. Ces deux groupes se réunissent quotidiennement autour d'un thé pour discuter et sociabiliser, et à l'occasion d'un "lab lunch" tous les vendredis, pour présenter de manière informelle leurs avancées au reste du pavillon. J'ai beaucoup aimé cette dynamique qui aide à la fois les numériciens à "garder les pieds sur Terre" et les expérimentalistes à vouloir vérifier les nouveaux résultats numériques. Le pavillon dans son entièreté se réunit de manière hebdomadaire pour le séminaire de mécanique des fluides.

1.2.3 Animation scientifique au laboratoire

A tous ces rendez-vous scientifiques, s'ajoutent une grande quantité de conférences et petits cours quotidiens proposés au DAMTP ainsi qu'au *Isaac Newton Institute*. Occasionnellement, une journée était dédiée aux étudiants en thèse pour qu'ils présentent leurs avancées. J'ai donc pu assister à de nombreuses conférences, que j'ai pour la plupart beaucoup appréciées, dont les sujets varient de l'explication de la forme des nuages à l'étude des tâches à la surface du soleil.

Il est aussi possible de suivre des conférences en dehors du CMS. J'ai par exemple assisté à un exposé présenté par un ingénieur de l'équipe de Nouvelle-Zélande de l'*America's Cup* (compétition internationale de voile) à propos des navires que cette équipe construit.

1.2.4 Vie sociale à Cambridge

J'ai particulièrement aimé la ville de Cambridge. C'est une ville plutôt petite organisée autour des *colleges* dont les bâtiments sont magnifiques et des étudiants de l'université. Cela présente l'avantage de faciliter la vie étudiante mais l'inconvénient de laisser la ville inactive quand les étudiants sont en vacances (en avril et à partir de juillet). Toutes les semaines, se tiennent des *formal dinners*, dîners formels comme son nom l'indique, durant lesquels les élèves s'habillent pour l'occasion (l'uniforme est le *gown*) et mangent dans les vieux bâtiments des *colleges* (voir figure 1).



FIGURE 1 – Photo au formal de Downing College avec des amis du CMS et du laboratoire Cavendish (je suis tout à droite en *gown*).

2 Contenu scientifique du stage

Durant ces quatre mois passés à Cambridge, j'ai travaillé sur un sujet proche de celui étudié par Dr. Miles Couchman : l'utilisation d'algorithmes non supervisés pour guider l'intuition du chercheur en mécanique des fluides. Nous sommes actuellement en train d'écrire un article que nous souhaitons publier dans le *Journal of Fluid Mechanics*. Celui-ci peut être retrouvé à la fin de ce rapport dans une version préliminaire. Avant cela, je souhaite présenter le cadre d'étude de l'article dans une introduction inspirée du livre [1] et de l'article [2].

2.1 L'importance du *turbulent mixing* ou mélange turbulent

Comprendre le fonctionnement des océans et en faire un modèle précis est crucial dans l'appréhension de la crise climatique. En effet, les océans captent une grande partie de la chaleur fournie par le Soleil.

On sait qu'à l'équateur, l'eau en surface emmagasine la chaleur fournie par le rayonnement du Soleil puis est transportée jusqu'aux pôles où elle refroidit et, devenant ainsi plus dense, plonge au fond de l'océan. Elle revient alors à son point de départ suivant une trajectoire comparable à celle d'un tapis roulant, comme schématisé dans la figure 2.

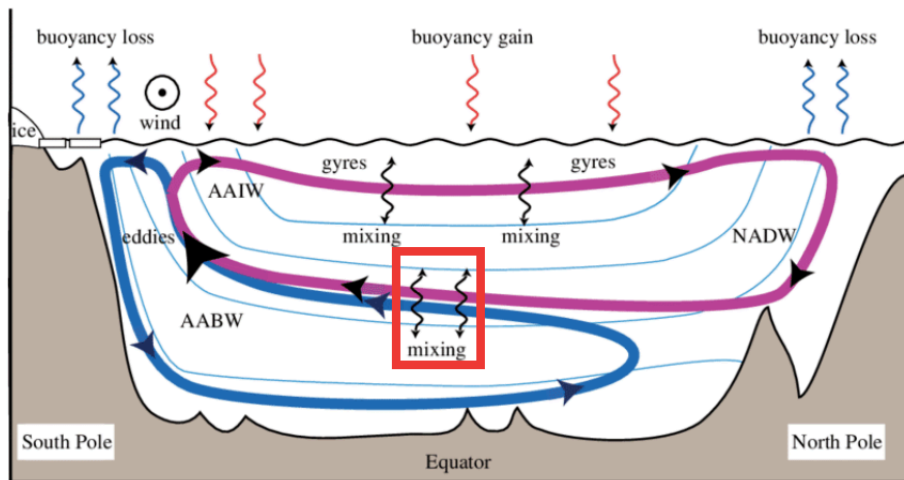


FIGURE 2 – Courants terrestres transportant l'eau chauffée en surface vers les pôles où elle refroidit avant de revenir à son point de départ (figure tirée de [3, 4]).

La manière dont l'eau froide située en profondeur remonte pour être chauffée à nouveau est encore mal comprise. Un premier mécanisme expliquant cela repose sur

l'implication du vent. Le second, qui nous intéresse ici, est la diffusion de la chaleur accélérée par la turbulence (*mixing* sur la figure 2). En effet, sans turbulence, seule la diffusion moléculaire permettrait à un écoulement d'eau chaude de communiquer sa chaleur à un écoulement d'eau froide, mais la turbulence accélère grandement ce processus par sa capacité à mélanger ces deux fluides.

Le défi est alors le suivant : pour créer un modèle des océans, la plus petite échelle à résoudre est centimétrique, c'est celle de la turbulence, les sources de turbulence étant nombreuses. Une synthèse des sources connues est représentée en figure 3. Malheureusement, les meilleurs modèles simulant l'entière des océans résolvent des zones de tailles kilométriques seulement. Il est donc nécessaire pour progresser sans attendre de plus grands moyens de calcul de comprendre la diffusion turbulente aux petites échelles des océans.

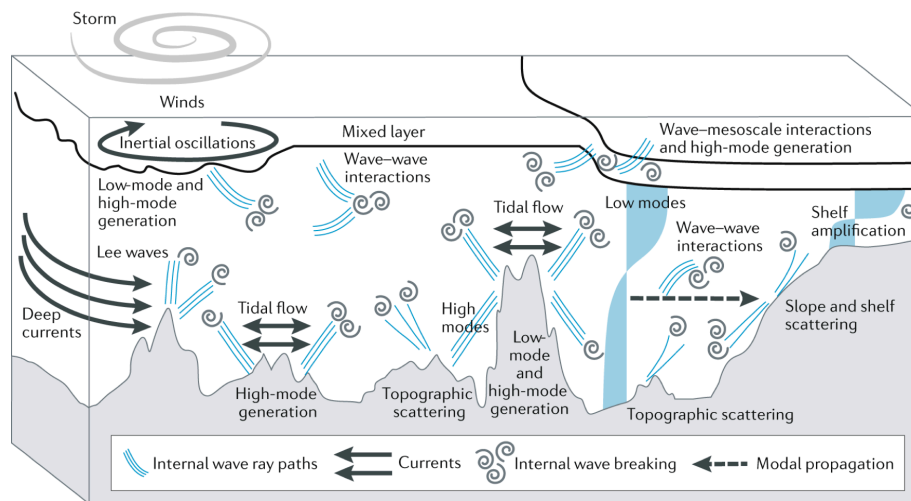


FIGURE 3 – Multiples sources connues de turbulence dans l'océan (figure tirée de [3, 5]).

2.2 Γ , l'efficacité de mélange

Il est important de garder en tête que l'océan est stratifié en densité (ainsi qu'en salinité mais nous négligeons ce phénomène ici), *ie* la densité de l'eau augmente avec la profondeur et ne dépend en moyenne que de celle-ci. Pour expliquer la notion de mélange turbulent, un raisonnement phénoménologique s'impose. Comme souvent pour décrire la dynamique de l'océan, utilisons la décomposition classique de Reynolds de la vitesse $u = \langle u \rangle + u' = U + u'$ et de la densité $\rho = \langle \rho \rangle + \rho' = \rho_0 + \rho'$,

séparant ainsi l'écoulement en une partie moyenne (au sens d'une moyenne qui serait faite sur de nombreuses répétitions de ce même écoulement) et une partie aléatoire, turbulente. Ces deux grandeurs respectent l'équation suivante sur l'énergie cinétique turbulente d'un fluide stratifié dans l'approximation de Boussinesq (elle repose sur une condition de faibles variations de la densité et de la température) :

$$\underbrace{\frac{\partial}{\partial t} \left(\frac{1}{2} \langle u'_i u'_j \rangle \right)}_{\mathcal{K}'} = - \underbrace{2\nu \langle s'_{ij} s'_{ij} \rangle}_{\varepsilon} - \underbrace{\frac{g}{\rho_0} \langle \rho' w' \rangle}_{\mathcal{B}} + \underbrace{(-\langle u'_i u'_j \rangle S_{ij})}_{\mathcal{P}} - \nabla \cdot J \quad (1)$$

avec $s'_{ij} = 1/2(\partial u'_i/\partial x_j + \partial u'_j/\partial x_i)$ et $S_{ij} = 1/2(\partial U_i/\partial x_j + \partial U_j/\partial x_i)$, w la composante verticale de u . En d'autres termes, cette équation nous dit que :

- l'énergie cinétique turbulente \mathcal{K}' est maintenue une source \mathcal{P} de turbulence (vues dans la figure 3).
- \mathcal{K}' est communiquée au reste du volume par le terme de transport $\nabla \cdot J$ et par l'effet de la stratification \mathcal{B} .
- \mathcal{K}' est dissipée par l'effet de la viscosité représentée dans le terme ε .

Une équation similaire est donnée pour la variance de la quantité scalaire ρ :

$$\frac{g^2}{\rho_0^2 N^2} \frac{\partial}{\partial t} \left(\frac{1}{2} \langle \rho'^2 \rangle \right) = - \underbrace{\frac{g}{\rho_0} \langle \rho' w' \rangle}_{\mathcal{B}} + \underbrace{\frac{g^2}{\rho_0^2 N^2} \kappa \left\langle \frac{\partial \rho'}{\partial x_i} \frac{\partial \rho'}{\partial x_i} \right\rangle}_{\chi} - \nabla \cdot J_\rho \quad (2)$$

où κ est la diffusivité thermique et N est la fréquence de Brunt-Väisälä. C'est la fréquence à laquelle une particule de fluide dans un milieu stratifié qui serait déplacée, mettons vers le haut, oscille autour de sa position d'origine du fait de la stratification. Elle serait en effet alternativement poussée vers le bas par le liquide situé précédemment au-dessus d'elle et qui est moins dense, puis vers le haut par le liquide situé en-dessous d'elle et qui est plus dense. La figure 4 présente schématiquement ce mouvement.

Dans ces deux équations, $\nabla \cdot J$ et $\nabla \cdot J_\rho$ sont des termes de transport. Le lecteur intéressé peut trouver leurs expressions complètes dans l'introduction de [2].

Ainsi, si les effets de transport $\nabla \cdot J$ et $\nabla \cdot J_\rho$ sont négligeables (par exemple dans un domaine périodique ou fermé) et si le régime est stationnaire, on obtient l'équation :

$$0 = -\varepsilon - \mathcal{B} - \mathcal{P} \quad (1')$$

$$0 = \mathcal{B} - \chi \quad (2')$$

ce qui donne :

$$\mathcal{P} \longrightarrow \varepsilon + \chi$$

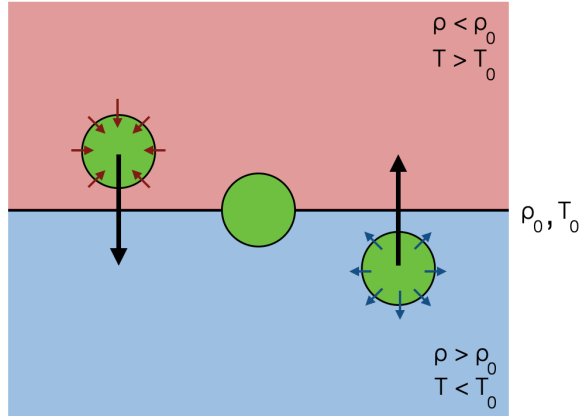


FIGURE 4 – En verts, trois particules de fluide dans un milieu stratifié. Celle de gauche a été déplacée vers le haut, elle est poussée vers le bas par le liquide qui l’entoure. Celle de droite présente le cas symétrique. Le particule au milieu montre la position d’origine. Enfin, considérons le liquide rouge plus chaud et le bleu plus froid que la particule initialement. Des transferts de chaleur irréversibles s’opèrent alors pendant l’oscillation, c’est le mélange turbulent.

Une quantité \mathcal{P} d’énergie turbulente introduite peut alors être dissipée de manière irréversible par seulement deux biais :

- par l’effet de la viscosité, via ε dans 1’ ;
- elle peut être transmise via \mathcal{B} qui intervient dans 1’ et 2’ puis dissipée par diffusion thermique, via χ .

De manière imagée, le terme \mathcal{B} correspond à une énergie potentielle de *buoyancy*, ie de flottabilité, qui est liée à la stratification. Comme expliqué plus haut, une particule de fluide poussée vers le haut par un évènement turbulent (voir figure 5) sera repoussée vers le bas par le fluide moins dense qui l’entoure alors. Si cette particule de fluide est initialement froide, elle capte de l’énergie du fluide moins dense (et plus chaud) qui l’entoure alors avant d’être repoussée vers sa position d’origine. Ce phénomène est irréversible : c’est le mélange turbulent.

Une manière naturelle de quantifier l’efficacité de ce mélange est de regarder le rapport de l’énergie fournie via \mathcal{P} donnant effectivement lieu à une diffusion de la chaleur contre la part d’énergie dissipée par viscosité. C’est l’efficacité de mélange Γ :

$$\Gamma = \frac{\chi}{\varepsilon}$$

Il est souvent considéré comme constant égal à 0.2 pour simplifier les simulations mais

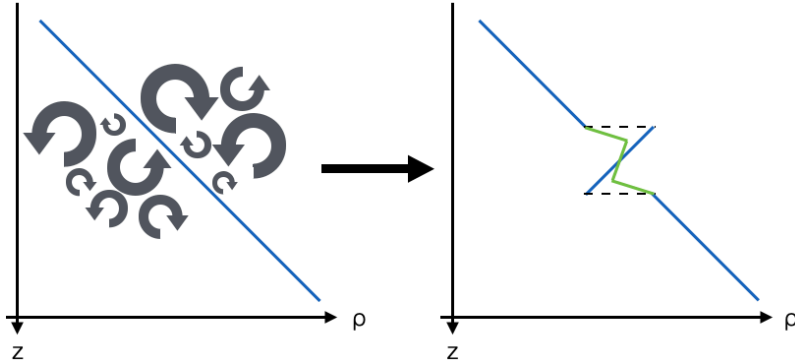


FIGURE 5 – Perturbation de la stratification par l'effet d'un évènement turbulent. En bleu, la courbe de densité ρ en fonction de la profondeur z . En vert, la stratification qui se rétablit diminuant ainsi \mathcal{B} (figure inspirée de [3]).

ceci n'est que sa valeur moyenne dans l'océan. La conclusion de la revue annuelle [2] est claire sur ce point : "it still remains extremely difficult to say anything generic about mixing".

2.3 L'analyse de données en turbulence stratifiée

Afin de faire des progrès dans la compréhension du comportement de cette efficacité de mélange, une piste de recherche repose sur l'utilisation d'analyse de données expérimentales ou numériques d'écoulements. Cette idée vise à déterminer différentes zones d'écoulements dynamiquement identifiables dans chaque situation. Cela a été proposé dans les articles [6] (sur des données numériques avec une identification supervisée) et [7] (sur des données expérimentales avec une identification non supervisée).

La problématique de mon projet de recherche s'inscrit dans cette lignée de pensée. Mon objectif était de trouver des indices permettant d'analyser une très grande simulation numérique ($2048 \times 2048 \times 1024$ points représentant $10 \times 10 \times 5 m^3$ de fluide [8]) d'une boîte périodiquement fermée de fluide stratifié et ce en utilisant une méthode de recherche de clusters. Cette idée repose sur un apprentissage non supervisée, caractéristique intéressante dans l'objectif de fournir des résultats non biaisés au mécanicien des fluides et de guider son intuition, souvent malmenée par le comportement imprévisible de la turbulence.

Pour connaître les détails de ce travail je renvoie le lecteur vers l'article en préparation ci-joint (la version finale soumise au journal sera transmise ultérieurement).

Dans cet article, je présente comment une méthode basée sur cette idée d'étude non supervisée d'une simulation numérique permet d'identifier un mécanisme de formation de *layers*, zones de l'écoulement dans lesquelles le cisaillement est fort et le mélange est donc important.

Remerciements

Je tiens particulièrement à remercier le Pr. Colm-cille Caulfield de m'avoir accueilli au DAMTP malgré les embûches administratives liées au Brexit. Merci de m'avoir donné l'opportunité de découvrir la recherche à Cambridge et merci pour les conversations scientifiques toujours enrichissantes.

Je souhaite bien sûr remercier le Dr. Miles Couchman de m'avoir encadré et aidé à travers ce projet de recherche et de m'avoir fait découvrir son sujet d'étude passionnant. Merci pour ton soutien au cours de ces nombreux *big days*.

Un grand merci à Alison, l'assistante du Pr. Caulfield, qui m'a beaucoup aidé à remplir tous les documents administratifs et chaleureusement accueilli avec une visite guidée du CMS qui s'est avérée très utile.

Merci à Emmanuel Dormy, mon tuteur à l'ENS, qui m'a donné le contact du Pr. Caulfield et qui a enquêté avec moi sur la possibilité de mener ce projet au Royaume-Uni.

Je souhaite enfin remercier tout le laboratoire du pavillon H qui a été très accueillant pendant ces quatre mois et parmi lequel j'ai trouvé des amis que j'espère retrouver à Cambridge dès que possible.

Références

- [1] P. A. DAVIDSON. « Turbulence in Rotating, Stratified and Electrically Conducting Fluids ». In : 2^e éd. 2013. Chap. 1, 2, 4.
- [2] C. P. CAULFIELD. « Layering, Instabilities, and Mixing in Turbulent Stratified Flows ». In : *Annual Reviews* (2021).
- [3] Miles M. P. COUCHMAN. *Data-driven identification of turbulent mixing in stratified flows*. Geophysical and Environmental Processes Seminar (February 2022).
- [4] Geoffrey K. VALLIS. « Geophysical fluid dynamics : whence, whither and why ? » In : *Proceed of the Royal Society A* (2016).
- [5] Jennifer A. MACKINNON et al. « Climate Process Team on Internal Wave–Driven Ocean Mixing ». In : *Bulletin of the American Meteorological Society* (2017).
- [6] Gavin D. PORTWOOD et al. « Robust Identification of Dynamically Distinct Regions in Stratified Turbulence ». In : *Journal of Fluid Mechanics* (2016).
- [7] Miles M. P. COUCHMAN et al. « Data-Driven Identification of Turbulent Oceanic Mixing From Observational Microstructure Data ». In : *Geophysical Research Letter* (2021).
- [8] Saba ALMALKIE et Stephen M. de BRUYN KOPS. « Kinetic Energy Dynamics in Forced, Homogeneous, and Axisymmetric Stably Stratified Turbulence ». In : *Journal of Turbulence* (2012).

Banner appropriate to article type will appear here in typeset article

Unsupervised Data-driven Identification of a Layering Mechanism around a Strong Vortex in a Stratified Flow

Vadim Limousin^{1,2}, Miles M. P. Couchman¹ and Colm-cille P. Caulfield^{1,3}

¹Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK,

²Département de Mathématiques et Applications, ENS Paris, 45 rue d'Ulm, 75005, Paris, France,

³BP Institute, University of Cambridge, Cambridge CB3 0EZ, UK

(Received xx; revised xx; accepted xx)

Layering appears in stratified fluid and plays a crucial role in the turbulent mixing of scalar quantities. Presumably, understanding how layers are formed can give insights on how scalars are irreversibly mixed in the ocean. We present a data-driven method that can split a stratified turbulent box in regions of interest using unsupervised clustering guided by the value of the strain $\partial_i u_j$ at each point of the box. We apply it to a snapshot of a direct numerical simulation of a box of the ocean where a strong vortex lies, thus helping us identify automatically the layered structures surrounding the swirl. This method also shows a mechanism by which those layers are formed. Interestingly, this method yields clusters that have different mixing properties even though ε and χ are not known by the algorithm.

1. Introduction

Understanding how a scalar quantity (e.g. temperature, salt concentration) is spread in a stratified fluid is a major point of interest for the geophysical community. It is indeed crucial to getting how heat, captured by the ocean's surface, is then transported around the globe's oceans. It is known that turbulence plays an leading role in this process as a means to transfer energy from large scales to smaller scales where irreversible scalar diffusion occur. It has also been shown that in the framework of oceans and stratified turbulence, layering is an important feature (Caulfield 2021). Within a box of the ocean, the turbulent flow is highly variable, it is not sufficient to just consider a box average of the mixing that would blur out the layered structures. In this article, mixing is defined as the irreversible diffusion of a scalar quantity, measured here by χ , and the mixing rate Γ is the ratio of this mixing against the kinetic energy dissipation rate ε .

Here we analyse a direct numerical simulation (DNS) of forced stratified turbulence in which lies a strong vortex injecting energy into the flow at large scales. This energy is eventually dissipated by diffusion and viscosity at small scales. Work has already been done on similar simulations to test our physical intuition and try identifying known behaviours (Portwood 2016; John Taylor 2019; Couchman 2022), but all these studies were manually operated. The objective here is to use an automated method for detecting a mechanism responsible for layer formation, which is relevant to understanding mixing. In this article, we

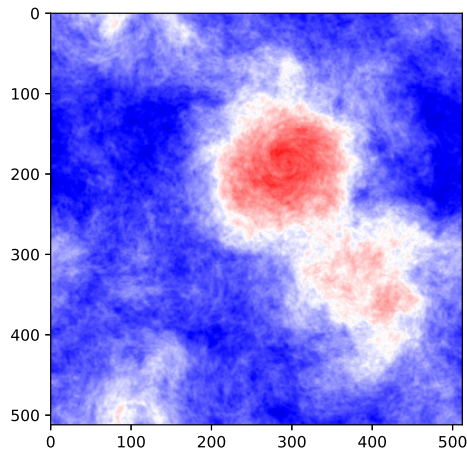


Figure 1: Map of the vertical mean of $\log(\varepsilon)$ used as a first map of the simulation. The red clouds are the most energetic ones and the blue area is the most quiescent one. A strong vortex lies in the domains and seems to only interact with itself through the periodic boundaries of the box.

38 show how clustering can identify such a process by which a strong vortex in a stratified flow
 39 gives way to a layered set-up throughout a domain.

40 2. Data and Methods

41 The data used here is a snapshot of a DNS of forced turbulence in a periodic box of
 42 strongly stratified fluid (with $Pe = 1$ and $Fr = 200$). The simulated grid originally has
 43 $2048 \times 2048 \times 1024$ points but we sparsed it down to $512 \times 512 \times 256$ points for memory
 44 management reasons and to keep computation times realistic and represents a $10 \times 10 \times 5 m^3$
 45 periodic box of the ocean. A vertically average map of $\log(\varepsilon)$ used as a view from above
 46 of the simulation is shown in figure 1. The red clouds are the most energetic ones and the
 47 blue area is the most quiescent one. It allows us to have a first look at these data, hinting a
 48 strong vortex (around the point of coordinate (300, 200)). This DNS yields the value of each
 49 component of the strain rate tensor each point of the volume, as well as the values of ε the
 50 dissipation rate of the kinetic energy and χ the dissipation rate due to heat diffusion. Going
 51 through these data by hand already hints some structures lying in the volume, encouraging
 52 the use of a clustering algorithm to automatically identify coherent areas in this flow. This
 53 idea has already proven to be insightful as presented in Couchman (2021).

54 Let us define the notion of cluster : for a set \mathcal{D} of points lying in a N -dimensional space
 55 \mathcal{S} , a cluster is a subset of \mathcal{D} characterised by a point in \mathcal{S} known as the cluster centre. Every
 56 point of this cluster is closer to its corresponding cluster centre than to the any other centres
 57 in \mathcal{S} . Clusters are also characterised by a particularly high point density around its centre.
 58 An example could be galaxies in the universe thought as clusters of stars. In our context,
 59 identifying a cluster in this dataset would gives two major information :

60 (i) each point of the cluster have similarities in their coordinates. In our case these
 61 coordinates hint the motion of the fluid at the corresponding gridpoint. Thus clusters should
 62 be coherent patches moving in harmony.

63 (ii) each cluster found can possibly represent multiple not connected regions in the real

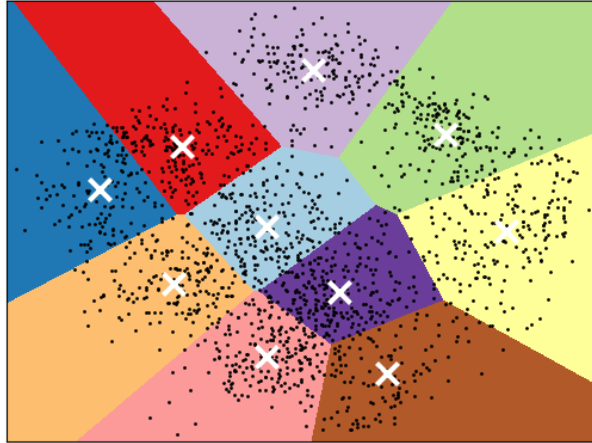


Figure 2: Example of how KMeans defines its estimated clusters. White crosses show the estimated cluster centres and the coloured region the estimated clusters surrounding them.
 . [FIGURE TO EITHER CHANGE OR ASK FOR PERMISSION TO USE]

64 flow. Indeed, not taking the spatial coordinates in count allows two points lying far away in
 65 the volume but behaving similarly to belong to the same cluster.

66 Here, the size of the simulation ($512 \times 512 \times 256$ points in a 7-dimensional space) rules
 67 out clustering algorithms that are too greedy. This is why we opted for the quickest clustering
 68 algorithm available : KMeans. KMeans works by placing n (n chosen by the user) points
 69 μ_1, \dots, μ_n in the space where the database lies and letting them converge as near as possible
 70 to the cluster centres. It defines the estimated clusters C_1, \dots, C_n as the Voronoi diagram
 71 generated by the μ_i (ie C_j is the set of points closer to μ_j than to any other μ_i) as shown in 2.
 72 To resolve the clusers, KMeans works iteratively, it alternatively updates μ_i as the mean point
 73 of C_i for each i and then defines the new C_i 's using the updated values of the corresponding
 74 μ_i 's,. This goes on for a chosen amount of iteration.

75 3. The Algotihm

76 As shown in figure 1, the simulation holds a strong vortex sustained by the forcing injecting
 77 energy at large scales. The rest of the domain seems quiescent at first sight but the clustering
 78 will show the layered set-up of this region.

79 We naturally chose to work with the value of the strain rate tensor $\partial_i u_j$ yielded by the
 80 DNS at each point of the simulated grid as the database to cluster. These coordinates bear
 81 information that are generally used as proxies to understand the behaviour of ε and χ . We
 82 intentionally don't include ε or χ as coordinates as we want our method to only compute on
 83 the strain values. Here, u , v and w are the components of the velocity along the x , y and z
 84 axis.

85 As a first try to know what coordinates are most relevant to represent the database, we ran
 86 principal component analysis on it. This quickly led us to reorganise the strain coordinates
 87 into shear and vorticity components. Let's then keep in mind that this simulation is built as
 88 a periodic box of the ocean. Thus the x and y coordinates are supposedly equivalent and
 89 the z coordinate is distinct *a priori*. Indeed, the z axis is the one parallel to \mathbf{g} , the gravity
 90 acceleration. We eventually decided to modify the coordinates again and chose to work with
 91 the new set of coordinates shown in table 1 at each point of the volume. This way, we have
 92 rotationally symmetric coordinates to cluster on.

$$\begin{array}{l}
\text{Vertical shear : } S_z = \frac{1}{2}(u_z^2 + v_z^2) \\
\text{Horizontal shear : } S_h^{(uv)} = \frac{1}{2}(u_y^2 + v_x^2) \quad S_h^{(w)} = \frac{1}{2}(w_x^2 + w_y^2) \\
\text{Strain : } s^{(uv)} = \frac{1}{2}(u_x^2 + v_y^2) \quad s^{(w)} = w_z^2 \\
\text{Rotation : } \omega_z = (v_x - u_y)^2 \quad \omega_h = (w_x - u_z)^2 + (w_y - v_z)^2
\end{array}$$

Table 1: Rotationally symmetric coordinates used at each point of the simulation.

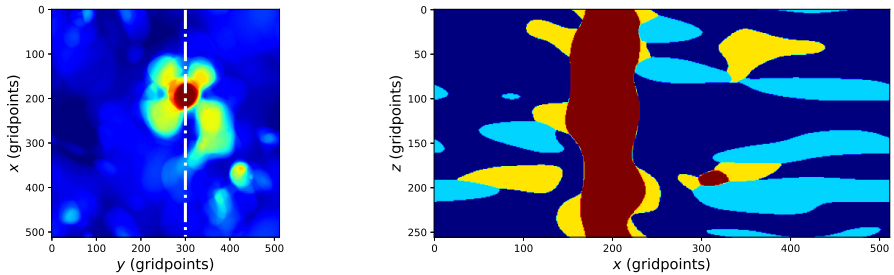


Figure 3: Two views of the result of the method. On the left, a vertical mean of the labels, as a view from above of the clustering. On the right, a vertical slice of the labels going through the swirl ($y = 300$). This shows the layered structure identified by the clustering.

93 The $\frac{1}{2}$ factor on some coordinates is only aesthetic to help our intuition but the clustering
 94 is ran on the normalised coordinates to avoid one dominating the others.

95 It is clear in table 1 that some coordinates have similar algebraic expression such as S_z
 96 and ω_h (the only differences are terms implying w the vertical component of the velocity)
 97 or $S_h^{(uv)}$ and ω_z . In fact, their distributions always proved to be strongly correlated. Besides,
 98 $s^{(w)}$ and $S_h^{(w)}$ were always distributed similarly in the clusters identified by our method. The
 99 fact that terms implying w are less significant is not surprising in the framework of stratified
 100 fluid dynamics as vertical movements are negligible with respect to horizontal ones. These
 101 reasons are why we chose to represent only ω_z , $s^{(uv)}$ and S_z in the following figures. These
 102 coordinates proved to be the most relevant to clustering.

103 4. Results

104 Running KMeans on the database yields a classification of the gridpoints that can be
 105 represented as a colouring of the volume. Here, the 4 clusters are shown as dark blue,
 106 light blue, yellow and red. These labels are shown in figure 3 viewed from above, as a vertical
 107 mean of the labels, and from the side, as a slice at $y=300$. To help us analyse how KMeans
 108 splits the dataset, we chose to represent in figure 4 distribution of the 3 coordinates we
 109 identified as most relevant in the estimated clusters.

110 On the left part of figure 4 we plotted the histograms of each of these coordinates coloured
 111 by cluster, and the total distribution in black. These histograms are normalised so we can
 112 better compare the differences in distribution between clusters ignoring their size. On the
 113 right side of this figure, we showed an other way to visualise the regions picked out by the
 114 algorithm : through the same vertical slice of the simulation as in 3 (this slice is at $y=300$
 115 and goes through the swirl). The contours of the clusters are drawn on these panels to help
 116 seeing how they lie on the hot-spots of S_z , ω_z and $s^{(uv)}$.

117 Among the 4 regions, the dark blue seems to contain all the background points. In every

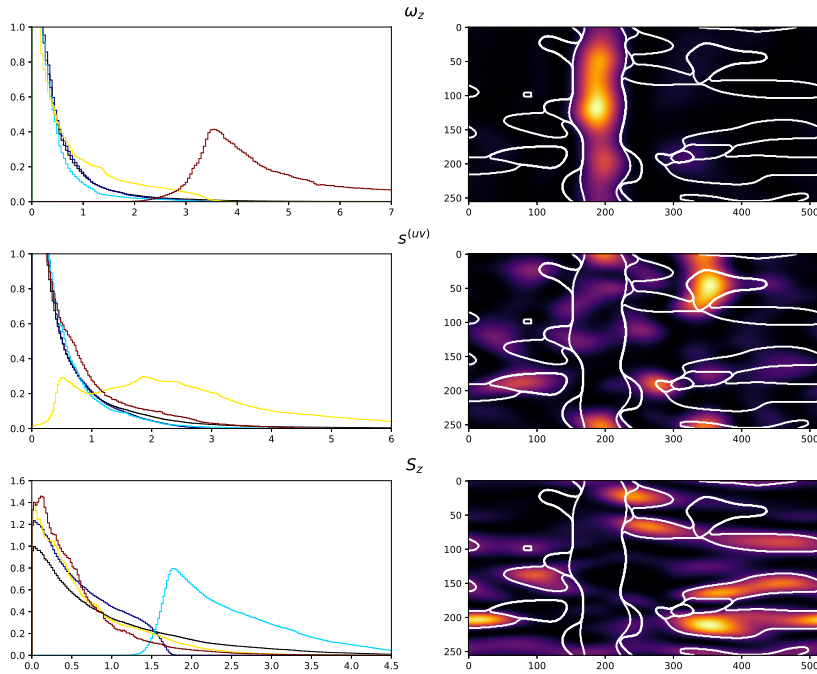


Figure 4: Understanding how each cluster is identified. On the left side, normalised histograms of the three most relevant coordinates. On the right side, the maps showing how the respective coordinates is distributed in a slice of the simulation. Represented in white are the border of each cluster shown in 3. In the histograms, the colour of each curve shows which cluster's distribution is plotted. The black curve shows the total distribution. These histograms are normalised to easily get which cluster stands out.

118 coordinates its distribution is similar to the total (black) one, only the three others stand out.
 119 These three clusters are :

120 • **the red cluster** : it contains what visually stands out as the core of the swirl, a strong
 121 vortex. It is characterised by its high values of vertical vorticity ω_z which fits our physical
 122 intuition.

123 • **the yellow cluster** : this region is mainly located around the red one and seems to be
 124 organised in four structures surrounding the swirl. It also include several patches throughout
 125 the volume. This cluster shows the highest values of horizontal strain $s^{(uv)}$ which hints that
 126 the four structures are stretching outward of the swirl.

127 • **the light blue cluster** : it is better seen in figure 3 that this cluster has a layered structure.
 128 This is here hinted as this is the region where the horizontal shear S_z reaches its highest
 129 values.

130 The right side of figure 4 confirms our understanding of the clustering as to which
 131 coordinates characterise which cluster.

132 This method allows us to better understand what happens in the simulation. The story
 133 it tells us seems to be what follows. The red cluster comprises the strong vertical vortex
 134 rotating around the z axis. Within this region, as seen in the slices in figure 4 and 5, some
 135 yellow patches appear. They are stretching outward the swirl as hinted by the high values of
 136 $s^{(uv)}$. These patches then extend into the light blue region with a pancake shape, these are
 137 the so-called layers we try to understand, where shear reaches its highest values.

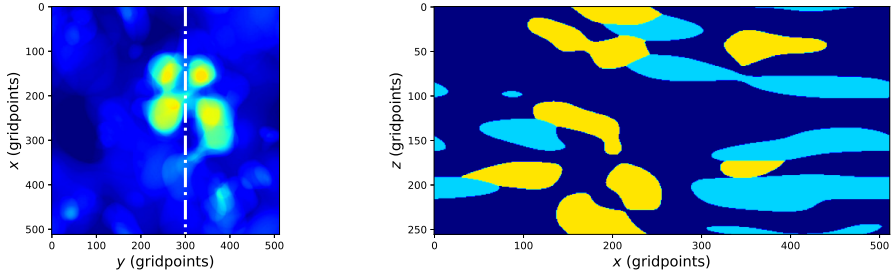


Figure 5: Applying the method to the same database putting away ω_z , and $S_h^{(uv)}$ that has similar algebraic expression as ω_z , as coordinates leaves every cluster but the red one nearly untouched. The rotating core of the swirl isn't identified anymore, but the forming layers, the layers and the background still lie in the same spots.

138 As shown in figure 4, some yellow and red patches may pop up between two layers. These
 139 are where the two layers interact.

140 5. Robustness

141 In our case, we chose $n=4$. Increasing n only led to splitting existing clusters into sub-clusters
 142 of the same size, which is how KMeans reacts to asking it to find more clusters than it can
 143 identify. Indeed, if two points converge to the same cluster centre but are not superimposed,
 144 they induce two such sub-clusters splitting the real cluster in halves.

145 Besides, we can easily check that the 4 clusters yielded by our method are robustly picked
 146 out. Simply ignoring the coordinate responsible for one cluster's identification should let
 147 every cluster but the one we test untouched and make the latter disappear. For example,
 148 in figure 5 are shown the 3 clusters obtained (n accordingly changed here) when ignoring
 149 ω_z (and $S_h^{(uv)}$ that has similar algebraic expression as ω_z as said before). As we could
 150 have guessed, the swirl is no longer seen but the forming layers, the formed layers and the
 151 background are unaltered.

152 6. Mixing properties

153 [TO BE WRITTEN]

154 7. Conclusions

155 [TO BE WRITTEN]

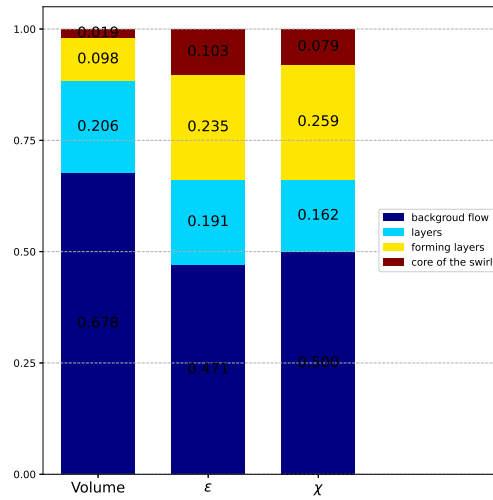


Figure 6: Each clusters absolute contribution to the mixing, meaning how much ε and χ they hold, compared to their volume. The three interesting clusters identified by our method represent only a third of the volume but half of the mixing. The core of the swirl in particular is responsible for a significant part of the mixing even though it is the smallest cluster in volume by far.

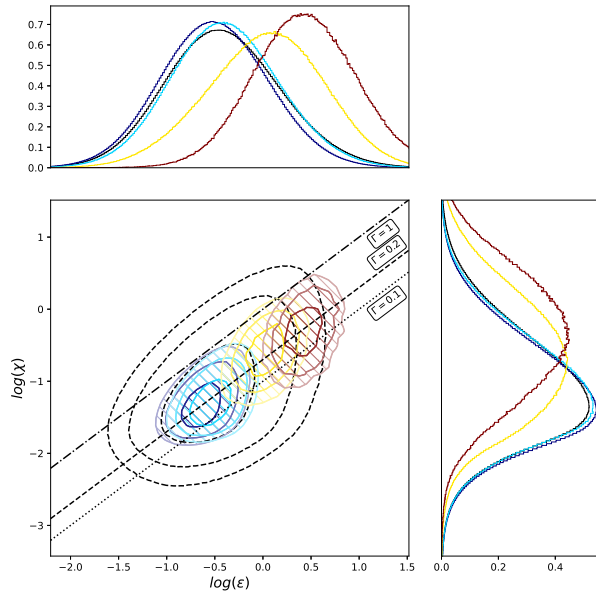


Figure 7: Overall joint PDFs of $\log(\varepsilon)$ and $\log(\chi)$ in black, and coloured by clusters. We only show the contours of those PDFs to keep a clear figure. Three straight lines show where the the values $\Gamma = 0.1$, $\Gamma = 0.2$ and $\Gamma = 0.3$ lie.

REFERENCES

- 156 CAULFIELD, C.P. 2021 Layering, Instabilities, and Mixing in Turbulent Stratified Flows, *Annual Review of*
157 *Fluid Mechanics* **53**, 113-145.
- 158 PORTWOOD, G.D., DE BRUYN KOPS, S., TAYLOR, J.R., SALEHIPOUR, H. & CAULFIELD, C.P. 2016 Robust
159 identification of dynamically distinct regions in stratified turbulence, *Journal of Fluid Mechanics*
160 **807**.
- 161 TAYLOR, J.R., DE BRUYN KOPS, S.M., CAULFIELD, C.P. & LINDEN, P.F. 2019 Testing the Assumptions
162 Underlying Ocean Mixing Methodologies Using Direct Numerical Simulations, *Journal of Physical*
163 *Oceanography* **49** (11), 2761-2779.
- 164 COUCHMAN, M.M.P., DE BRUYN KOPS, S.M. & CAULFIELD, C.P. 2022 Transport across stable density
165 interfaces in forced stratified turbulence, *in proceedings for Journal of Fluid Dynamics*.
- 166 COUCHMAN, M.M.P., WYNNE-CATTANACH, B., ALFORD, M.H., CAULFIELD, C.P., KERSWELL, R.R.,
167 MACKINNON, J.A. & VOET, G. 2021 Data-driven identification of turbulent oceanic mixing from
168 observations of microstructure data, *Geophy. Res. Lett* **48** (23), e2021GL094978.
- 169 ALMALKIE, S. & DE BRUYN KOPS, S. M. 2012 Kinetic energy dynamics in forced, homogeneous, and
170 axisymmetric stably stratified turbulence, *Journal of Turbulence* **13** (29).