AIX-MARSEILLE UNIVERSITÉ

École doctorale des Sciences de l'Environnement

THÈSE

pour obtenir le grade de

Docteur en Sciences de l'Université d'Aix-Marseille

Discipline : Océanographie Physique

Présentée par:

Andrea COSTA

Marine Connectivity: Exploring the Role of Currents and Turbulent Processes in Driving It.

Soutenance effectuée le 28 Avril 2017 devant le jury composé de:

Président du jury: Jean-Christophe POGGIALE - MIO (Marseille)

Rapporteurs: Pascale BOURUET-AUBERTOT - LOCEAN (Paris)

Sabrina SPEICH - LMD (Paris)

Examinateur: Patrick MARSALEIX - LA (Toulouse)

Directeurs de thèse: Anne PETRENKO - MIO (Marseille)

Andrea DOGLIOLI - MIO (Marseille)

To all the correlated fluctuations
that determined 쥬쥬.
And to her,
who doesn't give a diaper about science.
Yet.

Résumé : La connectivité marine traite du transfert de larves et/ou d'individus entre des habitats marins éloignés. Grâce à la connectivité, les populations marines éloignées peuvent faire face à la pression de l'habitat en s'appuyant sur le transfert qui vient des populations éloignées de la même espèce.

Le transfert entre des populations éloignées de l'océan est possible par le transport dû aux courants. Cependant, il n'est pas encore clair si le champ des courants détermine totalement la persistance des espèces marines ou si la démographie locale joue un rôle.

Les mesures in situ de la connectivité sont extrêmement difficiles. Par conséquence, notre connaissance de la connectivité est déduite des simulations numériques de dispersion. Mais on sait bien que les modèles numériques font plusieurs approximations dans la simulation du champ des courants. Le but de cette thèse est de préciser si nous pouvons déduire la persistance de la connaissance du champ des courants et étudier l'effet des paramétrisations numériques de turbulence dans l'estimation de la connectivité.

Premièrement, l'analyse en théorie des graphes et en modèle de métapopulation a permis de déterminer que les courants ont un rôle prédominant dans la determination de la connectivité. Dans notre cas d'étude d'une espece benthique, la quantité des connections d'un site de reproduction avec d'autres sites peut contenir jusqu'au 77% de l'information sur la persistence.

Deuxièmement, le comparaison avec des mesures de microstructure des simulations numériques a permis d'identifier la configuration de modèle numérique Symphonie qui reproduit le mieux l'activité turbulente observée in situ. Ce résultat a été obtenu avec une procédure de comparaison que j'ai developée et qui peut etre appliquée à tout modèle numérique. Un résultat novateur de cette analyse est la mise en évidence de l'importance de l'effet de la composante biologique, *via* la pénétration de la lumière, sur l'intensité de la turbulence.

Troisièmement, l'implémentation d'un modèle 2-D et sa comparaison avec les donnés in situ de dispersion d'un traceur passif a permis d'identifier un nouveau mécanisme de génération de turbulence au fond qui explique le mélange vertical observé dans l'océan profond. Cela a permis de déterminer le coefficient de frottement de l' écoulement au-dessus d'une topographie rugueuse. Plus important, je montre que, dans certaines conditions, les modèles numériques actuels sous-estiment le coefficient de frottement au fond par trois ordres de grandeur.

Mots clés : Connectivité, Persistence, Théorie des Graphes, Modèle de Métapopulation, Turbulence, Schémas de Fermeture de Turbulence, Mélange au Fond.

Abstract: Marine connectivity is the transfer of larvae and/or individuals between distant marine habitats. Thanks to connectivity, distant marine population can face habitat pressure by relying on the transfer from distant populations of the same species.

The transfer between distant populations in the ocean is made possible by the transport due to the currents. However, it is still not clear if the current field totally determines the persistence of the marine species or if the local demography plays a role.

In situ measurements of connectivity are extremely difficult. Therefore, our knowledge about connectivity is generally inferred from numerical dispersal simulations. However, numerical models make several approximations in simulating the current field. The aim of this thesis is to clarify if we can deduce the persistence from the knowledge of the current field and study the numerical modelling of surface and bottom turbulence mechanisms that can influence connectivity.

Firstly, the use of graph theory and metapopulation model permits to determine that currents have a predominant role in determining connectivity. In our case of the study of a benthic species, the amount of connections of a reproductive site with other reproductive sites could contain, with specific larval durations, up to 77% of the information about persistence. Furthermore, a specific graph theory tool identifying clusters in networks is able to identify sub-populations.

Secondly, the comparison of microstructure data with numerical simulations permits to identify the configuration of the numerical model Symphonie that best reproduces the observed turbulent activity. This result is achieved through a comparison procedure that I developed and that can be applied to all numerical models. A novel finding is the importance of the effect of the biological component, here *via* the light penetration, on the turbulence intensity.

Thirdly, the implementation of a 2-D numerical model and its comparison with in situ tracer dispersal data permits to identify a new generating mechanism of bottom boundary turbulence that accounts for the vertical mixing observed in the deep ocean. This allows to determine an effective drag coefficient due to flow over rough topography. More importantly, I show that -in some conditions- present numerical models underestimate by three orders of magnitude the bottom drag coefficient.

Keywords: Connectivity, Persistence, Graph Theory, Metapopulation Model, Turbulence, Turbulence Closure Schemes, Boundary Mixing.

Contents

1 Introduction **1**

1.1 Defining connectivity 1

1.1.1 Connectivity and marine populations persistence 1

1.1.2 Study of connectivity matrices and graph theory 3

1.2 Numerical models and turbulence 4

1.3 Bottom boundary turbulence and dispersion 8

1.4 Aims and strategy 10

2 Study of marine connectivity and marine populations persistence with graph theory **13**

2.1 On the Calculation of Betweenness Centrality in Marine Connectivity Studies Using Transfer Probabilities 16

2.1.1 Abstract 16

2.1.2 Introduction 17

2.1.3 Materials and Methods 18

2.1.4 Results 21

2.1.5 Conclusion 22

2.2 Tuning the interpretation of graph theory measures in analyzing marine larval connectivity: The Gulf of Lion study case. 25

2.2.1 Introduction 25

2.2.2 Materials and Methods 27

2.2.3 Results 39

2.2.4 Discussion and conclusions 43

2.3 Extra Material 47

2.3.1 Bridging centrality 47

2.3.2 Visual explanation of graph theory measures 50

3 Numerical Models and Turbulence **55**

3.1 Comparison of in situ microstructure measurements to different turbulence closure schemes in a 3-D numerical ocean circulation model 59

3.1.1 Abstract 59

3.1.2 Introduction 60

3.1.3 Materials and Methods 64

3.1.4 Results 73

3.1.5 Discussion 84

3.1.6 Conclusions 93

3.2 Appendices 97

3.2.1 Boundary conditions 97

3.2.2 Surface boundary conditions for k 97

3.2.3 Closure schemes 98

3.3 Segmentation and quality fit 101

3.4 Extra Materials for "Comparison of in situ microstructure measurements to different turbulence closure schemes in a 3-D numerical ocean circulation model" 104

3.5 Extra Material 109

3.5.1 Additional Simulations 109

3.5.2 Data analysis toolbox 109

4 Bottom boundary mixing and diffusion **117**

4.1 Spatially Inhomogeneous and Temporally Intermittent Boundary Mixing along the Northern Deepwater Gulf of Mexico 123

4.1.1 Abstract 123

4.1.2 Introduction 124

4.1.3 Data and Methods 128

4.1.4 Observations 133

4.1.5 Boundary Mixing 135

4.1.6 Summary and Discussion 140
4.2 Extra Material 154
5 Conclusions and perspectives 157
5.1 Conclusions 157
5.2 Perspectives 161
Bibliography 165

List of Figures

1.1 Upper panel - Example of Lagrangian trajectories idealizing larval dispersal (data from *Berline et al.* 2014). Middle panel - Example of connectivity matrix storing the transfer probability of larvae of benthic polychaete between the 32 reproductive sites in the Gulf of Lion (lower panel). Data from *Guizien et al.* (2014). . . 2

1.2 (a) *"Observe the motion of the surface of water, which resembles the behavior of hair, which has two motions, of which one depend on the weight of the strands, the other on the line of its revolving; thus water makes revolving eddies, one part of which depends upon the impetus of the principal current, and the other depends on the incident and the reflected motions."* (Leonardo da Vinci 1513) As *Davidson* (2004) asks himself: "Did Leonardo da Vinci foresee Reynolds' idea of dividing a turbulent flow into two components: a mean velocity and the turbulent fluctuations?". (b) This figure represents a one-dimensional simplification of the cascade process in which bigger eddies becomes unstable and break up in smaller eddies till when the scale of the eddies matches the scale at which dissipation dominates the dynamics. ε is the kinetic energy dissipation rate determining the energy flux from large to small scales. This figure was modified from *Frisch* (1995). 6

1.3 DNS: Very small scale flow (ex:turbulent boundary layers) but currently computationally intractable for most problems. LES: Aims to solve the computational cost that DNS poses and reveals the eddies hidden behind the mean in RANS. Good for small scale applications in both coastal and open ocean. RANS: The least computationally expensive method that is used for turbulent modeling. However, it is really not very good when certain phenomena cannot be averaged, such as instabilities. This figure is taken from André Bakker's lectures: `http://www.bakker.org/dartmouth06/engs150/10-rans.pdf` 7

1.4 Diapycnal diffusivity K_ρ estimated by microstructure measurements in the Gulf of Mexico (blue and black) versus the average K_ρ inferred from tracer release data during the GISR campaign by *Ledwell et al.* (2016) (magenta). The red line shows the depth at which the tracer was initially released. 9

1.5 2-D section of the zonal velocity field issued 3-D high resolution simulation. The black circle highlights a hydrostatic-like flow taking place in a 100m layer above rough topography. Figure provided by Max Nikurashin. 11

2.1 Representation of the connectivity matrix 7 and corresponding betweenness values using different metrics. The left side colorbars indicates betweenness values, the right side colorbars indicate the metric values. (a) Results obtained by using a_{ij} as edge weight, (b) results obtained by using d_{ij} as edge weight. In (a) the lowest 5% of edges weights are represented. In (b) the lowest 5% of edges weights are represented. Note the change in the colorbars' ranges. 23

2.2 Betweenness nodes in the Gulf of Lion for the 32 sites using 20 different connectivity matrices obtained with Lagrangian simulations by *Guizien et al.* (2014). (a) Results obtained by using a_{ij} as edge weight; (b) results obtained by using d_{ij}. Note the change in the colorbars' ranges. 23

2.3 Schematic representation of the typical circulation in the Gulf of Lion. The thick arrow represents the dominant alongshore Northern Current. The thinner arrow represents the eastward current that can be detected in stratified conditions or under particular wind field conditions. The positions of the 32 studied sites are plotted. The sites 1, 10, 18 and 32, used for the habitat loss scenario in *Guizien et al.* (2014), are highlighted by bigger gray dots. Node 21 is the smallest of the gray dots. The gray lines correspond to the 100 m, 200 m, 1000 m and 2000 m isobaths. 28

2.4 The red and yellow node has high bridging centrality. The red node has high betweenness centrality. Nodes are divided into one cyan and one green clusters. 49

2.5 Types of graph. 51

2.6 Degree: number of edges emanating from a single node. 51

2.7 Clustering coefficient: how many of your nearest neighbors are connected to one another. 52

2.8 Local efficiency: average shortest path connecting all neighbors of a given node. 52

2.9 Betweenness: number of shortest paths that pass through a given node. 53

3.1 Numerical model domain. The color code represents the water depth. The Gulf of Lion is magnified in the smaller box where the measurements sites are represented by red dots. Note that many profiles were taken at the same location over time. The black lines in the smaller box represent the 0, 50, 100, 500, 1000 and 1500 m isobaths. 70

3.2 The thick lines represent median values of ε estimated from in situ data (in black) and numerical experiments. The shades indicate 95% bootstrap confidence intervals. Water depth z is adimensionalised with respect to the mixed layer depth MLD for each profile. 75

3.3 The thick lines represent median values of N estimated from in situ data (in black) and numerical experiments. The shades indicate 95% bootstrap confidence intervals. Water depth z is adimensionalised with respect to the mixed layer depth MLD for each profile. 76

3.4 The thick lines represent median values of K_Z estimated from in situ data (in black) and numerical experiments. The shades indicate 95% bootstrap confidence intervals. Water depth z is adimensionalised with respect to the mixed layer depth MLD for each profile. 78

3.5 a) and b): ε probability density functions issued from all the in situ (black shade) and numerical data (thick lines) in the surface mixed layer for the KE closure scheme and the KL closure scheme respectively. c) and d): the same but below the surface mixed layer. 80

3.6 a) and b): N probability density functions issued from all the in situ (black shade) and numerical data (thick lines) in the surface mixed layer for the KE closure scheme and the KL closure scheme respectively. c) and d): the same but below the surface mixed layer. 82

3.7 a) and b): K_Z probability density functions issued from all the in situ (black shade) and numerical data (thick lines) in the surface mixed layer for the KE closure scheme and the KL closure scheme respectively. c) and d): the same but below the surface mixed layer. Note the difference in the range of the y axes in a) and b). 83

3.8 a) All in situ (in black) and numerical data of K_Z, N and ε. b) All in situ (in black) and numerical data of K_Z and ε. The color code is the same as in the other figures. 85

3.9 a) JP values distribution obtained with the II91 segmentation method, b) same as a) for S11, c) same as a) for C12, d) same as a) for M12. e) same as a) for a constant 512 points segmentation. f) same as a) for a 1024 points segmentation with 30% overlap. The red curve is the cumulative distribution of JP values for each case and the vertical line marks the point where the cumulative distribution equals 0.6. The database for this analysis comprehended 126 profiles with an average depth of 50m in various meteorological conditions in the Gulf of Lion. 102

3.10 Probability density function of the turbulence intensity parameter Re_b, obtained with all the in situ data (black), the in situ data above the MLD (green) and the in situ data below the MLD (blue). The dotted red lines indicate the separation between the turbulence regimes. 1 indicates the diffusive regime; 2 the buoyancy-controlled regime; 3 the transitional regime; and 4 the energetic regime. 104

3.11 The thick lines represent median values of temperature measured from in situ data (in black) and numerical experiments. The shades indicate 95% bootstrap confidence intervals. Water depth z is adimensionalised with respect to the mixed layer depth MLD for each profile. 105

3.12 The thick lines represent median values of density measured from in situ data (in black) and numerical experiments. The shades indicate 95% bootstrap confidence intervals. Water depth z is adimensionalised with respect to the mixed layer depth MLD for each profile. 106

3.13 The thick lines represent median values of turbulent kinetic energy measured from in situ data (in black) and numerical experiments. The shades indicate 95% bootstrap confidence intervals. Water depth z is adimensionalised with respect to the mixed layer depth MLD for each profile. . . . 107

3.14 In the upper panels the median profiles of kinetic energy dissipation rate for the two numerical experiments with $k_{min} = 10^{-6} m^2/s^2$ (suffix MINk2) are represented. In a) KLsetMINk2 has molecular diffusion set to zero. In b) KLsetMINk2MINKz has not null molecular diffusion. In the lower panels we represent the PDF of the ε values obtained for these numerical experiments both above (in c) and below (in d) the MLD. 108

3.15 Profiles of kinetic energy dissipation rate for all the numerical experiments and the in situ data. 110

3.16 Probability density functions of kinetic energy dissipation rate for all the numerical experiments and the in situ data both above the MLD (a) and below (b). 111

3.17 Folder tree of the mio_scaptool toolbox. Red scripts highlights where the changes detailed in the following can be found. 112

4.1 Meridional overturning circulation in the Atlantic. Different color highlight the different overturning cells forming the MOC. Image source `http://www.goes-r.gov/` 118

4.3 Zonally averaged meridional overturning streamfunction [4 Sverdrup (Sv; 1 Sv 5×10^6 m^3/s)] overlaid by isopycnals in the deep ocean. (top) A case with a constant diapycnal diffusivity of $K_Z = 3 \times 10^{-4}$ m^2/s; (bottom) a case with k decaying from $10^{-3} m^2/s$ at the bottom to $10^{-5} m^2/s$ at the top with a e-folding scale of 500 m. The arrows represent the direction of the circulation. Figure reproduced from *Mashayek et al.* (2015). 119

4.2 Diapycnal mixing followed by isopycnal dispersion. Isopycnal surfaces (of constant density) are shown as full lines and dots mark a region of very small neutrally buoyant particles. Initially, in (a), the density gradient is uniform and given by $\rho = \rho_0(1 - N^2 z/g)$, where z is the upward vertical measured from the mean depth of the mixed region and N is the buoyancy frequency of the water. Turbulence illustrated in (b) leads to the homogenization of a patch (c), so that the density in section A-A through the center of the patch is now as shown at the left. The density profiles outside the mixed patch is unchanged and as shown in (a). If the pressure at the top of the mixed region at $z = h$ is p_0, that within the mixed region is $p_I = p_0 + hzg\rho_0 dz = p_0 + g\rho_0(h - z)$, and that in the surrounding ambient water is $p_A = p_0 + hzg\rho dz = p_0 + g\rho_0[(h - z) - (N^2/2g)(h_2 - z_2)] = p_1 - g\rho_0(N^2/2g)(h_2 - z_2) < p_I$ at levels $-h < z < h$. There is therefore a horizontal pressure gradient that causes the mixed region to spread laterally whilst, to conserve volume, it collapses vertically as shown in (d), intruding, into the surrounding water along the isopycnal surface that has the density of the mixed region. During the turbulent mixing, and in the process of intrusion, waves may be radiated into the surrounding quiescent fluid. The spread of the patch leads to the dispersion of particles along the isopycnal surface. The effect of the Earth's rotation will eventually cause the divergent flow in the spreading patch to rotate anticyclonically under the action of Coriolis forces, whilst fluid converging to replace it will rotate cyclonically. This sketch is grossly oversimplified. In practice, turbulence may not homogenize a region (either in density or in particle distribution) and may be associated with momentum transfers that affect the subsequent horizontal dispersion. Image reproduced from *Thorpe* (2005). . . . 122

4.4 An example of shear (black solid) and strain (red) spectra illustrating signal-to-noise and choice of bandwidth. The green solid line represents ping-to-ping uncorrelated noise at 10 times the theoretical minimum rate. Dashed lines are an attempt to represent the effects of averaging and processing. Blue lines denote the high wavenumber asymptote of the Garrett and Munk spectrum. See *Polzin et al.* (2014a) for further details. In this study, the shear spectral level $\hat{E}$ (4.1.5) was estimated at wavelengths of 160 m and smaller under the criterion that the observed spectra (black solid) lie a factor of 3 above the estimate noise (green dashed). Noise in the CTD data at 0.2-0.4 cpm are related to ship heave in ocean swell being communicated down the lowering wire. 143

4.5 Bathymetry and station positions for the GISR experiment. Black squares denote the positions of the GISR moorings, large red circles locate the tracer injection, large green circles the initial 5-12 day sampling. Other dots depict tracer sampling stations at 4 and 12 months. Positions of Hurricane Isaac are indicated at 6 hourly intervals. Bathymetry is contoured every 200 m between 200-3000 m and every 50 m between 500-1500 m water depth. Black contours are used to emphasize the 500, 750, 1000 and 1250 m isobaths. 144

4.6 Station positions for the year 1 GISR tracer survey are rendered as circles. Column integrals of tracer concentration are indicated by relative shading of the circles. White indicates no tracer was found. Black lines depict isobaths at a 500 m contour interval. Boxes indicate binning into 'Boundary', 'South' and 'East' averages for the finestructure analysis, Fig. 4.7. 145

4.7 Diapycnal diffusivity estimates from the finescale parameterization applied to LADCP and CTD data. Results for 'Boundary' stations are depicted with the black trace. Results for 'South' and 'East' stations (Fig. 4.6) are rendered in blue. The red horizontal line indicates the depth of the injection surface at the boundary. 146

4.8 Vertical profile data from station 31. Left hand panel: potential density anomaly relative to 1200 dBars. Right hand panel: north-south velocity (red), east-west velocity (blue). Bottom track estimates and shipboard ADCP velocity estimates are rendered in black. Note the differing extent of the vertical axes. In both panels the vertical axis extends to the actual bottom ($\cong$ 1225 m). There are regions of significantly reduced stratification suggestive of either mixed layers or isopycnal separation in a partial hydraulic jump. Significant shear (0.1 m s^{-1} over 40 m) is noted beneath this feature, implying a gradient Richardson number smaller than one. These data were collected above a ridge saddle (Figure 4.9) and lend themselves to an expectation of strongly enhanced mixing associated with highly sheared flow on the lee side of the ridge. 147

4.9 The orographic context for Station 31. This station was taken at a ridge saddle with the near-bottom velocity being east-to-west. 148

4.10 Estimates of $\int U^3 dt$ for the bottommost current meters in the 6-element GISR moored array. The 4 grey panels indicate the time of the tracer injection (G02), the passage Hurricane Isaac, the 4-month and 12 month sampling (G03 and G05, respectively). Note that the U^3 statistics between injection and 4 month sampling are dominated by Isaac. The bottom current meter in the south-west quadrant of the array (M6), collocated with the injection site, was used to specify the near-bottom velocity U in the mixing scheme (4.1.10). 149

4.11 Four month column integrated tracer concentrations averaged as a function of distance from the 1500 m isobath from observations (blue) and model (black) with ($\beta = 2, u = 0.01, \frac{1}{2}C_d\frac{d}{\ell} = 0.06$ and $K_{\perp}^{int} = 3 \times 10^{-5}$ m^2 s^{-1}). Despite the observed variability there is an obvious requirement for the tracer center of mass to migrate off the boundary, which is addressed with the pseudo-advection parameter u (4.1.10). 150

4.12 Four month tracer distribution with ($\beta = 2, u = 0.01, \frac{1}{2}C_d\frac{d}{\ell} = 0.06$ and $K_\perp^{int} = 3 \times 10^{-5}$ m^2 s^{-1}). Green vertical lines delimit the horizontal extent of the 4 month tracer sampling and the extent of averaging for the numerically derived interior and boundary dispersion estimates. The black dot depicts the location of the initial condition. The vertical coordinate is the height above the target surface (hat). 151

4.13 Best fit boundary profile comparison for observations (blue) and model (black) with ($\beta = 2, u = 0.01, \frac{1}{2}C_d\frac{d}{\ell} = 0.06$ and $K_\perp^{int} = 3 \times 10^{-5}$ m^2 s^{-1}). Note the vertical asymmetry about the injection level implying a downward increase in diapycnal mixing. 152

4.14 Best fit interior profile comparison for observations (red) and model (black). Otherwise as in Figure 4.13. 152

4.15 An error map in (β, u, C_d and $K_\perp^{\rm int}$) space. There are two color panels with in each parameter square. The left panel represents the interior profile and the right the boundary. Each large square contains a constant C_d and β. Within each large square there is a three-by-four matrix of possible ($K_\perp^{\rm int}, u$) entries. Mean-square error estimates are normalized to 'best fit' parameters ($\beta = 2, u = 0.01$ m s^{-1}, $\frac{1}{2}C_d\frac{d}{\ell} = 0.06$ and $K_\perp^{\rm int} = 3 \times 10^{-5}$ m^2 s^{-1}), which is indicated as a green shading. Better fits can be obtained in one domain (interior or boundary) only at the expense of the other. 153

List of Tables

2.1 *Settings for the three simulation groups.* 32

2.2 Number of clusters found by modularity maximization for each PLD and nodes composing them. 41

2.3 $< \rho >$ is the mean of the Spearman correlation coefficients that have a significant (< 0.05) p-value. #vals indicates the number (out of 20) of correlation coefficient that are significant. 44

2.4 Swapping coefficients SC relative to graph theory-based site hierarchy and metapopulation-based one. The mode of the most important site according to graph theory (mGT) and metapopulation model (mMM) are also reported. 46

3.1 Set of numerical experiments performed for this study. 84

3.2 MLD ratio indicates the ratio of the numerical prediction of the MLD to the in situ value $< MLD >_{SCAMP} =$ 27.3 m. σ_ε indicates the mean value of the decadal standard deviation of the ε profiles. σ_N indicates the mean value of the decadal standard deviation of the N profiles. σ_{K_Z} indicates the mean value of the decadal standard deviation of the K_Z profiles. . . . 86

3.3 S_ε indicates the shift of the numerical distribution of ε in the surface layer respect to the in situ one. Δ^2_ε is the squared difference between the numerical and in situ distribution of ε values in the surface layer. Similar definitions apply for different subscript variables. Errors are calculated with a re-sampling procedure. 87

3.4 S_ε indicates the shift of the numerical distribution of ε below the surface layer respect to the in situ one. Δ^2_ε is the squared difference between the numerical and in situ distribution of ε values below the surface layer. Similar definitions apply for different subscript variables. Errors are calculated with a re-sampling procedure. 88

3.5 Free-parameters of the SYMPHONIE $k-\varepsilon$ closure scheme (*Warner et al.*, 2005). 99

4.1 Locations, instrument depths, total water depths of M1-M6 deepwater current moorings. *collected by S4 current meters. 130

Chapter 1

Introduction

Contents

1.1 Defining connectivity 1

1.1.1 Connectivity and marine populations persistence 1

1.1.2 Study of connectivity matrices and graph theory 3

1.2 Numerical models and turbulence 4

1.3 Bottom boundary turbulence and dispersion 8

1.4 Aims and strategy 10

1.1 Defining connectivity

1.1.1 Connectivity and marine populations persistence

Marine life is increasingly threatened by anthropogenic stressors. A broad range of ocean-based activities, anthropic pressure and climate change are rapidly modifying and/or degrading the ocean biodiversity (*Lubchenco et al.* 1999; *Duraiappah et Shahid* 2005).
A key weakness of marine populations is the habitat discontinuity due to both natural reasons (e.g. marine ridges, geographical distance) or either direct and indirect anthropogenic disturbances (e.g. coastal development, climate change). As a consequence, marine populations rely on the long-range exchange of larvae and/or individuals (i.e., connectivity) in order to maintain their spatial distribution (*Hanski*, 1999).

Connectivity permits the dispersion of larvae and the transfer of individuals across the network formed by distant populations of the same species. Previous studies have demonstrated how connectivity may enhance the persistence of a metapopulation after disturbances, by enhancing their recovery and/or recolonization (*Hastings et Botsford*, 2006; *Guizien et al.*, 2014). It follows that understanding the degree and patterns of connectivity between populations plays a central role in conservation planning, like the design of marine protected areas (*Sale et al.*, 2006) and spatial planning policies (*Boersma et Parrish*, 1999).

A widely used connectivity estimation method is the modeling of larvae and individual dispersal with biophysical modeling that couple ocean dynamics with a number of biological features of the species at study. The approach most commonly used in literature to study larval dispersal trajectories computed with biophysical models is the study of connectivity matrices (Figure 1.1). With the origin sites on the columns and destination sites on the rows, con-

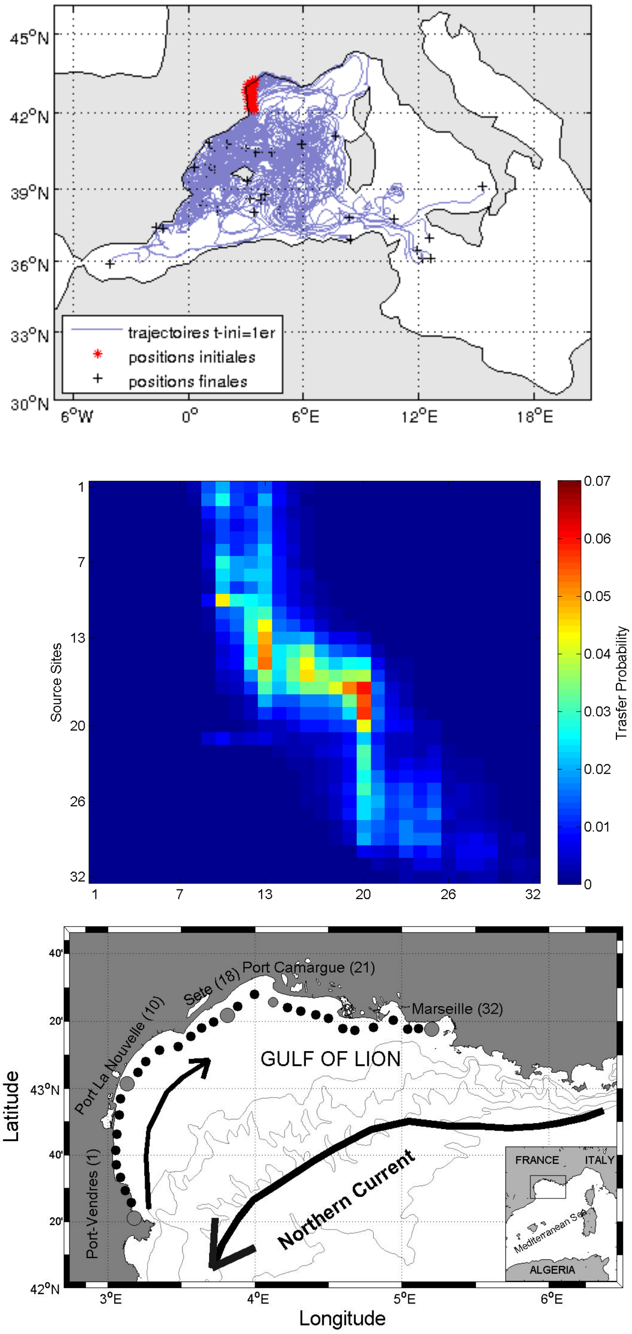


Figure 1.1: Upper panel - Example of Lagrangian trajectories idealizing larval dispersal (data from *Berline et al.* 2014). Middle panel - Example of connectivity matrix storing the transfer probability of larvae of benthic polychaete between the 32 reproductive sites in the Gulf of Lion (lower panel). Data from *Guizien et al.* (2014).

nectivity matrices stores either the average time that an individual from a given source site need to arrive to a destination (or sink) site (e.g., *Berline et al.*, 2014) or the probability that an individual disperses from a given origin site to a destination site (e.g., *Burgess et al.*, 2014).

1.1.2 Study of connectivity matrices and graph theory

A variety of techniques has been developed for the study of connectivity matrices: metapopulation models (*Caswel*, 2001), hierarchical clustering (*Jacobi et al.*, 2012), circuit theory (*McRae et al.*, 2008) and graph theory (*Urban et Keitt*, 2001) among others. In particular, graph theory has received an increasing amount of attention in the last decade (*Moilanen*, 2011). More and more classical graph theory concept has found an application to the study of marine connectivity: betweenness centrality (*Urban et Keitt*, 2001), eigenvalues centrality (*Andrello et al.*, 2014) and modularity (*Kininmonth et al.*, 2010a; *Thomas et al.*, 2014) among others. The use of such mathematical tools is appealing to the scientific community because, while other connectivity study methods, like metapopulation models, depend on the estimation of various biological parameters (see *Guizien et al.*, 2014, for example) and therefore need to be finely tuned in order to be applied to different species and geographical contexts, graph theory can be directly transposed to other study cases by just recalculating the connectivity matrices. Moreover, by relying on numerical simulation, graph theory can effectively be complementary to genetic studies and supply information on boundaries of genetically based communities when sampling of genetic diversity is sparse (*Kininmonth et al.*, 2010a).

However, the interpretation of the results obtained with graph theory still relies on the intuition we build on the study of social networks. In the scientific community, a definitive consensus on whether such intuition directly applies to the study of marine connectivity is still not achieved (*Moilanen*, 2011). I addressed this problem in the first part of my thesis.

A large part of the literature studying marine connectivity *via* biophysical modeling is focused on species with dispersal abilities directly proportional to their pelagic larval duration (PLD). Apart from horizontal currents, important physical factors that can influence the dispersal of the larvae during their PLD are temperature and vertical movements. While temperature can modify the duration of the PLD itself (*Houde* 1989; *Cowen et al.* 2000; *O'Connor et al.* 2007), vertical movements can expose the larvae to different ambient temperature conditions and to a more variable velocity field. Both these aspects can significantly modify the larvae's dispersal pattern (*sensu Butler et al.*, 2011; *Qiu et al.*, 2011).
Due to the sensitivity of the dispersal on such factors, it is important that the numerical models that we use in the biophysical models used to calculate connectivity matrices make accurate predictions of such factors.

1.2 Numerical models and turbulence

Many validations of numerical models have been done in the past about the accuracy of the prediction of temperature and velocity fields (e.g., *Peters et al.*, 2004, chapter 39; *Warner et al.*, 2005; *Peters et Baumert*, 2007). Less studies have assessed the effect of the numerous choices on which depend the correct prediction of vertical overturnings and turbulence modelling. In particular, their prediction and modelling depend on the reliability of the Turbulence Closure Scheme (TCS) employed to numerically integrate the Navier-Stokes equations and on its interplay with the type of surface boundary conditions that is specified. I address this question in the second part of my thesis.

While it is very easy to grasp the idea of the turbulent nature of geophysical flows, it is very difficult to give a precise definition of turbulence. In general, we can list some characteristics that are proper to turbulent flows (see for example *Tennekens et Lumley*, 1972, for a detailed discussion of them): i) randomness; ii) enhanced diffusivity of physical

properties; iii) three-dimensional fluctuations of vorticity; iv) decaying; and v) a continuous wide range of spatial scales. All these aspects can be linked to the nonlinearity of the equations describing fluid motion: the so-called Navier-Stokes Equations (NS Equations). From case to case, we are forced to make ad hoc assumptions in order to make the study of turbulence tractable. Two often-made assumptions in oceanographic studies are: i) the homogeneity of turbulence (*Kelvin*, 1887); and ii) its steadiness on temporal scales that permits to divide the fluid motion in an average component and a random one (Figure 1.2a): the so-called Reynold's decomposition (*Reynolds*, 1895). From these assumptions a number of results can be derived (see *Davidson*, 2004, for a comprehensive discussion). The first assumption permits to obtain the classical view of turbulence as an eddy cascade process (the so-called Richardson's cascade; Figure 1.2b) while the second one permits to define some length scales characterizing all the turbulent flows at high Reynolds number (*Kolmogorov*, 1942; *Obukhof* 1949; *Corrsin* 1951; *Kolmogorov* 1962). Kolmogorov's results can be further expanded to obtain an exact equation for the equilibrium scalar spectrum at very high spatial wave numbers by integrating the linear scalar diffusion equation for a typical small scalar Fourier element in pure strain (*Batchelor*, 1959).
The second assumption allows many different strategies in order to approximate the Navier-Stokes equations and solve them numerically (e.g., *Burchard et Bolding*, 2001).
This last result is maybe the most important for the modern studies of geophysical turbulence. In fact, a large part of the numerical models of oceanic and atmospheric circulation are built upon it; it is the so-called Reynolds Averaged Navier-Stokes (RANS) models.

Typically, one model cannot resolve the whole spectrum of the flow dynamics. Thus, it resolves only one portion of it while modelling the rest (Figure 1.3). Generally, we can distinguish between i) *Direct Numerical Simulations* that cover all the spatial scales but are limited in the range of Reynolds number (e.g., *Smyth et al.*, 2001); ii) *Large Eddy Simulations* that solve spatially filtered Navier-Stokes equations and are typically limited to the investigation of near-surface mixed layers (e.g., *Skyllingstad et al.*, 1998); and iii) *Statistical Turbulence Models* that solve Reynolds-Averaged Navier-Stokes equations usu-

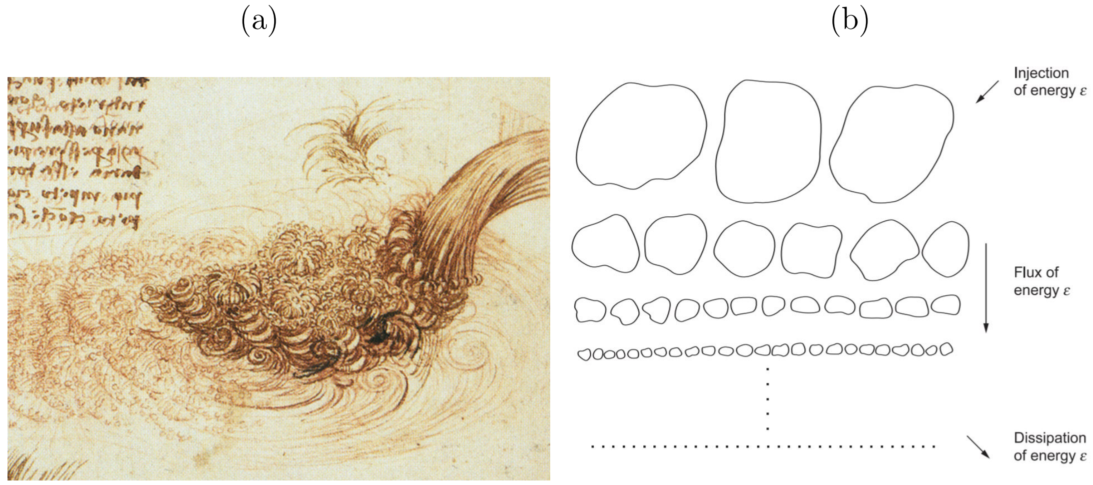

(a)
(b)
Injection of energy ε
Flux of energy ε
Dissipation of energy ε

Figure 1.2: (a) *"Observe the motion of the surface of water, which resembles the behavior of hair, which has two motions, of which one depend on the weight of the strands, the other on the line of its revolving; thus water makes revolving eddies, one part of which depends upon the impetus of the principal current, and the other depends on the incident and the reflected motions."* (Leonardo da Vinci 1513) As *Davidson* (2004) asks himself: "Did Leonardo da Vinci foresee Reynolds' idea of dividing a turbulent flow into two components: a mean velocity and the turbulent fluctuations?". (b) This figure represents a one-dimensional simplification of the cascade process in which bigger eddies becomes unstable and break up in smaller eddies till when the scale of the eddies matches the scale at which dissipation dominates the dynamics. ε is the kinetic energy dissipation rate determining the energy flux from large to small scales. This figure was modified from *Frisch* (1995).

ally simplified by local equilibrium assumptions and truncated at a certain order (e.g., *Simpson et al.*, 2000).

It is on some aspects of this last class of models that I focused part of my thesis. To implement this kind of models, the velocity in the NS equations is decomposed in mean and fluctuating part in order to express the viscous term as a function of the velocity. Nevertheless, once we time average the resulting equation in order to get rid of the fast fluctuating part, a nonlinear term containing the correlation of the velocity fluctuation does not vanish (e.g., *Davidson*, 2004). *Boussinesq* (1872) proposed to model this term

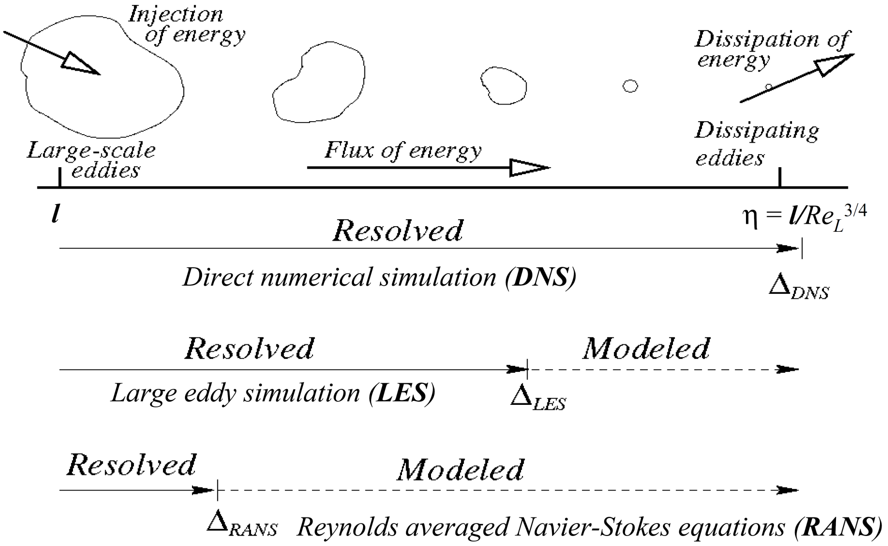

Injection of energy
Dissipation of energy
Large-scale eddies
Flux of energy
Dissipating eddies
l
η = l/Re_L^3/4
Resolved
Direct numerical simulation (DNS)
Δ_DNS
Resolved
Modeled
Large eddy simulation (LES)
Δ_LES
Resolved
Modeled
Δ_RANS
Reynolds averaged Navier-Stokes equations (RANS)

Figure 1.3: DNS: Very small scale flow (ex:turbulent boundary layers) but currently computationally intractable for most problems. LES: Aims to solve the computational cost that DNS poses and reveals the eddies hidden behind the mean in RANS. Good for small scale applications in both coastal and open ocean. RANS: The least computationally expensive method that is used for turbulent modeling. However, it is really not very good when certain phenomena cannot be averaged, such as instabilities. This figure is taken from André Bakker's lectures: `http://www.bakker.org/dartmouth06/engs150/10-rans.pdf`

in term of a turbulent viscosity so that the turbulence stresses can be related to the mean flow. This is still the modern approach to the problem. But we must notice that in this way we have just shifted the problem from modelling the velocity fluctuations to modelling the turbulent viscosity. In the literature, many different ways –in jargon called turbulence closure schemes (TCSs)– of doing this have been proposed. The most frequently found TCSs in the ocean modelling community's literature are the $k - k\ell$ by *Mellor et Yamada*, 1982; the $k - \varepsilon$ by *Rodi*, 1987; the $k - k\omega$ by *Wilcox*, 1988; the $k - \ell$ by *Gaspar et al.*, 1990 and the KPP by *Large et al.*, 1994.

Furthermore, different stability functions can be chosen in order to include the effect of the parametrized non-local moments and pressure strain correlations in the dynamical equations when employing certain TCSs (e.g., *Galperin et al.*, 1988; *Kantha et Clayson*, 1994; *Canuto et al.*, 2001). Also, additional complexity is added to the modelling by the interplay of the TCS and the choice of boundary conditions (b.c.). The choice of the surface and bottom boundary conditions can profit of a vast literature (e.g., *Craig et Banner*, 1994; *Stacey et Pond*, 1997; *Estournel et al.*, 2001; *Warner et al.*, 2005), aiming at modelling different forcing mechanisms.

The effectiveness of the different possible choices and of the different ways of implementing the RANS models really need to be experimentally verified. Thanks to the aforementioned theoretical results by *Batchelor* (1959), this has been made recently possible by microstructure shear and temperature probes. These lasts have permitted to develop an ever growing set of turbulence measurements to be compared to the models (see *Burchard et al.*, 2008, for a review). This is the approach that we used in this thesis relatively to the validation of the model Symphonie (*Marsaleix et al.*, 2008).

1.3 Bottom boundary turbulence and dispersion

Turbulence could also have an important role in the bottom boundary layer in determining the settling site of larvae. Moreover, deep ocean marine communities are currently viewed as a potential –if not the only one– candidate for sustaining the ones living in a coastal environment endangered by a variety of anthropogenic stressors. In order to establish the potential of the deep ocean communities to sustain the coastal ones through marine connectivity, we must improve our understanding of bottom boundary turbulence and develop new parameterizations of it in the numerical models used for the dispersion numerical experiments.

In fact, bottom boundary turbulence is an open field of research. Its relevance extends till to synoptic scales and turbulent mechanisms that were overlooked in the past and are gaining renewed attention since recent studies have highlighted the important role it can

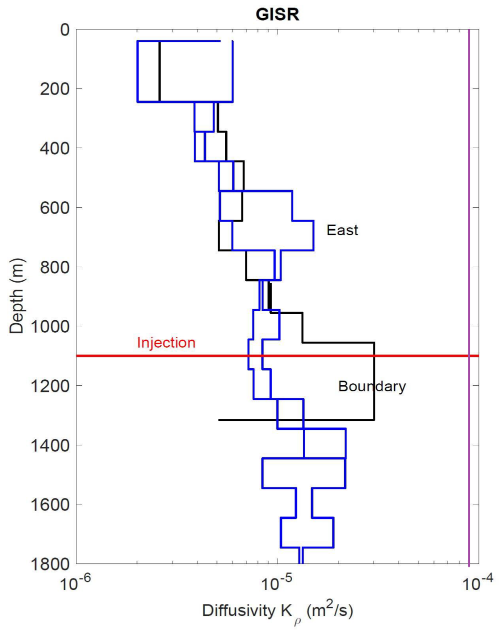


Figure 1.4: Diapycnal diffusivity K_ρ estimated by microstructure measurements in the Gulf of Mexico (blue and black) versus the average K_ρ inferred from tracer release data during the GISR campaign by *Ledwell et al.* (2016) (magenta). The red line shows the depth at which the tracer was initially released.

play in sustaining the global overturning circulation (*Naveira Garabato et al.*, 2004). More and more evidence is accumulating that a significant amount of mixing is generated in the bottom layer above rough topography (e.g., *Polzin et al.*, 1997; *Ledwell et al.*, 2016). Often, the differences between the amount of mixing that is needed to explain the ocean circulation and tracer vertical diffusion differs by more than one order of magnitude with respect to the amount of mixing measured with microstructure probes (Figure 1.4). However, the understanding of the processes generating near-bottom turbulence remain incomplete (*Polzin et al.*, Submitted) and the comprehension of how small scale processes translate into large-scale impacts on the fate and transport of pollutants in the ocean is still lacking. Moreover, microstructure probes are hardly deployed in the near bottom layer of the ocean (*Waterhouse et al.*, 2014). A major difficulty in studying these processes is their intrinsic patchiness and intermittency.

The principal physical mechanisms candidate to explain the mixing above topography are essentially all the possible interactions between ocean currents and internal wave field with the topography; principally lee waves and form drag. These kind of phenomena are well studied in the atmospheric community (*Palmer et al.*, 1986; *Alexander et al.*, 2010), but their comprehension is limited to the flow over isolated bumps (*Scinocca et McFarlane*, 2000). This is an idealization that is justified by the typical scales ratio between the atmospheric flows and land orography but is not the case of the oceanic flows above rough topography (*Polzin et al.*, Submitted). In the third part of my thesis I address the role of form drag as generating mechanism of bottom boundary mixing.

1.4 Aims and strategy

My thesis is divided in three parts. In Chapter 2, I present a new interpretation of graph theory in the framework of connectivity studies. My applied case study is the connectivity of a polychaete metapopulation in the Gulf of Lion (GoL). This is a benthic species that disperses mainly during its pelagic larval stage (that varies from 3 to 5 weeks). This characteristic of polychaete permits to model its larvae's dispersion as the dispersion of passive particles in a Lagrangian simulation. Note that the assumption of a passive dispersion is not completely realistic because polychaete's larvae display a diel vertical migration (e.g., *Hensler et Jude*, 2007). However, this behavior is not yet well understood and thus it is not considered by literature studies.

I then analyzed the connectivity matrices obtained with the Lagrangian simulations with different graph theory measures that in the literature are assumed to highlight reproductive sites important for persistence. Afterwards, I compared our results with other persistence and ecological studies in the literature. In particular, we exploited a metapopulation model study (*Guizien et al.*, 2014), a genetic structuring study (*Padron et Guizien*, 2015) and a sedimentary study (*Labrune et al.*, 2007) in the GoL. The rationale behind this comparison is that these studies have a much less debated interpretation than graph theory when studying connectivity. So that, if graph theory measures supposed to indicate sites important for persistence are in agreement with these precedent studies, I can conclude that the the current interpretation of graph theory is actually correct. Otherwise, I can obtain useful indications to start to re-think the interpretation of graph theory in the framework of connectivity studies.

In Chapter 3, I tried to answer the question about the limitations of current numerical schemes in modelling turbulence by comparing in situ data with the predictions made by the numerical model Symphonie when using different combinations of TCSs, boundary conditions and modeling of the light penetration. In particular, in situ measurements of kinetic energy dissipation rate ε and eddy viscosity K_Z were acquired with a Self-Contained Autonomous Microstructure Profiler (SCAMP) during various campaigns in the Gulf of Lion.

In Chapter 4, I focused on the bottom boundary layer mixing, studying a dataset issued from a tracer release experiment in the Gulf of Mexico (*Ledwell et al.*, 2016). These data highlighted the fact that a relevant portion of the observed diapycnal mixing takes place in the bottom boundary.

Following the indications of some high-resolution numerical simulations (Max Nikurashin personal communication, see Figure 1.5), I hypothesized that the mixing was generated by form drag due to the flow of deep currents above rough topography. In order to clarify the functional dependence of this mixing on various physical parameters, I proposed a functional dependence of the diapycnal diffusivity and implement a 2-D diffusion model that permits to vary the parameters until a best-agreement with the in situ data is achieved.

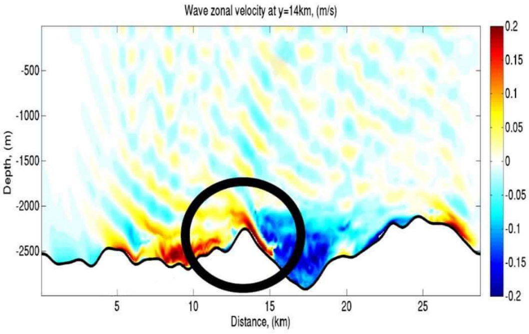


Figure 1.5: 2-D section of the zonal velocity field issued 3-D high resolution simulation. The black circle highlights a hydrostatic-like flow taking place in a 100m layer above rough topography. Figure provided by Max Nikurashin.

In particular, this approach demonstrates that bottom boundary mixing permits to close the mixing budget –the ratio of observed to required mixing ratio– that otherwise was off by more than an order of magnitude (Figure 1.4).

CHAPTER 2

Study of marine connectivity and marine populations persistence with graph theory

---

This work resulted in one publication : *On the Calculation of Betweenness Centrality in Marine Connectivity Studies Using Transfer Probabilities*, A. Costa, A.A. Petrenko, A.M. Doglioli, K. Guizien; PlosONE.
Another manuscript will be submitted to Ecological Modelling soon: *Tuning the interpretation of graph theory measures in analyzing marine larval connectivity*, A. Costa, A.M. Doglioli, K. Guizien, A.A. Petrenko.

Contents

2.1 On the Calculation of Betweenness Centrality in Marine Connectivity Studies Using Transfer Probabilities 16
2.1.1 Abstract 16
2.1.2 Introduction 17
2.1.3 Materials and Methods 18
2.1.4 Results 21
2.1.5 Conclusion 22
2.2 Tuning the interpretation of graph theory measures in analyzing marine larval connectivity: The Gulf of Lion study case. 25

2.2.1 Introduction 25
2.2.2 Materials and Methods 27
2.2.3 Results 39
2.2.4 Discussion and conclusions 43
2.3 Extra Material 47
2.3.1 Bridging centrality 47
2.3.2 Visual explanation of graph theory measures 50

Graph theory is a mathematical theory that idealizes a network of elements linked by pairwise relations, by representing the elements as nodes and the connections as links between the nodes. As I will show in more detail later, graph theory offers a variety of tools in order to identify nodes playing specific roles in the network under study.

A huge advantage of graph theory is that it is extremely flexible and can be adapted to a plethora of different problems by just redefining in a convenient way the value of the links (e.g., *Barrat et al.* 2008; *Barbasi* 2016). For this reason it is more and more applied in connectivity studies (*Moilanen*, 2011). However, its application is not straightforward and the scientific community still does not agree on how to interpret the results of the analysis of ecological networks with graph theory.

The first part of this chapter reports a study I conducted on developing a metric to quantify the distance between nodes when using connectivity matrices containing transfer probabilities. In particular, I point to two previous literature studies that employed incorrect metrics falsifying the physical interpretation of graph theory's results. Moreover, I propose a new metric that permits to correctly apply graph theory to connectivity matrices storing transfer probabilities.

In these two literature papers, the authors quantify the distance between nodes with the transfer probability between those two nodes. However, such a method does not allow to correctly calculate the probability of transfer between nodes more than one step apart. This latter probability is at the base of many graph theory measures (c.f. Section 2.3).

The error cannot be avoided without modifying the way of quantifying the distance between nodes. In fact, all the algorithms that permit to find the most probable path between a pair of nodes, calculate the value of the path by summing the values of the connections between nodes and look for the shortest one. However, when the connections are defined in term of probabilities, the path with the minimum values results from the sum of the lower probabilities. As it will be shown in more detail later, this invalidates the interpretation of different graph theory measures because these measures typically are interested in highlighting the most probable paths. Moreover, summing probabilities is a correct procedure when we consider the total probability of independent events. However, this is not the case of larval transfers between distant sites. Both the above problems can be avoided by carefully define a new metric to define the distance between nodes.

The second part of this chapter reports a study I am conducting in order to clarify the interpretation of graph theory measures in the context of marine connectivity and marine species persistence studies. The study takes advantage of the comparison of a metapopulation of benthic polychaete in the Gulf of Lion with graph theory. The data analysis is still being developed. Therefore, the results presented here are still temporary as well as the version of the manuscript that I report here. I will incorporate the final version as soon as it will be completed. We envision to submit the final version of the manuscript to Ecological Modeling.

2.1 On the Calculation of Betweenness Centrality in Marine Connectivity Studies Using Transfer Probabilities

2.1.1 Abstract

Herein we highlight a lack of methodological information we encountered in two literature papers concerning the application of graph theory to marine connectivity studies. We show the consequences of a possible error in those papers and propose the use of a metric for the node-to-node distance that solves the inconsistency at the base of it. Our argumentation is illustrated by the analysis of a literature data set.

2.1.2 Introduction

In the last decade, graph theory has increasingly been used in ecology and conservation studies *Moilanen* (2011) and particularly in marine connectivity studies (e.g., *Treml et al.* 2008; *Kininmonth et al.* 2010b; *Kininmonth et al.* 2010a; *Andrello et al.* 2013; *Rossi et al.* 2014). Graphs are a mathematical representation of a network of entities (called nodes) linked by pairwise relationships (called edges). Graph theory is a set of mathematical results that permit to calculate different measures to identify nodes, or set of nodes, that play specific roles in a graph. Graph theory application to the study of marine connectivity typically consists in the representation of portions of sea as nodes. Then, the edges between these nodes represent transfer probabilities between these different portions of sea.

Transfer probabilities estimate the physical dispersion of propagulae (*Jacobi et al.* 2012; *Andrello et al.* 2013; *Berline et al.* 2014; *Jonsson et al.* 2015), nutrients or pollutants (*Doglioli et al.*, 2004), particulate matter (*Mansui et al.*, 2015), or other particles either passive or interacting with the environment (see *Ghezzo et al.* 2015; *Bacher et al.* 2016 and references therein). As a result, graph theory already proved valuable in the identification of hydrodynamical provinces (*Rossi et al.*, 2014), genetic stepping stones (*Rozenfeld et al.*, 2008), genetic communities (*Kininmonth et al.*, 2010a), sub-populations (*Jacobi et al.*, 2012), and in assessing Marine Protected Areas (MPAs) connectivity (*Andrello et al.*, 2013).

In many marine connectivity studies, it is of interest to identify specific portions of the sea where a relevant amount of the transfer across the graph passes through. A well-known graph theory measure is frequently used for this purpose: betweenness centrality. In the literature, high values of this measure are commonly assumed to identify nodes sustaining the connectivity of the whole network. For this reason betweenness has been used in the framework of marine connectivity in order to identify migration stepping stones (*Treml et al.*, 2008), genetic gateways (*Rozenfeld et al.*, 2008), and MPAs ensuring a good MPAs connectivity (*Andrello et al.*, 2013).

Our scope in the present letter is to highlight some errors that can occur in implementing graph theory analysis. Especially we focus on the definition of edges when one is interested in calculating the betweenness centrality and other related measures. We also point out two papers in the literature in which this methodological inconsistency can be found: *Kininmonth et al.* (2010b) and *Andrello et al.* (2013).

In Materials and Methods, we introduce the essential graph theory concepts for our scope. In Results, we present our argument on the base of the analysis of a literature data set. In the last section, we draw our conclusions and provide perspectives.

2.1.3 Materials and Methods

A graph $\mathcal{G}$ is a couple of sets (V, E), where V is the set of nodes and E is the set of edges. The set V represents the collection of objects under study that are pair-wise linked by an edge a_{ij}, with $(i, j) \in V$, representing a relation of interest between two of these objects. If $a_{ij} = a_{ji}, \forall (i, j) \in V$, the graph is said to be 'undirected', otherwise it is 'directed'. The second case is the one we deal with when studying marine connectivity, where the edges values represent the transfer probabilities between two marine zones (e.g., *Kininmonth et al.* 2010a; *Kininmonth et al.* 2010b; *Andrello et al.* 2013; *Rossi et al.* 2014).

In a graph, there can be multiple ways (called paths) to go from a node i to a node j passing by other nodes. The value of a path is the sum of the weights of the edges composing the path itself (e.g., *Bondy et Murty* 1976). In general, it is of interest to know the shortest or fastest path σ_{ij} between two nodes, i.e. the one with the lowest value. But it is even more instructive to know which nodes participate to the greater numbers of shortest paths. This can be known by calculating the betweenness value for each node in the graph. The betweenness value of a node k, $BC(k)$, is defined as the fraction of shortest paths existing in the graph, σ_{ij}, with $i \neq j$, that effectively pass through k, $\sigma_{ij}(k)$, with $i \neq j \neq k$:

$$BC(k) = \sum_{i \neq k \neq j} \frac{\sigma_{ij}(k)}{\sigma_{ij}} \tag{2.1}$$

with $(i, j, k) \in V$. Betweenness is then normalized by the total number of possible connections in the graph once excluded node k: $(N-1)(N-2)$, where N is the number of nodes in the graph, so that $0 \leq BC \leq 1$.

Although betweenness interpretation is seemingly straightforward, one must be careful in its calculation. In fact betweenness interpretation is sensitive to the node-to-node metric one chooses to use as edge weight. If, as frequently the case of the marine connectivity studies (e.g., *Kininmonth et al.* 2010a; *Kininmonth et al.* 2010b; *Andrello et al.* 2013; *Rossi et al.* 2014), one uses transfer probabilities as edge weight, betweenness loses its original meaning. Based on the details given in their Methods sections, this seems to be the case of *Kininmonth et al.* (2010b) and *Andrello et al.* (2013). In this case, edge weight would decrease when probability decreases and the shortest paths would be the sum of edges with lowest value of transfer probability. As a consequence, high betweenness would be associated to the nodes through which a high number of improbable paths pass through. Exactly the opposite of betweenness original purpose.

We propose to solve this inconsistency by using a new metric for the edge weights. It is still based on transfer probabilities a_{ij}, but it transforms it in order to conserve the original meaning of betweenness, by ensuring that a larger transfer probability between two nodes corresponds to a smaller node-to-node distance. Therefore, the shortest path between two nodes effectively is the most probable one. So that high betweenness is associated to the nodes through which a high number of probable paths pass through.

In the first place, in defining the new metric, we need to reverse the order of the probabilities in order to have higher values of the old metric a_{ij} correspond to lower values of the new one. In the second place we also consider three other facts: (i) transfer probabilities a_{ij} are commonly calculated with regards to the position of the particles only at the beginning and at the end of the advection period; (ii) the probability to go from i to j does not depend on the node the particle is coming from before arriving in i; and (iii) the calculation of the shortest paths implies the summation of a variable number of transfer probability values. Note that, as the a_{ij} values are typically calculated on the base of the particles' positions at the beginning and at the end of a spawning period, we are dealing with paths whose values are calculated taking into account different numbers of generations. Therefore, the transfer probabilities between sites are independent from each other and should be multiplied by each other when calculating the value of a path. Nevertheless, the classical algorithms commonly used in graph theory analysis calculate the shortest paths as the summation of the edges composing them (e.g., the Dijkstra algorithm, *Djikstra* 1959 or *Brandes* 2006). Therefore, these algorithms, if directly applied to the probabilities at play here, are incompatible with their independence.

As a consequence, we define the weight of an edge between two nodes i and j as:

$$d_{ij} = \ln\left(\frac{1}{a_{ij}}\right) \tag{2.2}$$

This definition is the composition of two functions: $h(x) = 1/x$ and $f(x) = \ln(x)$. The use of $h(x)$ allows one to reverse the ordering of the metric in order to make the most probable path the shortest. The use of $f(x)$, thanks to the basic property of logarithms, allows the use of classical shortest-path finding algorithms while dealing correctly with the independence of the connectivity values. In fact, we are *de facto* calculating the value of a path as the product of the values of its edges.

It is worth mentioning that the values $d_{ij} = \infty$, coming from the values $a_{ij} = 0$, do not influence the calculation of betweenness values via the Dijkstra algorithm. Note that d_{ij} is both homogeneous: $\alpha d_{ij} = \ln\left(\frac{1}{a_{ij}^{\alpha}}\right)$, for any constant α and any $(i, j) \in V$; and additive:

$d_{il} + d_{lj} = \ln\left(\frac{1}{a_{il} \cdot a_{lj}}\right) = \ln\left(\frac{1}{a_{ij}}\right) = d_{ij}$, for any $(i, l, j) \in V$. So that it has the physical properties of a distance. Also, note that both a_{ij} and d_{ij} are both dimensionless.
We explicitly note that, a priori, other metrics permit to make the weight decrease when probability increases: $1 - a_{ij}$, $1/a_{ij}$, $\log(1 - a_{ij})$ and $-a_{ij}$ are the most intuitive. However, with some simple examples it is possible to show that: $1 - a_{ij}$ is biased towards paths with fewer edges; $1/a_{ij}$ can fail when calculating shortest paths as sum of their edges; and $\log(1 - a_{ij})$ and $-a_{ij}$ fail when there are negative cycles in the graph. See XXX Appendix A or the Supporting Information for a complete discussion of these metrics.

2.1.4 Results

The consequences of the use of the raw transfer probability (a_{ij}) rather than the distance we propose (d_{ij}) are radical. To show this, we used 20 connectivity matrices calculated for *Guizien et al.* (2014). They were calculated from Lagrangian simulations using a 3D circulation model (see *Marsaleix et al.* 2006) with a horizontal resolution of 750 m. Spawning was simulated by releasing 30 particles in the center of each of 32 reproductive sites (hereafter identified as nodes) for benthic polychaetaes alongshore the Gulf of Lion (NW Mediterranean Sea), on the 30 m isobath, every hour from January 5 until April 13 in 2004 and 2006. The proportions of particles coming from an origin node and arriving at a settlement node after 3, 4 and 5 weeks were weight-averaged to compute a connectivity matrix for larvae with a competency period extending from 3 to 5 weeks.

As an example, in Fig 2.1 we show the representation of the graph corresponding to the 7^{th} of the 20 connectivity matrices. The arrows starting from a node i and ending in a node j represent the direction of the element a_{ij} (in Fig 2.1a) or d_{ij} (in Fig 2.1b). The arrows' color code represents the magnitude of the edges' values. The nodes' color code indicates the betweenness values calculated using the metric a_{ij} (in Fig 2.1a) or d_{ij} (in Fig 2.1b).

In Fig 2.1a the edges corresponding to the lower 5% of the weights a_{ij} are represented. These are the larval transfers that, though improbable, are the most influential in determining high betweenness values when using a_{ij} as metric. In Fig 2.1b the edges corresponding to the lower 5% of the weights d_{ij} are represented. These are the most probable larval transfers that —correctly— are the most influential in determining high betweenness values when using d_{ij} as metric. While in figure Fig 2.1a the nodes with highest betweenness are the nodes 31 (0.26), 27 (0.25) and 2 (0.21); in Fig 2.1b the nodes with highest betweenness are nodes 21 (0.33), 20 (0.03) and 29 (0.03).

As we show in Fig 2.2, betweenness values of the 32 nodes calculated using the two node-to-node distances a_{ij} and $\log(1/a_{ij})$ are drastically different between each other. Moreover, in 10 out of 20 connectivity matrices, the correlation between node ranking based on betweenness values with the two metrics were not significant (p-value > 0.05). In the 10 cases it was (p-value < 0.05), the correlation coefficient was lower than 0.6 (data not shown). Such partial correlation is not unexpected as the betweenness of a node with a lot of connections could be similar calculated with a_{ij} or d_{ij} if among these connections there are both very improbable and highly probable ones, like in node 21 in the present test case. Furthermore, it is noticeable that if one uses the a_{ij} values (Fig 2.2a), the betweenness values are much more variable than the ones obtained using d_{ji} (Fig 2.2b). This is because, in the first case, the results depend on the most improbable connections that, in the ocean, are likely to be numerous and unsteady.

2.1.5 Conclusion

We highlighted a lack of methodological information in the works by *Kininmonth et al.* (2010b) and *Andrello et al.* (2013), applying graph theory to marine connectivity study. The methodological information in those work suggest that the authors did not reverse the probability when calculating shortest paths. Such an error would lead to consider the most improbable paths as the most probable ones. We showed the potential consequences

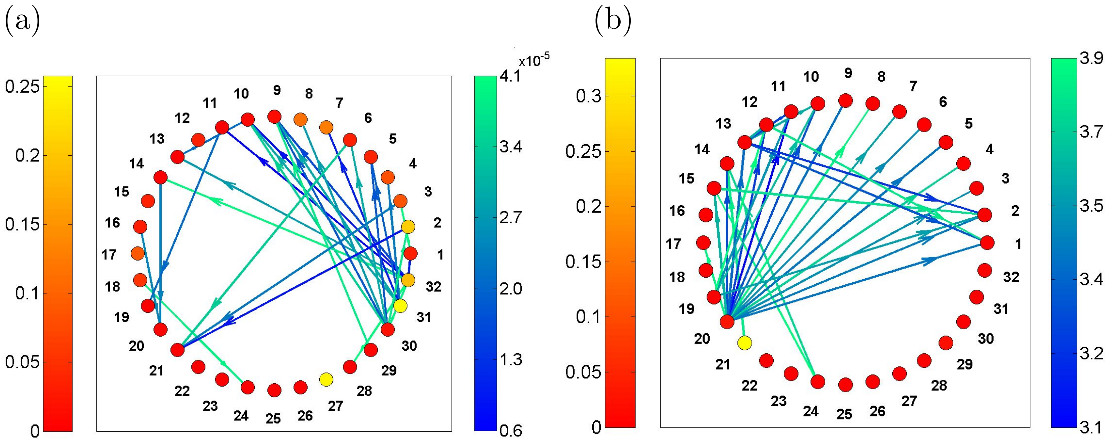

(a)
(b)
0.25
0.2
0.15
0.1
0.05
0
x10^-5
4.1
3.4
2.7
2.0
1.3
0.6
0.3
0.25
0.2
0.15
0.1
0.05
0
3.9
3.7
3.5
3.4
3.2
3.1
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32

Figure 2.1: Representation of the connectivity matrix 7 and corresponding betweenness values using different metrics. The left side colorbars indicates betweenness values, the right side colorbars indicate the metric values. (a) Results obtained by using a_{ij} as edge weight, (b) results obtained by using d_{ij} as edge weight. In (a) the lowest 5% of edges weights are represented. In (b) the lowest 5% of edges weights are represented. Note the change in the colorbars' ranges.

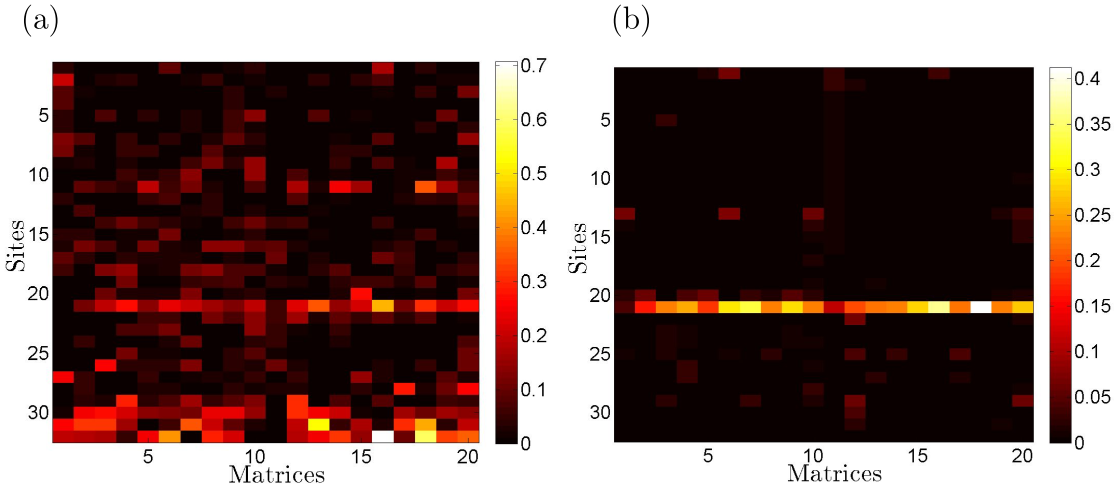


Figure 2.2: Betweenness nodes in the Gulf of Lion for the 32 sites using 20 different connectivity matrices obtained with Lagrangian simulations by *Guizien et al.* (2014). (a) Results obtained by using a_{ij} as edge weight; (b) results obtained by using d_{ij}. Note the change in the colorbars' ranges.

of this methodological error by analyzing a published data set of connectivity matrices for the Gulf of Lion *Guizien et al.* (2014). Therefore, we invite the reader to reconsider the part of methodology regarding the betweenness calculation proposed by *Kininmonth et al.* (2010b) and *Andrello et al.* (2013) and consequent results. In particular, the conclusions drawn by the authors about the importance of specific MPAs in maintaining the connectivity across the MPAs network in the Mediterranean Sea need to be revisited in order not to mislead eventual conservation policies.

Accordingly to these results, we proposed the use of a node-to-node metric for graph theory studies on marine connectivity that solves the inconsistency and provides a meaningful way to calculate shortest paths and —as a consequence— betweenness, when relying on transfer probabilities issued from Lagrangian simulations. In particular, this new metric permits to reverse the probability and to calculate the value of a path as the product of its edges. Moreover, this metric is not limited to the calculation of betweenness alone but is also valid for the calculation of every graph theory measure related to the concept of shortest paths: for example, shortest cycles and bridging centrality *Costa et al.* (In Preparation).

2.2 Tuning the interpretation of graph theory measures in analyzing marine larval connectivity: The Gulf of Lion study case.

2.2.1 Introduction

In the ocean, habitat fragmentation leads to the geographical separation of marine populations. However, most marine species exhibit a dispersive larval stage (*Horson*, 1946) which ensures the maintenance of the spatial coverage of these populations. This is made possible by the exchange of individuals during reproduction events, forming a so-called metapopulation (*Hanski*, 1999). Nowadays, oceans are increasingly threatened by multiple anthropogenic stressors (*Lubchenco et al.*, 1999), among which the increase of fragmentation due to habitat destruction. This process alters the connectivity within a metapopulation and may endanger the persistence of the latter. Indeed, connectivity is the mechanism that permits to a spatially isolated population to resist to local perturbations through the supply of individuals from other populations when self-recruitment is depleted. Moreover, connectivity promotes gene flow to balance genetic drift and avoid inbreeding (*Wright*, 1931). It follows that, setting up adequate spatial planning policies (*Boersma et Parrish*, 1999), with minimal impact on marine species persistence, should include connectivity assessment.

Consequently, in recent years, different methodologies have been developed for tackling marine connectivity assessment; each with their own advantages and shortcomings (see *Kool et al.*, 2013 and *Lagabrielle et al.*, 2014). Among these methodologies, graph theory has been increasingly employed due to its adaptability to different contexts (*Moilanen*, 2011). For example, it was applied to the study of landscape connectivity (*Urban et Keitt*, 2001); to the study of the connection between marine reefs (*Treml et al.*, 2008); to infer gene flux in marine populations networks (*Rozenfeld et al.*, 2008); to examine network topology of species (*Kininmonth et al.*, 2010b); for the identification of marine sub-populations (*Jacobi et al.*, 2012); and for the estimation of connectivity among marine Marine Protected Areas (*Andrello et al.*, 2013).

Moreover, in connectivity studies many different graph theory measures have been proposed to identify the sites playing certain functions within an ecological network (*Rayfield et al.*, 2010; and *Galpern et al.*, 2011). In the literature, some graph theory measures highlighting sites ensuring a large transfer in ecological networks have been proposed as relevant for conservation (e.g., *Treml et al.*, 2008; *Rozenfeld et al.*, 2008; *Andrello et al.*, 2013). However, this point is not bereft of controversies. For example, the equivalence between large transfer and conservation relevance remains unproven (*sensu Moilanen*, 2011; and *Lagabrielle et al.*, 2014).

Herein we want to reexamine the ecological interpretation of some graph theory measures in the context of marine metapopulations of three species of benthic invertebrate (sedentary adult stage with dispersion during the larval stage). For doing this, graph theory measures are compared with results of existing ecological studies assessing the regional persistence (*Guizien et al.*, 2014), the genetic structuring (*Padron et Guizien*, 2015) and regional assemblages structure of soft-bottom benthic invertebrates in the Gulf of Lion (*Labrune et al.*, 2007). Moreover, we extend the analysis in *Guizien et al.* (2014) to the multiple-species case.
The goal is to assess which of the graph theory measures that we consider in this study is in better agreement with the information provided by these studies. As a result, the conservation relevance of the different graph theory measures will be clarified.

The Gulf of Lion was selected because of the numerous studies, both physical and biological, already performed in this area that can be used to interpret and validate our results. In the 10 to 30 m bathymetric depth range, the GoL is characterized by a large continental margin dominated by soft-bottoms forming a continuous habitat for soft-bottom invertebrates. Its hydrodynamics is complex and highly variable (*Millot*, 1990). Depending on wind forcing, currents in the study zone can be either eastward or southwestward (*Estournel et al.*, 2003; *Petrenko et al.*, 2008). The circulation is strongly influenced by the Northern Current, which constitutes an effective dynamical barrier blocking coastal waters on the continental shelf (*Petrenko*, 2003) and delimits the regional scale of hydrodynamic connectivity. Exchanges between the GoL and offshore waters are mainly induced by processes associated with the Northern Current (*Petrenko et al.*, 2005). Moreover the recent study by *Rossi et al.* (2014) supports the choice of spatial scales of the size of the GoL as an appropriate closure scale for a metapopulation, meaning that the GoL populations are not fueled by populations outside the GoL. In fact, for month-long periods of larval dispersal, larval transfer was higher within the GoL that with areas outside it (*Rossi et al.*, 2014).

The paper is organized as follows. In the Materials section, we discuss the main characteristics of the graph theory input: 60 connectivity matrices issued from Lagrangian dispersal simulations. In the Procedures section, we detail the graph theory measures we tested. In the Assessment section, we present the systematic analysis of the hydrological connectivity matrices with graph theory measures. The graph theory measures' conservation meaning that descends from the comparison with previous ecological studies is finally discussed.

2.2.2 Materials and Methods

The common input to metapopulation model and graph theory analyses are the connectivity matrices specifying the larval transfer probabilities between 32 reproductive sites along the shore of the Gulf of Lion (Figure 2.3). These sites cover a substantial part of the available habitat in the GoL for soft-bottom benthic species *Aloisi et al.* (1973). Larval transfer probabilities were determined by means of Lagrangian simulations, that

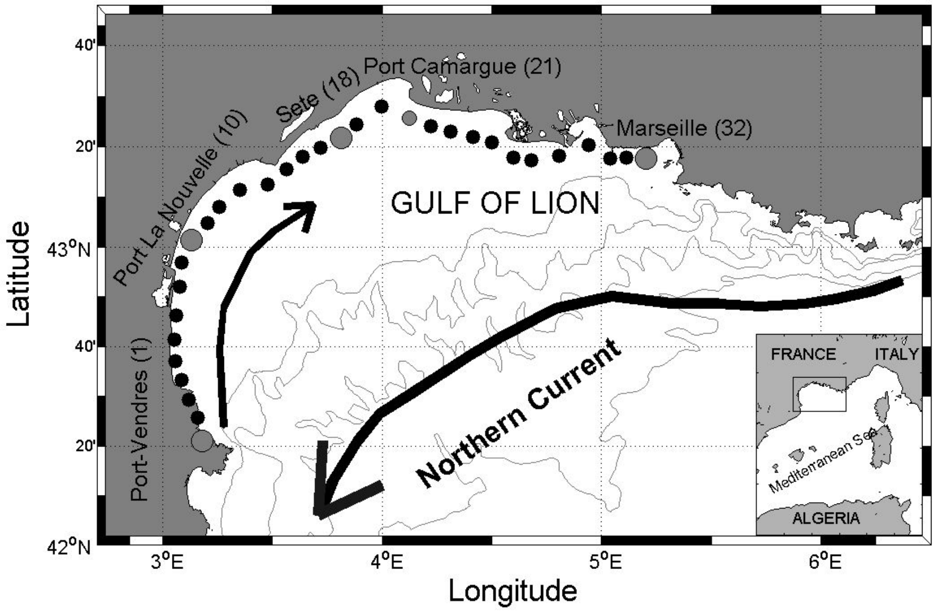

Port Camargue (21)
Sete (18)
Port La Nouvelle (10)
Marseille (32)
GULF OF LION
Port-Vendres (1)
Northern Current
FRANCE
ITALY
Mediterranean Sea
ALGERIA
40'
20'
43°N
40'
20'
42°N
3°E
4°E
5°E
6°E
Latitude
Longitude

Figure 2.3: Schematic representation of the typical circulation in the Gulf of Lion. The thick arrow represents the dominant alongshore Northern Current. The thinner arrow represents the eastward current that can be detected in stratified conditions or under particular wind field conditions. The positions of the 32 studied sites are plotted. The sites 1, 10, 18 and 32, used for the habitat loss scenario in *Guizien et al.* (2014), are highlighted by bigger gray dots. Node 21 is the smallest of the gray dots. The gray lines correspond to the 100 m, 200 m, 1000 m and 2000 m isobaths.

were obtained using a 3-D circulation model (see *Marsaleix et al.*, 2006) with a horizontal resolution of 750 m. Spawning was simulated by releasing 30 particles in the center of each of 32 sites alongshore the GoL, on the 30 m isobath, every hour from January 5 until May 16 in 2004 and 2006. The final positions of larvae after 3, 4 and 5 weeks (corresponding to the pelagic larval period) were processed to compute the proportion of the larvae coming from an origin site and arriving at a settlement site. Connectivity matrices were then built for ten consecutive 10-day spawning periods in each year and for each of the three different pelagic larval periods, for a total of 60 connectivity matrices.

2.2.2.1 Metapopulation Model

The spatio-temporal dynamics of benthic invertebrate populations with a dispersive pelagic larval stage were simulated with the spatially explicit metapopulation model of *Guizien et al.* (2014). The model describes explicitly, in discrete times and for a set of patches with different areas, the spatial density dynamics of only the sedentary adult stage resulting from local survivorship, and reproductive input potentially contributed by all the other patches forming the metapopulation. Reproductive input results from offspring production rate in source patches, its redistribution among the different patches through a larval flow connectivity matrix and recruitment success in destination patches, the latter being regulated by intra-specific competition for space.

The model was parameterized to describe the spatio-temporal dynamics of a 2-year life expectancy species dwelling in the soft-bottom habitat of the Gulf of Lion. The species was assumed to reproduce once a year with a dispersive pelagic larval phase lasting from 3 to 5 weeks. Such trait correspond to *Ditrupa arietina*, the dominant polychaetes species of the infralittoral soft-bottom benthic communities of the gulf (*Labrune et al.*, 2007). In the simulations, demographic parameters were assumed to be spatially uniform for the 10-30 m depth range to more clearly reveal the effect of connectivity on spatial structure. Propagule production rate was set to 500 and site saturating density was set to 25,000 ind m^{-2} Recruitment success was also considered to be spatially uniform, and used to quantify species persistence to metapopulation connections breakdown.

The continuous sand facies bed spanning the 10 to 30 m isobaths along the gulf was discretized into thirty-two contiguous sites as in *Guizien et al.* (2014) to describe larval transfer within the infralittoral soft-bottom habitat of the gulf. Connectivity matrices, with local retention rates on the diagonal, and transfer rates between distant sites off of the diagonal, were quantified from Lagrangian larval dispersal simulations. Simulations of larval dispersal in the gulf used three-dimensional currents and turbulent kinetic energy computed at a resolution of 750 m × 750 m following interpolation between the hourly outputs (*Guizien et al.*, 2006). Spawning was simulated by releasing thirty neutrally buoyant larvae in the center of each of the 32 sites, on the 20 m isobath, every hour during the reproductive season. As *Ditrupa arietina* reproduce in late winter-early spring in the NW Mediterranean Sea (*Charles et al.*, 2003), a 100-day reproductive season (from January 5 at 0h until April 13 at 23h) was considered and divided into ten consecutive 10-day spawning events in 2004 and 2006 (see details in *Guizien et al.*, 2012). In the absence of knowledge regarding sinking or swimming behaviour of the trocophore *Ditrupa arietina* larvae (no shell), larvae were considered neutrally buoyant. Three connectivity matrices corresponding to the three different pelagic larval durations (PLD) were built for each spawning period, and the proportion of larvae coming from the center of any site and arriving in each of the thirty-two sites was multiplied by the ratio of the areas of origin to the destination sites.

Three groups of simulations of *Ditrupa arietina* population density spatial distribution over a duration of 50 life-expectancy cycles were carried out, starting with a low uniform spatial density of 0.4% of the site's carrying capacity. Asymptotic stable equilibrium was defined as the average of the population density between the 45 and 50 life-expectancy cycles. Connectivity was either set deterministically (the same matrix repeated every year) or stochastically (a random matrix picked up every year among the twenty variants). When connectivity was set stochastically, ensembles of 500 simulations were performed for each set of the other parameters in order to evaluate the variability arising from the connectivity sequence. The average regional coverage was defined as the ratio between the regional average of population density and the saturating capacity.

Simulation settings are summarized in Table 2.1.

Simulations group 1 explored the resistance of a *Ditrupa arietina* to the removal of sets contiguous patches among the 32 patches forming the metapopulation for each of the 20 different connectivity variants. For each connectivity variant, the recruitment success value corresponding to the limit for regional persistence of *Ditrupa arietina* in the full network i.e. with all 32 habitat patches was determined by varying recruitment success from 10^{-6} to 1 (10 regular steps within each $\log_{10}$-interval), similarly as in (*Guizien et al.*, 2012). Then, applying this recruitment success value uniformly in all patches, the procedure consists in gradually increasing the number of patches removed from the network (lost habitat patches) starting from any location in the metapopulation. For each of the 32 initial patches, the minimum number of lost habitat patches around it causing average regional coverage at equilibrium of a 2-year lifespan species pass below 0.4% was detected.

Simulations group 2 explored the resistance of *Ditrupa arietina* to the removal of one patch among the 32 patches forming the metapopulation considering the inter-annual variability in connectivity patterns. For each patch removed, the recruitment success value was varied from 1.8% to 3% by step of 0.05%. The largest recruitment success value for which the average regional coverage at equilibrium of the ensemble of 500 simulations yielded pass below 0.4% (species disappearance) was seeked for.It is used to rank patches importance in *Ditrupa arietina* regional persistence: the larger the recruitment success, the more important the site is.

Simulations group 3 explored the resistance of a *Ditrupa arietina* to the removal of sets contiguous patches among the 32 patches forming the metapopulation, similarly to simulations group 1, except considering the inter-annual variability in connectivity patterns. Thus, connectivity was no more the same every year but varied randomly from year to year (stochastic connectivity case). The fixed and uniform recruitment success of

3%, that was determined as the limit for regional persistence of *Ditrupa arietina* in the 32 patches network (*Guizien et al.*, 2012) was applied. The same procedure as in simulations group 1 was applied that is gradually increasing the number of patches removed from the network (lost habitat patches) starting from any location in the metapopulation. For each of the 32 initial patches, the minimum number of lost habitat patches around it causing average regional coverage at equilibrium of a 2-year lifespan species pass below 0.4% in 10%, 50% and 90% of 500 simulations were seeked for.

	Group 1	Group 2	Group 3
Connectivity	D	S	S
Propagule production rate	500	500	500
Life expectancy	2 years	2 years	2 years
Disturbance	varying number of habitat loss	one habitat loss	varying number of habitat loss
Number of simulations	20	500	500
Recruitment success	fixed to persistence threshold with all habitat patches	[1.8:0.05:3]%	3%

Table 2.1: *Settings for the three simulation groups.*

2.2.2.2 Graph Theory

A graph $\mathcal{G}$ is a couple of sets (V, E) where V is the set of nodes and E is the set of edges. The set V represents the collection of objects under study that are pair-wise linked by an edge representing a relation of interest between these two objects. When the relation is symmetric, the graph is said to be 'undirected', otherwise it is 'directed'. An example of an undirected graph in the context of ecological networks is the genetic distance among populations used in *Rozenfeld et al.* (2008), while an example of directed graph is the probability of connections due to the current field between two Marine Protected Areas as in *Andrello et al.* (2013). If every existing edge has the same importance as the others, the graph is said to be 'binary' (e.g, *Rozenfeld et al.*, 2008): the edges can exist or not. If each edge has a specific relative importance, a weight can be associated to each of them and the graph will then be called 'weighted' (e.g, *Andrello et al.*, 2013). In the present study, we deal with directed weighted graphs. The nodes of our graphs represent the 32 sites in the GoL. Thus, henceforth we will only use the term ‘sites’. The edges of our graphs represent the not null probability that a Lagrangian particle released in one site is transported to another site, after a certain amount of time corresponding to its larval duration period. It follows that the edge weight quantifies the amount of larvae exchanged between two sites in one generation.

Typically, one is interested to identify the nodes that are more important in a network. This can be done by calculating different so-called centrality measures, each targeting nodes with different characteristics.
A simple centrality measures is strength centrality. The total weight of the connections of a node $i \in V$ is called strength k_i. In a directed graph, it is possible to distinguish between in-strength, out-strength and total strength. The first one is the sum of the values of the edges terminating in the node: $k_i^{in} = \sum_j a_{ji}$ with $j \in V$ and $i \neq j$, where the values a_{ij} are the terms of the connectivity matrix A. The second one is the sum of the values of the edges starting from the node: $k_i^{out} = \sum_j a_{ij}$. The total strength of a node i is $k_i = k_i^{in} + k_i^{out}$. Nodes with high strength values are said to have high strength centrality. In classical graph theory studies, . In connectivity studies,.. andrello or kinni.

A more refined version of the same concept is eigenvector centrality. Where strength centrality gives a simple count of the intensity of connections a site has, eigenvector centrality is based on the idea that not all connections are equal. In general, connections to sites who are themselves important will give to a site more influence than connections to less important nodes. If the centrality of a site i is x_i, we can allow for this effect by prescribing x_i to be proportional to the average of the centralities of i's network neighbors:

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{N} a_{ij} x_j$$

where λ is a constant. Defining the vector of centralities $\mathbf{x} = (x_1, x_2, \dots)$, we can rewrite this equation in matrix form as

$$\lambda \mathbf{x} = A\mathbf{x}$$

Hence, **x** is an eigenvector of the adjacency matrix with eigenvalue λ. In order to have **x** positive, λ must be the largest eigenvalue of the adjacency matrix and **x** the corresponding eigenvector. The eigenvector centrality defined in this way accords each site a centrality that depends both on the number and the intensity of its connections: having a large number of connections is still relevant, but a site with a smaller number of high-quality contacts may be more important than one with a larger number of middling contacts. In classical graph theory studies, eigenvector centrality has been used to identify the most influential people in a social network, key nodes in the Internet, electrical or urban networks, and diseases super-spreaders (see *Barrat et al.*, 2008, and references therein). Also, a modified eigenvector centrality is employed by web search engines (*Altman et Tennenholtz*, 2005).
In connectivity studies, *Andrello et al.* (2014) uses eigenvector centrality to identify sites contributing the most to population persistence.

The k-core is the largest subnetwork comprising nodes of degree at least k. A vertex i has coreness c if it belongs to a c-core in which all sites have strength at least equal to c but not to any $(c + \Delta c)$-core in which sites have strength at least equal to $c + \Delta c$.
The concept of a k-core was introduced to study the clustering structure of social networks (*Seidman*, 1983); to describe the evolution of random graphs (*Bollobás*, 1984); and to identify large functional assemblages in proteins (*Bader et Hogue*, 2003).

The clustering coefficient CC, as defined by *Fagiolo* (2007), is the the geometric mean of all triangular linkages within the set of neighbors for a selected site divided by the maximum possible neighbor linkages that the site could possibly form. For a directed weighted graph the clustering coefficient of a node i, $CC(i)$, can be written as:

$$CC(i) = \frac{\left[A^{1/3} + (A^T)^{1/3}\right]^{1/3}_{ii}}{2\left[k_i(k_i - 1) - 2f_i\right]} \tag{2.3}$$

where A^T is the transpose of the connectivity matrix and d_i counts the pair of edges both pointing to the same node (e.g., $i \leftarrow j$ and $i \leftarrow j$) that does not generate triangles:

$f_i = \sum_{j \neq i} a_{ij} a_{ji}$.
Intuitively, the clustering coefficient measures the local cohesiveness of the network. In classical graph theory studies, CC is normally used to infer the behavior of sites with certain CC characteristics (e.g., trading countries, *Fagiolo*, 2007). In marine connectivity it has been proven useful in determining the small-world behavior of the dispersion in the Great Barrier Reef: describe how community formation can facilitate the effective interaction of distant individuals across an entire population where there are only interactions between small groups of neighbors (*Kininmonth et al.*, 2010b).

In a connected directed-unweighted graph (i.e., directed-unweighted with no disconnected parts), it is possible to define the shortest path σ_{ij} connecting two sites $i \in V$ and $j \in V$ as the shortest possible alternating sequence of sites and edges, beginning with site i and ending with site j, such that each edge connects the preceding site with the succeeding one. The definition can be extended to directed weighted graphs: the shortest path has the lowest cost between two sites. The most frequent choice to define the cost of a path is the sum of its edges' weights. Nonetheless, when defining edges as larval transfer probabilities, one has to define the cost of a path as the product of its edges' weights (*Costa et al.*, In Preparation). So that, when calculating shortest paths and other measures based on them, we transform the connectivity matrix according to:

$$d_{ij} = \log \left(\frac{1}{a_{ij}} \right) \tag{2.4}$$

Cycles are defined as those paths that, starting from site $i \in V$, end up to the site i itself, after a certain number L of steps. In order to neglect the effect of the particles remaining at the same site with respect to the effect of the ones leaving the site and coming back, we only consider cycles with $L > 2$. In connectivity studies, one of the essential requisites for ensuring the persistence of a species in a given zone is the high probability to see the larvae returning home after a certain number of generations (see *Hastings et Botsford*, 2006, for details). This means that the more probable the cycle starting from a given site, the more likely the site is important for persistence. In fact, in this case, the site survival would be quite independent from the import of larvae from other sites. Thus it can act as a source in our graph (*Hastings et Botsford*, 2006). Once transformed the connectivity matrices according to Equation 2.4, we can identify the more probable cycles for each node by looking for the shortest cycles of a certain length L departing from each node. In this study we considered $L = 4$.

The efficiency of a network measures how efficient is the flow over the network by calculating the average shortest path of a network:

$$E(V) = \frac{1}{N(N-1)} \sum_{\substack{i \neq j \\ (i,j) \in V}} \sigma_{ij} \tag{2.5}$$

One can also restrict the calculation of E to the neighborhood of a given node. In this case we refer to the concept of local efficiency:

$$E_{loc}(V_i) = \frac{1}{N} \sum_{i \in V} E(V_i) \tag{2.6}$$

where V_i is the neighborhood of i.
Local efficiency reveals how much the system is fault tolerant (ref), thus it shows how efficient the communication is between the first neighbors of i when i is removed. In classical graph theory studies, .

Betweenness centrality $BC(i)$, $i \in V$, is based on the concept of shortest path. Betweenness estimates the importance of a site i by counting the fraction of existing shortest paths in the graph $\sigma_{lj}, \forall (l,j) \in V \mid l \neq j$ that effectively pass through this site $\sigma_{ij}(i), \forall (l,i,j) \in V \mid i \neq l \neq j$:

$$BC(i) = \sum_{l \neq i \neq j} \frac{\sigma_{lj}(i)}{\sigma_{lj}} \tag{2.7}$$

We then normalized the betweenness value by the total number of possible connections in the graph: $(N-1)(N-2)$, where N is the number of sites in the graph so that

$0 \leq BC \leq 1$.

In classical graph theory studies, High values of the betweenness indicate that a site can be linked with others on short paths. If one removes a site with high betweenness value it will worsen the efficiency of the transfer across a graph by lengthening the paths between many pairs of sites. For this reason, in connectivity studies it has been considered to be a proxy for migration stepping-stones (*Treml et al.*, 2008), genetic gateways (*Rozenfeld et al.*, 2008) and dispersion gateways (*Andrello et al.*, 2013).

The mean average path $\overline{\sigma}$ is commonly used to characterize the topology of a network (*Albert et Barábasi*, 2002):

$$\overline{\sigma} = \frac{1}{N-1} \sum_{i \neq j} \sigma_{ij} \tag{2.8}$$

A low mean average path means that the network is more concentrated and sites are not far from each other. On the other hand a high value indicate that on average there are many steps (i.e., generations in our case) between distant sites.

A widely used method for identifying clusters in physical networks is the maximum modularity criterion introduced by *Newman et Girwan* (2004). Modularity Q is defined, up to a multiplicative constant, as the difference between the number of edges falling within given groups of sites and the expected value in a network that conserves the strength values of the sites but with randomly placed edges (further details can be found in *Newman*, 2006). Intuitively, random connections between groups of sites are unlikely to be a meaningful cluster structure. Rather, we expect sites of a cluster to communicate more with other sites belonging to the same cluster rather than with external ones. Therefore, we can investigate the cluster structure of a graph by looking for the divisions of the graph that most differ from a random pattern.

Given a network, let c_i be the cluster in which site i is assigned. For a directed weighted graph the modularity assumes the form (see *Nicosia et al.*, 2009, for details):

$$Q = \frac{1}{m} \sum_{i,j \in V} \left[a_{ij} - \frac{k_i^{out} k_j^{in}}{m} \right] \delta(c_i, c_j) \tag{2.9}$$

where k_i and k_j are the strengths of the sites i and j, $m = \sum_i k_i$ and $\delta(c_i, c_j)$ is the Kronecker δ-function.

In particular, the first term in the summation indicates the real value of the connection between two sites. Instead, the second term indicates the value we expect by chance. Therefore positive values of modularity will indicate the possible presence of cluster structures in our network. In classical graph theory studies, modularity has been used to . In connectivity studies, it has been used by *Kininmonth et al.* (2010a).

The within-module degree z-score is a within-module version of degree centrality. If k_i is the strength of a site i to other nodes in its module s_i, k_{s_i} is the average of k over all the nodes in s_i, and $\sigma_{k_{s_i}}$ is the standard deviation of k in s_i, then:

$$z_i = \frac{k_i - \overline{k}_{s_i}}{\sigma_{k_{s_i}}} \tag{2.10}$$

The within-module degree z-score measures how well-connected node i is to other nodes in the module.

2.2.2.3 Comparing metapopulation model and graph theory

In order to determine the graph theory measures that give the most reliable identification of sites important for persistence, we calculated the correlation coefficient between the minimum recruitment success of each site with the value of the different graph theory measures. Indeed, the higher the minimum recruitment success R_{min} needed in a certain site to preserve the metapopulation after the removal of a sites around it, the more important the site is. So that, if a graph theory measure actually gives information about persistence we expect a high value of that measure for the same node as well. Therefore, a high correlation between R_{min} and the graph theory measure is expected. The only exception are the shortest cycles. Indeed, the shortest a cycle the most probable it is; so that a low value of the shortest cycles is expected to correlate with high values of R_{min}. Also, in order to calculate R_{min} in a way consistent with the shortest path calculation, the metapopulation model analysis removed each time the all the nodes composing the shortest cycle (with L=4) starting from each nodes.

We calculated the correlation on 20 connectivity matrices for each PLD. Successively, we calculated the mean value of the Spearman correlation coefficient (or rank correlation coefficient) associated to a significant p-value (< 0.05) for each PLD. In the following, the number of significant p-values (out of 20) is denoted as #vals. The rank correlation coefficient aims to identify the graph theory measures that better reproduces the hierarchy calculated by the metapopulation model analysis. However, for conservation policies it could be of greater interest to just identify the most important sites and eventually discard the information on the relative importance between sites. In order to identify the graph theory measures that best identifies the most important nodes we calculate, for each metric, the swapping score between the hierarchy of the most important site for metapopulation modelling and the hierarchy based on the different graph theory measures. In particular, we define the swapping score as the number of differences between a couple of hierarchies.

2.2.3 Results

The connectivity matrices calculated with a PLD of 1 and 2 weeks correspond to networks with a significantly higher average path length $\overline{\sigma}$ than the networks obtained with a PLD of 3, 4 and 5 weeks. in fact, the average path length for a PLD of 1 and 2 weeks were $\overline{\sigma}_{1w} = 0.25$ and $\overline{\sigma}_{2w} = 0.15$ respectively. Instead, for PLDs of 3, 4 and 5 weeks we obtain: $\overline{\sigma}_{3w} = 0.05$, $\overline{\sigma}_{4w} = 0.10$ and $\overline{\sigma}_{5w} = 0.05$. We can see that the latter ones are lower than the former. Moreover, while for PLDs of 1 and 2 weeks we found –by modularity maximization– that the sites in the GoL are organized in three clusters, for PLDs of 3, 4 and 5 weeks we found only two clusters (see Table 2.2). These facts indicate that while for PLDs of 1 and 2 weeks the network is more sparse and divided, for PLDs of 3,4 and 5 weeks the network is more dense and interconnected. Accordingly, the modularity value decreases as PLD augments. Modluarity values is 0.22 for a 3-weeks PLD, 0.17 for 4 weeks and 0.14 for 5 weeks. The reason for this is we can expect a particle in the GoL to travel a maximum 300km per week in the GoL (considering an average velocity of $50cm/s$ in Northern current, see Figure 2.3). Therefore, for more than 2 weeks a particle can easily travel across the GoL even if it travels at a lower average speed. As a consequence, the network is more connected than in the previous case.

The correlation coefficient and the number of significant p-values obtained for each graph theory measure and each PLD are reported in Table 2.3.
Strength is always significantly correlated in the most part of cases. However, the number of significant correlation coefficients is lowest for a 1 week-PLD. This last fact seems to be valid in general for all the graph theory measures. The mean correlation coefficient ranges between 0.45 for a 1 week-PLD and augment with the PLD up to 0.77. In particular, for PLDs of 3, 4 and 5 weeks, strength is the better correlated graph theory measure.
Eigenvector centrality is significantly correlated with the metapopulation model analysis in few cases and its mean correlation coefficient is always below 0.54.
The core number is well correlated for all PLDs even though, for a 1 week-PLD its correlation is significant in just two cases out of 20. However, the minimum mean correlation coefficient (0.57) is obtained for a 2 weeks-PLD and the maximum one (0.70) for a PLD of 5 weeks.
Clustering coefficient's mean correlation coefficient ranges between the 0.49 and the 0.58. Also, it is significantly correlated in more than half of the cases for PLDs of 2, 3, 4 and 5 weeks. Only in the 1 week-PLD case it is correlated in less than half (9) of the cases.
Shortest cycles overall not correlated with R_{min} and its correlation coefficient is signifi-

PLD [weeks]	# clusters	Sites
1	3	1-15 16-24 25-32
2	3	1-12 13-19 20-32
3	2	1-17 18-32
4	2	1-16 17-32
5	2	1-17 18-32

Table 2.2: Number of clusters found by modularity maximization for each PLD and nodes composing them.

cant in half or less of the half of the cases independently of PLD.
Betweenness' mean correlation coefficient ranges between 0.45 for a 1 week-PLD to 0.62 for a 5-weeks PLD. Its correlation coefficient is significant in more than half of the cases for all PLD, but the 1 week one.
Z-score's mean correlation coefficient increases from 0.45 to 0.70 with the PLD. Also, it is significantly correlated in most of cases for all PLDs, but the 1 week one.

The swapping scores between the hierarchy based on the different graph theory measures and the one based on metapopulation model analysis are reported in Table 2.4. Therein, we also report the most important site that was most frequently found by the different graph theory measures and by metapopulation model.
In the case of a 1 week-PLD, betweenness is the graph theory measure that, on average, better identifies the first three most important sites for persistence. Also, it is the only graph theory measure that most frequently identifies the site 21 as more important for betweenness as the metapopulation model analysis does.
In the case of a 2 weeks-PLD, core number is the graph theory measure that better identifies the first three most important sites for persistence. However, betweenness is still the only graph theory measure that most frequently identifies the site 21 as more important for betweenness as the metapopulation model analysis does.
In the case of a 3 weeks-PLD, strength and clustering coefficient are the graph theory measures that better identifies the first three most important sites for persistence. However, betweenness is the only graph theory measure that most frequently identifies the site 21 as more important for betweenness as the metapopulation model analysis does.
In the case of a 4 and 5 weeks-PLD, betweenness both better identifies the first three most important sites for persistence and identifies the site 21 as more important for betweenness as the metapopulation model analysis does.

Overall, the graph theory measures that are better correlated with R_{min} for PLDs of 1 and 2 weeks are core number, clustering coefficient and strength. For longer PLDs, they are strength, Z-score and core number. The graph theory measures that more reliably identifies the most important site for persistence is betweenness. Betweenness is also the graph theory measure that generally better identifies the first three most important sites for persistence; even though, occasionally, other metrics (core number for a 2 weeks-PLD; strength and clustering coefficient for a 3-weeks PLD) perform slightly better.

2.2.4 Discussion and conclusions

Connectivity estimation is crucial in order to set up effective conservation policies to face the deleterious effects of anthropic pressure and climate change on coastal marine populations. Graph theory has increasingly been applied in the context of connectivity studies but a general consensus on its interpretation has not yet been reached. In order to clarify which among the different graph theory measures that can be applied to the study of connectivity, we compared graph theory analysis' results with the results issued from a metapopulation model. Indeed, metapopulation model have been extensively used in the context of connectivity estimation and its interpretation is settled.
Our study, considered species with a passive PLD ranging from 1 to 5 weeks, the study area being the Gulf of Lion. We considered two spawning events in two different years and calculated the connectivity matrices between 32 reproductive sites by means of a numerical ocean circulation model. These connectivity matrices were used as common input of metapopulation models and graph theory analyses.

For short PLDs (1 and 2 weeks) the communities in the Gol are not solidly connected over the whole GoL. Indeed, we observe three clusters with higher modularity values with respect to the modularity value of the two clusters we observe with longer PLDs. For short PLDs we observe rather low correlations between the graph theory measures and the metapopulation model analysis' results. A reason for this is probably the fact that the communities are not well connected. Therefore, the effects of local demography

	1 week		**2 weeks**		**3 weeks**		**4 weeks**		**5 weeks**	
	$< \rho >$	#vals.	$< \rho >$	#vals.	$< \rho >$	#vals.	$< \rho >$	#vals.	$< \rho >$	#vals.
Strength	0.45	11	0.50	16	0.69	18	0.70	19	0.77	19
Eigenvector Centr.	0.26	7	0.54	11	0.37	9	0.28	8	0.18	11
Core Number	0.62	2	0.57	13	0.64	17	0.61	18	0.70	18
Clustering Coeff.	0.51	9	0.58	15	0.60	13	0.49	13	0.58	13
Local Efficiency	-0.19	3	-0.40	9	-0.62	13	-0.59	15	-0.70	18
Shortest Cycles	0.03	5	0.01	8	-0.06	10	-0.05	9	-0.05	8
Betweenness	0.45	6	0.50	12	0.55	17	0.58	18	0.62	14
Z-score	0.45	6	0.49	13	0.61	18	0.65	18	0.70	18

Table 2.3: $< \rho >$ is the mean of the Spearman correlation coefficients that have a significant (< 0.05) p-value. #vals indicates the number (out of 20) of correlation coefficient that are significant.

are much more important than in the case of long PLDs. Indeed, the untold hypothesis at the base of connectivity estimation by mean of graph theory analysis is often that the local demographic effect are ignored and that connectivity is maintained by means of the sole marine currents driving the transport of propagulae between distant sites.

For longer PLDs we observe higher correlations between graph theory and metapopulation model results. This suggests that in the GoL, persistence of species dispersing passively is likely more due to the transport by current than to local demographic aspects.

Overall, the strength of a site results the most reliable indicator for persistence among the graph theory measures that we considered. Its correlation can attain high values and be significant in the vast majority of cases. Therefore, strength can be reliably used to evaluate the relative importance for persistence of different sites. However, the graph theory measure that more reliably identifies the single most important site for persistence is betweenness. This peculiarity of betweenness could likely be important in the frame of conservation policies that –due to obvious limitations– cannot preserve all the reproductive habitats but rather focus their efforts on the most important sites.

	1 week			**2 weeks**			**3 weeks**			**4 weeks**			**5 weeks**		
	SC	mGT	mMM	SC	mGT	mMM	SC	mGT	mMM	SC	mGT	mMM	SC	mGT	mMM
Strenght	2.80	20	21	2.60	20	21	2.55	20	21	2.60	20	21	2.50	20	21
Eigenvector Centr.	2.75	20	21	2.90	17	21	2.95	15	21	2.85	15	21	2.85	15	21
Core Number	2.95	27	21	2.30	25	21	2.75	25	21	2.45	20	21	2.50	20	21
Clustering Coeff.	2.75	20	21	2.60	13	21	2.55	14	21	2.65	20	21	2.70	20	21
Local Efficiency	2.80	32	21	2.95	32	21	3.00	32	21	3.00	32	21	2.90	32	21
Shortest Cycles	3.00	5	32	2.90	30	32	2.90	3	29	2.85	1	31	2.95	3.00	18
Betweenness	2.60	21	21	2.50	21	21	2.65	21	21	2.45	21	21	2.40	21	21
Z-score	2.95	20	21	2.80	20	21	2.70	20	21	2.50	20	21	2.55	20	21

Table 2.4: Swapping coefficients *SC* relative to graph theory-based site hierarchy and metapopulation-based one. The mode of the most important site according to graph theory (mGT) and metapopulation model (mMM) are also reported.

2.3 Extra Material

2.3.1 Bridging centrality

When exploring the graph theory literature looking for measures that could potentially be applied to the connectivity study, I encountered the bridging centrality (*Hwang et al.*, 2008). The available studies in some undirected cases seemed promising. Thus I generalized to the directed case and applied to our study case. The results were reported in a first version of our study that was rejected due to the the fact that the metapopulation model study with which I compared our results did not rank all the 32 reproductive sites in the GoL. In the new version, we did not report the results relative to the bridging centrality because we significantly augmented the number of graph theory measures employed and because bridging centrality was never used before in connectivity studies. Moreover, some other measures resulted to be better correlated with persistence than bridging centrality.

Bridging centrality (BR) was first proposed by *Hwang et al.* (2008) for undirected unweighted graphs (see Figure 2.4 for an example). For our analysis we reformulated it in order to extend its use to directed weighted graphs.

Bridging centrality highlights those sites that connect different clusters of a network (see *Hwang et al.*, 2008). It is derived both from the betweenness value of a site and from the bridging coefficient that accounts for the probability of leaving the direct neighborhood of the site by starting from one of the sites composing it. Intuitively, sites with a high number of such edges fall on the boundary of clusters. In *Hwang et al.* (2008), for a site $i \in V$, the bridging coefficient is defined as:

$$\Psi^{uu}(i) = \frac{1}{k_i} \sum_{v \in V_i} \frac{\Delta(v)}{k_v - 1} \tag{2.11}$$

where V_i is the direct neighborhood of i: the set of sites reachable from i in one step. $\Delta(v)$ is the out-strength of sites $v \in V_i$ once deleted the edges going from v to other sites in V_i.

I classical graph theory studies, it has been used identify critical nodes that are important for the robustness of the network (*Nanda et Kotz* 2008; *Zhang et al.* 2010).

We propose the way to generalize the bridging coefficient to directed weighted graphs by accounting for the weight of the edges and by checking which edges are effectively leaving the neighborhood of the site. Then, we correct the out-strength of i via the term $-a_{vi}$ and the strength of v via the term $-(a_{iv} + a_{vi})$. Note that, for this calculation, all the terms a_{vv} on the diagonal of the connectivity matrix are irrelevant. Hence, in the directed weighted case, we redefine the bridging coefficient as:

$$\Psi^{dw}(i) = \frac{1}{k_i} \sum_{v \in V_i} \frac{\Delta(v) - a_{vi}}{k_v - (a_{iv} + a_{vi})} \tag{2.12}$$

In this way, we retain both the information on the flux of information through a site (given by the betweenness) and the topological information on the position of this site relative to clusters (given by the bridging coefficient). In fact, a site falling on the border of a cluster and channeling a high flux of information will have both high bridging coefficient and high betweenness values. As a result, the removal of such a high bridging centrality site would have a much more disruptive effect than the removal of a site having only either a high betweenness value or a high bridging coefficient (see *Hwang et al.*, 2008, for an analysis and discussion of this phenomenon in the undirected case). An important aspect to pay attention to, when calculating the betweenness centrality and the bridging coefficient of a site, is the different orders of magnitude in play. While the betweenness' one is normalized to one, the bridging coefficient's one is not: its value depends upon the particular metric used to define the distance between the sites. In order to give to the two parameters equal importance in characterizing the centrality of a site, we follow the suggestions of *Hwang et al.* (2008), and (i) calculate the betweenness centrality and the bridging coefficient for each site, (ii) calculate the rank vector of the sites on the base of their value of betweenness and bridging values, and (iii) calculate the normalized bridging centrality as:

$$BR(i) = \Gamma_{BC(i)} \cdot \Gamma_{\Psi(i)} \tag{2.13}$$

where $\Gamma_{BR(i)}$ is the rank of a site i in the betweenness vector and $\Gamma_{\Psi(i)}$ is the rank of a site i in the bridging coefficient vector. We then normalize bridging centrality by dividing it by the product of the maximum possible ranks N^2.

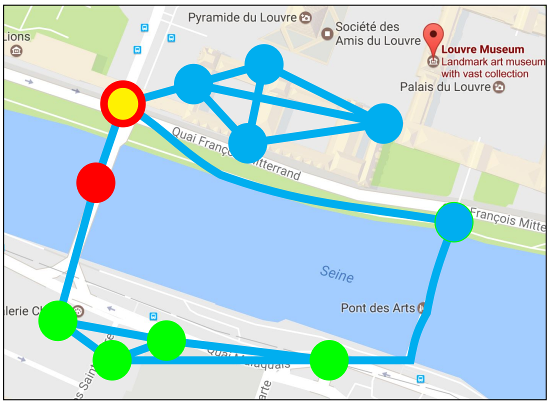
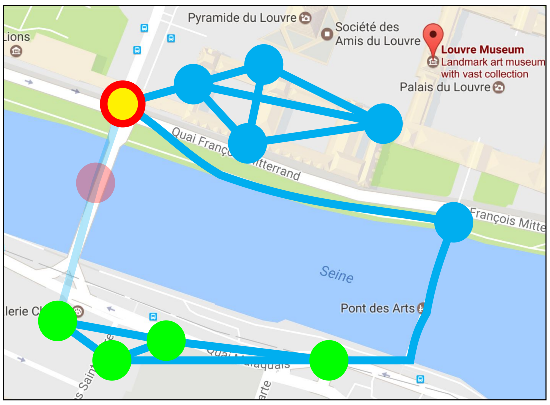
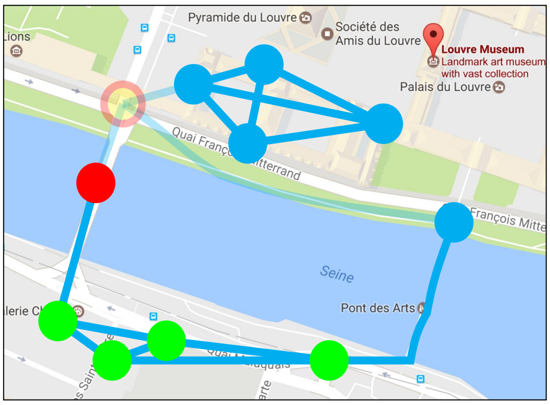

Pyramide du Louvre
Lions
Société des Amis du Louvre
Louvre Museum
Landmark art museum with vast collection
Palais du Louvre
Quai François Mitterrand
François Mitte
Seine
Pont des Arts
lerie C

Figure 2.4: The red and yellow node has high bridging centrality. The red node has high betweenness centrality. Nodes are divided into one cyan and one green clusters.


Pyramide du Louvre
Lions
Société des Amis du Louvre
Louvre Museum
Landmark art museum with vast collection
Palais du Louvre
Quai François Mitterrand
François Mitte
Seine
Pont des Arts
lerie C

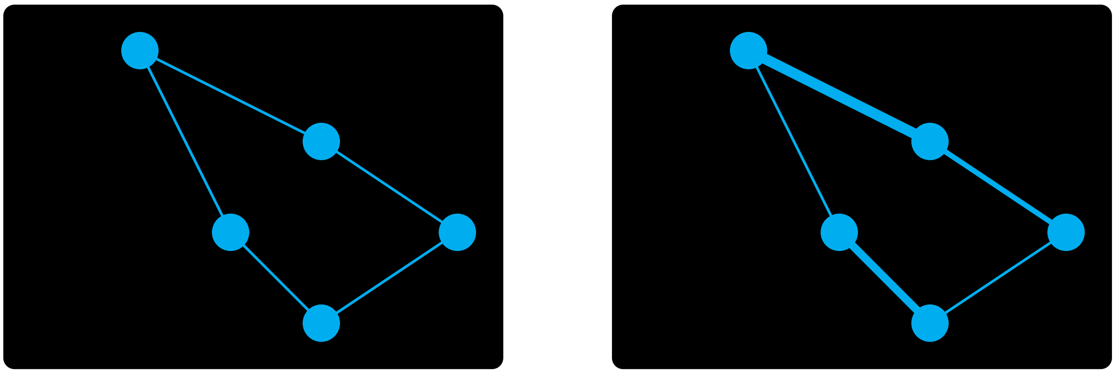
If we eliminate the high-betweenness node, communication between the two clusters is still possible even if less efficient.


Pyramide du Louvre
Lions
Société des Amis du Louvre
Louvre Museum
Landmark art museum with vast collection
Palais du Louvre
Quai François Mitterrand
François Mitte
Seine
Pont des Arts
lerie C

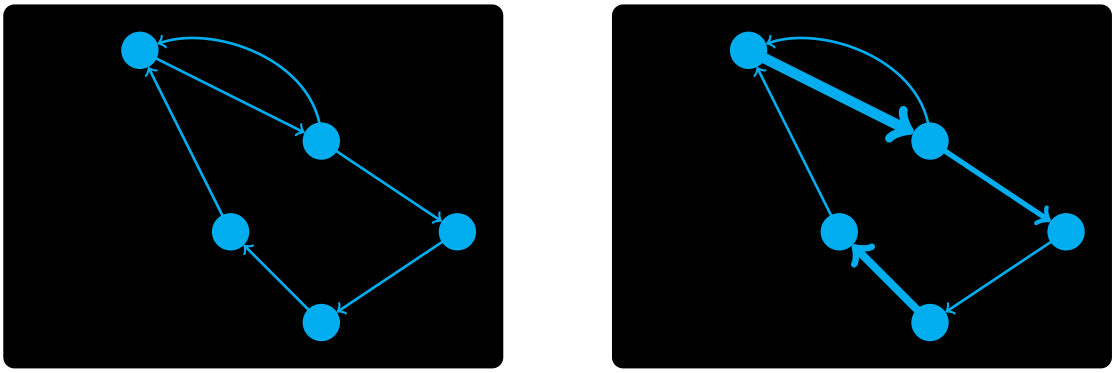
But if we remove the high bridging centrality-node, we fragment the network in a more severe way.

2.3.2 Visual explanation of graph theory measures

Hereafter, some graph theory measure are clarified in a visual way. For simplicity, we will restrict the representation mostly to undirected unweighted networks. The captions give a short definition of each measure, refer to Section 2.2.2.2 for the mathematical details. Moreover, the sub-captions, the considerations about the value of the different measures is always relative to the red dot.

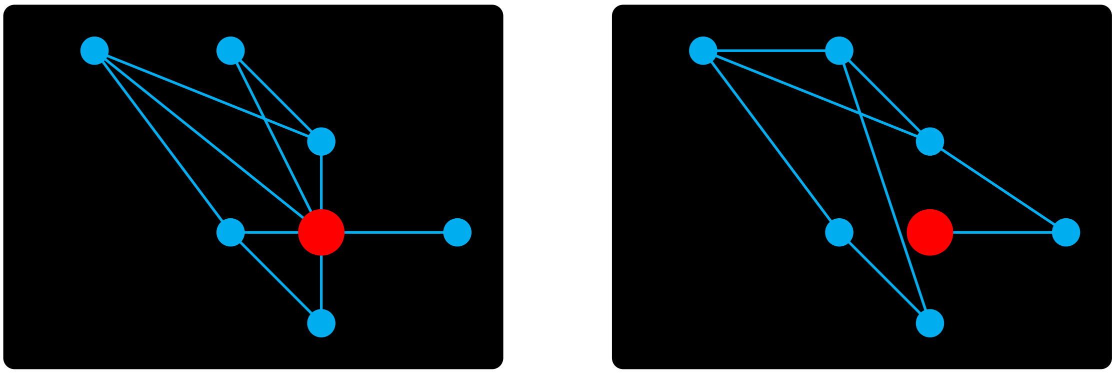


Unweigted undirected graph. Weigted undirected graph.



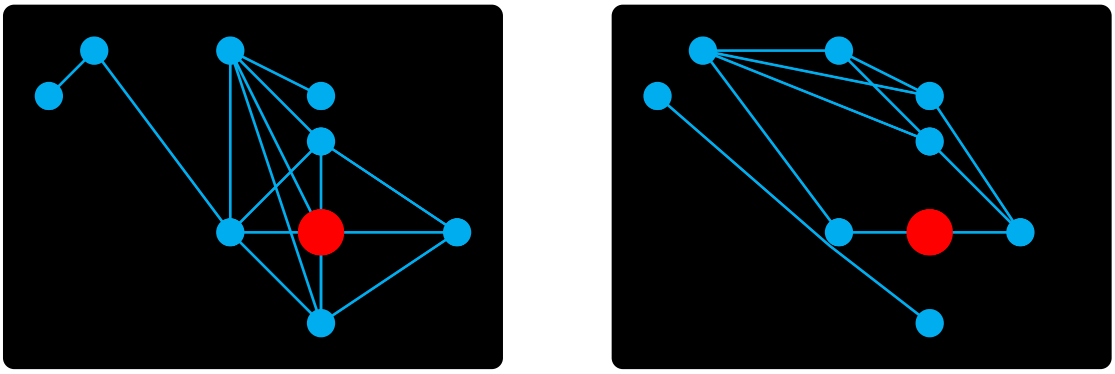
Unweigted directed graph. Weigted directed graph.

Figure 2.5: Types of graph.



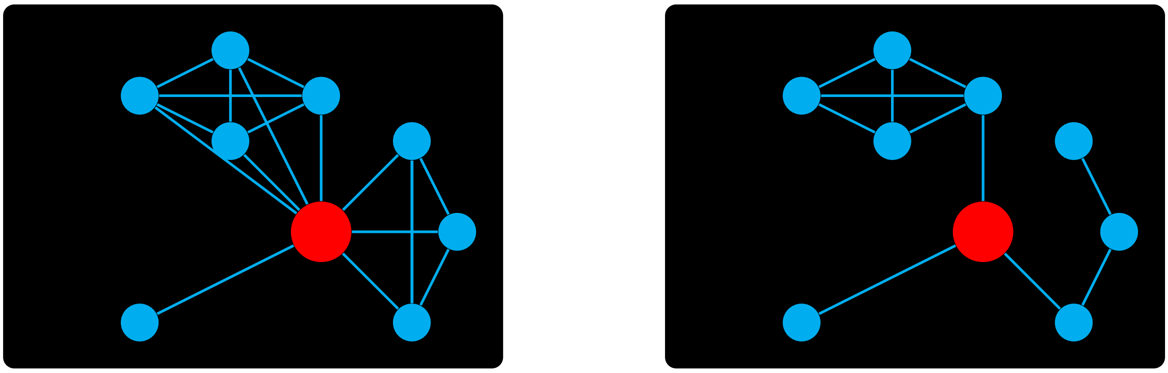
High. Low.

Figure 2.6: Degree: number of edges emanating from a single node.



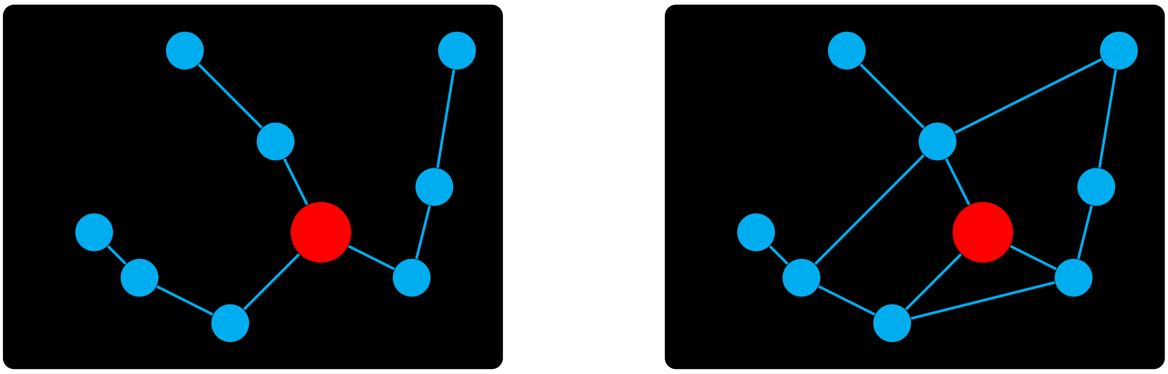
High. Low.

Figure 2.7: Clustering coefficient: how many of your nearest neighbors are connected to one another.



High. Low.

Figure 2.8: Local efficiency: average shortest path connecting all neighbors of a given node.



High. Low.

Figure 2.9: Betweenness: number of shortest paths that pass through a given node.

Chapter 3

Numerical Models and Turbulence

This work has been submitted to Ocean Modelling: *Comparison of in situ microstructure measurements to different turbulence closure schemes in a 3-D numerical ocean circulation model*, A. Costa, A.M. Doglioli, P. Marsaleix, A.A. Petrenko - CEMOD-D-16-00210

Contents

3.1 Comparison of in situ microstructure measurements to different turbulence closure schemes in a 3-D numerical ocean circulation model . . . 59
  3.1.1 Abstract . . . 59
  3.1.2 Introduction . . . 60
  3.1.3 Materials and Methods . . . 64
  3.1.4 Results . . . 73
  3.1.5 Discussion . . . 84
  3.1.6 Conclusions . . . 93
3.2 Appendices . . . 97
  3.2.1 Boundary conditions . . . 97
  3.2.2 Surface boundary conditions for k . . . 97
  3.2.3 Closure schemes . . . 98
3.3 Segmentation and quality fit . . . 101

3.4 Extra Materials for "Comparison of in situ microstructure measurements to different turbulence closure schemes in a 3-D numerical ocean circulation model" 104
3.5 Extra Material 109
3.5.1 Additional Simulations 109
3.5.2 Data analysis toolbox 109

In this chapter I present a study I conducted in order to establish if the Reynold Averaged Navier-Stokes numerical model Symphonie (*Marsaleix et al.*, 2008) performs better with certain turbulence closure schemes rather than other ones. This work has been submitted to Ocean Modelling. Symphonie is a widely used model for both physical and biochemical studies and it has been extensively validated (see Section 3.1.3.1 for references). However, here the focus is on the modelling of turbulent quantities like the kinetic energy dissipation rate and the turbulent diffusivity. A validation of this kind has never been conducted for Symphonie and it is also one of the few examples in the literature (e.g., *Burchard et Bolding* 2001; *Ilicak et al.* 2008; *Reffray et al.* 2015). The importance of reliably modelling the turbulence is due to the sensitivity of the large scale dynamics to the small scale energy dissipation and transfer. In particular, a modified large scale circulation could have non negligible effects on the dispersion of propagulae between distant reproductive sites.

The first aspect that could make a difference in implementing an ocean circulation numerical model is the turbulence closure scheme (TCS). Let's take a step back in order to understand what it is.

Given the nonlinearity and consequent high sensitivity of the Navier-Stokes equations on the initial conditions, a statistical approach to turbulent flows is often exploited. Within this framework the Reynolds decomposition ($u = \overline{u} + u'$) permits to obtain the equations for steady shear flows. By shear flow, we mean flows in which velocity is predominantly one dimensional in nature like in large scale geophysical flows. However, the quadratic nonlinear advective term in the Navier-Stokes equations gives rise to new terms involving the correlation of turbulent velocity fluctuations that couples the mean flow to the turbulence. These terms are called Reynolds stresses ($\tau_{ij} = -\rho\overline{u'_i u'_j}$) and represent the mean momentum fluxes induced by the turbulence. So that, in order to describe the evolution of the mean flow, we need to describe these stresses. Unfortunately writing a dynamical equation for the turbulent components u' does not help because it would involve higher-order correlations like $\overline{u'_i u'_j u'_k}$. We thus understand that in order to take a statistical approach we paid a price. Before we had a deterministic –yet chaotic– set of equations. Now, with a statistical approach, the quantities we are interested in, $\overline{u'_i u'_j}$ are perfectly reproducible but unknown: we do not have a closed set of equations to describe them. It is at this point that different approximations enter into play to add *ad hoc* extra information and close the equations. The different "tactics" to parameterize the velocity fluctuations' correlations as a function of some other property of the flow are called turbulence closure schemes.

The first proposed TCS dates back to *Boussinesq* (1872) who proposed a shear-stress strain-rate formulation for the Reynolds stresses in mean flows: $\tau_{xy} = \rho\nu_T \frac{\partial \overline{u_x}}{\partial y}$, where ν_T is the eddy viscosity. The general idea behind this formulation is that the turbulent mixing for momentum is analogous to the molecular transport of momentum, which leads to the laminar stress τ_{xy} (see *Tennekens et Lumley*, 1972, for a derivation). Note that this formulation states that the mean current is the energetic source for the turbulent fluctuations. We now have a situation in which there is a continual two-way interaction between the mean flow that generates and maintain the turbulence and the turbulence itself that shapes the mean velocity distribution. It is precisely this last interaction that we need to parameterize but this cannot be done without having some information about the intensity of turbulence.

One family of TCSs, the so called $k - \ell$, basically proceeds in complete analogy to the kinetic theory of gases and assumes that the eddy viscosity is proportional to the eddy mean free path (or length scale) ℓ and to their characteristic velocity scale V: $\nu_T = \ell V$,

where $V = k^{1/2} = \sqrt{\frac{1}{2}\overline{u'^2}}$. This leads to a form for the eddy viscosity along the lines of $\nu_T = \ell^2 \frac{\partial \overline{u'_x}}{\partial y}$ (see *Davidson*, 2004, for a derivation). So that now the problem is shifted again and we need to estimate ℓ on the base of some property of the flow. In the present work we considered a formulation proposed by *Gaspar et al.* (1990) that links the mixing length to the stratification of the flow (see Appendix 3.2.3 for more details).
Another family of TCSs, the $k - \varepsilon$, models the mixing length as function of $u \sim \sqrt[3]{\varepsilon\ell}$ exploiting the results of *Kolmogorov* (1942). We thus have a description of the turbulent flow through an exact dynamical equations for k and an empirical dynamical equation for ε.
More details about the particular $k - \varepsilon$ and $l = \ell$ models that we exploited in the study presented in this chapter can be found in Appendix 3.2.3.

We note at this point that the analogy between diffusion of momentum through turbulence and the diffusion through molecular motion has some severe theoretical limitations. For instance, turbulent eddies are distributed entities which continually interact, whereas molecules are discrete and collide intermittently. Moreover, the mean-free path of molecular motion is small compared to the macroscopic dimensions of the flow. The same cannot be said of the meanderings of turbulent eddies. Indeed, the large scale, energy containing eddies frequently have a size comparable to the characteristic scale of the mean flow (see *Davidson* 2004 for a more in depth-discussion). Nevertheless, this formulation of turbulent motion classically performs well in the context of geophysical flows and provides realistic results.

Among many other aspects that can affect the implementation of ocean circulation numerical models, we study the influence of the choice of the surface and bottom boundary conditions and of the stability functions parameterizing the unresolved higher moments in the dynamical equations. Specific details are given in the rest of this chapter.

The first part of this chapter coincides with the manuscript I submitted to Ocean Modelling. I am currently revising the manuscript in order to address some comments of the reviewers. Overall, the revision consist in adding some more numerical experiments to the analysis. The new numerical experiments are already listed in Section 3.5.1.

In addition to the paper I submitted to Ocean Modelling, at the end of the chapter, I also resume the preparatory work for this study. In particular, I will discuss some more details of the physical theory at the base of the SCAMP measurements; the analysis of the error on these measurements; the modification I made to the data analysis code; the tests of the data analysis code that I conducted on numerically generated data.
Furthermore, in the extra materials I show some results that expands the results of the submitted paper: four extra simulations that try to investigate the sensibility of the closure schemes to additional factors respect to the ones reported in the manuscript.

3.1 Comparison of in situ microstructure measurements to different turbulence closure schemes in a 3-D numerical ocean circulation model

3.1.1 Abstract

In situ measurements of kinetic energy dissipation rate ε and estimations of eddy viscosity K_Z from the Gulf of Lion (NW Mediterranean Sea) are used to assess the ability of $k-\varepsilon$ and $k-\ell$ closure schemes to predict microscale turbulence in a 3-D numerical ocean circulation model. Two different surface boundary conditions are considered in order to investigate their influence on each closure schemes' performance. The effects of two types of stability functions and optical schemes on the $k-\varepsilon$ scheme is also explored. Overall, the 3-D model predictions are much closer to the in situ data in the surface mixed layer

as opposed to below it. Above the mixed layer depth, we identify a model's configuration that outperforms all the other ones. Such a configuration employs a $k - \varepsilon$ scheme with Canuto A stability functions, surface boundary conditions parameterizing wave breaking and an appropriate photosynthetically available radiation attenuation length. Below the mixed layer depth, reliability is limited by the model's resolution and the specification of a hard threshold on the minimum turbulent kinetic energy.

3.1.2 Introduction

Turbulence is an essential mechanism for the transport of energy, salinity, suspended and dissolved matter. Turbulent fluxes of such quantities are the result of correlated, small-scale fluctuations of the velocity field and of the transported quantity itself. The prevalent turbulence production mechanisms in coastal ocean are: mean shear, unstable stratification, Langmuir circulation (*Farmer et Li*, 1995) and breaking surface waves (*Agrawal et al.*, 1992). For coastal ocean, mean shear is mainly generated by the action of winds and tides, but also by surface waves and baroclinic flows (e.g., *Thorpe*, 2005), including nonlinear internal waves (*Toole et Schmitt*, 1987). Unstable stratification results from surface processes such as surface cooling, evaporation or differential advection (e.g., *Kantha et Clayson*, 2000). Destruction of turbulence occurs by transformation into potential energy during stable stratification or viscous dissipation into heat (e.g., *Kantha et Clayson*, 2000). The complexity of these processes by themselves and of their interactions requires numerical models to cover a wide range of spatio-temporal scales and Reynolds number (e.g, *Burchard et al.*, 2008). This is especially true in the upper ocean where all the above phenomena concur together to generate turbulence.

Upper ocean connects –through various turbulent mechanisms– the surface forcing from the atmosphere with the quiescent deeper ocean where heat and fresh water are sequestrated and released on longer time and global scales (*Ferrari et Wunsh*, 2009). Also, upper ocean turbulence plays an important role in biological phenomena by, for example, determining phytoplankton growth rate (*Thomas et Gibson*, 1990), influencing primary production (*Flierl et Davis*, 1993) and the onset of blooms (*Taylor et Ferrari*, 2011). The complexity of modelling such mechanism within ocean circulation numerical models gave rise to several approaches. In particular, many turbulence closure schemes have been proposed. The ones most frequently found in the ocean modelling community's literature are the $k - k\ell$ by *Mellor et Yamada*, 1982; the $k - \varepsilon$ by *Rodi*, 1987; the $k - k\omega$ by *Wilcox*, 1988; the $k - \ell$ by *Gaspar et al.*, 1990 and the KPP by *Large et al.*, 1994. Following recent numerical modelling literature (*Ilicak et al.*, 2008; *Reffray et al.*, 2015), in the present study, we consider the $k - \varepsilon$ and $k - \ell$ second moments closure (SMC) schemes. Note that other kinds of closure schemes such as the KPP *Large et al.*, 1994) are not considered here being not as well suited as the other two schemes for a comparison with in situ data of kinetic energy dissipation rate ε.

Additional complexity is added to the modelling by the interplay of the SMC and the choice of boundary conditions (b.c.). Also the choice of surface and bottom boundary conditions can profit of a vast literature (e.g., *Craig et Banner*, 1994; *Stacey et Pond*, 1997; *Estournel et al.*, 2001; *Warner et al.*, 2005), aiming at modelling different forcing mechanisms. Furthermore, different stability functions can be chosen in order to include the effect of the parametrized non-local moments and pressure strain correlations in the dynamical equations (e.g., *Galperin et al.*, 1988; *Kantha et Clayson*, 1994; *Canuto et al.*, 2001). The choice of the optical scheme is particularly important considering the high number of studies coupling Symphonie to biochemical models as it can influences turbulent fluxes and nutrient availability.

Thus, the in situ validation of the closure schemes, boundary conditions, stability functions, optical scheme and their interplay is fundamental for assessing the reliability of numerical models (*Warner et al.*, 2005; *Peters et Baumert*, 2007; *Arneborg et al.*, 2007; *Ilicak et al.*, 2008).

The current study presents the comparison of kinetic energy dissipation rates ε measurements and vertical eddy viscosity K_Z estimations issued from a Self Contained Microstructure Profiler (SCAMP) with the predictions of a 3-D numerical ocean circulation model (Symphonie; *Marsaleix et al.*, 2008) obtained with different model's setup. The aim is to gain some insights on which scheme and/or boundary conditions permit to have the representation of turbulence activity closer to the observations.

Microstructure measurements with the SCAMP profiler have already been used for turbulence estimations in lakes and ocean (e.g., *Ruddick et al.*, 2000; *Sharples et Moore*, 2001; *Burchard et al.*, 2002; *Anis et Singhal*, 2002; *Sharples et al.*, 2003; *Peters et al.*, 2009; *Steinbuck et al.*, 2010; *Steinbuck et al.*, 2011; *Cuypers et al.*, 2012; *Jurado et al.*, 2012; *Bouffard et Boegman*, 2013). The dataset we exploit is described in Section 3.1.3. It consists in measurements taken in a coastal environment in the Gulf of Lion (GoL).

The GoL is located in the northwestern Mediterranean Sea and is characterized by a large continental margin (Figure 3.1) and a complex hydrodynamics (*Millot*, 1990). Its circulation is strongly influenced by the southwestward along-slope Northern Current. This density current flows in a cyclonic way and constitutes a barrier between the coastal waters of the continental shelf from the open northwestern Mediterranean Sea (*Alberola et Millot*, 1995; *Sammari et al.*, 1995; *Petrenko*, 2003). Cross-shore exchanges between the GoL and offshore waters are regulated by wind induced dynamics (*Estournel et al.*, 2003; *Hauser et al.*, 2003; *Petrenko et al.*, 2017) and by processes associated with the Northern Current, such as intrusions into the continental shelf and barotropic and baroclinic instabilities (*Conan et Millot*, 1992; *Flexas et al.*, 1997; *Petrenko et al.*, 2005; *Barrier et al.*, 2016). The Gulf of Lion is a suitable case study because of the high number of physical (*Qiu et al.*, 2010; *Hu et al.*, 2011), sediment dispersion (*Bourrin et al.*, 2011) and biochemical (*Pinazo et al.*, 2001; *Herrmann et al.*, 2014) numerical studies carried out there.

Symphonie has already been validated on a variety of different aspects like current modelling and eddy generation (*Rubio et al.*, 2009; *Hu et al.*, 2011; *Kersalé et al.*, 2013), river plume dynamics (*Reffray et al.*, 2004; *Gatti et al.*, 2006) and dense water formation

(*Dufau-Julliand et al.*, 2004; *Estournel et al.*, 2016). But a study of the different SMC that the user can implement in the Symphonie code has not yet been done. In particular, the modelling of the near-surface physical processes has been proved to be sensible to the choice of SMC as well as the physical-biogeochemical models to the K_Z values (*Fraysse et al.*, 2014).

In general, we can regard all modeled large-scale circulation features in an integrated fashion as they result from successive calculation steps and approximations. Hence, a major difficulty in validating numerical models –beside the high number of variables at play– is the possible compensation of different errors between each other. This fact makes difficult to attribute a specific amount of the total error on a certain quantity to a specific step in its calculation, in the present case the turbulence scheme. Here, our goal is to assess the model predictions focusing on turbulence modelling in the most realistic configuration we can achieve: 3-D dynamics with realistic forcing. Indeed, mixing has a primary role in influencing the large-scale circulation motion (*Rhines*, 1988; *Ferrari*, 2014). Nevertheless, the 3-D nature of the model brings in play an augmented number of numerical issues, among which spurious numerical diffusion (*Marsaleix et al.*, 2008; *Marchesiello et al.*, 2009; *Hu et al.*, 2009) that are usually neglected in similar, but 1-D, studies (*Gaspar et al.*, 1990; *Burchard et al.*, 2002; *Reffray et al.*, 2015).

The manuscript is organized as follows. In Section 3.1.3 we describe the properties of the numerical model, the microstructure in situ data and how we carry out the comparison between them and the numerical data. In Section 3.1.4 we report the results of our analysis and we discuss them in Section 3.1.5. In Section 3.1.6 we summarize the conclusions of our study.

3.1.3 Materials and Methods

3.1.3.1 Numerical modelling

The numerical model Symphonie is a 3-D primitive equations, free surface, sigma coordinate ocean model, based on Boussinesq and hydrostatic approximations (*Marsaleix et al.*, 2008; *Marsaleix et al.*, 2009; *Marsaleix et al.*, 2012). Components of current, temperature and salinity are computed on a C-grid (*Arakawa et Lamb*, 1977) using a classic finite difference method detailed in *Marsaleix et al.* (2006) and *Marsaleix et al.* (2008). This model has been extensively used in studies of the Mediterranean Sea, mostly at the scale of the continental shelves (*Ulses*, 2005; *Estournel et al.*, 2003; *Estournel et al.*, 2005), generally comparing satisfactorily with available in-situ observations of classical hydrological quantities. Symphonie has also been coupled to biochemical models for studies that demonstrated the impact of the turbulence level on determining the vertical flux of nutritive salts, the nutricline depth and –as a consequence– the results given by the biochemical models (*Ulses et al.*, 2016; *Herrmann et al.*, 2013). However, a study of the consequences of choosing a certain model's implementation of the Symphonie code has not yet been done. To fill this gap, we compare the model predictions of ε and K_Z with the values measured with the SCAMP profiler.

The model domain we use, shown in Figure 3.1, is that of *Estournel et al.* (2016). Note that all the measurements sites are far from the open boundaries. The horizontal resolution of the model grid is 1/110 degree (about 1 km). All the numerical experiments we perform cover the whole period in which in situ data are available: from 1 July 2010 to 13 March 2014, plus ten weeks of spin up.

In the vertical the model exploits a generalized sigma coordinate with 50 levels. Surface fluxes are computed using the bulk formulae by (*Large et Yaeger*, 2004) and the 3-hours ECMWF by *Estournel et al.* (2016). The boundary condition for ε is deduced with a length scale reasoning from the value of the Richardson number (*Estournel et Guedalia*, 1987; *Michaud et al.*, 2012), see Appendix 3.2.1 for more details. In order to simulate the limiting effect of stable stratification, following *Galperin et al.* (1988), the minimum of ε

is linked to the minimum turbulent kinetic energy value k_{min} through:

$$\varepsilon_{min} = k_{min} \frac{0.55^3 N}{0.53\sqrt{2}} \tag{3.1}$$

As a default in Symphonie, $k_{min} = 10^{-8}\ kg\ m^2/s^2$; this implies $\varepsilon_{min} = 10^{-12}\ m^2/s^3$. The choice of this threshold follows from underestimating the more standard value of $k_{min} = 10^{-6}\ kg\ m^2/s^2$ (*Gaspar et al.* 1990; *Burchard et al.* 2002), that is based on the estimation of the internal wave activity.

Similar low-frequency buoyancy conditions are maintained for all numerical experiments using a nudging procedure on temperature and salinity toward the corresponding MERCATOR fields (product PSY2V4R4). The nudging time scale is 30 days, enabling the free development of higher frequencies (including those of the turbulence closure scheme), and, at the same time, ensuring that the different turbulence schemes are tested in similar general conditions of stratification.

We choose the a $k - \varepsilon$ (*Burchard et Bolding*, 2001) and a $k - \ell$ (*Gaspar et al.*, 1990) closure schemes because, other than being the more exploited by Symphonie's users, they are also widely used in the wider scientific community. *Reffray et al.* (2015) –in a 1-D case– showed that $k - \varepsilon$ gives more reliable mixing estimation with respect to other schemes widely used in the literature: $k - kl$, $k - \omega$ and $k - \ell$. The $k - \ell$ scheme in *Reffray et al.* (2015) is based on *Gaspar et al.* (1990) but simplified for 1-D applications. We want to test if the original scheme by *Gaspar et al.* (1990) performs better than $k - \varepsilon$ in a 3-D case.

Moreover, our questioning the numerical results' sensitivity on the value of k_{min} follows from the study by *Gaspar et al.* (1990). Therein the authors encouraged (but not implemented) the use of a parametrization k_{min} rather than fixing a hard value. Herein we test if a good result can be achieved in a simpler way by specifying a different value of k_{min}.

A study by *Burchard et Bolding* (2001) showed that, in a 1-D study of temperature and mixed layer depth data of the well-known dataset OWS Papa (northern Pacific), the $k - \varepsilon$ closure scheme performs better when employing the stability functions proposed by *Canuto et al.* (2001) (commonly called Canuto A), rather than the ones by *Kantha et Clayson* (1994), *Rodi* (1980) and *Hossain* (1980). On the other hand, *Ilicak et al.* (2008) showed that –in a 3-D study of the Red Sea outflow– the stability functions of both *Canuto et al.* (2001) and *Kantha et Clayson* (1994) perform similarly. We want to further investigate the different performance of these two stability functions in a 3-D case. Note also that this study –as opposed to *Burchard et Bolding* (2001) and *Ilicak et al.* (2008)– has a strong focus on microstructure measurements and not only on more standard quantities like mixed layer depth, temperature and salinity.

In the literature there are different formulations for K_Z. In particular, it can include the molecular diffusivity D_T (so that we always have $K_Z > D_T$; e.g., *Burchard et Bolding* 2001) or not (e.g., *Han*, 2014). Here we want to clarify the differences (if any) between the two approaches.

Not having the necessary computational power to explore all the possible combinations of these factors in a 3-D model (as done for example in a similar study in the 1-D case by *Reffray et al.* (2015), and in a 3-D case by *Ilicak et al.* (2008) but on a shorter time span), we restrict the study to a subset of combinations.

In particular, nine different numerical experiments employing different combinations of turbulent closure schemes, boundary conditions, stability functions, values of the minimum of turbulent kinetic energy and optical schemes are analyzed here (see Table 3.1 for a concise summary).

Five of these numerical experiments employ a closure scheme based on a $k - \varepsilon$ approach (*Burchard et Bolding*, 2001) —hereafter marked by the prefix **KE**. The other four numerical experiments employ a $k - \ell$ scheme based on *Gaspar et al.* (1990) (hereafter marked by the prefix **KL**). Details of these nine numerical experiments can be found in Appendix 3.2.3. We test the effect of two possible surface boundary conditions. The first one (marked by a suffix **set**) supposes equilibrium between the production and dissipation terms in the dynamic equation for k (see Equation 3.2.2). The second one (marked by a suffix **flu**), takes into account the effect of breaking waves of all scales, as suggested by

Craig et Banner (1994); see Equation A.1 and Appendix 3.2.1 for details. The numerical experiments exploiting the **KE** scheme and using the stability by *Canuto et al.* (2001) –instead of the ones by *Kantha et Clayson* (1994)– are marked by the suffix **CAN**. Simulations with a higher minimum TKE –10^{-7} $kg\ m/s^2$ instead of 10^{-8} $kg m/s^2$– are marked with a suffix **MINk**.

All the numerical experiments with a higher threshold on k are such that $K_Z > D_T$. Therefore, in the other numerical experiments the total diffusivity could be smaller than the molecular one.

One numerical experiment (marked by the suffix **Opt**) investigates the effect of the attenuation length of the penetrative solar radiation Q_{sr}. Q_{sr} is parametrized as a two-band exponential scheme (*Maraldi et al.*, 2013):

$$Q_{sr}(z) = Q_{sr}(0)\left[Re^{-z/l} + (1-R)e^{-z/l_{PAR}}\right] \tag{3.2}$$

where the first right-hand term parameterizes the attenuation of red and near-infrared radiation (whose attenuation length is $l = 0.35$ cm); and the second right-hand term is the one of the visible and ultra-violet radiation; l_{PAR} is the photosynthetic available radiation (PAR) diffuse attenuation length. $Q_{sr}(0)$ is the fraction of the available penetrative solar radiation assuming a constant albedo of 6.6%. Following *Maraldi et al.* (2013), the l_{PAR} default in Symphonie is set to 11 m. However, this value has to be considered as an annual climatological estimation of the PAR. With this numerical experiment we test the effect of the seasonality of l_{PAR} by setting its value to 23 m coherently with the fact that most of the in situ measurements were acquired in September when we expect a low biological activity in the surface boundary layer of the GoL.

With this set of numerical experiments we can answer three principal questions: 1) which SMC between $k-\varepsilon$ and $k-\ell$ performs better with respect to our dataset? 2) what is the effect of the b.c. on the results of the numerical numerical experiments? and 3) what is the effect of the two stability functions?

3.1.3.2 SCAMP measurements

An in situ estimate of ε can be derived from high-resolution vertical profiles of temperature T. *Batchelor* (1959) derived the spectral shape of a conserved scalar field that is passively advected by an incompressible turbulent fluid with a molecular Prandtl number $Pr = \nu/D$ greater than 1 (seawater has $Pr = 7$ at $20°C$), where ν and D are respectively the molecular viscosity and the molecular diffusivity of the scalar. In the present case the scalar is the temperature. *Gibson et Schwarz* (1963) derived the one-dimensional Batchelor spectrum $E(K)$ of temperature gradient as a function of the rate of dissipation of temperature variance χ_T, the kinetic energy dissipation rate ε, the molecular diffusivity of temperature D_T and the circular wavenumber K:

$$E(K) = f(\chi_T, D_T, K_B) \tag{3.3}$$

where K_B is the inverse of the Batchelor length scale describing the length scales at which fluctuations in scalar concentration (temperature in this case) can still exist before being evened out by molecular diffusion. Therefore, (see *Ruddick et al.*, 2000; *Luketina et Imberger*, 2000 and *Steinbuck et al.*, 2009 for details) once measured the temperature vertical gradient at the millimeter scale and derived χ_T and K_B by fitting the Batchelor spectrum, a measure of ε follows from (*Batchelor*, 1959):

$$\varepsilon = \frac{\nu D_T^2}{K_B^4} \tag{3.4}$$

The temperature gradient profiles were measured with a SCAMP profiler. This instrument is equipped with a 100 Hz FP07 glass rod microthermistors (sensitivity of $0.001°C$). Our SCAMP was deployed in an upward configuration. After deployment, it sinks to a predetermined depth following an oblique trajectory and then rises up vertically at an approximately constant velocity $U = 10^{-1}\ m/s$. This type of trajectory permits the SCAMP to get away from the ship and be free from the influence of the ship's wake when rising up in an undisturbed water column. This allows to have reliable measurements of ε and K_Z near the sea surface (*Anis*, 2006).

The dataset (spanning the period 2010-2014) for this analysis consisted in 126 profiles of variable vertical extent —between 1 m below the surface and 100 m depth— collected in different sites in the Gulf of Lion (Figure 3.1). The profiles were collected during various oceanographic campaigns in the GoL conducted by the Mediterranean Institute of Oceanography (MIO - Marseille, France). Because the SCAMP measurements were opportunistically taken in cruises mainly not dedicated to turbulence measurements, repeated casts were often not possible. Data were collected mostly during summer and with meteorological conditions favorable to operations with a small boat, generally with wave heights less than half meter. Precipitation was always absent or negligible. Surface buoyancy flux was positive for 115 profiles indicating a gain of buoyancy by the ocean surface.

Temperature gradient spectra were computed from 128 points ($\approx$ 13 cm) windows without overlap. The choice of this segmentation resulted from a sensitivity analysis using different segmentation methods proposed in the literature (see Appendix 3.3).

The vertical eddy viscosity coefficient K_Z can be derived on the basis of the turbulence intensity parameter $Re_b = \varepsilon/\nu N$. Re_b expresses the ratio of the destabilizing effect of turbulence to the stabilizing effect of stratification and viscosity. In different Re_b regimes, the mixing efficiency, expressing the portion of the energy produced by shear which is dissipated by viscosity, assumes different values and determines different vertical turbulent diffusivity of density K_ρ. Here we use a recent field-validated parametrization of K_ρ as function of Re_b proposed by *Bouffard et Boegman* (2013) based on a previous parametrization derived by *Shih et al.* (2005). At very low Re_b ($Re_b < 10^{2/3}/\sqrt{Pr} \approx 1.7$), the turbulent regime is regarded as diffusive and K_ρ is set equal to the temperature molecular diffusivity $D_T = 1.4 \times 10^{-7} m^2 s^{-1}$. At low Re_b ($10^{2/3}/\sqrt{Pr} < Re_b < (3 \ln \sqrt{Pr})^2 \approx 8.5$) turbulent mixing tends to be controlled by buoyancy effects with incomplete mixing favoring up-gradient fluxes reducing the mixing efficiency. Here K_ρ can be expressed as

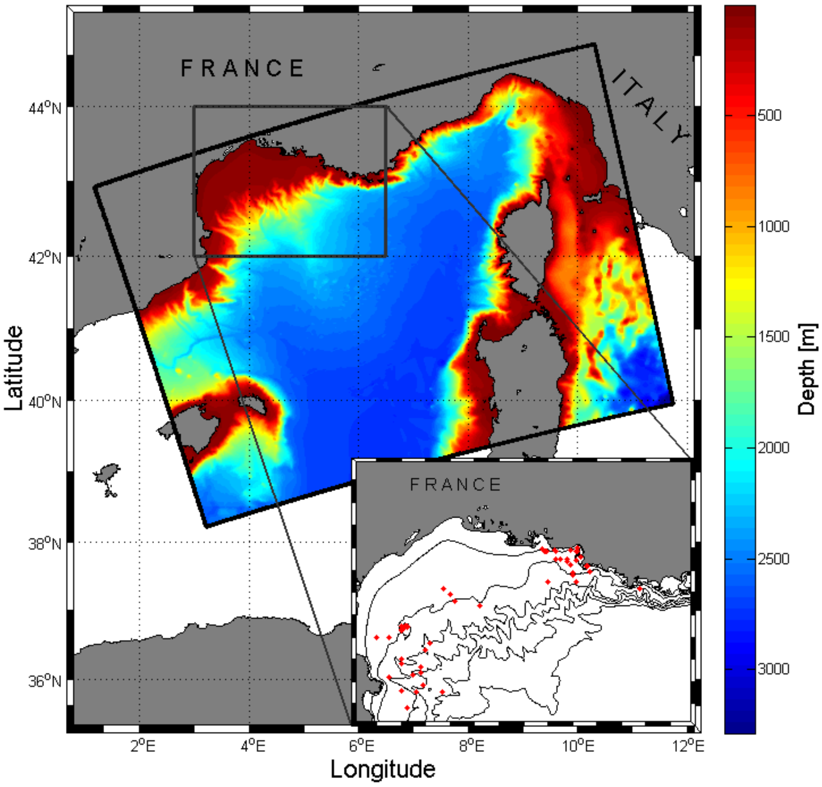

FRANCE
ITALY
FRANCE
44°N
42°N
40°N
38°N
36°N
Latitude
2°E
4°E
6°E
8°E
10°E
12°E
Longitude
500
1000
1500
2000
2500
3000
Depth [m]

Figure 3.1: Numerical model domain. The color code represents the water depth. The Gulf of Lion is magnified in the smaller box where the measurements sites are represented by red dots. Note that many profiles were taken at the same location over time. The black lines in the smaller box represent the 0, 50, 100, 500, 1000 and 1500 m isobaths.

$K_\rho = 0.1 Pr^{-1/4} \nu Re_b^{3/2}$. For moderate Re_b $(3 \ln \sqrt{Pr})^2 < Re_b < 400$) the mixing efficiency has the classical form derived by *Osborn* (1980): $K_\rho = 0.2 \nu Re_b$. For high Re_b ($Re_b > 400$), turbulence is in an energetic regime, where $K_\rho = 4\nu\sqrt{Re_b}$. Therefore, for high Re_b the buoyancy flux $K_\rho N$ tends to vanish together with stratification: $K_\rho N = \sqrt{\nu \varepsilon} N \xrightarrow{N \to 0} 0$, as we expect in weakly stratified fluids (*Osborn*, 1980).

As in the case of the Mediterranean Sea (e.g., *Cuypers et al.*, 2012), when the density variations are dominated by those of temperature, the density vertical eddy diffusivity coefficient is assumed to be equal to the temperature vertical eddy diffusivity coefficient (*Peters et al.*, 1988). Assuming then a turbulent Prandtl number $Pr_t = K_Z/K_\rho = 1$ (*Hogg et al.*, 2001), it is possible to estimate the momentum eddy viscosity (or eddy viscosity) K_Z to be equal to K_ρ.

3.1.3.3 Comparison of numerical and in situ data

While the vertical resolution of the in situ profile of ε and K_Z is constant (13 cm), the vertical resolution of the model is variable and generally coarser ($\approx 1 - 2\ m$) than the in situ one. In order to compare the numerical data to the measurements, the SCAMP data are grouped in windows centered on each sigma level. Then, the median of each window is calculated and compared to the numerical data at that sigma level. The choice of the median permits to give less importance to outliers and reduce the error due to the fact that the profiles were mainly single casts (*Lozovatsky et al.*, 2005).

For both the in situ and numerical data, we define the surface mixed layer (MLD) as the depth at which the temperature is smaller than the surface value by 0.5°C (*Anis*, 2006; *Jurado et al.*, 2012). Then we separate the data in the MLD from the ones below it. The data in the bottom layer are not considered for this analysis because of the low number of in situ profiles near the sea bottom.

Following *Burchard et al.* (2002), we define the decadal standard deviation of the numerical values of ε as:

$$\sigma_\varepsilon = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left[\log\left(\frac{\varepsilon_i^{mod}}{m^2 s^{-3}}\right) - \log\left(\frac{\varepsilon_i^{obs}}{m^2 s^{-3}}\right)\right]^2} \tag{3.5}$$

where M is the number of points in a given profile. Similar definitions hold for the decadal standard deviation of N (σ_N) and K_Z (σ_{K_Z}). To calculate these quantities, no division between data above and below the MLD is made because the number of in situ and numerical values in the MLD can differ for a given profile. Hereinafter, the mean values of σ_ε, σ_N and σ_{K_Z} on all the profiles are denoted σ_ε, σ_N and σ_{K_Z} for ease of reading, instead of $< \sigma_\varepsilon >$, $< \sigma_N >$ and $< \sigma_{K_Z} >$.

We also compare the probability density functions (PDFs) of the in situ and numerical values of kinetic energy dissipation rate ε, Brunt-Väisälä frequency N and eddy viscosity K_Z. In particular, the PDFs are calculated with a kernel density estimation (*Bowman et Azzalini*, 1997) on the base of the frequency distributions of values from all the in situ and numerical profiles.

In order to compare a distribution of numerical data $g(x)$ with the distribution of the in situ data $f(x)$, we compute the squared difference between the empirical cumulative distribution functions (F_n and G_n) of the two distributions:

$$\Delta^2(S) = \int_0^\infty \left(F_n(x) - G_n(x+S)\right)^2 dx \tag{3.6}$$

We define the shift S between two distributions as the shift of $G_n(x)$ that permits to minimize Δ^2. A subscript will tell to which distribution the values of Δ^2 and S refer to: Δ^2_ε and S_ε for the ε's PDF; Δ^2_N and S_N for the N's PDF; $\Delta^2_{K_Z}$ and S_{K_Z} for the K_Z's PDF.An estimation of the error on Δ^2 values is estimated by re-sampling the in situ empirical distribution function (*Nerini et Ghattas*, 2007) in order to further account for in situ measuring incertitude.

3.1.4 Results

To quantify the agreement between the numerical and the in situ data we use two methods. One takes into account the deviation of the numerical data from the in situ profiles over the whole column and expresses it by the decadal standard deviations (Table 3.2). The other approach looks at the shape of the probability density functions of the numerical values of ε and K_Z and expresses it by the squared differences and shift values above and below the MLD (Tables 3.3 and 3.4).

Median in situ and numerical profiles of ε are shown in Figure 3.2 vs. nondimensional depth z/MLD. Note that, the average value of the in situ MLD value is 27.3 m. As reported in Table 3.2, all the numerical MLDs are systematically lower by $\approx 10-15\%$ compared to in situ MLDs. Overall, the KE numerical experiments predict a slightly deeper MLD closer to the observations.

The thick lines in Figure 3.2 represent the median dissipation profiles calculated with all the profiles we analyzed and the shades represent the 95% bootstrap confidence interval. At the surface, all the numerical experiments agree within one order of magnitude with the in situ data. However, KEflu is the numerical experiment with values the closest to the observations. We also observe a difference between the effects of the equilibrium b.c. and the flux b.c. on the KE and KL closure schemes. While near the surface the numerical experiment KEset has a value of ε lower than KEflu, the numerical experiment KLset has a greater value than KLflu.
As we descend the water column, beyond $z/MLD = 1$, the in situ data tend to become more variable due to the lower number of profiles reaching depths greater than the MLD. On the contrary, the numerical data tend to become much less variable. Moreover, there

is an evident difference between the SMCs: when the lower threshold on k is applied, the KE numerical experiments always have lower levels of turbulence with respect to the KL numerical experiments with KEs' median values of ε one order of magnitude lower than the KLs' ones. When a higher threshold on the kinetic energy k is applied, the difference is slightly reduced.

At $z/MLD = 0.5$ the in situ data exhibit a kink that separates higher mixing at the surface from lower mixing in the rest of the MLD. KEflu, KLflu, KLsetMINk and KLfluMINk are the numerical experiments that better reflect this behavior. Note also that the in situ data show a clear increment of ε at $z/MLD = 1.1$ that is only slightly reflected in some numerical data.

We do not show values for $z/MLD > 2$ because of the large bootstrap error bars due to the lower number of profiles in that depth range.

Overall, looking at the column as a whole, the numerical experiments with a higher minimum value of kinetic energy (KEfluCANMINk, KEfluCANMINkOpt KLsetMINk and KLfluMINk) have the lower average error with respect to the measurements (see values of σ_ε in Table 3.2) because of the better agreement with in situ data below the MLD.

Median in situ and numerical profiles of N and their error are shown in Figure 3.3 vs. nondimensional depth z/MLD. Near the surface, the in situ data are considerably more variable than the numerical data. However, the numerical data are of the same order of magnitude as the measurements. As we go below the MLD, the median value of the in situ data increases in the depth range $1 < z/MLD < 1.6$. Overall, all the numerical experiments do not show such a marked increase and are rather similar one to another (as we expect because all of them are nudged to the MERCATOR fields). However, KEfluMINkOpt seems to be the numerical experiment that better reproduce the in situ data's trend right below the MLD.

Looking at the column as a whole, KEsetMINk has the lower average error with respect to the measurements (see values of σ_N in Table 3.2).

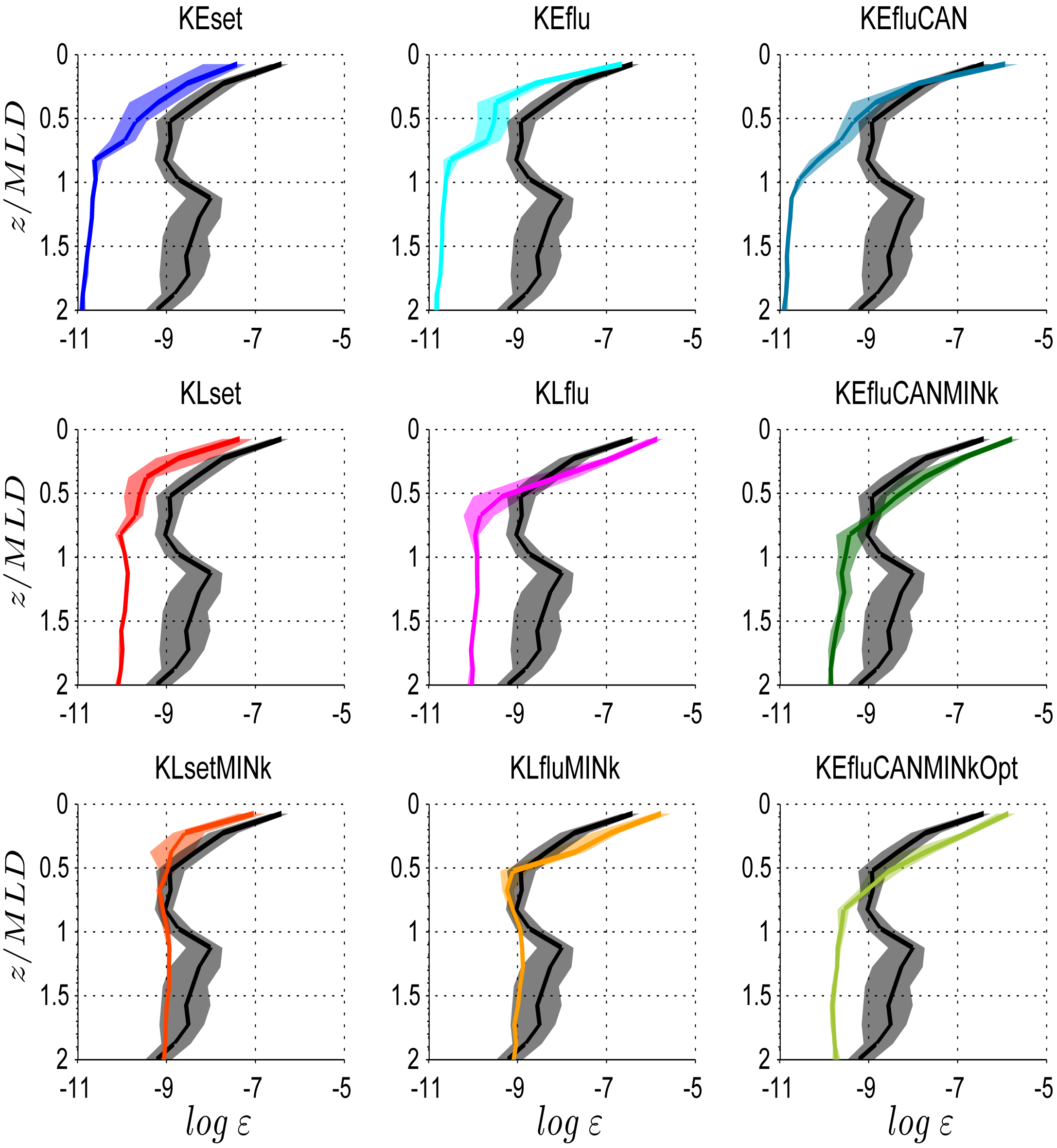


Figure 3.2: The thick lines represent median values of ε estimated from in situ data (in black) and numerical experiments. The shades indicate 95% bootstrap confidence intervals. Water depth z is adimensionalised with respect to the mixed layer depth MLD for each profile.

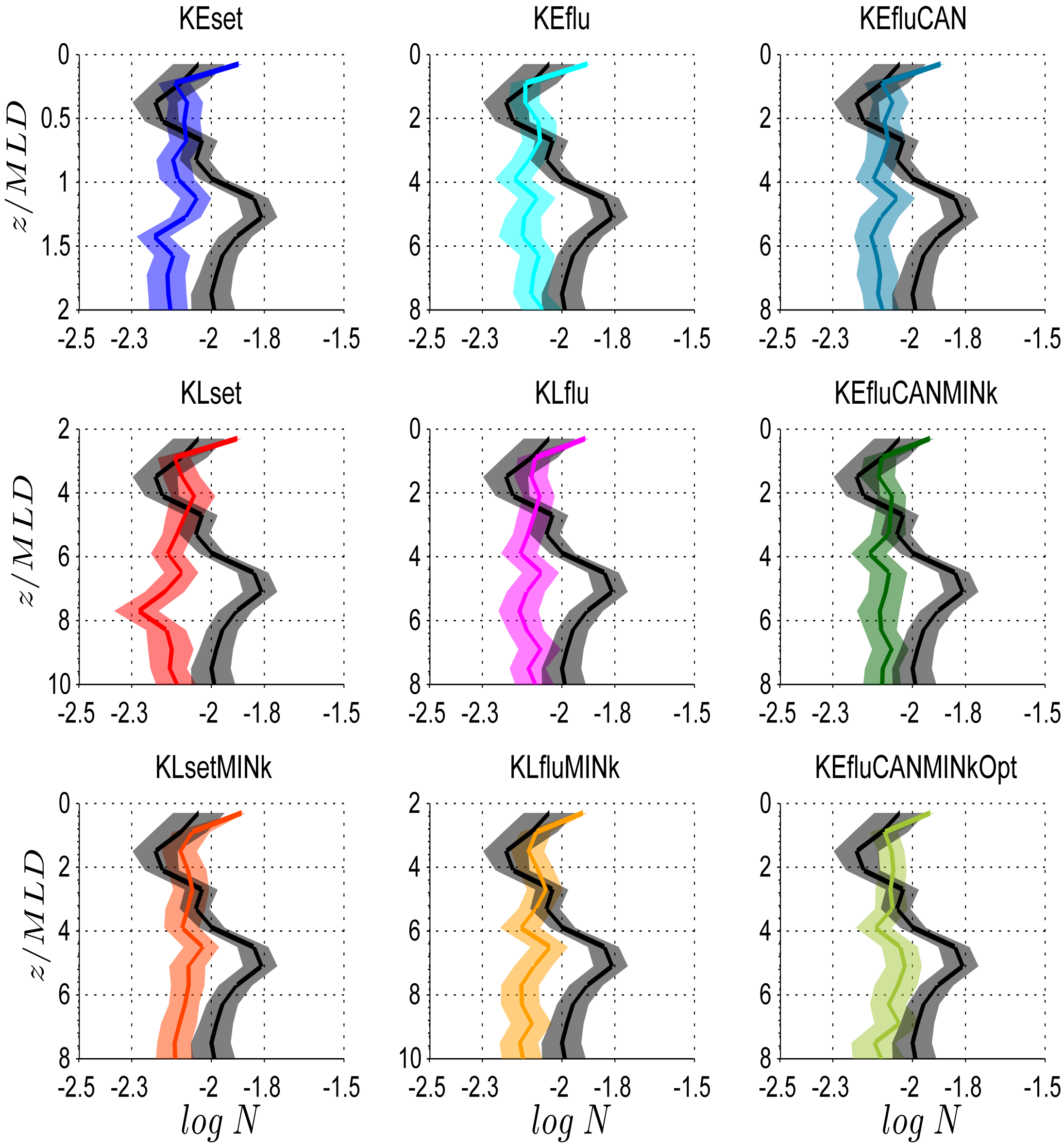


Figure 3.3: The thick lines represent median values of N estimated from in situ data (in black) and numerical experiments. The shades indicate 95% bootstrap confidence intervals. Water depth z is adimensionalised with respect to the mixed layer depth MLD for each profile.

The measurements of ε and N permitted to calculate the turbulence intensity parameter Re_b. The in situ probability density functions of the in situ Re_b are reported in Figure 4.11 in the Extra Materials. In particular, in our dataset, 30% of the in situ Re_b values fell in the diffusive regime, 15% in the buoyancy-controlled regime, 29% in the transitional regime and 26% in the energetic regime. If only the data above the MLD are considered these numbers respectively become 17%, 10%, 31% and 42%; or 36%, 16%, 28% and 20%, if only the data below the MLD are considered.

Median in situ and numerical profiles of K_Z and their error are shown in Figure 3.4 vs. nondimensional depth z/MLD. In the surface layer the numerical values are much more variable than the in situ ones. The opposite is observed below the MLD.
Near the surface, KEflu shows the best agreement with the observations among the simulations with the lower threshold on k. KEset is not as good as KEflu, but it significantly improves when the Canuto A stability functions. KLset and KLflu perform similarly but have an error in opposite directions. Near the surface, KLset is higher than KLflu. The same holds for KLsetMINk and KLfluMINk.
Below the MLD, the numerical experiments appear to be grouped on the base of the threshold on k: no differences between different SMC and b.c. appear obvious. Noticeably, below the MLD, the numerical experiments with a higher threshold on the kinetic energy k are more compatible with the in situ data than the other numerical experiments. The median value of K_Z below the MLD appears to have a seasonal behavior. In particular, it is equal to $5.90 \times 10^{-6} m^2/s^2$ in spring, 3.02×10^{-6} in summer, 3.27×10^{-6} in autumn and 5.57×10^{-7} in winter.
Overall, looking at the whole column, the numerical experiments that have the lower average error with respect to the measurements are the four numerical experiments with a high threshold on k, as the values of σ_{K_Z} in Table 3.2 indicate.

The probability density functions of the in situ dissipation rate ε data (in black) and of the numerical data are shown in Figure 3.5. Above the MLD (Figures 3.5a-b), the in

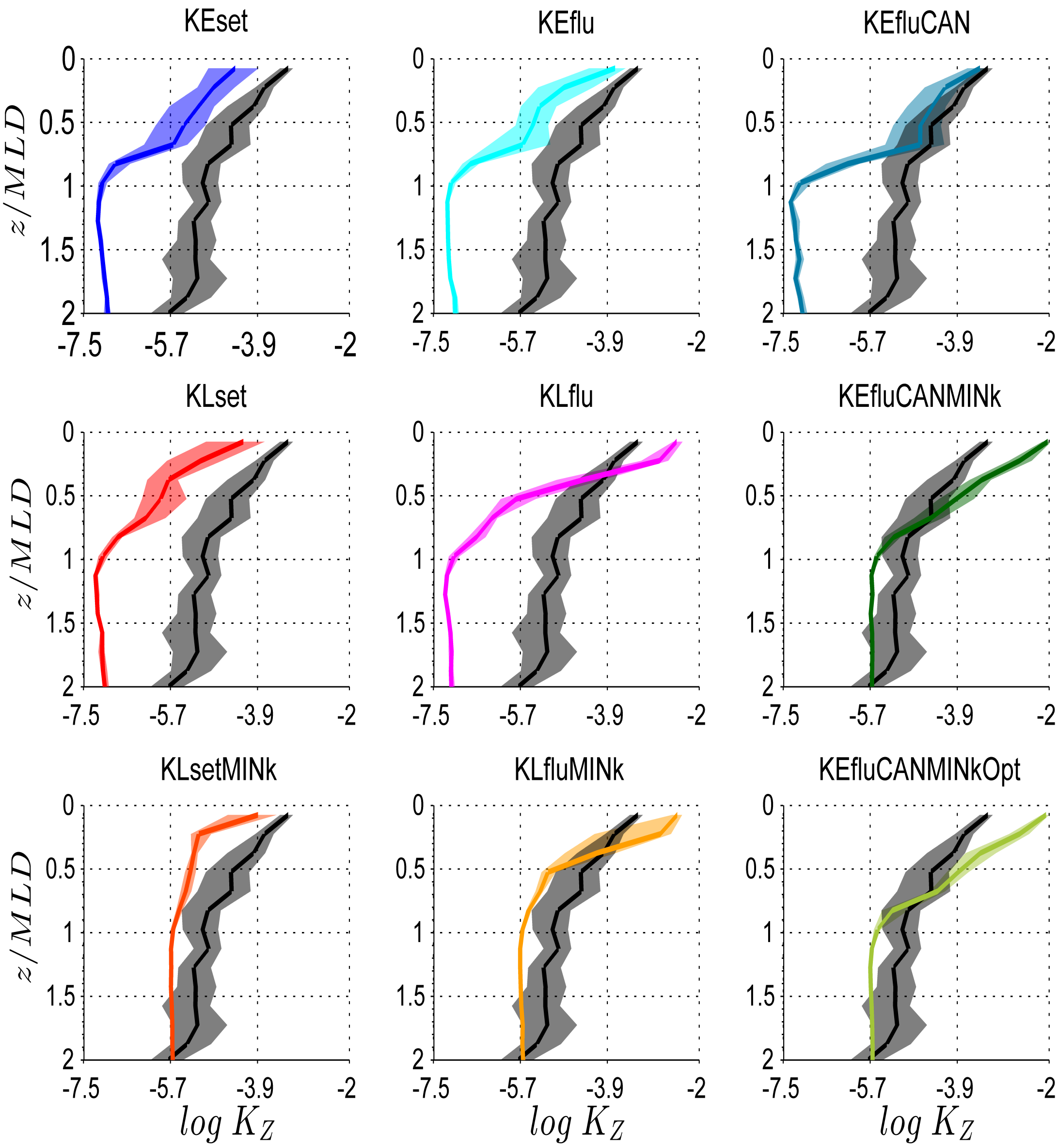


Figure 3.4: The thick lines represent median values of K_Z estimated from in situ data (in black) and numerical experiments. The shades indicate 95% bootstrap confidence intervals. Water depth z is adimensionalised with respect to the mixed layer depth MLD for each profile.

situ values show a clear bimodal behavior with one peak at $\varepsilon \approx 5 \times 10^{-10} m^2/s^3$ and one at $\varepsilon \approx 5 \times 10^{-7} m^2/s^3$. The numerical distributions show this feature in a less marked way and are generally low-biased with respect to the in situ data's PDF. As values of S_ε and Δ^2_ε in Table 3.3 show, above the MLD, the distribution more aligned to the in situ data one are KEfluMINk and KEfluMINkOpt, while the distribution with the shape more resembling the in situ data one is KEfluMINkOpt.

Below the MLD (Figures 3.5c-d), all the numerical PDFs have mean values similar to the ones of the corresponding PDF in the surface layer, but are much more narrow than the in situ data. KL numerical experiments have higher mean values than KE ones and the numerical experiments with a threshold on k have higher mean values than the corresponding numerical experiments without threshold. Below the MLD, none of the model PDFs has a behavior resembling the in situ one: they are all peaked and not spread out like the in situ distribution. As we see in Table 3.4, below the MLD, KLsetMINk and KLfluMINk have the same shift with respect to the in situ distribution, while KEfluCANMINkOpt and KLfluCANMINkOpt have the lower value of Δ^2_ε. But we note that the differences between the values of Δ^2_ε below the MLD are due to small differences in the low right-end tails that are unlikely to be significative.

The probability density functions of the in situ stratification N data (in black) and of the numerical data are shown in Figure 3.6. Both above and below the MLD, there is no drastic difference between the numerical distributions in both shape and peak. The similar behavior of all the numerical experiments can be due to the low frequency nudging to the MERCATOR fields. Overall, above the MLD (Figures 3.6a-b), the numerical experiments overestimate the stratification. Note that this is not evident by just looking at the median profiles in Figure 3.3. As values of Δ^2_N in Table 3.3 indicate, the distribution more resembling the in situ one is KEfluCANMINkOpt.

Below the MLD (Figures 3.6c-d) the peaks of the numerical PDFs are less aligned than above the MLD. In this layer, the numerical distributions show a peak at $N \approx 5 \times 10^{-3.3} s^{-2}$ that is not present in the measurements. As values of Δ^2_N in Table 3.4 show, KEfluCAN-

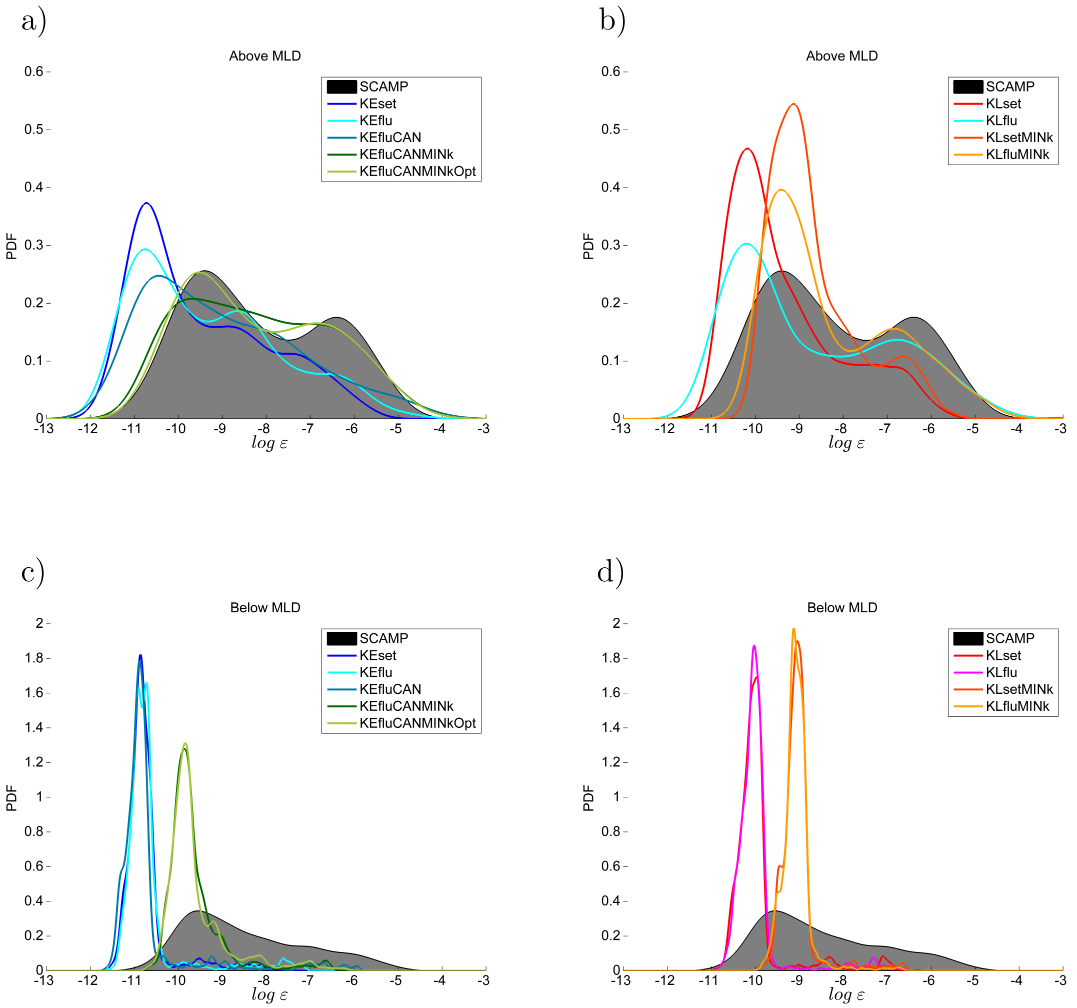


Figure 3.5: a) and b): ε probability density functions issued from all the in situ (black shade) and numerical data (thick lines) in the surface mixed layer for the KE closure scheme and the KL closure scheme respectively. c) and d): the same but below the surface mixed layer.

MINk is the distribution with the shape more resembling the in situ one above the MLD, while KEfluCAN is the more resembling below the MLD. We note that when the secondary peak is removed the numerical experiment with the better N's PDF are KEfluCANMINk and KEfluCANMINkOpt.

The probability density functions of the in eddy viscosity K_Z in situ data (in black) and of the numerical data are shown in Figure 3.7. Above the MLD (Figure 3.7a-b), the in situ values have three peaks with a higher peak at $K_Z \approx 5 \times 10^{-4} m^2/s^2$. None of the numerical experiment well reproduces this behavior. As values of σ_{K_Z} in Table 3.2 suggest, above the MLD the distribution more aligned to the in situ one is KLsetMINk. Besides, $\Delta^2_{K_Z}$ values in Table 3.3 indicate that KEfluCANMINkOpt is the distribution with the shape more resembling the in situ one. The effect of raising k_{min} is similar on the two SMC families: it shifts the distribution towards higher values. However, KLsetMINk and KLfluMINk show a prominent first peak and thus perform less well than KEfluCANMINk and KEfluCANMINkOpt.

Below the MLD (Figures 3.7c-d) –similarly to what seen for ε– the numerical distributions are clearly different from the in situ one. $\Delta^2_{K_Z}$ values in Table 3.4 indicate that KEfluCAN is the distribution with the shape more resembling the in situ one. But, as we already noted for the distributions of ε below the MLD (Figure 3.5), in this layer there is not a real difference between the numerical experiments.

All the measurements of ε, N and K_Z and the corresponding numerical data are shown in Figure 3.8. It turns out that the point clouds reveal that the in situ values are much less disperse than the model ones (not evident from the picture; see Supplementary Data in the on-line version of this paper). In Figure 3.8a-b we can see six main planes corresponding to: KEfluCAN, KEset and KEflu, KLset, KLflu and some in situ data at low K_Z. Indeed, these last points lie on a plane with constant K_Z as prescribed by the parametrization we used (see Methods and Discussion). Data belonging to the numerical experiments with a higher threshold on k (KEfluCANMINk, KEfluCANMINk,

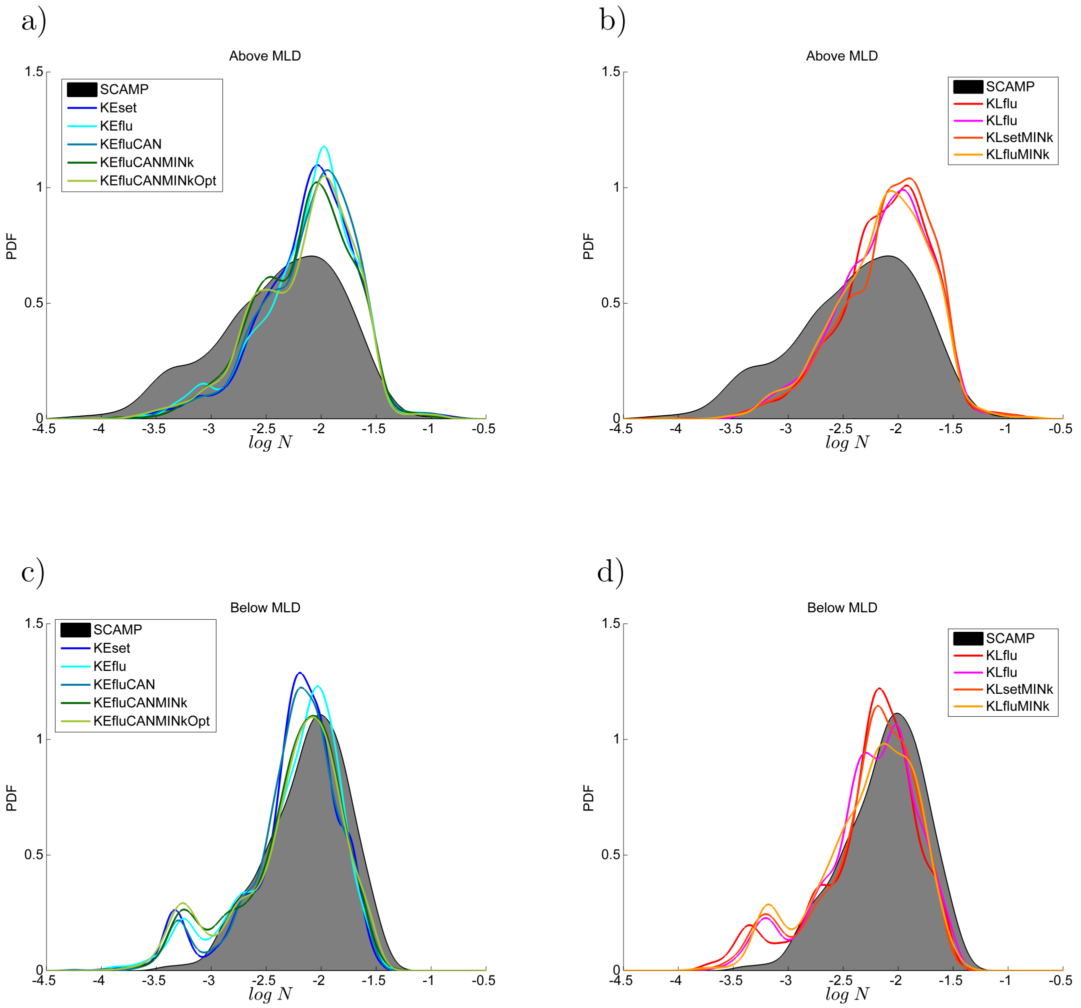


Figure 3.6: a) and b): N probability density functions issued from all the in situ (black shade) and numerical data (thick lines) in the surface mixed layer for the KE closure scheme and the KL closure scheme respectively. c) and d): the same but below the surface mixed layer.

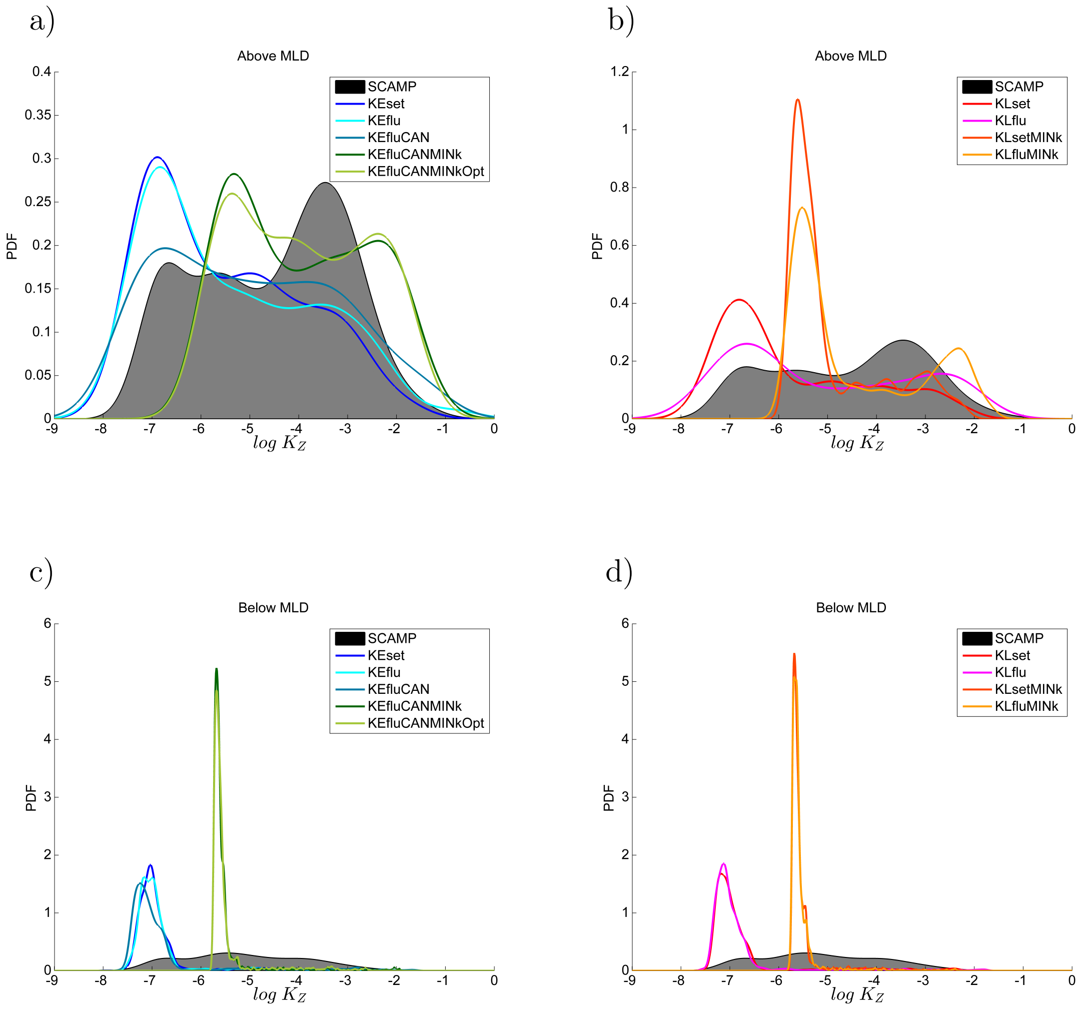


Figure 3.7: a) and b): K_Z probability density functions issued from all the in situ (black shade) and numerical data (thick lines) in the surface mixed layer for the KE closure scheme and the KL closure scheme respectively. c) and d): the same but below the surface mixed layer. Note the difference in the range of the y axes in a) and b).

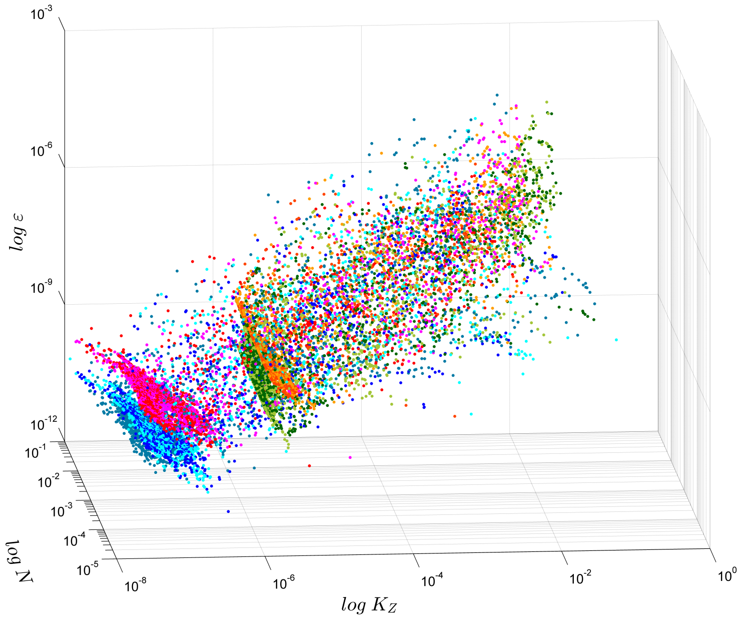
KLsetMINk and KLfluMINk) are first found at $K_Z \approx 2 \times 10^{-6}$ (Figure 3.8b). The KE numerical experiments have planes with an average value of ε that is lower than the average value of ε of the KL numerical experiments. These last ones have lower average values of ε than the numerical experiments with higher threshold on k. This closely mirrors what we already noted in Figure 3.5c-d. But it turns out that the points lying on the planes in Figure 3.8 do not belong exclusively to depths greater than the MLD (the differentiation between above and below the MLD is not shown in Figure 3.8).

	SMC	b.c.	k_{min} $[m^2/s^2]$	Stability Functions	l_{PAR}
KEset	$k-\varepsilon$	Equilibrium	10^{-8}	*Kantha et Clayson* (1994)	11m
KEflu	$k-\varepsilon$	Flux	10^{-8}	*Kantha et Clayson* (1994)	11m
KEfluCAN	$k-\varepsilon$	Flux	10^{-8}	*Canuto et al.* (2001)	11m
KEfluCANMINk	$k-\varepsilon$	Flux	10^{-7}	*Canuto et al.* (2001)	11m
KEfluCANMINkOpt	$k-\varepsilon$	Equilibrium	10^{-7}	*Canuto et al.* (2001)	23m
KLset	$k-\ell$	Equilibrium	10^{-8}	none	11m
KLflu	$k-\ell$	Flux	10^{-8}	none	11m
KLsetMINk	$k-\ell$	Equilibrium	10^{-7}	none	11m
KLfluMINk	$k-\ell$	Flux	10^{-7}	none	11m

Table 3.1: Set of numerical experiments performed for this study.

3.1.5 Discussion

By considering the values of decadal standard deviation, the shape of the probability density functions and the median profiles, we can progress towards the identification of a)



b)

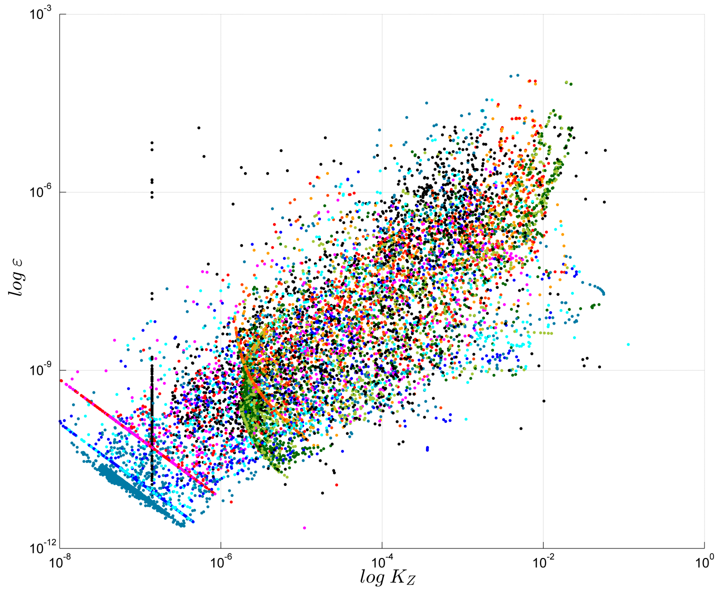


Figure 3.8: a) All in situ (in black) and numerical data of K_Z, N and ε. b) All in situ (in black) and numerical data of K_Z and ε. The color code is the same as in the other figures.

	MLD ratio	σ_ε	σ_N	σ_{K_Z}
KEset	0.91	2.48	1.16	2.19
KEflu	0.91	2.43	1.15	2.16
KEfluCAN	0.85	2.42	1.14	2.11
KEfluCANMINk	0.94	1.88	1.16	1.63
KEfluCANMINkOpt	0.89	1.82	1.16	1.56
KLset	0.88	2.07	1.14	2.20
KLflu	0.87	2.10	1.14	2.21
KLsetMINk	0.85	1.59	1.12	1.46
KLfluMINk	0.86	1.67	1.14	1.54

Table 3.2: MLD ratio indicates the ratio of the numerical prediction of the MLD to the in situ value $< MLD >_{SCAMP} =$ 27.3 m. σ_ε indicates the mean value of the decadal standard deviation of the ε profiles. σ_N indicates the mean value of the decadal standard deviation of the N profiles. σ_{K_Z} indicates the mean value of the decadal standard deviation of the K_Z profiles.

`Above MLD`	S_ε	Δ^2_ε	S_N	Δ^2_N	S_{K_Z}	$\Delta^2_{K_Z}$
KEset	1.31	0.010±0.002	-0.47	0.018±0.001	0.96	0.032±0.003
KEflu	1.21	0.008±0.001	-0.47	0.015±0.001	0.83	0.039±0.004
KEfluCAN	0.87	0.006±0.001	-0.50	0.016±0.001	0.46	0.021±0.003
KEfluCANMINk	0.62	0.0031±0.0009	-0.45	0.013±0.001	0.49	0.023±0.002
KEfluCANMINkOpt	1.07	0.0008±0.0003	-0.47	0.011±0.001	1.19	0.022±0.003
KLset	0.03	0.012±0.002	-0.41	0.014±0.001	-0.66	0.053±0.005
KLflu	0.34	0.041±0.005	-0.52	0.017±0.001	0.31	0.060±0.005
KLsetMINk	0.07	0.082±0.006	-0.43	0.017±0.001	-0.15	0.139±0.009
KLfluMINk	0.03	0.014±0.003	-0.41	0.014±0.001	-0.68	0.022±0.003

Table 3.3: S_ε indicates the shift of the numerical distribution of ε in the surface layer respect to the in situ one. Δ^2_ε is the squared difference between the numerical and in situ distribution of ε values in the surface layer. Similar definitions apply for different subscript variables. Errors are calculated with a re-sampling procedure.

`Above MLD`	S_ε	Δ^2_ε	S_N	Δ^2_N	S_{K_Z}	$\Delta^2_{K_Z}$
KEset	1.97	0.180±0.006	0.22	0.004±0.0005	0.96	0.032±0.004
KEflu	1.93	0.163±0.008	0.19	0.005±0.0006	0.83	0.039±0.004
KEfluCAN	1.98	0.143±0.008	0.22	0.002±0.0003	0.46	0.021±0.004
KEfluCANMINk	0.91	0.136±0.007	0.22	0.006±0.0006	-0.66	0.023±0.003
KEfluCANMINkOpt	0.88	0.11±0.006	0.22	0.008±0.0009	-0.68	0.022±0.002
KLset	1.21	0.186±0.008	0.22	0.003±0.0005	0.50	0.053±0.004
KLflu	1.2	0.168±0.007	0.23	0.004±0.0005	1.19	0.060±0.005
KLsetMINk	0.26	0.203±0.006	0.22	0.004±0.0006	0.31	0.080±0.007
KLfluMINk	0.29	0.22±0.004	0.24	0.008±0.0009	-0.15	0.022±0.002

Table 3.4: S_ε indicates the shift of the numerical distribution of ε below the surface layer respect to the in situ one. Δ^2_ε is the squared difference between the numerical and in situ distribution of ε values below the surface layer. Similar definitions apply for different subscript variables. Errors are calculated with a re-sampling procedure.

an optimal Symphonie's configuration with respect to our dataset.

The decadal standard deviation values are suited for whole-column comparisons between numerical experiments and in situ data, when an idea on the general behavior of the numerical experiment is of interest (e.g., *Burchard et al.*, 2002). However, we are also interested to details of the model's configuration that likely have a major effect only in the MLD and cannot be appreciated by this metric.

The comparison of the shape of the numerical and in situ probability density functions permits to better investigate the numerical experiments' behavior above the MLD. In particular, it permits to appreciate the model's response to the different configuration choices on the full range of values of ε, N and K_Z. For example, Figure 3.5a-b shows that the effect of raising the threshold value of k is not just a mere shift as it is suggested by the σ_ε values and the median profiles (Figure 3.2). In fact, the numerical experiments with a higher threshold (KEfluCANMINk, KEfluCANMINkOpt, KLsetMINk, KLfluMINk) are not merely shifted towards higher values but also show significant differences in the peaks' heights.

The visual inspection of the median profiles permits to appreciate details that are not immediately captured by the decadal standard deviation and the spectral shape. For instance, the ε median profiles of the numerical experiments with the default threshold on k (KEset, KEflu, KLset, KLflu) suggests that we are underestimating the turbulent activity (3.5a-b). By using Equation 3.1 and a value of $\varepsilon \approx 10^{-4} m^2/s^3$ that we can infer from the in situ data below the MLD (Figure 3.2), we estimated the value of a more appropriate threshold ($k \approx 10^{-7} m^2/s^2$) that was used for the numerical experiments with the higher threshold. As Figures 3.2 and 3.6 and the values of σ_ε and σ_{K_Z} in Table 3.2 indicate, this new threshold permitted to get better estimates of both ε and K_Z.

The analysis of ε data above the MLD permits to get an idea of the effect of b.c., stability functions and optical scheme. The effect of the surface boundary conditions is readily seen by visually inspecting the ε median profiles (Figure 3.2). The KEset, KE-

flu, KLset and KLflu profiles highlight that, at the surface, the equilibrium boundary condition determines ε values lower than both the observations and the values of the corresponding numerical experiments employing a flux b.c. Note that the values of ε close to the surface that one obtains using a certain b.c. also depend on the employed SMC. In fact, the $k-\varepsilon$ and the $k-\ell$ schemes have different values of wind drag coefficients.
As a consequence, an influence of the boundary conditions is also found in the ε's PDFs (Figure 3.5). In fact, the value of ε at which the PDFs die out is different for the schemes employing the equilibrium b.c. and the ones employing the flux b.c. and does not depend on the threshold on k. Indeed, the right peak at higher ε values is mainly due to values from the depth range $0 \leq z/MLD \leq 0.5$ (data not shown).
Moreover, the $k-\varepsilon$ numerical experiments with a flux boundary condition seem to better follow the in situ ε median profile at $z/MLD \approx 0.5$. At this depth, the in situ data show an inflection that we interpret as an average mixing layer depth. This can be justified by inspecting the median profiles of turbulent kinetic energy (see Figure 4.12 in the Extra Materials). In fact, these lasts show that the depth of the inflection in the numerical profiles coincides with the depth at which the kinetic energy has a similar inflection or reaches its minimum value. This depth, thus, identifies a layer in which turbulence is active. Likely, the better performance of the flux b.c. in estimating the mixing layer depth indicates that wave breaking plays an important role in determining the mixing layer depth.

Secondly, the effect of the stability functions is made clear by both the median profiles and the PDFs of ε. In Figure 3.2 it appears that the Canuto A stability functions augment the agreement with respect to the observations (cf. KEflu vs. KEfluCAN), especially near the surface. The same information is given by the PDFs (Figure 3.5; Table 3.3): KEfluCAN's PDF has just two peaks as the in situ PDF and, thus, it is in better agreement with the observations.
While KEfluCAN and KEfluCANMINk show better PDFs, the inflection at the mixing layer depth (that was reproduced by KEflu) is lost. This is likely due to the predominance

of the the effect of the Canuto A stability functions over the effect of the flux b.c. rather than to the threshold on k. In fact, the profile KLfluMINk reproduces well the mixing layer depth.

Thirdly, the effect of the optical scheme is best seen by looking at the ε PDF (Figure 3.5). Indeed, we see that the modified optical scheme enables KEfluCANMINkOpt to reproduce remarkably well the in situ distribution. We attribute this significant improvement to the better estimation of the temperature and salinity profiles (see Figures 4.13 and 4.14 in the Extra Materials) that, in turn, permits to have a better stratification. Indeed, the profile of N (Figure 3.6) of KEfluCANMINkOpt is slightly better than the one of KEfluCANMINk. Even if this is not captured by the values of S_N (Table 3.3), we note that –right below the MLD– KEfluCANMINkOpt better reproduces the bump that we observe in the in situ stratification. However, it is not straightforward to say how this can influence the dynamics above the MLD (indeed the KEfluCANMINkOpt's N PDFs are totally comparable with the ones of KEfluCANMINk).

The analysis of the N data indicate that all the numerical experiments overestimate the stratification above in the surface layer (Figures 3.6a-b). On the other hand the stratification is well reproduced by all the numerical experiments below the MLD. The only absent feature in the numerical data of N is a bump of high values right below the MLD where the entrainment takes place and where we also observe higher levels of turbulence (see Figure 3.2). Overall, below the MLD the numerical PDFs of ε and K_Z markedly differ from the in situ ones: they are more peaked than the in situ PDFs. A possible cause for this behavior could be the too coarse resolution of the model that does not permit to adequately resolve turbulent processes below the MLD or to the different natures of mixing mechanisms.

Also the analysis of the K_Z probability density functions leads to the conclusion that KEfluCANMINkOpt is the better configuration. Even though the values of $\Delta^2_{K_Z}$ in Table

3.3 indicate that there is not a clear difference between KEfluCANMINk and KEfluCANMINkOpt, we note that KEfluCANMINkOpt K_Z PDF has three peaks as the in situ PDF whereas KEfluCANMINk's PDF has only two. In general, the $k-\varepsilon$ scheme performs significantly better in reproducing the K_Z's in situ PDF. This is likely due to the advantage of $k-\varepsilon$ in estimating ε with respect to $k-\ell$.

Overall, the effect of the formulation of K_Z including the molecular diffusivity D_T did not cause significative differences, probably because the turbulence level is too high to allow to appreciate the difference. Here we do not show a numerical experiment in which we change only this aspect of the model's implementation. However, one example is given in the Extra Materials (Figure S5).

The analysis of the data below the MLD shows a sharp disagreement between the numerical and the in situ values of ε and K_Z. However, we can observe three main features: i) the typical value of numerical ε depends on the SMC that one employs; ii) the model values of ε and K_Z are much less variable than the measurements; iii) setting an hard threshold on the kinetic energy limits the reliability of the modeling this layer.
The first point is made clear by Figure 3.2. In fact, we can see that when the same threshold on k is applied, the ε values obtained with the $k-\varepsilon$ scheme are one order of magnitude smaller than the ones obtained with the $k-\ell$ scheme.
The second point is illustrated by Figures 3.5c-d and 3.6c-d. Therein, the numerical PDFs in the lower layer are much less spread out than the in situ ones. A possible cause for this could be the resolution of the model that does not permit to adequately resolve the turbulent processes in this layer.
Also, a possible cause for the sharp discrepancy between numerical and in situ PDFs below the MLD could be the use of a hard threshold for k. In fact, Figure 3.8 suggests that this choice leads to a non-physical result, with many low-energy points on a single plane in the space (K_Z, N, ε). This behavior descends directly from Equations B.6 and B.7 implying $\varepsilon = c_k c_\varepsilon k^2/K_Z$ for the KL scheme and from Equations B.3 and B.4

implying $\varepsilon = \sqrt{2} c k^{5/6} / K_Z$ for the KE scheme (where c_k, c_ε and c are constants). This is exactly the behavior we observe in Figure 3.8b, with the planes following the line $\log \varepsilon = -\log K_Z + const$. Thus, the points on the planes in Figure 3.8 are values with a minimal kinetic energy. In situ data do not show this behavior, suggesting that specifying a hard value for k_{min} is likely not the best choice. A more suitable approach would probably be a parametrization of k_{min} on the base of the different turbulent processes at play. For example, *Gaspar et al.* (1990) hypothesized the use of a parametrization on the internal wave field activity when dealing with data in a depth range similar to the one in our analysis. Note here that such a solution is likely to depend on the amount of numerical mixing and to the processes that are resolved by each model.

3.1.6 Conclusions

Vertical mixing in the surface layer of the ocean plays an essential role in both physical and biochemical phenomena. Therefore, it must be correctly estimated. Different turbulence closure schemes have different performances in predicting mixing. Moreover, the problem is made more complicated by the interplay of the SMC with other aspects of ocean dynamics numerical modelling like boundary conditions, stability functions and optical scheme. The influence of the minimum value of kinetic energy that is allowed in has also been investigated.

We performed nine numerical experiments and compared their estimates of the turbulent quantities ε and K_Z to in situ microstructure measurements in the Gulf of Lion. In particular, two SMCs were considered: a $k - \varepsilon$ scheme proposed by *Burchard et Bolding* (2001) (herein KE); and a $k - \ell$ based on *Gaspar et al.* (1990) (herein KL). We considered two surface b.c.: one supposing equilibrium between the production and dissipation terms in the dynamic equation for k (herein set); and one taking into account the effect of breaking waves of all scales based on *Craig et Banner* (1994) (herein flu). We considered the stability functions proposed by *Kantha et Clayson* (1994) and *Canuto et al.* (2001). In addition, we also considered two different attenuation lengths for the photosynthetically available radiation in the model's optical scheme. The combinations of these factors that we explored are resumed in Table 3.1.

A recent study by *Reffray et al.* (2015) showed that, in a 1-D situation, the KE closure scheme gives better mixing estimates than other widely used SMCs . However, that study did not compare KE to a mixing length scheme based on *Gaspar et al.* (1990). Our study fills this gap. Moreover, our study exploits a 3-D model that, by definition, considers more terms in the dynamical equations than what 1-D models do. Also, our study compares the modeled turbulence activity directly to microstructure turbulence measurements rather than with derived quantities like the mixed layer depth.
The two SMCs do not show relevant differences when estimating K_Z both above and below the MLD. The fact that the two SMC perform similarly in estimating K_Z when KL has an advantage in estimating ε, is not explained by differences in estimating N and need further investigation in the future.
Previous studies, in 1-D and 3-D numerical simulations, gave conflicting indications on the effect of the stability functions proposed by *Kantha et Clayson* (1994) and *Canuto et al.* (2001) on the KE scheme. In our study, the effect of the Canuto A stability functions on the KE is to improve the performance of the closure scheme when the ε's and K_Z's PDFs are considered.

Our study shows that choosing the value of the kinetic energy threshold plays a pivotal role in approaching the observations below the MLD (see Figures 3.2 and 3.4). Nonetheless, we found a non physical behavior of the numerical experiments for low kinetic energy levels. This supports the idea that the minimum of kinetic energy should rather be parametrized as function of different turbulent mechanisms rather than being a hard threshold (as already hypothesized by *Gaspar et al.*, 1990). Our study also suggests that such a parameterization should also depend on the SMC that is employed. Indeed, we found that the two SMC predict different turbulence levels below the MLD even when they employ the same threshold on the turbulent kinetic energy.

We found that the attenuation length of the photosynthetically available radiation plays an important role in determining the stratification and, as a consequence, the performance of the model in predicting ε and K_Z. This result highlights the importance of biological activity in influencing physical processes.

Our study shows that the comparison with in situ microstructure data can effectively help in setting up the implementation of a SMC for an ocean numerical model. In the future, we should expect more studies of this kind as new automated instrumentation becomes available. In particular, microstructure and wave height probes mounted on autonomous platforms, such as drifting profilers or gliders, will permit to have new and bigger datasets, especially during adverse meteorological situations. Moreover, this kind of data will provide essential data to developers of future ocean-wave coupled models, as these are expected to significantly advance the ocean modelling state of the art by reducing the need of general surface boundary conditions.

Acknowledgements

The authors want to thank B.Ruddick for helping verifying the implementation of the SCAMP data analysis code. Also we thank F.Cyr, L.Jullion, O.Ross and D.Nerini for helpful discussions. Surface buoyancy fluxes were calculated with the WHOI air-sea toolbox. Data were collected during the LAgrangian Transport EXperiment (PIs A.Petrenko and F.Diaz); the RHOMA2 campaign (PI I.Pairaud); the SPECIMED cruises (PI B.Queguiner); the SUNMEX campaign (PI R.Sempere); the SUBCORAD campaign (PI P.Fraunie) and some one-day teaching cruises of AMU Master of Oceanography specialty OPB. We especially thank D.Malengros for the data acquisition.

This study was carried out in the framework of the project NUMEROFIX (NUmerical ModElling of Real Ocean surFace mIXing), funded by LEFE/GMMC. Simulations were performed using HPC resources from CALMIP (Grant 2016-P1325).

3.2 Appendices

3.2.1 Boundary conditions

3.2.2 Surface boundary conditions for k

Equilibrium

The k surface boundary conditions can be obtained hypothesizing equilibrium between production and dissipation. Using the surface momentum boundary conditions:

$$K_Z\left(\frac{\partial u}{\partial x},\frac{\partial v}{\partial y}\right)=\frac{(\tau_x,\tau_y)}{\rho_0}$$

and the wind stress $\tau=\sqrt{\tau_x^2+\tau_y^2}$, we obtain:

$$k_s=\frac{\tau/\rho_0}{\sqrt{2^{1/2}c_0^3S_M}}$$

Flux

Alternatively, the boundary conditions can be specified as surface flux conditions, namely:

$$K_Z\frac{\partial k}{\partial z}=F \tag{3.2.1}$$

where the surface flux can be computed as $F=100(\tau/\rho_0)^{3/2}$ (*Craig et Banner*, 1994) or directly prescribed from the 'wave to ocean' turbulence flux computed by a wave model when available.

3.2.2.1 Bottom and surface boundary conditions for ε

The ε surface and bottom conditions are computed on the first level under the surface and above. Let z_1 denotes the distance between this level and the considered boundary. Boundary conditions for ε are obtained from k and Equation 3.2.3.2, using the latter with some appropriate hypothesis for l_B a boundary length scale value. A simple formulation (*Warner et al.*, 2005) is eventually given by $l_B=0.4(z_1+z_0)$, where z_0 is a length scale representing the roughness of the bottom boundaries. Unfortunately, this formulation potentially leads to unrealistic high values when the underlying hypothesis of neutral stratification is no longer valid (a situation that is more likely to occur in deep zone where the vertical grid resolution near the bottom is generally coarse). One way to solve this problem is to introduce a dependency on the Richardson number (*Estournel et Guedalia*, 1987; *Michaud et al.*, 2012):

$$l_B = \begin{cases} 0.4(z_1 + z_0) & \text{if} \quad Ri < 0 \\ \\ 0.4(z_1 + z_0)(1 - 5Ri) & \text{if} \quad 0 \leq Ri \leq 0.16 \\ \\ 0.4(z_1 + z_0)(1 + 41Ri)^{-0.8} & \text{if} \quad Ri > 0.16 \end{cases}$$

3.2.3 Closure schemes

3.2.3.1 $k - \varepsilon$ scheme

Following *Burchard et Bolding* (2001), the equations describing the dynamics for k and ε are:

$$\frac{dk}{dt} = \frac{\partial}{\partial z}\left(K_Z \frac{\partial k}{\partial z}\right) + P + B - \varepsilon \tag{3.2.2}$$

$$\frac{\partial \varepsilon}{\partial z} = \frac{\partial}{\partial z}\left(\frac{K_Z}{\sigma}\frac{\partial k}{\partial z}\right) + \frac{\varepsilon}{k}(c_1 P + c_3 B - c_2 \varepsilon) \tag{3.2.3}$$

where $P = K_Z \left\{ \left(\frac{\partial u}{\partial z}\right)^2 + \left(\frac{\partial v}{\partial z}\right)^2 \right\}$ and $B = \frac{g}{\rho_0} K_T \frac{\partial \rho}{\partial z}$ are the production terms due to shear and buoyancy respectively. Parameters values issued from *Warner et al.* (2005) are given in Table 3.5.

The eddy viscosity K_Z and the temperature eddy diffusivity K_T used in Equations 3.2.2 and 3.2.3 are given by:

$$K_Z = \sqrt{2k} l S_M, \quad K_T = \sqrt{2k} l S_H \tag{3.2.4}$$

Table 3.5: Free-parameters of the SYMPHONIE $k-\varepsilon$ closure scheme (*Warner et al.*, 2005).

Free parameters values	
sigm	$\sigma = 1.3$
sigm	$c_0 = 0.5544$
sigm	$c_1 = 1.44$
sigm	$c_2 = 1.92$
sigm	$c_3 = 1 \quad B \geqslant 0$ $c_3 = -0.52 \quad B < 0$

The turbulent length l is related to k, TKE and ε according to

$$l = c_0^3 k^{3/2} \varepsilon^{-1} \tag{3.2.5}$$

S_M and S_H , the quasi-equilibrium stability functions defined by *Kantha et Clayson* (1994), depend on the Richardson flux number.

3.2.3.2 $k-\ell$ scheme

We used the $k-\ell$ closure scheme proposed by *Gaspar et al.* (1990). Therein the authors assume ε to be (following *Kolmogorov*, 1942):

$$\varepsilon = c_\varepsilon k^{3/2} / l_\varepsilon \tag{3.2.6}$$

where k is the TKE, $k = \sqrt{u'^2 + u'^2 + w'^2}$ and $c_\varepsilon = 0.7$ (following *Bougeault et Lacarrere*, 1989).

The eddy momentum diffusivity is related to TKE according to:

$$K_Z = c_k l_k \sqrt{k}$$

where c_k has to be determined.

The dissipation and mixing length scales l_k and l_ε are the ones proposed by *Bougeault et Lacarrere* (1989):

$$l_\varepsilon = (l_u l_d)^{1/2}$$

$$l_k = \min(l_u, l_k)$$

that are determined through

$$\frac{g}{\rho_0}\int_z^{z+l_u}[\overline{\rho}(z)-\overline{\rho}(z')]dz' = k(z)$$

$$\frac{g}{\rho_0}\int_z^{z-l_d}[\overline{\rho}(z)-\overline{\rho}(z')]dz' = k(z)$$

where ρ is the water density.
As highlighted by *Gaspar et al.* (1990), these length scales have a straightforward physical interpretation: they are the distances traveled upward/downward by a fluid particle by converting all of its kinetic energy into potential energy.

Hypothesizing the turbulence to be stationary and homogeneous *Gaspar et al.* (1990) show that, in stably stratified regions, the model parameterizations yields:

$$K_Z = \sqrt{2}Pr^{-1}c_k kN^{-1} \tag{3.2.7}$$

$$\varepsilon = \frac{1}{\sqrt{2}}c_\varepsilon kN \tag{3.2.8}$$

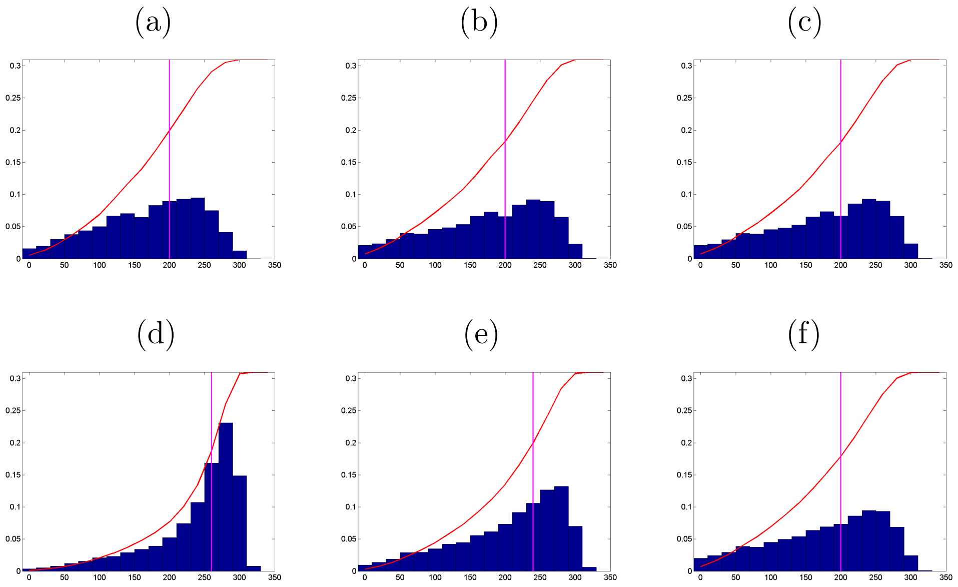
from which it follows that $c_k = 0.1$.

3.3 Segmentation and quality fit

In order to apply the Batchelor's theory to the temperature gradient one has to divide the data into segments in which turbulence can be considered homogeneous, viz. the turbulent motion can be regarded as a random motion which average properties are independent of position in the fluid. Given the fact that the first aim is to fit the data to a theoretical Batchelor's spectrum, arguably the best segmentation method is the one giving the fits with the best quality.

To determine the fit quality we use a Maximum Likelihood Estimation technique. Following *Ruddick et al.* (2000), we calculated the goodness of a fit by maximizing the joint probability of the measured spectrum with respect to the theoretical Batchelor spectrum. Hence, the higher the joint probability, the higher the fit goodness. In the following we call this joint probability JP (corresponding to $C11$ in *Ruddick et al.*, 2000).
To make the fit goodness criteria more rigorous, following *Sanchez et al.* (2011) we also require that: i) the mean absolute deviation (MAD) of the ratio between the in situ and theoretical spectra within the fitting domain be lower than 1.1; ii) the signal to noise ratio (SNR) be lower than 1.3; iii) the likelihood ratio (LHR) –which quantifies if the measured spectrum fits the Batchelor's spectra or a power-law spectrum better– be lower than 2.

We tested which of the following segmentation methods permitted to have the better fits in our dataset: a segmentation based on an eight order AR model by *Imberger et Ivey* (1991) (II91); a constant segmentation of 1024 data points in which the values of temperature gradient variance in 7 sub-segments are in the same order of magnitude proposed by *Sanchez et al.* (2011) (S11), a constant segmentation of 1024 data points with no overlap suggested by *Cuypers et al.* (2012) (C12); a constant segmentation of 128 data points with 50% overlap suggested by *Moniz et al.* (2012) (M12); a constant 512 data points segmentation with no overlap; and a constant 1024 data points segmentation with 30% overlap. In practice, we fitted the segmented data to the Batchelor's spectrum following the fitting procedure of *Steinbuck et al.* (2009) and we compared the results searching for the distribution of JP values with the greater proportion of high values. The results are depicted in Figure 3.9.


(a) (b) (c)

(d) (e) (f)

Figure 3.9: a) JP values distribution obtained with the II91 segmentation method, b) same as a) for S11, c) same as a) for C12, d) same as a) for M12. e) same as a) for a constant 512 points segmentation. f) same as a) for a 1024 points segmentation with 30% overlap. The red curve is the cumulative distribution of JP values for each case and the vertical line marks the point where the cumulative distribution equals 0.6. The database for this analysis comprehended 126 profiles with an average depth of 50m in various meteorological conditions in the Gulf of Lion.

In Figure 3.9 the histogram of the JP values is plotted with the cumulative distribution of the latter. Note that each histogram is normalized to the total number of segments obtained with the corresponding segmentation method. As it is evident by eye, the M12 is the segmentation method that gives the highest percentage of good profiles, as highlighted by the higher histogram bars on the right end of the distribution with respect to the other cases. Moreover this is the method that has the vertical line lying at the highest value, meaning that a higher percentage of segments have a better fit. Furthermore cases b), e) and d) show that progressively reducing the segmentation window gives better results. Case f) also shows, when compared to c), that adding overlap, though permitting to have a higher resolution, does not augment the percentage of good fits. These evidences suggest to apply a 128 points ($\approx$ 12.5 cm) segmentation with no overlap. Anyway, not needing such a high resolution in order to compare the in situ data with the numerical data, we avoided the overlap. The distribution of JP values of this segmentation method did not significantly differ from the one of M12 (data not shown).

3.4 Extra Materials for "Comparison of in situ microstructure measurements to different turbulence closure schemes in a 3-D numerical ocean circulation model"

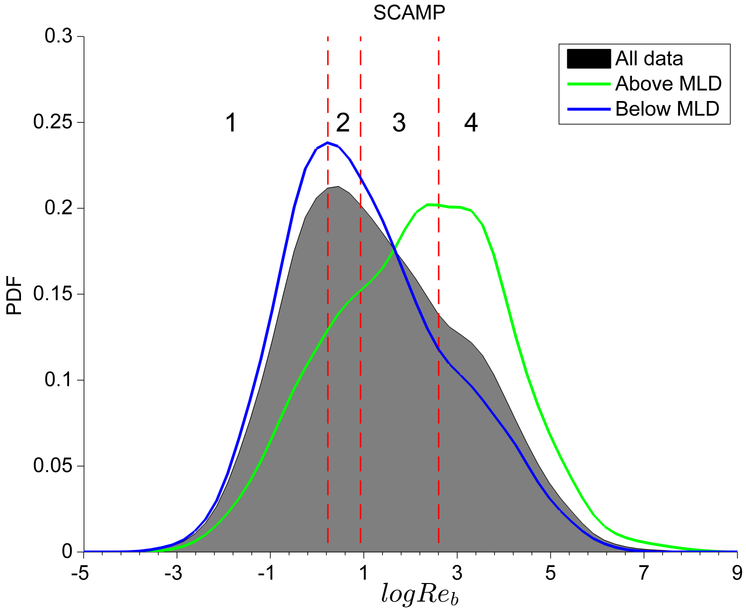


Figure 3.10: Probability density function of the turbulence intensity parameter Re_b, obtained with all the in situ data (black), the in situ data above the MLD (green) and the in situ data below the MLD (blue). The dotted red lines indicate the separation between the turbulence regimes. 1 indicates the diffusive regime; 2 the buoyancy-controlled regime; 3 the transitional regime; and 4 the energetic regime.

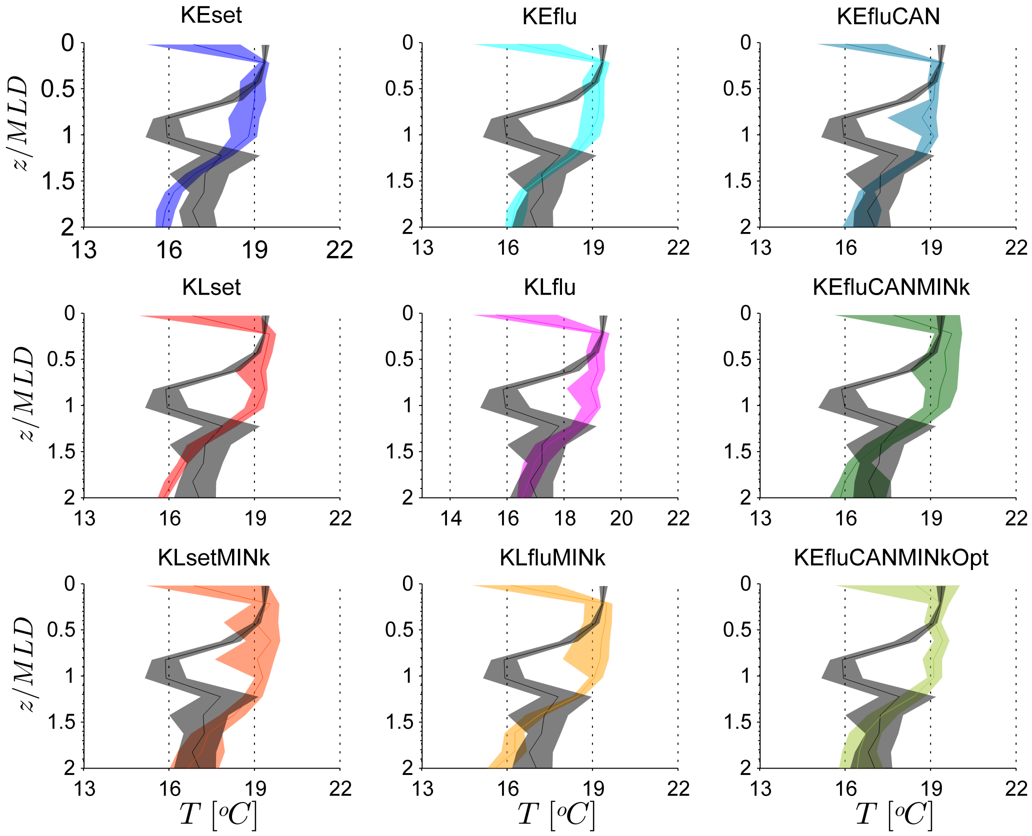


Figure 3.11: The thick lines represent median values of temperature measured from in situ data (in black) and numerical experiments. The shades indicate 95% bootstrap confidence intervals. Water depth z is adimensionalised with respect to the mixed layer depth MLD for each profile.

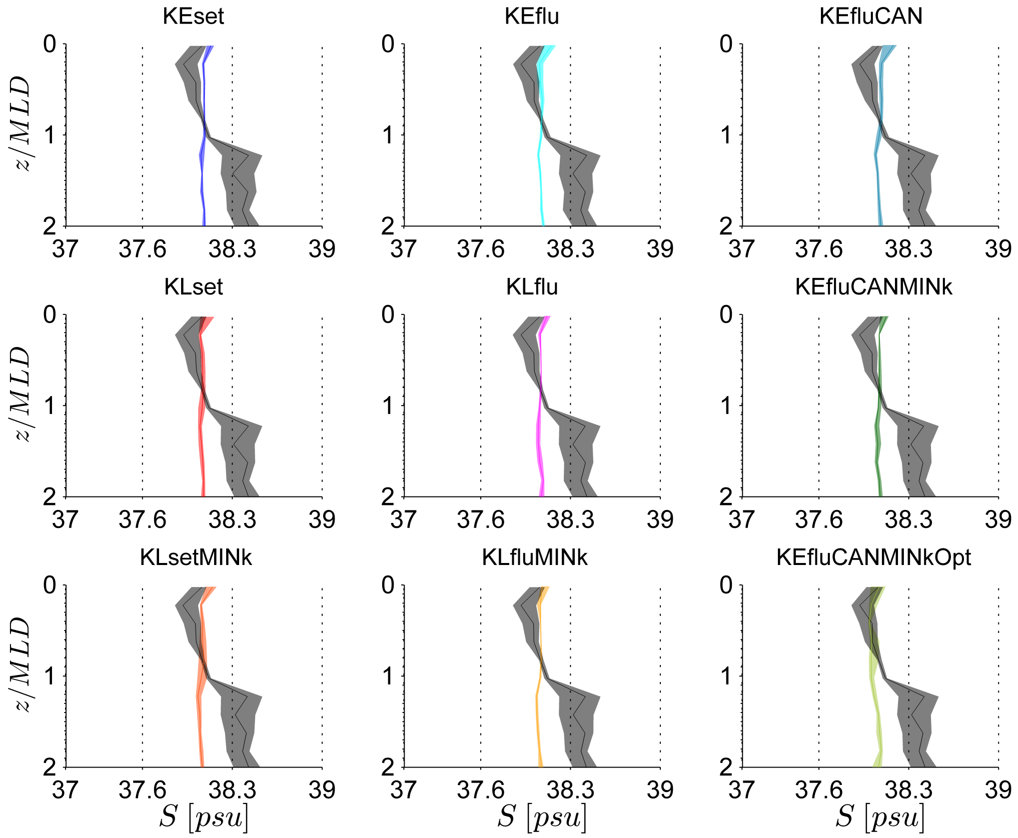
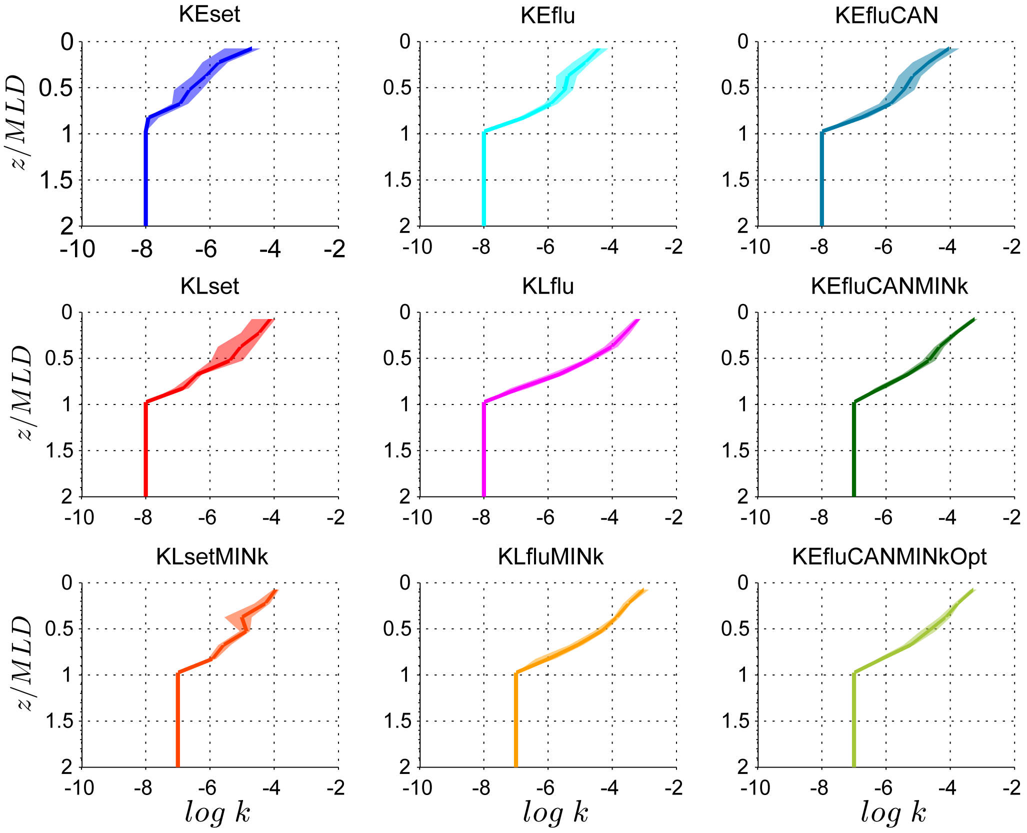


Figure 3.12: The thick lines represent median values of density measured from in situ data (in black) and numerical experiments. The shades indicate 95% bootstrap confidence intervals. Water depth z is adimensionalised with respect to the mixed layer depth MLD for each profile.



KEset KEflu KEfluCAN

KLset KLflu KEfluCANMINk

KLsetMINk KLfluMINk KEfluCANMINkOpt

Figure 3.13: The thick lines represent median values of turbulent kinetic energy measured from in situ data (in black) and numerical experiments. The shades indicate 95% bootstrap confidence intervals. Water depth z is adimensionalised with respect to the mixed layer depth MLD for each profile.

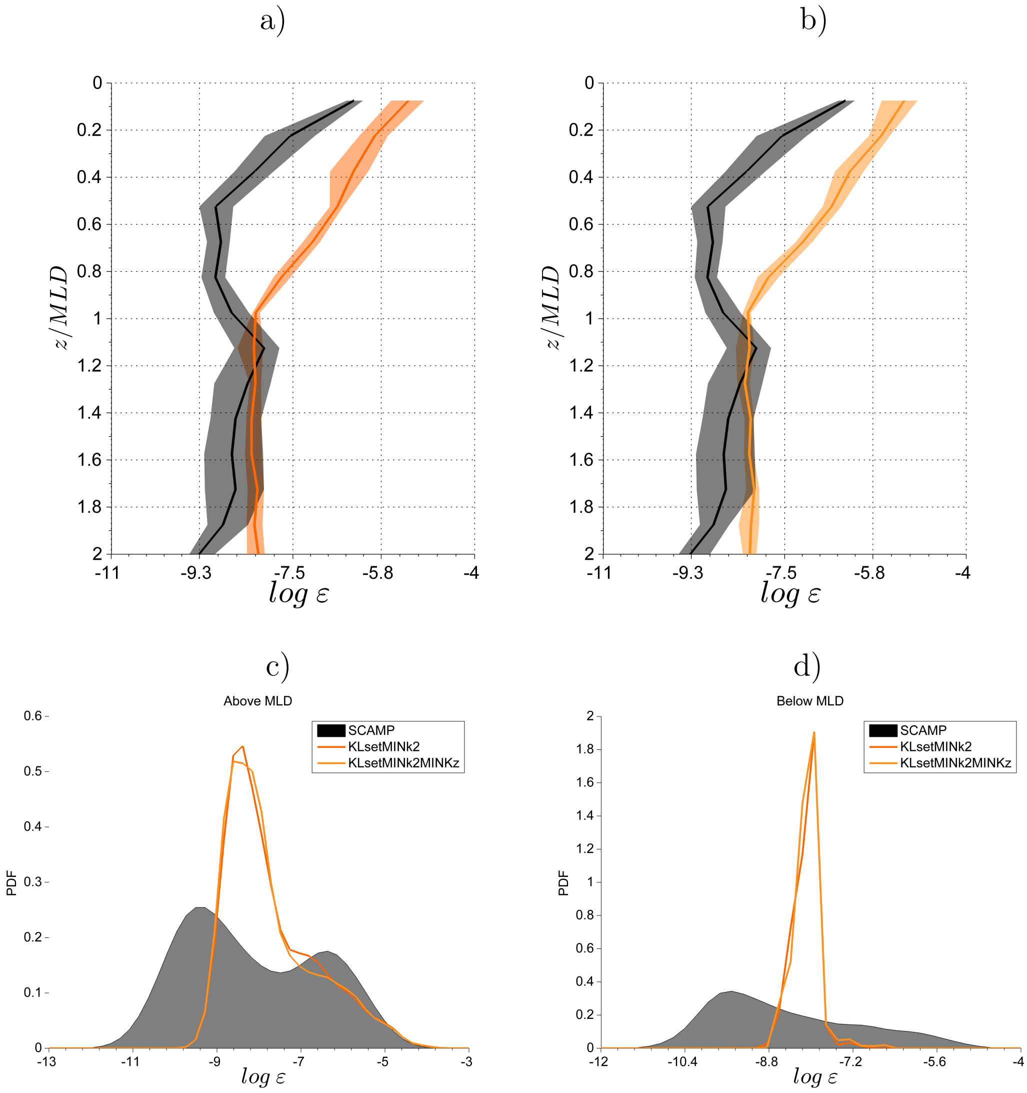


Figure 3.14: In the upper panels the median profiles of kinetic energy dissipation rate for the two numerical experiments with $k_{min} = 10^{-6} m^2/s^2$ (suffix MINk2) are represented. In a) KLsetMINk2 has molecular diffusion set to zero. In b) KLsetMINk2MINKz has not null molecular diffusion. In the lower panels we represent the PDF of the ε values obtained for these numerical experiments both above (in c) and below (in d) the MLD.

3.5 Extra Material

3.5.1 Additional Simulations

In the first version of the paper submitted to Ocean Modelling, I included different numerical experiments. Some were not included in the new version of the paper. These numerical experiments exploited a higher threshold on the turbulent kinetic energy ($k_{min} = 10^{-6} m^2/s^2$). This value was used on the base of previous literature (e.g., *Gaspar et al.*, 1990). This particular value improved the agreement between the numerical model and the in situ data in the lower layer. However, the agreement within the mixed layer was lower due to the too high level of turbulence induced by this threshold. This is showed in Figures 3.15 and 3.16 on pages 110 and 111. Therein, KEsetMINk is a $k - \varepsilon$ scheme with equilibrium boundary conditions and $k_{min} = 10^{-6} m^2/s^2$; KLsetMINk is a $k - \ell$ scheme with equilibrium boundary conditions and $k_{min} = 10^{-6} m^2/s^2$; and KLsetMINKz is a $k - \ell$ scheme with equilibrium boundary conditions and $k_{min} = 10^{-6} m^2/s^2$ and such that $K_Z > D_T$.

3.5.2 Data analysis toolbox

A considerable part of my work on the SCAMP data consisted in checking and further developing the data analysis toolbox. Main changes (in red in Figure 3.17) consisted in the data segmentation technique, error estimation, bug correction. Moreover, I changed the calculation of K_Z following (*Bouffard et Boegman*, 2013), coded the calculation of adimensional Kraichnan profiles and modified the code in order to automatically analyze multiple profiles. The toolbox is freely available at `http://mio.pytheas.univ-amu.fr/~costa.a/SCAMP/scritps/`.

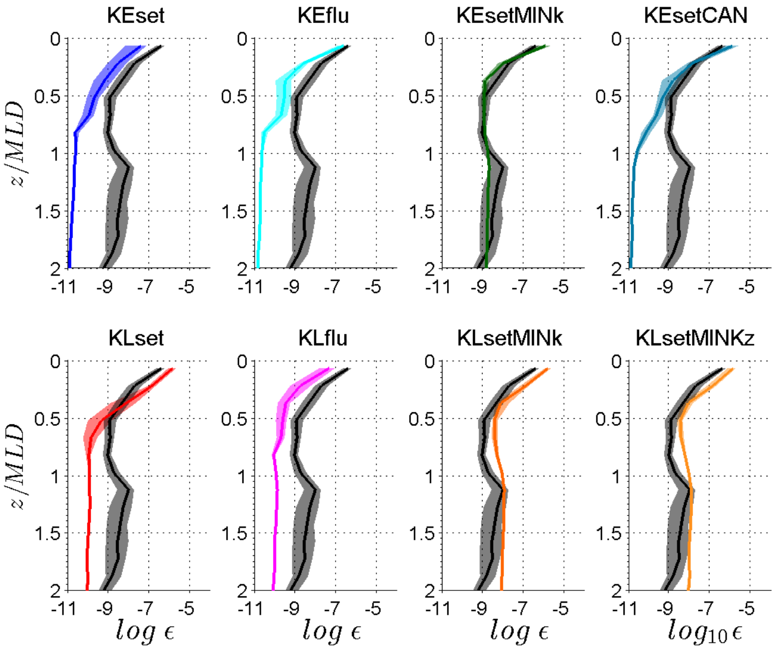


Figure 3.15: Profiles of kinetic energy dissipation rate for all the numerical experiments and the in situ data.

Data segmentation

In order to apply the Batchelor's theory to the temperature gradient one have to divide the data into segments in which turbulence can be considered homogeneous, viz. the turbulent motion can be regarded as a random motion which average properties are independent of position in the fluid.

The segmentation problem is not trivial due to the dependency of the measure on four coordinates: three spatial variables $\vec{x}$ and time t. Consistently with the Batchelor's theory one can use the hypothesis of horizontal homogeneity to reduce the number of variables to only two: z and t.

A first approach was suggested by *Imberger et Boashash* (1986), *Imberger et Ivey* (1991) and *Chen et al.* (2002). They ignored the dependence on space and assessed the stationarity of the temperature gradient time series by exploiting Wigner-Ville distribution or AR models properties or wavelet analysis respectively. A second approach is to neglect the

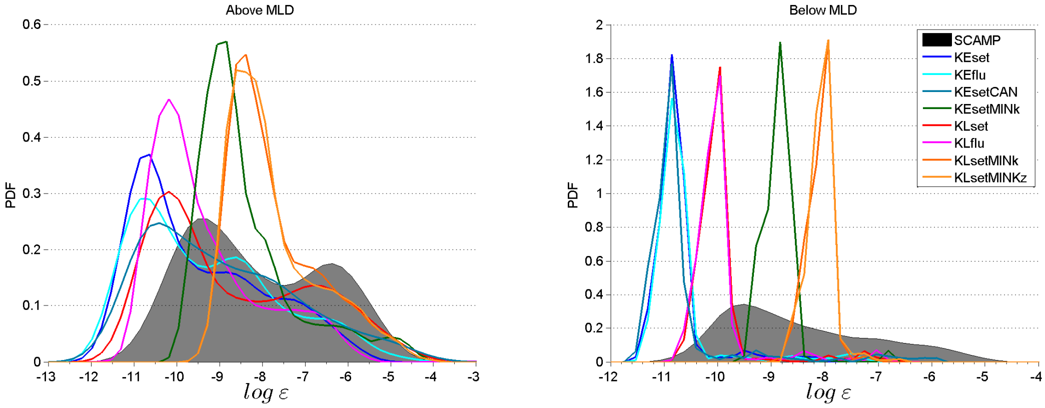
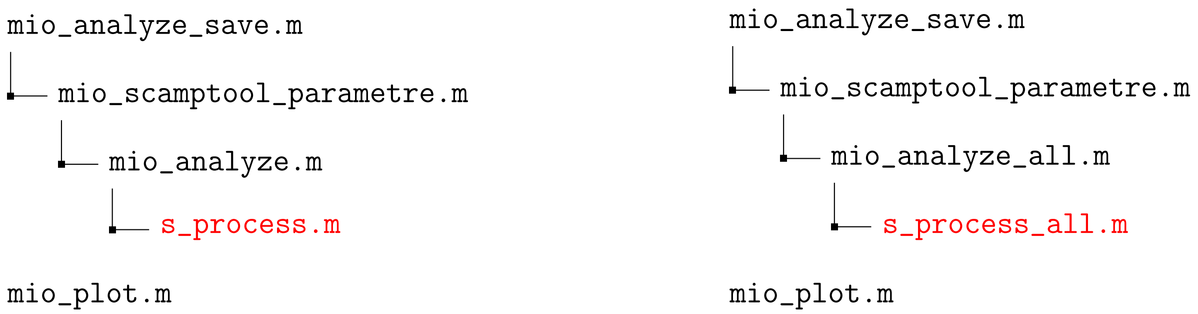


Figure 3.16: Probability density functions of kinetic energy dissipation rate for all the numerical experiments and the in situ data both above the MLD (a) and below (b).

dependency on time. *Sanchez et al.* (2011) considered satisfying the segments in which the values of temperature gradient variance in 7 subsegments are in the same order of magnitude. A third approach neglecting both spatial and temporal dependency of the turbulent process was more recently proposed by *Cuypers et al.* (2012) and *Moniz et al.* (2012). The authors used a constant segmentation of 1024 data points ($\approx 1m$) with no overlap and 100 data points ($\approx 0.1m$) with a 50% overlap respectively.
All these segmentation techniques are now coded in the s_process.m script. The comparison of their performance on our dataset is detailed in Section 3.3.

Error estimation

The SCAMP measures the temperature T with a millimeter scale resolution. From this it calculates the temperature gradient using windows of 128 values. From the gradient we can derive the Batchelor spectrum linking T and the kinetic energy dissipation rate ε. In fact, the spectrum is a function of the temperature variance dissipation rate χ_T and the Batchelor wavelength k_B. However, from the temperature we can derive χ_T but k_B remains unknown. To determine it, a Maximum Likelihood Estimation approach is used. This approach returns the k_B that maximises the probability of the measures (*Ruddick*


mio_analyze_save.m
mio_scamptool_parametre.m
mio_analyze.m
s_process.m
mio_plot.m
mio_analyze_save.m
mio_scamptool_parametre.m
mio_analyze_all.m
s_process_all.m
mio_plot.m

(a) Folder tree analyzing SCAMP profiles one by one.

(b) Folder tree analyzing all SCAMP profiles in a folder.

Figure 3.17: Folder tree of the mio_scaptool toolbox. Red scripts highlights where the changes detailed in the following can be found.

et al., 2000). The error on k_B is given by the Cramer-Rao inequality:

$$V = var(k_B) \geq -\frac{1}{E[\partial^2 \ln P/\partial k_B^2]} \tag{3.5.1}$$

Given k_B we can derive ε (*Gibson et Schwarz*, 1963):

$$\varepsilon = \nu D^2 \cdot k_B^4 \tag{3.5.2}$$

where ν is the water molecular viscosity and D the temperature molecular diffusivity.

Therefore, we have a value of ε every 13 centimeter with an error coming from the error on k_B. In the literature the error on ε is calculated using the relative standard deviation $RSD = \sigma/\mu$, where σ and μ are the standard deviation and the mean of the measurements (*Jurado et al.*, 2012). In particular, with error propagation, it can be shown that:

$$RSD(\varepsilon) = 4 \cdot RDS(k_B) = 4 \cdot RSD \tag{3.5.3}$$

In fact, if we define ε like this:

$$\varepsilon = \alpha k^4 \tag{3.5.4}$$

with $\alpha = \nu D^2$ (cf. Equation 3.5.4) and $k = k_B$ for ease of notation; and the standard deviation σ_k for k:

$$\sigma_k = \sqrt{\frac{\sum_{i=1}^{N}\left(k_i - \overline{k}\right)^2}{N-1}} \ , \tag{3.5.5}$$

where $\overline{k} = \sum_{i=1}^{N} k_i$.

By defining the N individual errors on k as

$$\delta_{k_i} = k_i - \overline{k} \ , \tag{3.5.6}$$

we can relate them to the N errors in ε making use of partial integration:

$$\delta_{\varepsilon_i} = \frac{\partial \overline{\varepsilon}}{\partial \overline{k}} \delta_{k_i} = 4\alpha \overline{k}^3 \delta_{k_i} \ . \tag{3.5.7}$$

Therefore the standard deviation of ε is equal to:

$$\sigma_\varepsilon = \sqrt{\frac{\sum_{i=1}^{N} \delta_{\varepsilon_i}^2}{N-1}} = \sqrt{\frac{\left(4\alpha\overline{k}^3\right)^2 \sum_{i=1}^{N} \delta_{k_i}^2}{N-1}} = 4\alpha\overline{k}^3 \sigma_k \ . \tag{3.5.8}$$

Dividing this by the average of ε:

$$\frac{\sigma_\varepsilon}{\overline{\varepsilon}} = \frac{4\alpha\overline{k}^3\sigma_k}{\alpha\overline{k}^4} = 4\frac{\sigma_k}{\overline{k}} \ . \tag{3.5.9}$$

Now, we want to compare the ε values measured with the SCAMP with the one predicted by the model. One problem is that the model has a much coarser resolution than the SCAMP. In particular, we compare the model value with the median experimental value M of the values between the model depths.

How can we derive the error on the median from the errors on the ε values? One appealing way that I developed for doing it is reported here below. Unfortunately, it turns out it cannot be applied to ε data because they are distributed lognormally (e.g.,*Moum et al.*

1995). I report it hoping it will potentially be useful for other data.

The median of ε that we compare to the model is the median of 10-13 in a portion of the water column. In the literature, it is consolidated that these values are distributed lognormally:

$$p(\varepsilon) = \frac{1}{\sigma\sqrt{2\pi}} \exp -(\ln \varepsilon - \mu)^2/2\sigma^2 \tag{3.5.10}$$

where:

$$\mu = \int_0^\infty \varepsilon p(\varepsilon) d\varepsilon \tag{3.5.11}$$

and

$$\sigma = \int_0^\infty (\varepsilon - \mu)^2 p(\varepsilon) d\varepsilon \tag{3.5.12}$$

Given these relations, it is possible to calculate the standard deviation and the median of ε^S (σ_M).

In particular, if we can solve this integral:

$$\sigma_M = \int_0^\infty (\varepsilon - M)^2 p(\varepsilon) d\varepsilon \tag{3.5.13}$$

where M is the median of $p(\varepsilon)$; and noting that

$$\sigma_M = \int_0^\infty (\varepsilon^2 - 2\varepsilon M - M^2) p(\varepsilon) d\varepsilon \tag{3.5.14}$$

we can find an analytic relationship between M and μ. As a consequence, we can find the error on M to be something along the lines of $\delta M = \alpha\delta\mu = \alpha\sigma$.

The exact relationship can be found by solving the following equations:

$$P(\varepsilon \leq M) = \int_0^M p(\varepsilon) d\varepsilon = \cdots = 1/2 \tag{3.5.15}$$

$$P(\varepsilon > M) = \int_M^\infty p(\varepsilon)d\varepsilon = \cdots = 1 - 1/2 = 1/2 \tag{3.5.16}$$

However, this technique is not analytically solvable if we assume a lognormal probability distribution of ε. It certainly would by assuming a normal distribution.

An alternative is calculate the error on the SCAMP measures with bootstrap. In particular, if we estimate the error on the whole water column between the model values ε_M and the in situ measurements ε_S like:

$$S = \sqrt{\frac{\sum_{i=1}^{N} (\log \varepsilon_i^M - \log \varepsilon_i^S)^2}{N}} \tag{3.5.17}$$

To calculate the error on S we can use the classic error propagation technique:

$$\delta S = \sqrt{\left(\frac{\partial S}{\partial \varepsilon^M}\right)^2 (\delta\varepsilon^M)^2 + \left(\frac{\partial S}{\partial \varepsilon^S}\right)^2 (\delta\varepsilon^S)^2} \tag{3.5.18}$$

However, the errors $\delta\varepsilon^M$ and $\delta\varepsilon^S$ are not constant so that we must rewrite the equation:

$$\delta S = \sqrt{\sum_{j=1}^{N} \left[\left(\frac{\partial S}{\partial \varepsilon_j^M}\right)^2 (\delta\varepsilon_j^M)^2\right] + \sum_{j=1}^{N} \left[\left(\frac{\partial S}{\partial \varepsilon_j^S}\right)^2 (\delta\varepsilon_j^S)^2\right]} \tag{3.5.19}$$

where $\delta\varepsilon_j^S$ can be calculated by comparing the measured median value and the bootstrap estimate calculated by extracting 10-13 ε values from a lognormal distribution whose parameters are determined on the dataset at study.
mbox Some exploratory analysis of this technique gave a 6% error estimate on the SCAMP median value. $\delta\varepsilon_j^M$ can also be calculated with bootstrap by, for example, looking at successive profiles.
Note that this error estimation could be useful in a profile-profile comparison between in situ and numerical data. However, we did not need to estimate the error in the above way for the study in Section 3.

Bug in the code

We noticed that in the section $IX.DETERMINE\quad QUALITY\quad FACTOR$ in the file $s\_batfit.m$ supplied by PME, the fit quality factor R (see Section 3.1.3) is badly defined. The user should substitute the line $R = std((PSD\_bat./Batchelor\_spect)) * dof\hat{\ }0.5;$ with $R = var(PSD\_bat./Batchelor\_spect);$ in order to be coherent with Ruddick et al. (2000). We communicated the issue to PME who will correct the bug in its next release.

Chapter 4

Bottom boundary mixing and diffusion

This work has been submitted to the Journal of Geophysical Research: *Spatially Inhomogeneous and Temporally Intermittent Boundary Mixing along the Northern Deepwater Gulf of Mexico*, K. Polzin, A. Costa, A. Ruiz-Angulo, A. Thurnherr, J. Ledwell, Z. Wang, S. DiMarco - 2016JC012310

Contents

4.1 Spatially Inhomogeneous and Temporally Intermittent Boundary Mixing along the Northern Deepwater Gulf of Mexico 123

4.1.1 Abstract 123

4.1.2 Introduction 124

4.1.3 Data and Methods 128

4.1.4 Observations 133

4.1.5 Boundary Mixing 135

4.1.6 Summary and Discussion 140

4.2 Extra Material 154

On a global scale, the ocean deep circulation can be viewed in terms of interconnected overturning cells that transport cold dense abyssal waters formed at high latitudes back to the surface (see Figure 4.1). This meridional overturning circulation (MOC) has been known for a long time (*Sverdrup* 1933; *Stommel* 1958) but still the mechanisms that provide the energy to sustain it are not well understood. We know that the main energy source for the MOC is the wind (*Ferrari et Wunsh*, 2009). However, there still is not a consensus on how this energy reaches the deep ocean and how it goes up through mixing. Indeed, the mechanism causing mixing are the ones that permit water motion across isopycnals (see Figure 4.2 on page 122). Classically, *Munk* (1966) estimates that an average value of turbulent viscosity $K_\rho = 10^{-4} m^2/s^2$ is needed to explain the observed overturning circulation. Such a value is hardly measured in the open ocean. Therefore, in the recent years, the question has shifted on where this "missing" mixing takes place.

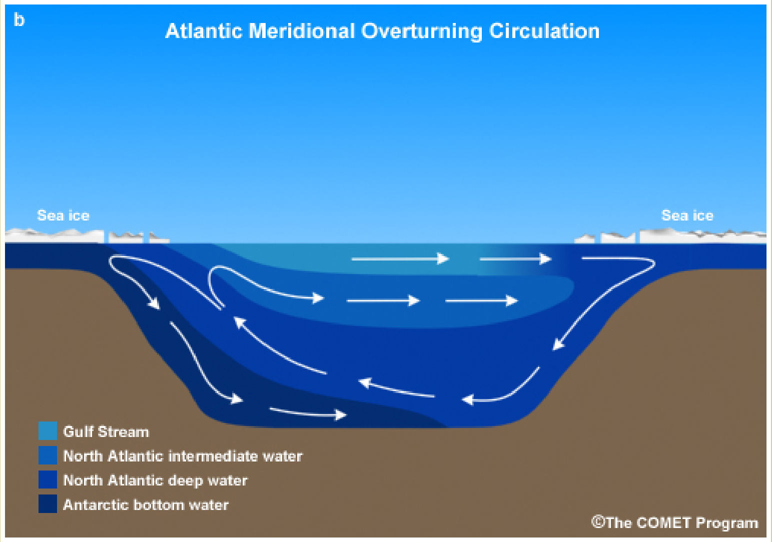

b
Atlantic Meridional Overturning Circulation
Sea ice
Sea ice
Gulf Stream
North Atlantic intermediate water
North Atlantic deep water
Antarctic bottom water
©The COMET Program

Figure 4.1: Meridional overturning circulation in the Atlantic. Different color highlight the different overturning cells forming the MOC. Image source `http://www.goes-r.gov/`

Polzin et al. (1997) showed that near the bottom –above rough topography– there is a sufficient amount of mixing to close the energetic budget proposed by *Munk* (1966). Afterward, many confirmations to that study has been found (e.g., *Ledwell et al.* 2000; *Watson et al.* 2013).

Furthermore, *Mashayek et al.* (2015) showed that a vertical decrease of mixing is essential in order to successfully reproduce the macroscopic characteristics of MOC with numerical models. A variety of phenomena has been proposed to explain the enhanced bottom boundary mixing (e.g., internal waves breaking, *Nikurashin et Ferrari* 2009; tides, *Finnigan et al.* 2002; topographically trapped waves *Kunze et Smith* 2002). Recent high-resolution numerical simulations (Figure 1.5) pointed to hydrostatic phenomena (such as form drag) as possible generating mechanism of bottom boundary mixing. In order to

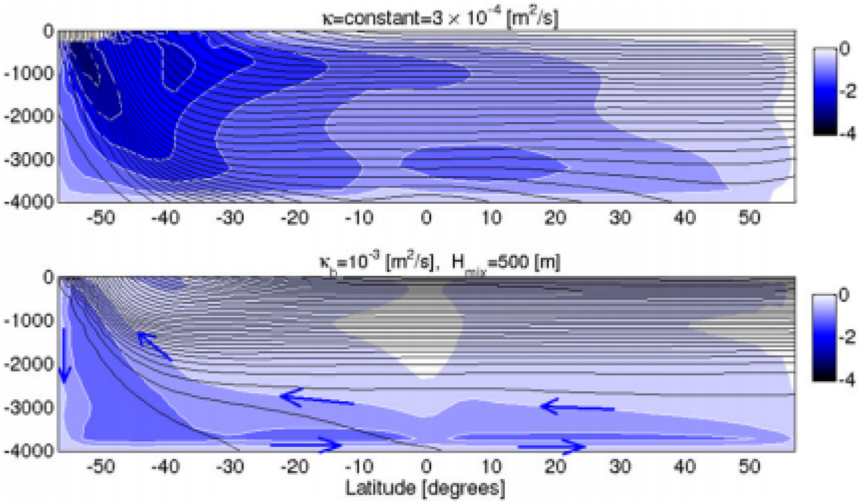


Figure 4.3: Zonally averaged meridional overturning streamfunction [4 Sverdrup (Sv; 1 Sv 5×10^6 m^3/s)] overlaid by isopycnals in the deep ocean. (top) A case with a constant diapycnal diffusivity of $K_Z = 3 \times 10^{-4}$ m^2/s; (bottom) a case with k decaying from $10^{-3} m^2/s$ at the bottom to $10^{-5} m^2/s$ at the top with a e-folding scale of 500 m. The arrows represent the direction of the circulation. Figure reproduced from *Mashayek et al.* (2015).

understand if form drag can potentially permit to close the energy budget of the MOC, we exploit a dataset issued from a tracer release experiment (*Ledwell et al.*, 2016) that highlights enhanced mixing above topography and prescribe a vertical variation of K_ρ in order to reach an agreement between the observations and a 2-D model. As a result, the parameters characterizing form drag-generated bottom mixing are derived.

A further complication of the problem comes from the measure of the amount of turbulent kinetic energy that is irreversibly converted to background potential energy as a result of turbulent mixing in a stably stratified fluid as the ocean. As well summarized by *Venayagamoorthy et Koseff* (2006): the conventional method for relating estimates of turbulent dissipation, ε, to turbulent diapycnal diffusivity, K_ρ, is based on a simplified turbulent kinetic energy formula first proposed by *Osborn* (1980), namely: $K_\rho = \Gamma \frac{\varepsilon}{N^2}$, where N is the stratification and Γ is the so-called mixing efficiency. Major inherent assumptions of indirect methods are that the turbulent flow is statistically stationary and homogeneous. These assumptions are used to simplify the energetics of the turbulent flow field. A value of $\Gamma = 0.2$ was used by Osborn based on some controlled laboratory experiments by *Britter* (1974), but many studies over the last few decades have attempted to parameterize Γ as function of the strength of the stratification, which is commonly quantified either in terms of the gradient Richardson number $Ri_g = N^2/S^2$, where $S = dU/dz$ is the mean shear rate (e.g., *Itsweire et al.* 1993; *Imberger et Ivey* 1991); or in terms of the turbulence intensity parameter $Re_b = \frac{\varepsilon}{\nu N^2}$, where ν is the water kinematic viscosity (e.g., *Shih et al.* 2005; *Bouffard et Boegman* 2013). However, a universal parameterization of Γ remains elusive.

For the comparison of the SCAMP measurements with the numerical model Symphonie I chose to use the parameterization proposed by *Bouffard et Boegman* (2013) due to the similarities of the turbulent dynamics in our two studies: shear induced turbulence in the surface mixed layer.

However, the turbulent mechanisms in the bottom boundary layer are not yet well understood and it is still not clear which parameterization is more appropriate. As a consequence, even if there are good leads pointing to the fact that the mixing efficiency is variable also in these regions of the ocean (Pascal Bouruet-Aubertot, personal communication), for this study we chose to use the classical constant value.

As regards my thesis, diapycnal mixing plays also a pivotal role in the vertical dispersion of larvae and pollutants. Regarding these last ones, a very active field of research is the understanding of how small scale processes translate into large-scale impacts on the fate and transport of hydrocarbons in the ocean.

But the study of the transport by mixing of larvae in the deep ocean is also gaining visibility. In fact, marine populations in the coastal ocean are diminishing as a result of the intense anthropic pressure on the coastal environment. Moreover, the deep ocean communities are the ones that are expected to replenish or even recolonize the despoiled coastal environment. The study of these communities and of their ability to possibly sustain the coastal marine communities heavily relies on numerical dispersion studies. As already put forward in the previous chapter, the numerical circulation models have some limitations that must be understood and possibly overcome in order to make reliable predictions of the larvae spawning from the deep ocean.

In the previous chapter the focus was on the numerical models' limitations in predicting mixing in the water column. This was possible as the mixing processes in the water column are relatively well understood. However, as explained above, our knowledge of bottom boundary mixing processes is still incomplete. As a consequence, in the present chapter, we focus on a new parameterization of diapycnal mixing that –if implemented– could ameliorate the dispersal prediction of the numerical models.

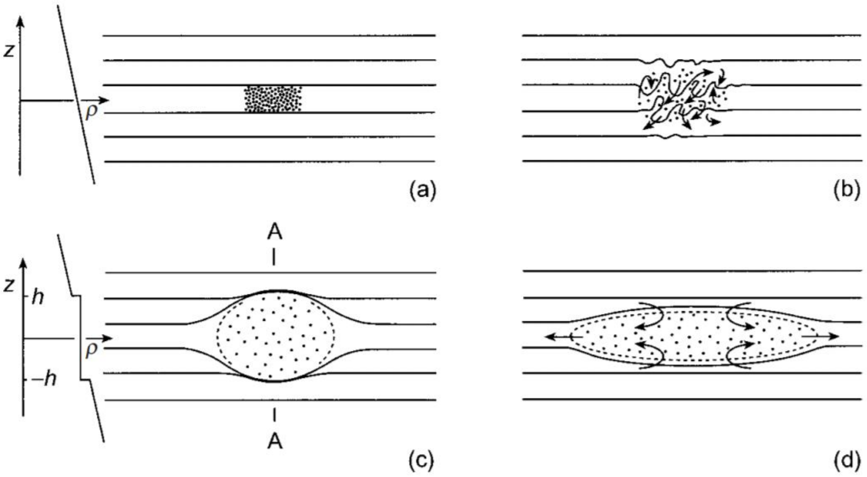


Figure 4.2: Diapycnal mixing followed by isopycnal dispersion. Isopycnal surfaces (of constant density) are shown as full lines and dots mark a region of very small neutrally buoyant particles. Initially, in (a), the density gradient is uniform and given by $\rho = \rho_0(1 - N^2 z/g)$, where z is the upward vertical measured from the mean depth of the mixed region and N is the buoyancy frequency of the water. Turbulence illustrated in (b) leads to the homogenization of a patch (c), so that the density in section A-A through the center of the patch is now as shown at the left. The density profiles outside the mixed patch is unchanged and as shown in (a). If the pressure at the top of the mixed region at $z = h$ is p_0, that within the mixed region is $p_I = p_0 + hzg\rho_0 dz = p_0 + g\rho_0(h - z)$, and that in the surrounding ambient water is $p_A = p_0 + hzg\rho dz = p_0 + g\rho_0[(h - z) - (N^2/2g)(h_2 - z_2)] = p_1 - g\rho_0(N^2/2g)(h_2 - z_2) < p_I$ at levels $-h < z < h$. There is therefore a horizontal pressure gradient that causes the mixed region to spread laterally whilst, to conserve volume, it collapses vertically as shown in (d), intruding, into the surrounding water along the isopycnal surface that has the density of the mixed region. During the turbulent mixing, and in the process of intrusion, waves may be radiated into the surrounding quiescent fluid. The spread of the patch leads to the dispersion of particles along the isopycnal surface. The effect of the Earth's rotation will eventually cause the divergent flow in the spreading patch to rotate anticyclonically under the action of Coriolis forces, whilst fluid converging to replace it will rotate cyclonically. This sketch is grossly oversimplified. In practice, turbulence may not homogenize a region (either in density or in particle distribution) and may be associated with momentum transfers that affect the subsequent horizontal dispersion. Image reproduced from *Thorpe* (2005).

4.1 Spatially Inhomogeneous and Temporally Intermittent Boundary Mixing along the Northern Deepwater Gulf of Mexico

4.1.1 Abstract

An anthropogenic tracer reported in *Ledwell et al.* (2016) was released along an isopycnal at approximately 1100 m water depth, 150 m above local bottom on the continental slope of the Northern Gulf of Mexico. Sampling after 4 months returned an average interior diapycnal diffusivity estimate of 1.3×10^{-4} m^2 s^{-1}. Sampling at one year returned a background interior mixing rate an order of magnitude smaller, $1.5 \pm 0.5 \times 10^{-5}$ m^2 s^{-1}, and an inference that the 4 month estimate is dominated by a boundary process. In this work we examine ancillary evidence collected during the field program that might explain the observed diapycnal dispersion. We find only background (1×10^{-5} m^2 s^{-1}) mixing levels in the interior Gulf of Mexico, diagnosed from the application of finescale parameterizations to LADCP/CTD data. If these interior estimates are representative of climatological mixing rates, it also implies the 4 month tracer dispersion estimate is dominated by a boundary mixing process. However, estimates of boundary mixing from LADCP/CTD data at the time of the 12 month survey are not sufficient to explain the 4 month tracer dispersion. Plausibility that the inferred boundary mixing could result from non-propagating form drag is explored with a 2-dimensional time dependent advection diffusion model. The model is used with current meter data to establish the expectation that flow over complex topography situated on the continental slope provides a turbulent energy source sufficient to explain the tracer dispersion. The model estimates indicate a significant sensitivity to sub-inertial forcing associated with Hurricane Isaac.

4.1.2 Introduction

The catastrophic blowout of the Deepwater Horizon (DWH) oil well in April 2010 released about 5 million barrels of oil and 2.5×10^8 standard m^3 of natural gas, and has so far cost over $40 billion in cleanup and fines. The research program initiated by the Gulf of Mexico Research Initiative (GoMRI) has stimulated research in a number of fields, but there remains a critical need to understand how small scale processes translate into large-scale impacts on the fate and transport of hydrocarbons in the ocean. Motivation for the proposed study comes from an anthropogenic tracer release funded as part of the GISR (Gulf of Mexico Integrated Spill Response Consortium) response to GoMRI-III which was directed at documenting the linkages between small scale processes and large scale effects.

The tracer release (*Ledwell et al.*, 2016) consisted of a 25 km streak of CF_3SF_5 injected on an isopycnal surface at about 1100 m depth approximately following the 1250 m isobath. The tracer was sampled 5-12 days after release, at 4 months and lastly at the 1 year point. Differences in the vertical tracer distribution between injection and the 4 month time interval reveal a cross-isopycnal (diapycnal) diffusivity of $K_{\text{tracer}} = 1.3 \times 10^{-4}$ m^2 s^{-1}. This is a relatively large number, some 20 times larger than that associated with the background internal wavefield in the ocean interior (*Polzin et al.*, 2014a).

Lowered Acoustic Doppler Current Profiler (LADCP) and CTD profiles obtained coincident with the tracer sampling at the 1 year point enable an independent estimate of diapycnal mixing through the use of 'finescale parameterizations' (*Polzin et al.*, 2014a). The finescale parameterization seeks to estimate the net transfer of energy to smaller scales as high frequency waves refract in near-inertial shear at vertical wavelengths of tens to hundreds of meters. This energy transport shows up at vertical scales smaller than 1 m as internal wave breaking and turbulence, with a fraction of this (about 15%) appearing as diapycnal mixing. An extensive review of the methodology and uncertainties is provided in *Polzin et al.* (2014a). Results from this study indicate background mixing rates in the Gulf of Mexico interior and thus locate the proximate cause for the observed

4 month tracer dispersion at the northern boundary.

There are two broad categories of forcing mechanisms for the inferred boundary mixing. One class emphasizes that the large scale inclination of the Gulf of Mexico (GOM) continental slope approximately matches that of ray trajectories associated with energy containing frequencies of the internal wavefield: near-inertial (periods of approximately 1 day) and tidal motions (diurnal and semi-diurnal). Reflection of internal waves under these approximate matching (critical) conditions is understood to result in significant velocity gradients and mixing (e.g. *Eriksen*, 1985; *Garrett et Gilbert*, 1988). While GOM tides are relatively weak (*Kantha*, 2005) there may be a notable near-inertial source associated with sea-breeze forcing at diurnal periods (*Zhang et al.*, 1991a,b). A second class of turbulence forcing mechanisms emphasizes topographic variability superimposed upon the large-scale continental slope. Topographic irregularities will impose their horizontal scales on the flow field, potentially leading to significant vertical gradients and mixing. A mechanism that recommends itself as an explanation for the intense boundary mixing inferred from the GISR tracer and LADCP/CTD measurements is that of low frequency currents associated with Loop Current Eddies and Topographic Rossby Wave variability (*Hamilton et Lugo-Fernandez*, 2001; *Oey et Lee*, 2002; *Hamilton et Lugo-Fernandez*, 2009) and local atmospheric forcing. When the amplitude of the topographic variability is small and has length scales permitting the radiation of internal waves, internal lee waves (*Bretherton*, 1969) are generated. When the topographic variability is large, however, near-boundary wave breaking, hydraulic like effects, flow separation and vortex shedding arise and dominate the radiating response (e.g. *Baines*, 1995). The distinction between 'large' and 'small' is quantified in this context by a steepness parameter

$$s = N h_{\rm rms}/U \tag{4.1.1}$$

in which N is the buoyancy frequency, h_{rms} represents the root-mean-squared topographic perturbation and U the impinging flow. There is a significant literature in the atmospheric science community (*Alexander et coauthors*, 2010) as it was noted quite early (e.g. *Palmer et al.*, 1986) that inclusion of the associated drag was necessary to eliminate westerly wind biases in atmospheric general circulation models. This drag arises as pressure differences across topographic obstacles, i.e. 'form drag'. While that body of literature (e.g. *Lott et Miller*, 1997; *Scinocca et McFarlane*, 2000; *Garner*, 2005) provides a wealth of information, the underpinning of that literature is one of isolated topography in the hydrostatic non-rotating approximation and the issue of momentum, i.e. drag. The oceanic paradigm needs to address the issues of non-isolated (complex) topography and rotational effects. The oceanic observational constraints differ as well, as they concern diapycnal mixing, an issue of turbulent energetics, not momentum.

As a rule of thumb, the non-propagating form drag that arises in $s \gtrsim O(1)$ high drag states is parameterized with a bluff drag formula (e.g. *Baines*, 1995; *Scinocca*, 2001),

$$\tau/\rho = \frac{1}{2} C_d U^2 \frac{\ell d}{L^2} \tag{4.1.2}$$

with a drag coefficient $1 < C_d < 10$, U representing the upstream flow, $d(s)$ a steepness dependent depth of the blocked flow and ℓ the width of the obstacle. The formulation is for isolated topography and the factor L^2 represents the total area of the domain. Rotation is not a part of the discussion of (4.1.2). The relevant paradigm on the northern continental slope is likely that topography is interconnected and 'complex' rather than isolated, with rotational effects being important on the dominant scales of topographic variability (1-10 km).

This scale of variability and much of the phenomena are familiar to us in our environment on the Earth's surface. On the other hand, (4.1.2) could be used to describe the effects of roughness elements such as trees, shrubs, coral reefs, etc. What is intended here is form drag associated with hills, mountains, ridges and mud waves which dominate h_{rms} and occasionally finds the label of 'orographic form drag'. We use this terminology to avoid confusion that might occur with topographic 'roughness' being interpreted as a 'roughness length' in the context of log-layer formulations akin to (4.1.2).

Here we postulate that diapycnal mixing associated with such orographic drag can be cast as

$$K_\rho H N^2 = R_f [\frac{C_d}{2} \frac{d}{\ell}] U^3 \tag{4.1.3}$$

in which the flux Richardson number R_f represents the proportion of turbulent production that gives rise to a turbulent buoyancy flux, turbulent production is cast as τU and H represents the height scale over which the turbulent production is dissipated.

If we had looked at this from the perspective of potential energy anomalies ΔPE associated with mixed layers in uniform stratification, for which

$$\Delta PE = \frac{1}{12} N^2 H^3 \ ,$$

asserted that the mixed layer height scales as $H = U/N$ and then assumed that such overturns occur at a rate of U/L where L is the scale associated with the dominant scales of the complex topography, we would arrive at

$$K_\rho H N^2 = \frac{1}{12} U^3 \frac{H}{L} \ . \tag{4.1.4}$$

Thus (4.1.3) is difficult to distinguish from (4.1.4) if $C_d = 1$. Our approach is to engage in idealized 2D modeling of the tracer evolution, treating $[C_d \frac{d}{\ell}]$ and H as tunable parameters. Comparison of numerical results with observations provide best fit parameters of $\frac{C_d}{2}\frac{d}{\ell} \approx 0.06$ and $H \approx 2U/N$. If we interpret d/ℓ as $d = h_{\mathrm{rms}} \cong 100$ m and $\ell \cong 2\pi/20$ km the dominant topographic length scale, best fit drag coefficients are $C_d \cong 3.6$, similar to $1 < C_d < 10$ found in the literature (*Baines*, 1995).

Section 4.1.3 presents our methods, Section 4.1.4 the results of our data based analysis that demonstrates the observed finescale variability is insufficient to explain the observed tracer dispersion. These results are interpreted in an imaginative way, subjected to appropriate observational constraints, using an idealized 2-D advection/diffusion scheme in Section 4.1.5. Our conclusions, summarized in Section 4.1.6, are based only indirectly on the observations, as no measurements are yet available in the boundary region which wee argue dominates the mixing.

4.1.3 Data and Methods

4.1.3.1 LADCP/CTD

Simultaneous LADCP/CTD data were collected during the R/V Pelican 14-05 cruise, one year after the tracer release. The LADCP system consisted of an upward- and a downward-pointing Teledyne/RDI Workhorse 300-kHz ($WH300$) ADCP, as well as a battery case, mounted on a LUMCON rosette frame. Except on profiles 24-29, when the bin size was halved in an attempt to increase the effective vertical resolution, all LADCP data were collected with a bin size of 8 m, and 2 m blanking distance. Out of the three different WH300 ADCPs used on the cruise, one had all four transducers misconnected. The affected ADCP data files from profiles 1–43 had to be repaired for processing with the LDEO IX_10 software to obtain horizontal velocity profiles. CTD data were obtained from sensor pairs on a Sea-Bird SBE 9plus instrument. The CTD data were affected by ship motion, introducing noise at 5 m vertical wavelengths (Fig. 4.4) and limiting the ability to identify overturns and estimate Thorpe scales. Typical closest bottom approaches are 10-15 meters above the seabed.

4.1.3.1.1 Finescale Parameterization The contribution of internal wave breaking to K_ρ is assessed using the finescale parameterization formula for turbulent production, $\mathcal{P}$ (*Polzin et al.* (2014a), their equation (40)), given by

$$\mathcal{P} = 8 \times 10^{-10} \frac{f}{f_o} \frac{N^2 \cosh^{-1}(N/f)}{N_o^2 \cosh^{-1}(N_o/f_o)} \hat{E}^2 \frac{3(R_\omega + 1)}{4R_\omega} \sqrt{\frac{2}{R_\omega - 1}}. \tag{4.1.5}$$

The factor $\hat{E}$ represents a band-limited estimate of the shear spectral density relative to the high-wavenumber asymptote of the Garrett and Munk (1976) spectrum, $2\pi N^2/10$, and R_ω is the ratio of shear to strain spectral density. Shear is the gradient analogue of kinetic energy and strain represents the rarification and compression of isopycnals as a gradient analogue of potential energy. The factor $\sqrt{\frac{2}{R_\omega - 1}}$ is a single-frequency interpretation of the expected value of the aspect ratio, k_h/m, in a broadband wavefield, in which k_h is the magnitude of the horizontal wavenumber and m is the vertical wavenumber. The factors

$f_o = 2\pi\Omega \mid \sin(32.5 \text{ degrees latitude})|$ and $N_o = 3$ cph are normalization constants, with Ω equal to twice the Earth's rotation rate. The diapycnal buoyancy flux associated with this turbulent production is $R_f\mathcal{P}$, giving a diapycnal diffusivity $K_\rho = R_f\mathcal{P}/N^2$, where the flux Richardson number $R_f \cong 0.15$ and N^2 is the buoyancy gradient.

An extensive review of this finescale parameterization is presented in *Polzin et al.* (2014a), which documents the parameterization's use and its potential for bias. The most important of these relating to instrumental limitations is the bandwidth for estimating the spectral density at high wavenumber. The bandwidth was chosen here on the basis of avoiding high wavenumber noise, Figure 4.4. No attempt was made to subtract noise or correct for suppression of oceanic signal by the data processing algorithms. Our analysis relies on a strain-only algorithm below approximately 1500 m, due to decreasing signal-to-noise ratios in the LADCP measurements.

A significant physical limitation of the finescale parameterization is that it represents a closure for wave-wave interactions, not for hydraulic-like processes. A practical limitation is that the finescale parameterization relies upon Fourier methods to isolate signal from noise; the associated spectral windowing process will not capture the full dynamic range of internal wave driven mixing if it is strongly enhanced at the bottom boundary.

As a further quality control measure we also require a neutral density (potential temperature) anomaly three times larger than a lower bound, $\gamma^n_{rms} = (0.5 - 1.0) \times 10^{-3}/3$ ($\theta_{rms} = 1 \times 10^{-4}$ C).

4.1.3.2 Current Meters

As part of the Gulf Integrated Spill Response (GISR) project, six deep current meter moorings (M1-M6, black squares in Fig. 4.5) were deployed around the BP Deepwater Horizon spill site in the northern Gulf of Mexico. The moorings were initially deployed for a period of one year from 10 July 2012 to 09 July 2013. They were serviced in July 2013 and five were redeployed for one more year. Each mooring contained an upward-looking TRDI 75-kHz ADCP on the top, three Aanderaa Recording current meters (RCM11), dual acoustic releases and an anchor. On M1, four additional InterOcean S4-A current meters were installed between 790 and 1090 m depth. Table 1 describes the locations, instrument depths and total water depths of those moorings. Currents were sampled at an hourly rate and stored internally. The ADCPs used are Teledyne RDI workhorse long ranger acoustic profilers with a beam angle of 20 degrees and an acoustic center frequency of 76.8 kHz. The depth cell (bin) size of the ADCPs was set to be 16 m with blanking distance of 24 m. The RCM11s are single point recording current meters using the Aanderaa Doppler Current Sensor (DCS) technology. RCM11s were configured with additional sensors for temperature, conductivity, pressure and dissolved oxygen.

Table 4.1: Locations, instrument depths, total water depths of M1-M6 deepwater current moorings. *collected by S4 current meters.

Mooring	Latitude	Longitude	ADCP depth (m)	Single-point CM depths (m)	Water depth (m)
M1	28.4999	-88.5001	690	790*, 890*, 990*, 1090* 1290, 1490, 1677	1690
M2	28.7484	-88.7463	535	635, 835, 1020	1035
M3	28.7499	-88.2501	837	937, 1137, 1317	1337
M4	28.4998	-89.0001	336	436, 636, 816	836
M5	28.2503	-88.7499	1150	1250, 1450, 1630	1650
M6	28.0001	-89.0001	812	912, 1112, 1297	1312

4.1.3.3 Orography

Topographic variability (orographic variability) on scales of 100 m to 10 km is required to characterize the hydrodynamic response. Measurements at a resolution of better than 100 m are available from multibeam surveys for the northern Gulf from the NOAA National

Geophysical Data Center, U.S. Coastal Relief Model, at the web site: `http://www.ngdc.noaa.gov/mgg/coastal/crm.html`.

The steepness parameter $s = Nh_{\mathrm{rms}}/U$ (*Baines*, 1995) is a fundamental non-dimensional variable quantifying the behavior of continuously stratified flow in a hydrostatic, non-rotating parameter regime. It represents the ratio of topographic slope to that of internal lee waves arising in the linear limit. Identifying typical vertical and horizontal topographic scales can be straightforward for isolated topography, but complex topography such as that situated on the GoM continental slope can exhibit a multiplicity of scales that introduces ambiguity into the interpretation of s.

One interpretation (*Nikurashin et al.*, 2014) is to consider only those horizontal scales contributing to freely propagating lee waves ($f < kU < N$) within the definition of h_{rms}. We used a spectral approach to define the topographic variance in this bandwidth and for characteristic stratification rates and horizontal velocities of 0.05-0.20 m s^{-1} find characteristic steepness values of $s \sim O(1)$ (0.5-3). This strongly suggests a significant role for non-propagating orographic form drag.

4.1.3.4 Tracer

Details of the tracer experiment are described in *Ledwell et al.* (2016). Here we highlight the findings that are important to the argument here.

A map of the column integrals of tracer found at the stations occupied four months after the tracer release is shown in Figure 12 in *Ledwell et al.* (2016), together with the location of the tracer injection, the mooring locations and the path of a RAFOS float deployed with the tracer. The tracer was sampled between 5 and 12 days after the release just to the southwest of the injection. The rms spread of the tracer about the target isopycnal surface was approximately 15 m, which is quite narrow compared with later surveys.

The 4-month survey found the vertical distribution of the tracer to be considerably broader near the continental slope than in the interior. Thus, the stations were divided into two sets: those shoreward, and those seaward of the 1500-meter isobath, a boundary which seemed to separate the broader profiles from the narrower, for the most part. The two sets of profiles were averaged separately to place limits on the turbulent diffusivity, as follows.

A turbulent diapycnal eddy diffusivity of 1.3×10^{-4} m^2 s^{-1} was estimated using a model of the one-dimensional diffusion equation taking the initial profile to the 4-month mean interior profile. The diffusivity required to bring the initial profile to the 4-month boundary profile, on the other hand, was estimated to be greater than 4×10^{-4} m^2 s^{-1}.

These results indicated that the diffusivity in the boundary region, K_{boundary}, is much greater than the interior diffusivity K_{interior}. However, since it could not be known how much time any of the tracer had spent in the interior versus the boundary region (e.g. 4.1.6), it was not possible to assign an accurate value to the diffusivity in either region, from the 4-month survey. There is also arbitrariness in the choice of isobath used to distinguish boundary from interior.

The mean vertical profile from the survey at 12 months (*Ledwell et al.* (2016), their Figure 14) was nearly indistinguishable from the mean interior profile at 4 months. Applying the model to the transition between these two profiles yielded a diffusivity of $(0.15 \pm 0.05) \times 10^{-4}$ m^2 s^{-1}, much smaller than either diffusivity from the first period. This must be an upper limit on K_{interior}, since the dispersion of the tracer would still have been influenced by processes near the boundaries, although it had spent relatively little time there. The inference is that K_{boundary} must have been much greater than 1.3×10^{-4} m^2 s^{-1}.

4.1.3.5 Numerical Scheme

Numerical solutions of the advection-diffusion equation are obtained using a Forward-Time Central-Space (FTCS) scheme. Velocities and isopycnal and diapycnal diffusivities are functions of y, z and t. A Von Neumann stability analysis determines the time step Δt of the simulation through the condition $\Delta t \leq \left(\frac{2 \cdot max\{K_{\parallel}\}}{(\Delta y)^2} + \frac{2 \cdot max\{K_{\perp}\}}{(\Delta z)^2} + \frac{V}{\Delta y} + \frac{W}{\Delta z} \right)^{-1}$, where $K_{\|}$ is the isopycnal diffusivity, $K_{\perp}$ the diapycnal diffusivity, V the maximum value of horizontal velocity and W the maximum value of vertical velocity. Given this condition, computational constraints determined the grid spacing we used. For the present study the grid sizes are $\Delta y = 500$ m in the horizontal and $\Delta z = 5$ m in the vertical.

4.1.4 Observations

4.1.4.1 Finescale Parameterization

Stations having significant tracer concentrations (Figure 4.6) were binned into three groups: a collection of 'boundary' stations in water depths of 1400 m and shallower, a group of open ocean stations immediately to the south, and a group in a region occupied by the Loop Current to the southeast. Statistical uncertainties in diffusivity estimates based on averaging a collection of 10-20 LADCP profiles are approximately a factor of ±2 and thus the results (Figure 4.7) at the injection level are unequivocal: in the open ocean $K_{\text{interior}} \cong 1 \times 10^{-5}$ m^2 s^{-1}. This diffusivity estimate is an order of magnitude smaller than that inferred from the overall dispersion rate of the tracer, including the boundary regions, though it agrees well with the diffusivity experienced by the tracer between 4 and 12 months when most of the tracer was in the interior. An enhancement to some 3×10^{-5} m^2 s^{-1} is noted in the bottommost 300 m of the boundary stations. However, this enhancement is not sufficient to explain the vertical dispersion of the tracer, which requires a spatial/temporal *mean* value over most of the GOM several times larger.

There are a number of pertinent caveats about this finescale parameterization. First, it is not an all purpose turbulent closure; it addresses only interactions between internal waves. It would not, for example, detect non-wave-induced turbulence associated with hydraulic effects, flow separation and vortex shedding from complex topography. Second, it is based upon a spectral method that limits the ability to capture anomalous enhancements of mixing in the bottommost 50 m. Here we have used 320 m vertical intervals with a Hanning window. Despite such caveats, the finescale parameterization is fully capable of detecting internal-wave-induced mixing at a rate of $K \cong 1 \times 10^{-4}$ m^2 s^{-1} in the GOM interior. It is not there.

Experience based upon observations (e.g. *Polzin et al.*, 1997) and modeling of internal tides above rough topography (*Polzin*, 2004a,b) strongly suggests that *if* K_{boundary} were associated with an internal wave process, the finestructure parameterization method should have returned diffusivity estimates that come much closer to balancing the tracer budget than the estimated 3×10^{-5} m^2 s^{-1}.

4.1.4.2 Internal waves?

The ADCP data from the GISR moorings were examined by lowpass filtering velocity data at frequencies smaller than $2f$ to limit ping-to-ping uncorrelated noise and then first differencing on scales of 48 m. There are occasional incidences of shear variance exceeding N^2, which implies diffusivities 2-3 orders of magnitude above background values by a number of methodologies (either the finescale parameterization (Section 4.1.3.1.1) or a shear-instability closure (*Polzin*, 2014)). However, the persistence of such features was not sufficient to imply annual averaged values in excess of 1×10^{-4} m^2 s^{-1}. In short, internal wave diagnostics applied to the GISR moored ADCP data do not explain the high values of diffusivity affecting the tracer.

There is a near-inertial response to Hurricane Isaac. However, it is not the largest in the moored ADCP record. Greater internal wave shear is apparent with enhanced eddy velocities in the later 6 months of the records.

4.1.4.3 If not Internal Waves, then Non-propagating Orographic Form Drag?

Our investigations of this data set included estimates of a buoyancy deficit at the boundary relative to the interior and a characterization of near-boundary density profiles in the context of mixing. Buoyancy anomalies of both signs could be discerned at the boundary relative to the interior over vertical scales of hundreds of meters. But there was nothing to suggest that this could be anything other than the thermal wind signature associated with a time varying eddy field. Occasional density profiles exhibited detached mixed layers of

O(10's) of m height, reminiscent of features described by *Armi et Jr.* (1976) and *Armi* (1978). A particularly suggestive example appears in Figure 4.8 with the orographic context appearing in Figure (4.9). However, the emphasis is on the word occasional, rather than dominant. Investigation of near boundary data via a Thorpe scale analysis was limited by noise in the CTD profiles at 5 m vertical wavelengths, Fig. 4.4, as mentioned earlier.

Form drag comes in both propagating and non-propagating forms, with internal lee waves representing the former and hydraulic-like effects, etc., representing the latter. In the $s \sim O(1)$ high drag parameter regime applicable to the GoM, the non-propagating component dominates the total drag (e.g. *Baines* 1995). While there is an outstanding question of how efficiently the non-propagating drag is converted to a turbulent buoyancy flux, the non-propagating drag is deposited near the boundary whereas this need not be the case for internal lee waves. Both issues serve to focus attention upon the non-propagating component as the most likely agent of near boundary diapycnal tracer mixing.

4.1.5 Boundary Mixing

For the purpose of defining a back-of-the-envelope mixing estimate related to this boundary process, one can crudely regard the tracer dispersion as a tracer weighted average of an intense boundary process characterized by tracer amount $\overline{C}_{\text{boundary}}$ and a weak interior process acting upon $\overline{C}_{\text{interior}}$:

$$\overline{C}_{\text{interior}}(t)K_{\text{interior}} + \overline{C}_{\text{boundary}}(t)K_{\text{boundary}} = \overline{C}_{\text{total}}K_{\text{tracer}}$$

If boundary contributions dominate,

$$K_{\text{boundary}} \cong \frac{\overline{C}_{\text{total}}}{\overline{C}_{\text{boundary}}(t)} \times K_{\text{tracer}} \tag{4.1.6}$$

The implication here is that K_{boundary} needs to be in excess of $K_{\text{tracer}} = 1.3 \times 10^{-4}$ m^2 s^{-1} by a factor of $\overline{C}_{\text{total}}/\overline{C}_{\text{boundary}}$ with the complication that $\overline{C}_{\text{boundary}}$ and $\overline{C}_{\text{int}}$ can be quite time dependent. In order to avoid defining boundary and interior regions and time spent in there to arrive at a weighted mean prediction for K_{tracer}, we turn to a numerical assessment.

4.1.5.1 Numerics

In this section we pursue a pattern matching exercise in which we ask whether plausible distributions of isopycnal stirring and diapycnal mixing result in tracer dispersion that is consistent with the observations. Isopycnals are assumed to be flat and the northern Gulf of Mexico continental slope is represented as a planar sloping boundary.

We solve the time dependent advection-diffusion equation for tracer concentration C,

$$\partial_t C + \partial_y K_{\|}(y,z)\partial_y C + \partial_z K_{\perp}(y,z,t)\partial_z C + v^* \partial_y C + w^* \partial_z C = 0 \tag{4.1.7}$$

with $K_{\|}(y)$ an isopycnal stirring coefficient and $K_{\perp}(y,z,t)$ a diapycnal mixing coefficient. The diapycnal velocity w^* arises from the divergence of the turbulent buoyancy flux and v^* is the corresponding isopycnal velocity required by continuity.

We report boundary and interior concentration profiles by averaging over the model domain so as to mimic the tracer sampling as accurately as possible. Interior profiles represent an average from the 1500 m model depth to 255 km offshore. Boundary profiles are an average over depths of 1000-1500 m.

4.1.5.1.1 Diapycnal Mixing We utilize a production-dissipation-buoyancy flux decomposition of the turbulent kinetic energy equation to specify the diapycnal dispersion rate $K_{\perp}$. We will assume that the buoyancy flux $K_{\perp}N^2$ can be related to turbulent production $\mathcal{P}$ using a constant flux Richardson number R_f, $K_{\perp}N^2 = R_f\mathcal{P}$ and assume $R_f = 1/6$. An entirely *ad hoc* approach is taken by modeling the vertical structure of $K_{\perp}$ using an exponential depth dependence. The depth integrated production rate $\mathcal{W}$ is then equated with a quadratic drag law,

$$\mathcal{W} = [\frac{1}{2}C_d\frac{d}{\ell}]U^3 \ , \tag{4.1.8}$$

with $\frac{1}{2}C_d\frac{d}{\ell}$ treated as a tunable parameter. Thus:

$$\begin{aligned}
K_\perp(y,z) &= K_\perp(y_0,z_0)e^{-(z-z_0)/h(z)} + K_\perp^{int} \\
K_\perp(y_0,z_0)N^2(z_0) &= R_f\frac{1}{2}C_d\frac{d}{\ell}U^3/h(z_0) \\
h(z) &= \beta U/N(z) \\
K_\perp^{int} &= \text{background diffusivity} \\
w^*(y,z) &= \frac{\partial_z(K_\perp(y,z)N^2(z))}{N^2(z)}
\end{aligned} \tag{4.1.9}$$

Line two in (4.1.10) arises from integrating the buoyancy flux while assuming $N^2(z)$ to be slowly varying relative to $h(z)$: $\frac{\partial_z N^2(z)}{N^2(z)} \ll h^{-1}$. The height scale h in line 3 is taken to be a multiple β of the height scale characterizing the hydrostatic, non-rotating result, (*Baines*, 1995). Our initial results, summarized in an error map presented below, assume $N^2(z)$ to be slowly varying in lines two and five. Our estimates of β are consistent with this assumption. Use of $N^2(z)$ in line 5 (results not shown) returns best fit parameter estimates that vary by less than 50%. The boundary is specified by $z_0 = \alpha y_0 + b$ with slope $\alpha = 1/50$. The factor U is constrained with time series of near-bottom velocity measurements from GISR mooring M6, Fig 4.5. Linear fits to the 12 month, near-boundary buoyancy profile are used to specify $N^2(z)$.

4.1.5.1.2 Isopycnal Stirring The effects of isopycnal stirring are parameterized using a specification that acknowledges eddy length scales are limited by the distance to the boundary (the 'law of the wall'):

$$K_\parallel = \begin{cases} u(y-y_0) \ \text{ if } K_\parallel < 1000 \text{ m}^2 \text{ s}^{-1} \\ 1000 \text{ m}^2 \text{ s}^{-1} \text{ otherwise} \end{cases} \tag{4.1.10}$$

This formulation gives rise to a pseudo-advection (*Davis*, 1991) at a rate u which will tend to move the tracer center of mass away from the boundary region, as suggested by observations of column integrated tracer as a function of distance from the boundary, Fig. 4.11. A value of $u = 0.01$ m s^{-1} and a limiting value of $K_\parallel = 1000$ m^2 s^{-1} returns an eddy length scale of 100 km. This representation of isopycnal stirring is likely the most unrealistic part of the model. Much of the tracer is likely swept off the boundary in a single event after Isaac (*Ledwell et al.*, 2016) and the pseudo-advection prescribes a more continuous connection with the interior.

4.1.5.1.3 Initial Conditions Initial conditions are taken as the vertical structure from the 5-12 day sampling. A Gaussian distribution with rms width of 100 m defines the lateral distribution.

4.1.5.2 Results

The numerical assessment was set up with four tunable parameters: $\beta, u, \frac{1}{2}C_d\frac{d}{\ell}$, and $K_\perp^{int}$. Best fit values are of $(\beta, u, \frac{1}{2}C_d\frac{d}{\ell}$, and $K_\perp^{int}) = (2$, 0.01 m s^{-1}, 0.06, $1 - 3 \times 10^{-5}$ m^2 s^{-1}) were determined using an error metric that consists of the mean square difference between observed and predicted 4 month concentration profiles, Figures 4.12 – 4.14, for both interior and boundary regimes. An error map in the $(\beta, u, \frac{1}{2}C_d\frac{d}{\ell}$, and $K_\perp^{int})$ parameter domain (Fig. 4.15) depicts the results of our parameter search, which includes only qualitative information about isopycnal stirring, as variability in the isopycnal coordinate, Fig. 4.11, is greater than in the diapycnal coordinate. Better fits are possible for interior profiles at the expense of agreement with the boundary profiles, and *vice-versa*. Best fit parameters are $\frac{C_d}{2}\frac{d}{\ell} \approx 0.06$ and $H \approx 2U/N$. If we interpret d/ℓ as $d = h_{\mathrm{rms}} \cong 100$ m and $\ell \cong 2\pi/20$ km the dominant topographic length scale, inferred drag coefficients are $C_d = 3.6$, similar to $1 < C_d < 10$ found in the literature (*Baines*, 1995).

The most serious drawback of this 2-d modeling approach is revealed in the numerical solutions as the interplay between the pseudo-advection of tracer away from the boundary and the time dependence of diapycnal mixing via U that can be understood through the tracer weighted estimate (4.1.6). The pseudo-advection in (4.1.10) is required to move the tracer center of mass away from the northern boundary, as evidenced in the four month observations (Figure 4.12). Near boundary work estimates (Figure 4.10) are dominated by Isaac, so that the cross slope advection at the time differential between injection and Isaac is crucial to the model results as that cross slope pseudo-advection translates directly into an increasing height above boundary.

We tried to address this by

- adding a constant 50-100 $\mathrm{m}^2\ \mathrm{s}^{-1}$ to $K_\parallel$ (4.1.10), thinking that such a representation could be rationalized by the presence of topographically trapped motions along the slope. Overall it was not possible to reach as good agreement as with a straight law-of-the-wall representation. There was no change in best fit drag coefficient C_d.

- excluding Hurricane Isaac. Replacing the time series of U^3 with its mean value ($[0.08\ \mathrm{m\ s}^{-1}]^3$) resulted in significant degradation of the boundary error metric. Replacing the extreme values of U associated with Hurricane Isaac (in excess of 0.18 $\mathrm{m\ s}^{-1}$) with 0.12 $\mathrm{m\ s}^{-1}$ has the most effect upon the ability of the numerics to replicate the observations. Agreement with the boundary profile was especially poor.

- a Gaussian depth profile, $e^{-\pi(z-z_0)^2/4h^2}$, changes the proximity of downwelling relative to the diapycnal buoyancy flux. It was possible to reach a good agreement with the observations using both Gaussian and exponential formulations. The best-fit Gaussian drag coefficients were a factor of two larger.

An interpretation suggested by results in *Ledwell et al.* (2016) is that the tracer is swept off the boundary in a single event related to the confluence of *along*-slope eddy velocity some 50 days after the tracer injection and 25-30 days after Isaac, resulting in the wholesale ejection of tracer from the boundary.

There is one direct observational constraint given by a single float released to track the tracer injection. This float grounded on the continental slope immediately after Hurricane Isaac (*Ledwell et al.*, 2016), see also a movie in Supplemental Material. In contrast, with $u = 0.01\ \mathrm{m\ s}^{-1}$, approximately 1/7 of the tracer remains at the boundary to be mixed during Isaac. The afore mentioned episodic nature of off slope transport could imply drag coefficients C_d an order of magnitude smaller with boundary diffusivities of 10^{-3} rather than 10^{-2}.

4.1.6 Summary and Discussion

In this study we have examined ancillary data collected in conjunction with a tracer release in which the space/time dependence of the tracer dispersion strongly implies a boundary mixing process. The ancillary data provide no evidence for a sufficiently intense boundary mixing process, but also suggest that interior mixing processes are an order of magnitude smaller than required to explain the tracer dispersion. Noting that our metrics of boundary mixing are based upon internal wave dynamics, we propose that the required boundary mixing is associated with non-propagating form drag that arises in response to flow over rough topography on the northern continental slope of the Gulf of Mexico.

In order to be one step more quantitative than a simple back-of-the-envelope argument, we forward the conjecture that a fraction (R_f) of the energy conversion associated with a quadratic drag formulation (4.1.10) appears as a diapycnal buoyancy flux. Our estimates utilize observed near-bottom current measurements and observed stratification profiles. Empirical fits to the tracer distribution suggest $\frac{1}{2}C_d\frac{d}{\ell}R_f \cong 0.06$ and $K_{\perp}^{int} = 1-3\times10^{-5}$ m^2 s^{-1}. We have used a law-of-the-wall prescription in order to affect off-boundary tracer transport along isopycnals and have noted a substantial dependence upon the isopycnal off-boundary tracer transport and the timing of energetic diapycnal mixing at the boundary. Our interpretation is that the isopycnal stirring prescription (4.1.10) draws down the boundary concentration level too quickly relative to the passage Hurricane Isaac. Effective drag coefficients C_d could be smaller by an order of magnitude if all the tracer remained at the boundary for Isaac.

In terms of practical implications, we note that with a strong dependence of diapycnal mixing on near-bottom velocity, a direct connection of diapycnal mixing to near-bottom velocity as a response to extreme atmospheric events and the propensity that such events could be correlated with the destruction of petroleum industry related infrastructure together imply that the subsequent fate of blow-out related hydrocarbons could be highly dependent upon the parameterizations laid out in this paper.

Our proposed hypothesis to explain the observed tracer dispersion has roots in a discussion between *Garrett* (1979) and *Armi* (1979), relating to *Armi et Jr.* (1976) & *Armi* (1978), and draws its motivation for a 2-d model from that historical context. Armi interprets detached mixed layers in CTD casts from the western North Atlantic as providing evidence for ventilation of the abyssal ocean at rates of $O(1 \times 10^{-4}$ m^2 s$^{-1})$. *Garrett* (1979) criticizes this on energetic grounds, suggesting that the required energy source is not to be found with a quadratic drag law (4.1.2), $U \sim O(0.1$ m s$^{-1})$ and $C_d = 1 \times 10^{-3}$. This represents the standard closure for viscous stresses associated with a no slip boundary condition. Regional estimates of subinertial velocities characterizing the data examined by Armi peak to values of $O(0.2)$ m s^{-1} only within a limited area under the Gulf Stream (*Schmitz*, 1984). A second criticism leveled by *Garrett* (1979) is that boundary layer turbulence will result in the mixing of mixed water in the boundary layer, implying a small buoyancy production relative to the turbulent production rate, i.e. a small flux Richardson number, R_f. A third criticism is that, if considering the two dimensional problem encapsulated within his figure, there is a remaining issue of how to transport mixed fluid away from the boundary. One sees these questions developed further in *Trowbridge et Lentz* (1991) and *Garrett et al.* (1993), etc.

Our view is that near boundary mixing is significantly more intense than envisioned in these earlier papers and we emphasize the role of non-propagating form drag associated with orography on scales of 0.1-10 km, not viscous stresses associated with a no slip boundary condition. With orography, the effective drag coefficients are orders of magnitude larger and mixing is associated with highly sheared regions in well stratified regions of $O(U/N)$ height above the bottom. We conjecture that the 3-dimensional character of orography may, as well, provide more efficient off-boundary buoyancy transport than the configuration of a 2-d, planar sloping boundary.

Acknowledgments

The skill and cooperation of the masters and crew of R/V Brooks McCall, operated by TDI Brooks, Inc. and of R/V Pelican, operated by the Louisiana Universities Marine Consortium (LUMCON) were essential to the success of the cruises from which the data reported herein were obtained. KP gratefully acknowledges conversations with Larry Armi. Funding for GISR is provided by BP through the Gulf of Mexico Research Initiative (GoMRI), and is administered by the Consortium for Ocean Leadership. The GISR data used in this study are available via the GoMRI GRIIDC data base (https://data.gulfresearchinitiative.org).

AA, SD , JL and ZW recognize funding from GISR/GoMRI contract # SA12-09/GoMRI-006; AC recognizes funding by the Ministére de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche; and KP from WHOI's ISP program.

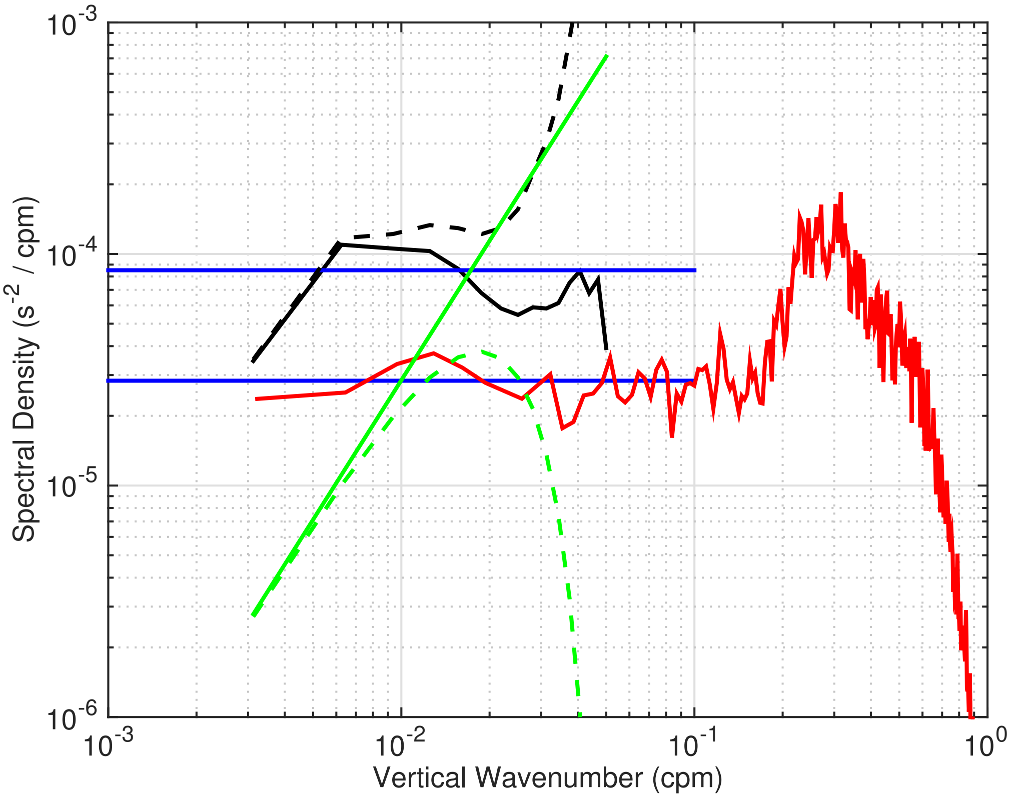


Figure 4.4: An example of shear (black solid) and strain (red) spectra illustrating signal-to-noise and choice of bandwidth. The green solid line represents ping-to-ping uncorrelated noise at 10 times the theoretical minimum rate. Dashed lines are an attempt to represent the effects of averaging and processing. Blue lines denote the high wavenumber asymptote of the Garrett and Munk spectrum. See *Polzin et al.* (2014a) for further details. In this study, the shear spectral level $\hat{E}$ (4.1.5) was estimated at wavelengths of 160 m and smaller under the criterion that the observed spectra (black solid) lie a factor of 3 above the estimate noise (green dashed). Noise in the CTD data at 0.2-0.4 cpm are related to ship heave in ocean swell being communicated down the lowering wire.

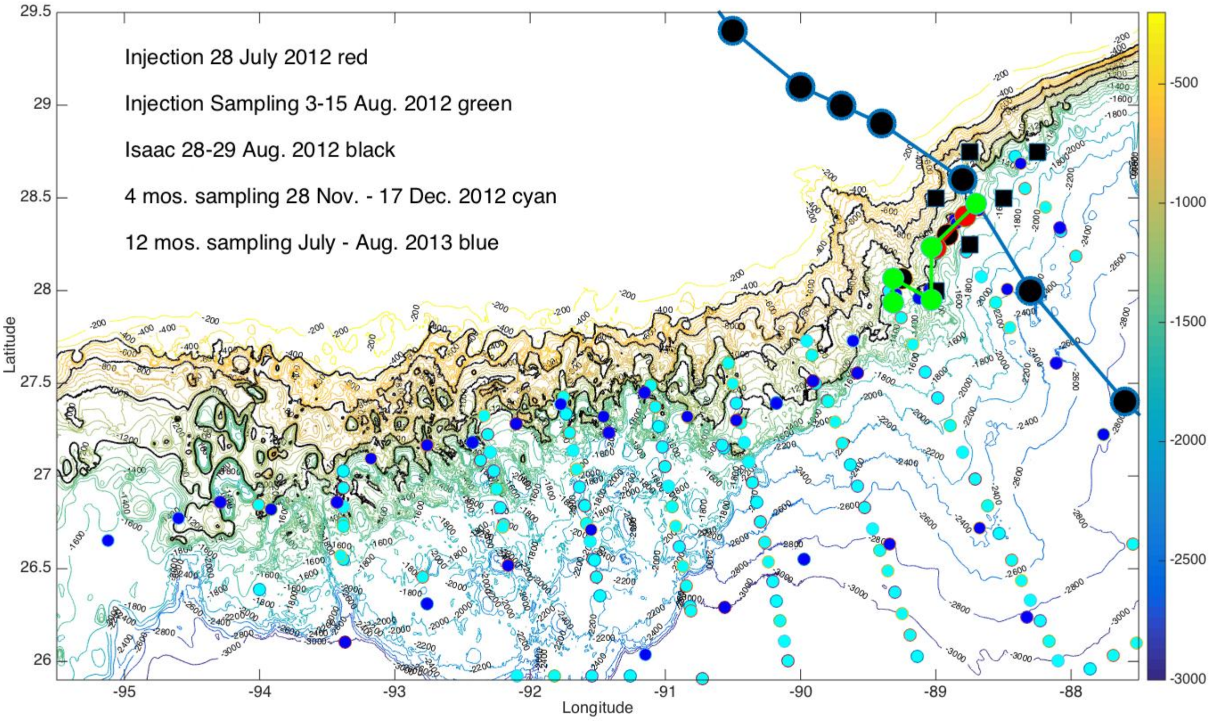

Injection 28 July 2012 red
Injection Sampling 3-15 Aug. 2012 green
Isaac 28-29 Aug. 2012 black
4 mos. sampling 28 Nov. - 17 Dec. 2012 cyan
12 mos. sampling July - Aug. 2013 blue
Latitude
Longitude
29.5
29
28.5
28
27.5
27
26.5
26
-95
-94
-93
-92
-91
-90
-89
-88
-500
-1000
-1500
-2000
-2500
-3000

Figure 4.5: Bathymetry and station positions for the GISR experiment. Black squares denote the positions of the GISR moorings, large red circles locate the tracer injection, large green circles the initial 5-12 day sampling. Other dots depict tracer sampling stations at 4 and 12 months. Positions of Hurricane Isaac are indicated at 6 hourly intervals. Bathymetry is contoured every 200 m between 200-3000 m and every 50 m between 500-1500 m water depth. Black contours are used to emphasize the 500, 750, 1000 and 1250 m isobaths.

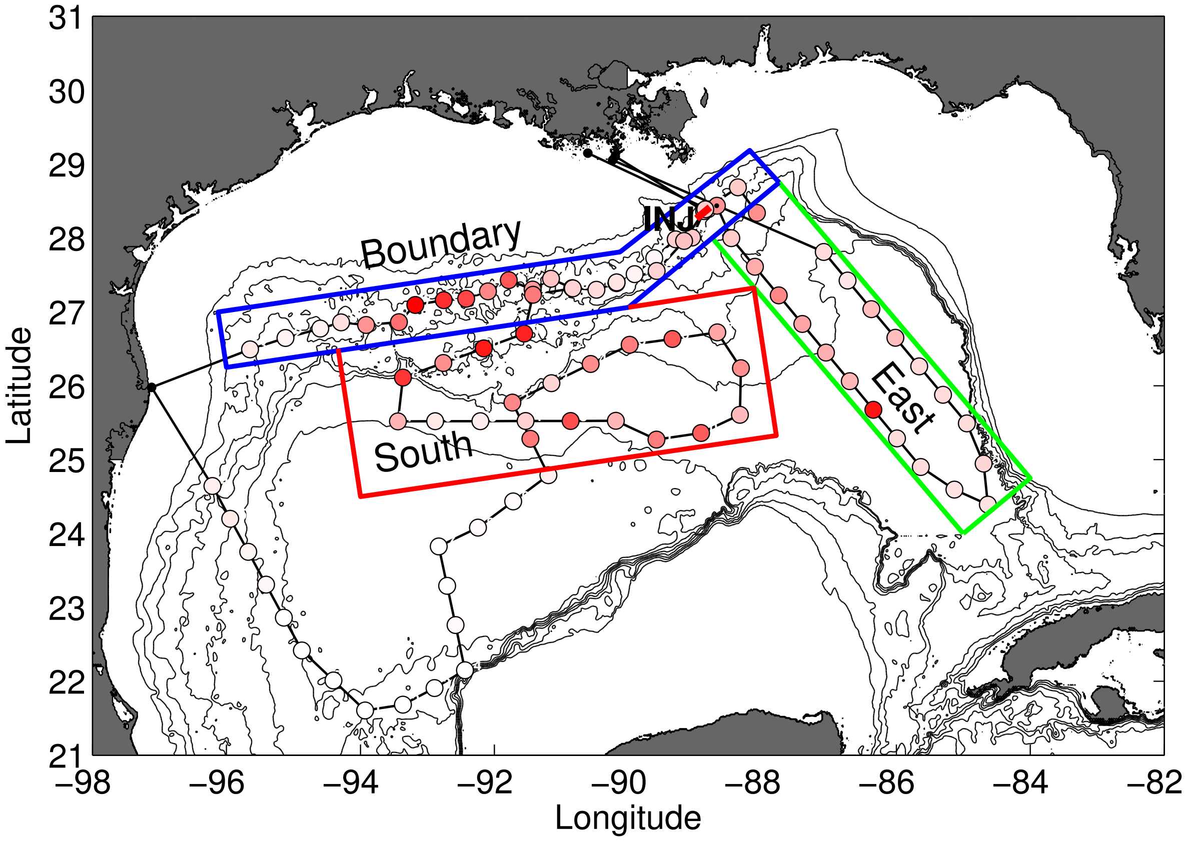

31
30
29
28
27
26
25
24
23
22
21
Latitude
–98
–96
–94
–92
–90
–88
–86
–84
–82
Longitude
Boundary
INJ
South
East

Figure 4.6: Station positions for the year 1 GISR tracer survey are rendered as circles. Column integrals of tracer concentration are indicated by relative shading of the circles. White indicates no tracer was found. Black lines depict isobaths at a 500 m contour interval. Boxes indicate binning into 'Boundary', 'South' and 'East' averages for the finestructure analysis, Fig. 4.7.

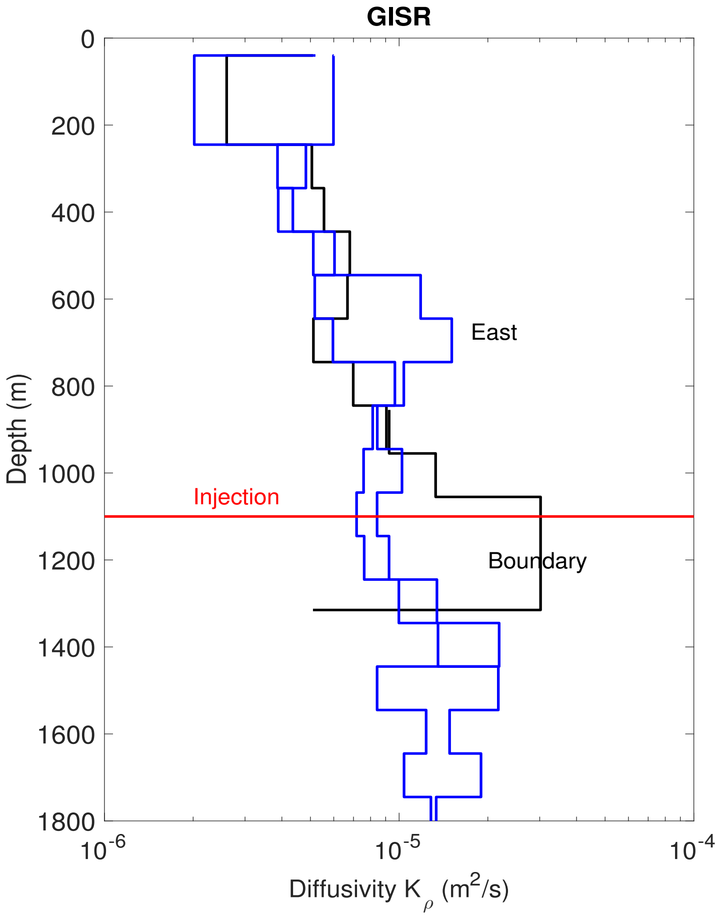


Figure 4.7: Diapycnal diffusivity estimates from the finescale parameterization applied to LADCP and CTD data. Results for 'Boundary' stations are depicted with the black trace. Results for 'South' and 'East' stations (Fig. 4.6) are rendered in blue. The red horizontal line indicates the depth of the injection surface at the boundary.

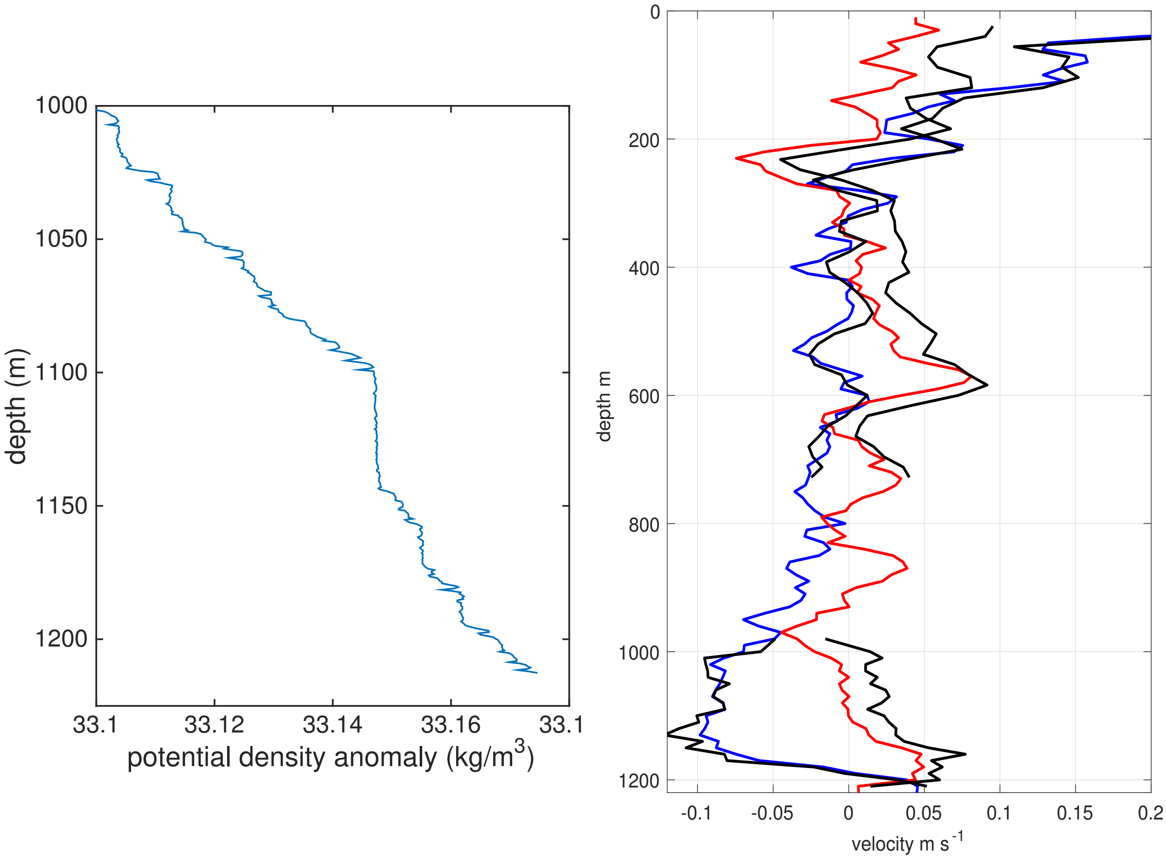


Figure 4.8: Vertical profile data from station 31. Left hand panel: potential density anomaly relative to 1200 dBars. Right hand panel: north-south velocity (red), east-west velocity (blue). Bottom track estimates and shipboard ADCP velocity estimates are rendered in black. Note the differing extent of the vertical axes. In both panels the vertical axis extends to the actual bottom ($\cong$ 1225 m). There are regions of significantly reduced stratification suggestive of either mixed layers or isopycnal separation in a partial hydraulic jump. Significant shear (0.1 m s^{-1} over 40 m) is noted beneath this feature, implying a gradient Richardson number smaller than one. These data were collected above a ridge saddle (Figure 4.9) and lend themselves to an expectation of strongly enhanced mixing associated with highly sheared flow on the lee side of the ridge.

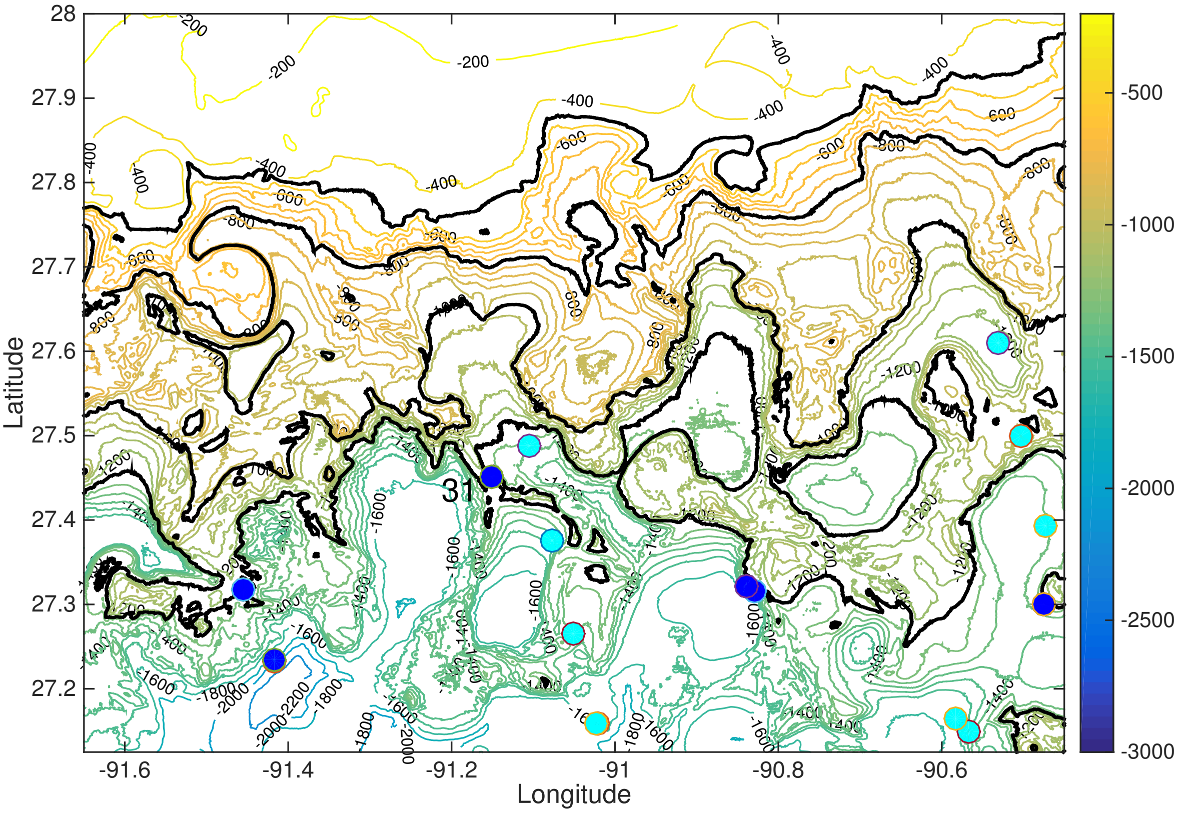

Latitude
28
27.9
27.8
27.7
27.6
27.5
27.4
27.3
27.2
Longitude
-91.6
-91.4
-91.2
-91
-90.8
-90.6
31
-500
-1000
-1500
-2000
-2500
-3000
-200
-400
-600
-800
-1200
-1400
-1600
-1800
-2000
-2200

Figure 4.9: The orographic context for Station 31. This station was taken at a ridge saddle with the near-bottom velocity being east-to-west.

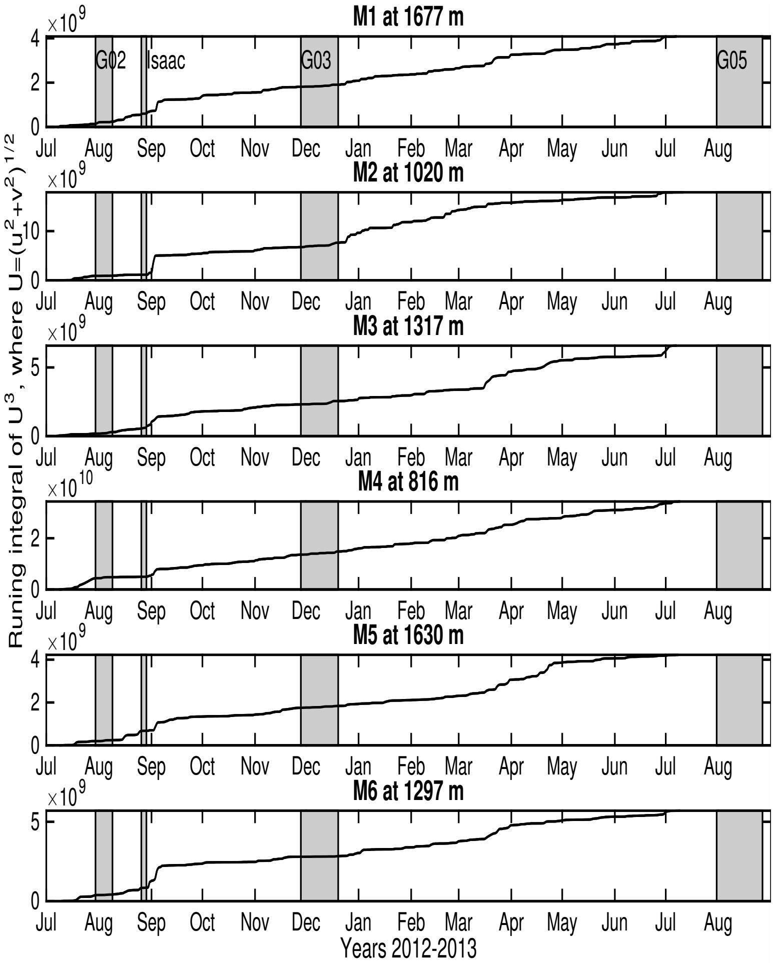


Figure 4.10: Estimates of $\int U^3 dt$ for the bottommost current meters in the 6-element GISR moored array. The 4 grey panels indicate the time of the tracer injection (G02), the passage Hurricane Isaac, the 4-month and 12 month sampling (G03 and G05, respectively). Note that the U^3 statistics between injection and 4 month sampling are dominated by Isaac. The bottom current meter in the south-west quadrant of the array (M6), collocated with the injection site, was used to specify the near-bottom velocity U in the mixing scheme (4.1.10).

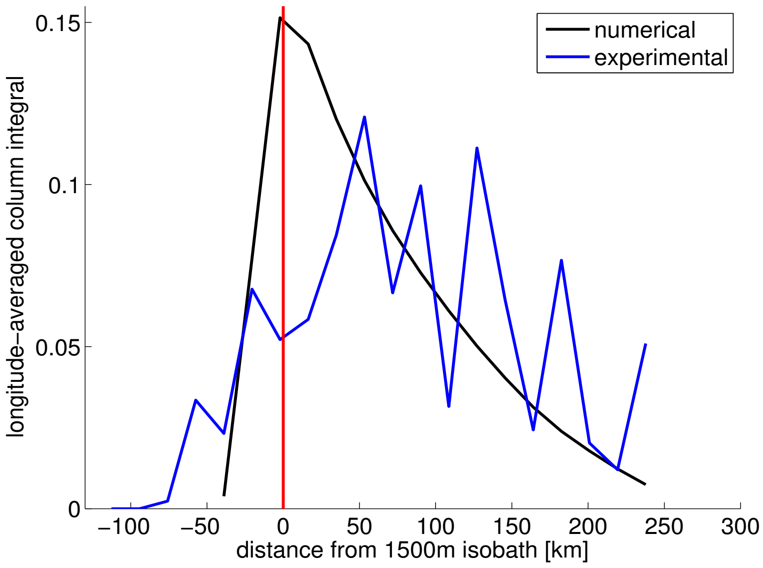


Figure 4.11: Four month column integrated tracer concentrations averaged as a function of distance from the 1500 m isobath from observations (blue) and model (black) with ($\beta = 2, u = 0.01, \frac{1}{2}C_d\frac{d}{\ell} = 0.06$ and $K_\perp^{int} = 3 \times 10^{-5}$ m^2 s^{-1}). Despite the observed variability there is an obvious requirement for the tracer center of mass to migrate off the boundary, which is addressed with the pseudo-advection parameter u (4.1.10).

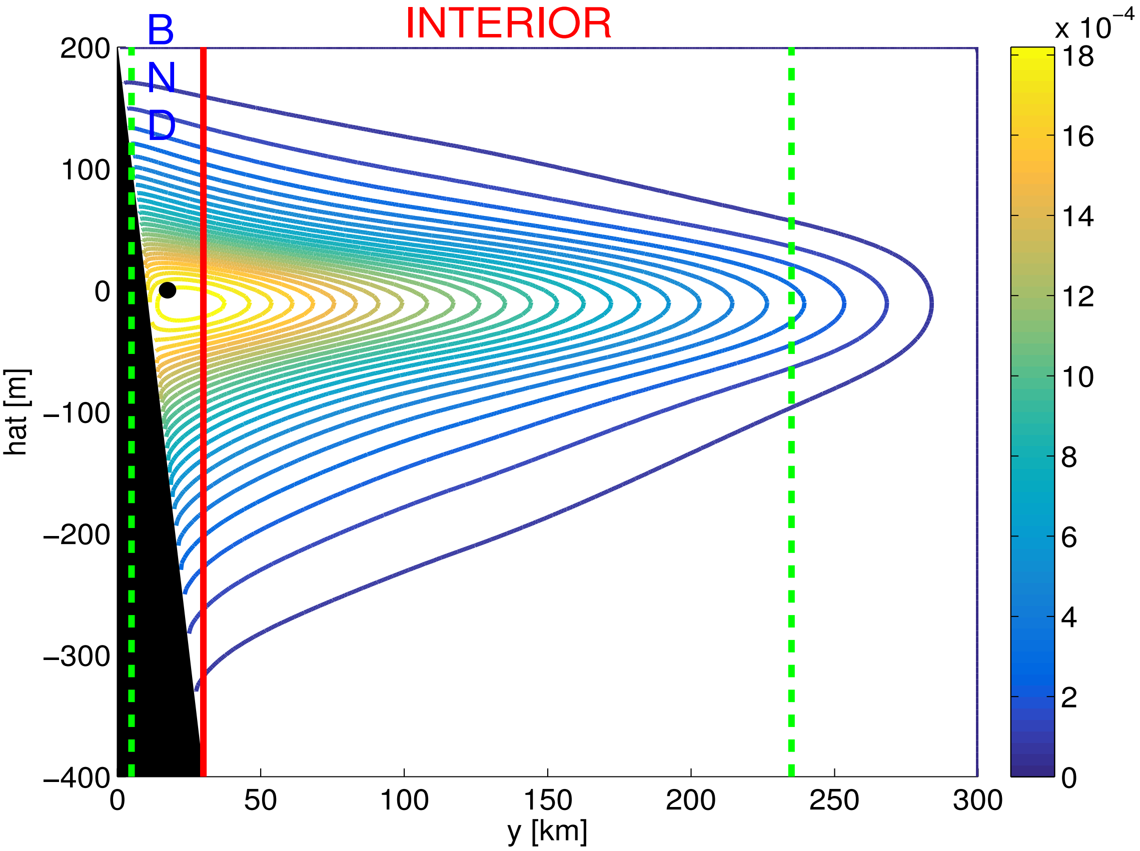


Figure 4.12: Four month tracer distribution with ($\beta = 2, u = 0.01, \frac{1}{2}C_d\frac{d}{\ell} = 0.06$ and $K_{\perp}^{int} = 3 \times 10^{-5}$ m^2 s^{-1}). Green vertical lines delimit the horizontal extent of the 4 month tracer sampling and the extent of averaging for the numerically derived interior and boundary dispersion estimates. The black dot depicts the location of the initial condition. The vertical coordinate is the height above the target surface (hat).

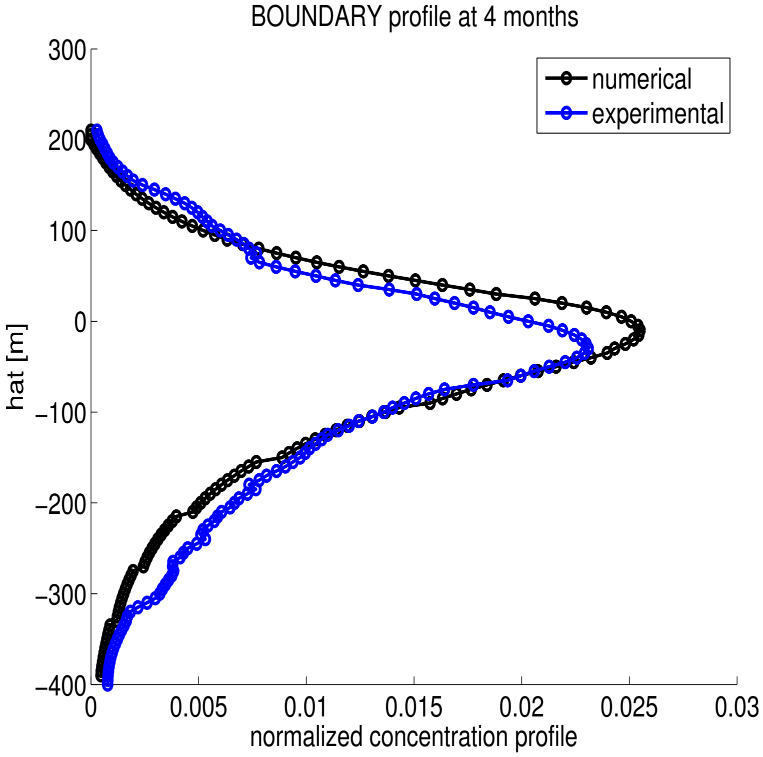


Figure 4.13: Best fit boundary profile comparison for observations (blue) and model (black) with ($\beta = 2, u = 0.01, \frac{1}{2}C_d\frac{d}{\ell} = 0.06$ and $K_\perp^{int} = 3 \times 10^{-5}$ m^2 s^{-1}). Note the vertical asymmetry about the injection level implying a downward increase in diapycnal mixing.

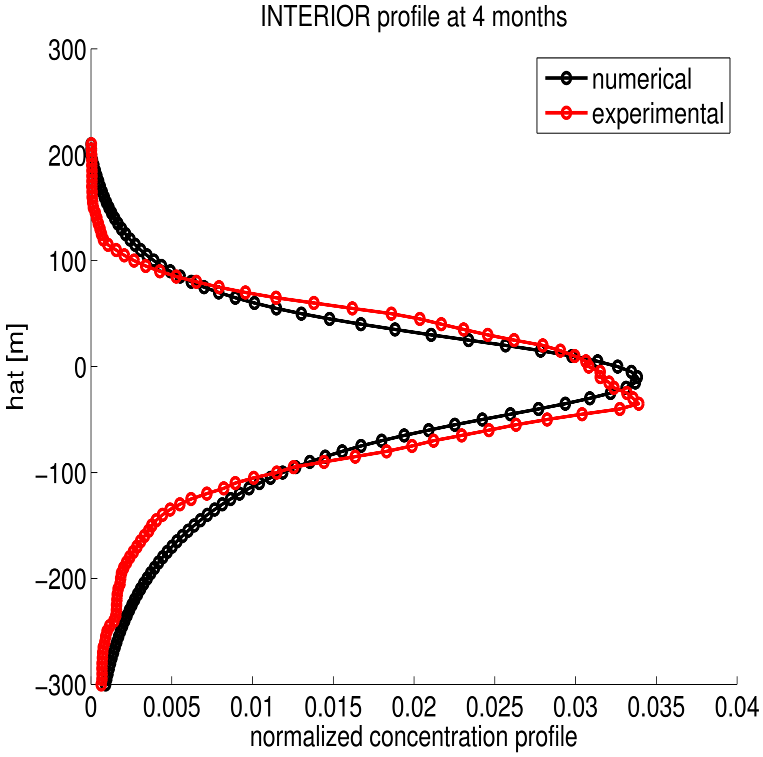


Figure 4.14: Best fit interior profile comparison for observations (red) and model (black). Otherwise as in Figure 4.13.

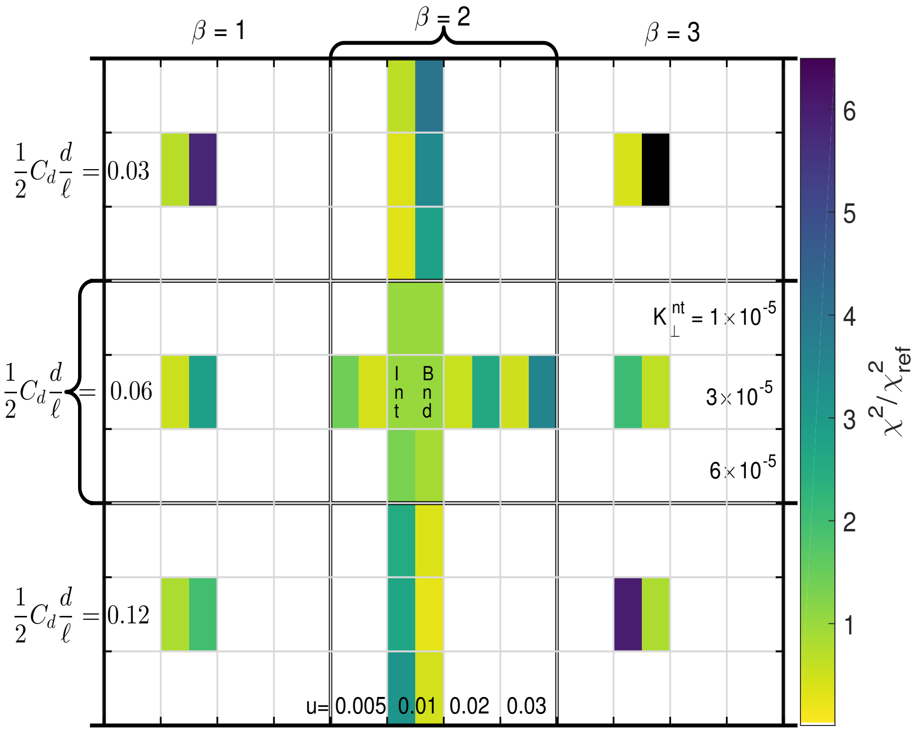


Figure 4.15: An error map in (β, u, C_d and $K_\perp^{\rm int}$) space. There are two color panels with in each parameter square. The left panel represents the interior profile and the right the boundary. Each large square contains a constant C_d and β. Within each large square there is a three-by-four matrix of possible ($K_\perp^{\rm int}, u$) entries. Mean-square error estimates are normalized to 'best fit' parameters ($\beta = 2, u = 0.01$ m s^{-1}, $\frac{1}{2}C_d\frac{d}{\ell} = 0.06$ and $K_\perp^{\rm int} = 3 \times 10^{-5}$ m^2 s^{-1}), which is indicated as a green shading. Better fits can be obtained in one domain (interior or boundary) only at the expense of the other.

4.2 Extra Material

Here, I briefly report the main mathematical details to replicate the results reported in the paper.

Exponential profile

First, we impose the following form of the diapycnal diffusivity:

$$K_Z(z) = K_Z(0)e^{-zN/\beta U}$$

Then we have to determine $K_Z(0)$ by hypothesizing that the energy for the mixing is supplied by a flow with velocity U:

$$6N^2 \int_0^\infty K_Z(0)e^{-zN/\beta U} dz = C_d U^3$$

By imposing $K_Z(\infty) = 0$, this implies:

$$K_Z(z) = \frac{C_d U^2}{6\beta N} e^{-zN/\beta U} + K_Z^{bk} \tag{4.2.1}$$

with K_Z^{bk} being the background diffusivity.

It can be of interest to derive the form of the diffusivity as function of the distance x from the slope. To do this we transform the vertical length scale $z = \beta U/N$ into $x = \frac{\beta U}{N} \frac{\cos\alpha}{\sin\alpha} \sim \frac{1}{\alpha} z$, where α is the slope of the continental shelf ($\alpha = 0.002$ for the Gulf of Mexico). By considering that $z/\alpha = x$ it follows $dz = \alpha dx$ and we obtain:

$$K_Z(x) = \frac{C_d U^2}{6\beta N} e^{-xN\alpha/\beta U} + K_Z^{bk}$$

We now calculate the vertical velocity due to mas conservation. Given Equation 4.2.1 and the following form for the diapycnal velocity:

$$w^* = \frac{\frac{\partial}{\partial z}\left(K_Z \frac{\partial \rho}{\partial z}\right)}{\frac{\partial \rho}{\partial z}} = \frac{\frac{\partial}{\partial z}\left(K_Z N^2(z)\right)}{N^2}$$

it follows:

$$w^* = \frac{1}{N^2}\frac{C_d U^2}{6\beta}\left[\frac{\partial N(z)}{\partial z}e^{-zN/\beta U} + N\frac{\partial e^{-zN/\beta U}}{\partial z}\right] + \frac{K_Z^{bk}}{N^2}\frac{\partial N^2(z)}{\partial z}$$

Given:

$$\begin{cases} N^2(z) = c_1 z + c_0 \rightarrow \frac{\partial N^2}{\partial z} = c_1 \\ \\ N(z) = \sqrt{c_1 z + c_0} \rightarrow \frac{\partial N}{\partial z} = \frac{c_1}{2\sqrt{c_1 z + c_0}} \\ \\ \frac{\partial}{\partial z}e^{-zN/\beta U} = \frac{\partial}{\partial z}\left(\frac{-zN}{\beta U}\right)e^{-zN/\beta U} = \left[\frac{-N}{\beta U} - \frac{z}{\beta U}\frac{\partial N}{\partial z}\right]e^{-zN/\beta U} \end{cases}$$

it follows:

$$w^* = \frac{C_d U^2}{6\beta N^2}\left[\frac{c_1}{2\sqrt{c_1 z + c_0}} - \frac{N^2}{\beta U} - \frac{z c_1 N}{2\beta U\sqrt{c_1 z + c_0}}\right]e^{-zN/\beta U} + \frac{K_Z^{bk}}{N^2}c_1 \tag{4.2.2}$$

Chapter 5

Conclusions and perspectives

Contents

5.1 Conclusions 157

5.2 Perspectives 161

Connectivity is crucial for the persistence of marine species populations facing anthropic pressure and climate change. The aim of this thesis was to investigate the physical mechanisms that drive connectivity between distant marine populations. In particular, I investigated the role of currents in sustaining persistence of marine populations and the role of turbulence in influencing the modeling of passive larvae dispersion.

Before outlining the future research pathways identified by this thesis, we briefly recapitulate the research strategy that we chose and its outcomes.

5.1 Conclusions

Due to contingent limitations, legislators cannot protect all marine sites. In order to set up effective conservation policies they need to identify the most important sites for persistence for a species or a set of species.

The advantage to apply graph theory to connectivity studies is due to its flexibility and independence from demographic parameters that are difficult to estimate. For this reason, the use of graph theory in connectivity studies has grown significantly in the last decade (*Moilanen*, 2011) but its interpretation is still mainly based on studies that are intrinsically different from networks of marine areas. For example, those studies either investigate relations between individuals that, by definition, have a personal behavioral preferences (e.g., *Mastrandrea et al.*, 2015) or relations between elements (e.g., proteins, *Vishveshwara et al.* 2002) that do not have a behavior but are single entities not ecological communities.

In order to clarify the ecological interpretation of graph theory, we set up a comparison between graph theory and metapopulation model analysis. We chose as reference the metapopulation model analysis because such analysis has been widely used for a much longer time than graph theory. Moreover, its interpretation is more straightforward. We wanted to assess whether the persistence of marine species that disperse passively can be trustworthy obtained from information concerning only ocean currents and the basic biological characteristics (e.g., PLD) of the species at study.

In particular, we considered different species in the Gulf of Lion that distinguish themselves by a different PLD. For each PLD, we calculated the probability of connection between different reproductive zones in the GoL with Lagrangian simulations. The connectivity matrices were then used as input for both metapopulation analysis and graph theory analysis. Metapopulation model analysis estimated the importance of a site for ensuring species' persistence in the GoL by calculating the minimum recruitment success required in that site in order to ensure the persistence when other nodes are deleted. In fact, the higher the minimum recruitment success, the more important the node is. In order to understand which graph theory measure could provide the same information, we looked for correlations between the minimum recruitment success and different graph theory measures.

The results suggests that the importance of the graph theory measures depends on the topology of the network. In general, we can distinguish a first kind of topology (corresponding to pelagic larval durations of 1 and 2 weeks) in which the density of the network is lower and another kind of topology (corresponding to pelagic larval durations of 3, 4 and 5 weeks). The first topology is characterized by more dispersed networks with longer average shortest paths between the nodes. On the other hand, the second topology is characterized by a more clustered networks with a lower mean shortest path. The correlations of these graph theory measures with persistence change on the base of the topology. Core number and clustering coefficient are the more reliable measures in the first case; z-score and core number are the more reliable measures in the second case. The only graph theory measure that results to be reliable in both topologies is the strength of the connections of a site. Another important result of this study is that betweenness centrality is not the graph theory measure that better correlates with persistence as previously assumed in the literature (e.g., *Andrello et al.*, 2013). Our results also show that modularity -in contrast to what stated in previous literature (*Thomas et al.*, 2014)- can reliably identify sub-populations.

A key aspect for an effective numerical simulation of the larvae dispersion concerns the turbulence modelling, in particular in the surface and in the bottom boundary layers. The comparison of in situ turbulence measurements with the numerical model Symphonie permitted to investigate the sensibility of numerical models to the turbulence parameterizations in the surface boundary layer. In particular, we exploited in situ measurements of kinetic energy dissipation rate ε and turbulent diffusivity K_Z acquired with an autonomous microstructure profiler (SCAMP) in the Gulf of Lion. Measurements were acquired during different campaigns and, consequently, well adapted for a statistical comparison with the predictions made by different TCSs integrated in the high-resolution numerical model.

Previous studies in the literature (e.g., *Burchard et Bolding* 2001; *Ilicak et al.* 2008; *Reffray et al.* 2015) also focused on different closure schemes and boundary conditions. However, those studies mainly exploited in situ measurements of temperature and mixed layer depths. From a turbulence point of view, those quantities can be viewed as "integrated" quantities that are the final result of more than one process. Although the dynamics of those quantities directly depends on turbulence, their modelling in a numerical way mingles the effect of various approximations at different levels. As a consequence, it is more difficult to directly point out which factor causes one TCS to perform better or worse than another one. Therefore, the originality of our approach consists in developing a series of numerical experiments to identify more directly the different elements of a TCS that could be changed in order to improve the numerical predictions of turbulence. This is a particularly important point especially for the correct estimation of turbulent fluxes.

Our study suggests that the modelling of ε is quite sensible to the choice of the TCS, b.c. and stability functions. In particular, the $k - \ell$ scheme performs better, in the surface layer, when employing surface boundary conditions specifying an equilibrium between the wind stress and the turbulent production. Instead, the $k - \varepsilon$ scheme performs better when the surface boundary conditions link the turbulent production to the energy flux through the surface, thus better accounting for the effect of breaking waves with respect to the equilibrium boundary conditions. Furthermore, the $k - \varepsilon$'s performance considerably improves when the Canuto A stability functions (*Canuto et al.*, 2001) are used. This improvement seems to be linked to the fact that less numerical values of ε are at the threshold. However, this advantage does not extend to the prediction of K_Z. In fact, the two schemes seem to predict the turbulent diffusivity comparably well. One important result of our analysis is that there is a bias between the two schemes' predictions of ε below the surface layer. In fact, when employing the same threshold on the kinetic energy k, the $k - \ell$ values of ε in this layer are roughly one order of magnitude greater than the ones predicted by the $k - \varepsilon$ scheme. This result allows current or future Symphonie's users to be aware that the threshold on k has to be chosen also on the base of the TCS that they are using. However, another important hint that our study provides is the fact that the threshold on k should not be a fixed value but rather some kind of parameterization dependent on the ocean's interior dynamic. A phenomenon, that could fit as a suitable candidate, on which such a parameterization could be based is the interior wave activity (*Gaspar et al.*, 1990).

The study of the generating mechanisms of bottom boundary turbulence is a necessary start point to investigate the capability of deep ocean communities to sustain and possibly

repopulate coastal communities. Moreover, it is pivotal in order to improve the prediction of the evolution of contaminants (*Ledwell et al.*, 2016, *Polzin et al.*, Submitted). Our study is based on the accumulated evidence (*Polzin et al.* 2014b; *Ledwell et al.* 2016) that the mixing levels measured with microstructure probes are often not sufficient to explain the ocean overturning circulation. In particular, this fact is highlighted by one tracer release experiment's data (GISR campaign, *Ledwell et al.* 2016). We exploited this data set in order to estimate the values of the parameters characterizing form drag-induced boundary mixing. To accomplish this, I implemented a 2-D numerical dispersion model that was initialized with the data from the tracer release experiment and compared the results with the tracer concentration measurements at the end of the GISR campaign.
Overall, our results suggest that the effective drag coefficients associated with highly sheared regions in well stratified regions of $O(U/N)$ height above the bottom, where U is the velocity of the flow above topography and N the Brunt-Vaisala frequency, are orders of magnitude larger than classical literature values ($C_d \sim 3.6$). Our results potentially imply that the standard closure for viscous stresses associated with a no-slip boundary condition ($C_d \sim 1 \times 10^{-3}$) is a substantial underestimation of the effective drag coefficient.

5.2 Perspectives

The approach we adopted to understand the information that can be gained applying graph theory to connectivity assessment could profit from a comparison with metapopulation models that incorporate genetic spreading (e.g., *Padron et Guizien*, 2015). This would further enhance the reliability of graph theory in connectivity studies.
An important aspect of larvae's dispersion that is currently ignored in dispersion modeling is their swimming behavior and their PLD's dependence on environmental variables like temperature. In fact, a great number of larvae possesses some diel vertical migration in response to light variations in the water column. Moreover, temperature can significantly shorten the PLD thus affecting connectivity (*Houde* 1989; *Cowen et al.* 2000; *O'Connor*

et al. 2007). This last aspect is particularly relevant if we reason in the framework of climate change. In the near future, an urgent need is to assess how a rise in the water temperature can negatively affect marine connectivity.

Furthermore, our analysis highlights a high temporal variability of the transfer patterns and, as a consequence, of the graph theory measures (Figure 2.2). A natural question is whether this variability plays a role in determining persistence. In other words, is it a a winning strategy to permanently protect some well-specified sites ? Or should we consider more closely the possibility to implement time-varying protection strategies?

Such protection strategies would require the knowledge in advance of the circulation patterns. This can be done -to some extents- with numerical circulation models. An open question is to what extent the knowledge of the current field must be known in order to predict the likelihood for a certain site to have a high value of a certain graph theory measure of importance.

On a practical point of view, protection policies of this kind still do not exist. However, the need for their implementation is suggested by an increasing number of studies (*Hobday et Hartmann* 2006; *Maxwell et al.* 2015; *Penna et al.* In Press).

As we have already said, a natural continuation of this thesis would be the investigation of the effect of different turbulence parameterizations on the simulation of larval dispersion. Indeed, as we observed significant differences between the performances of different TCSs below the MLD, some differences in larvae dispersion have to be expected.

Further insights on the different performances of the possible turbulence parameterizations and other modelling aspects could be gained with dedicated experiments. In fact, our SCAMP measurements were opportunistically taken and just a statistical comparison with the model's predictions was possible. Therefore, a study of the response of the TCSs to variations of turbulence intensity resulting from the evolution of meteorological conditions was not possible. Consequently, it is still not understood if the response of the different TCSs is equally fast (*Burchard et al.*, 2002). Hence, it would be interesting to have a dedicated project envisioning some one-day cruises -or, even better, a moored setting - targeting different meteorological situations. This would permit, through the acquisition of repeated profiles, to follow the evolution of the turbulent activity and assess the behavior of the model.

Moreover, multi-model comparisons could further highlight modelling limitations. By comparing outputs of various models using the same TCS, eventual differences would highlight the effect of, for example, vertical resolution, spatio-temporal resolution of the forcing, bottom boundary conditions. This will shortly be accomplished by a study, in the framework of the project Numerofix (PI P.Marsaleix, funded by the LEFE/GMMC program of CNRS), in which a 1-D version of NEMO will be compared to the SCAMP measurements that I analyzed.

A natural consequence of this kind of study is the investigation of the effect of different model implementations on geochemical models. Both small differences in K_Z and the MLD values could potentially have sizable impacts on biogeochemical models' predictions.

The effect of bottom boundary turbulence on connectivity is still difficult to quantify. First, its parameterizations in current ocean models are inaccurate to model some mixing mechanisms of the kind we studied in Chapter 4.

The mixing mechanisms in the bottom boundary region over rough topography that we considered are form drag and hydraulic-like effects due to flow over rough topography. However, the numerical model we implemented was an approximation of the actual ocean dynamics. In 3-D flows, form drag may act in concert with boundary trapped waves and instabilities arising from either wind forcing or eddies flow over topography. We must investigate the relative importance and interplays of these different mechanisms. One possible way that we have envisioned in a Marie-Curie proposal was to exploit high-resolution non-hydrostatic numerical models. In fact, these models allow to explore in detail the effects of different mechanisms; especially when used with unstructured mesh grids that enable very high resolutions (~ 10 cm) where and when it is needed (near the bottom in this case). From 4-D comparisons of these model outputs and in situ measurements, insights on the parameterization of the aforementioned phenomena would follow.

Moreover, I will study the importance of such mechanisms in the polar regions during my post-doc at the Korean Institute for Basic Science. Indeed, increased turbulent fluxes induced by flows above rough topography could have crucial effects in augmenting sea ice's and ice shelves' melting rates.

Bibliography

Agrawal, Y., E. Terray, M. Donelan, P. Hwang, A. W. III, W. Drennan, K. Kahma, et S. Krtaigorodskii (1992), Enhanced dissipation of kinetic energy beneath surface waves., *Nature*, *359*, 219–220. (Cited page 60.)

Alberola, C., et C. Millot (1995), On the seasonal and mesoscale variabilities of the Northern Current during the PRIMO-0 experiment in the western Mediterranean Sea., *Oceanol. Acta*, *2*, 163–192. (Cited page 62.)

Albert, R., et A.-L. Barábasi (2002), Statistical mechanics of complex networks., *Rev. Mod. Phys.*, *74*, 47. (Cited page 37.)

Alexander, M., et coauthors (2010), Recent developments in gravity-wave effects in climate models and the global distribution of gravity-wave momentum flux from observations and models., *Quart. J. Roy. Meteor. Soc.*, *136*, 1103–1124. (Cited page 125.)

Alexander, M., M. Geller, C. McLandress, S. Polavarapu, P. Preusse, F. Sassi, K. Sato, S. Eckermann, M. Ern, A. Hertzog, Y. Kawatani, M. Pulido, T. Shaw, M. Sigmond, R. Vincent, et S. Watanabe (2010), Recent developments in gravity-wave effects in climate models and the global distribution of gravity-wave momentum flux from observations and models., *Q. J. Roy. Metor. Soc.*, *136*(650), 1103–1124. (Cited page 9.)

Aloisi, J., H. Got, et A. Monaco (1973), *Carte gologique du pre-continent languedocien au 1/250000ime, International Institute for Aerial survey and Earth sciences (ITC), Neth- erlands, Enschede, Netherlands.* (Cited page 27.)

Altman, A., et M. Tennenholtz (2005), "ranking systems: The pagerank axioms., *Proceedings of the 6th ACM conference on Electronic commerce (EC-05).* (Cited page 34.)

Andrello, M., D. Mouillot, J. Beuvier, C. Albouy, W. Thuiller, et S. Manel (2013), Low connectivity between Mediterranean Marine Protected Areas: a biophysical modeling approach for the dusky grouper: *Epinephelus Marginatus.*, *PLoS ONE*, *8*(7). (Cited pages 17, 18, 19, 22, 24, 26, 32, 37, and 159.)

Andrello, M., M. Jacobi, S. Manel, W. Thuiller, et D. Mouillot (2014), Extending networks of protected areas to optimize connectivity and population growth rate., *Ecography*, *38*, 273–282. (Cited pages 3 and 34.)

Anis, A. (2006), Similarity relationships in the unstable aquatic surface layer., *J. Marine Syst.*, *33*(19), l19609. (Cited pages 68 and 71.)

Anis, A., et G. Singhal (2002), Mixing in the surface boundary layer of a tropical freshwater reservoir., *J. Marine Syst.*, *33*, 225–243. (Cited page 62.)

Arakawa, A., et V. Lamb (1977), Computational design of the basic dynamical processes of the UCLA general circulation model., *Meth. Comp. Phys.*, *17*, 173–275. (Cited page 64.)

Armi, L. (1978), Some evidence for boundary mixing in the deep ocean., *J. Geophys. Res.*, *83*, 1972–1979. (Cited pages 135 and 141.)

Armi, L. (1979), Reply to comments by c. garrett., *J. Geophys. Res.*, *84*, 5097–5098. (Cited page 141.)

Armi, L., et R. M. Jr. (1976), The bottom boundary layer of the deep ocean., *J. Geophys. Res.*, *81*, 4983–4990. (Cited pages 135 and 141.)

Arneborg, L., V. Fiekas, L. Ulmaf, et H. Burchard (2007), Gravity current dynamics and entrainment - a process study based on observation in the Arkona basin., *J. Phys. Oceanogr.*, *37*, 2094–2113. (Cited page 61.)

Bacher, C., R. Filgueira, et T. Guyondet (2016), robabilistic approach of water residence time and connectivity using Markov chains with application to tidal embayments., *J. Marine Syst.*, *153*, 25–41. (Cited page 17.)

Bader, G., et C. Hogue (2003), An automated method for finding molecular complexes in large protein interaction networks., *BMC Bioinformatics*, *4*. (Cited page 34.)

Baines, P. (1995), *Topographic Effects in Stratified Flows*, 482 pp., Cambridge University Press, Cambridge. (Cited pages 125, 126, 127, 131, 135, 137, and 138.)

Barbasi, A. (2016), *Network Science*, Cambridge University Press. (Cited page 15.)

Barrat, A., M. Barthelemy, et A. Vespignani (2008), *Dynamical processes on complex networks.*, Cambridge University Press. (Cited pages 15 and 34.)

Barrier, N., A. Petrenko, et Y. Ourmieres (2016), Strong intrusions of the Northern Mediterranean Current on the eastern Gulf of Lion: insights from in-situ observations and high resolution numerical modelling., *Ocean Model.*, *66*, 313–327. (Cited page 62.)

Batchelor, G. (1959), The spectrum of small-scale variation of convected quantities like temperature in turbulent fluid. 1. general discussion and the case of small conductivity., *J. Fluid Mech.*, *5*, 113–133. (Cited pages 5, 8, and 68.)

Berline, L., A.-M. Rammou, A. Doglioli, A. Molcard, et A. Petrenko (2014), A connectivity-based eco-regionalization method of the Mediterranean Sea., *PLoS One*, *9*. (Cited pages xv, 2, 3, and 17.)

Boersma, P. D., et J. Parrish (1999), Limiting abuse: marine protected areas, a limited solution., *Ecol. Econom.*, *31*, 287–304. (Cited pages 2 and 25.)

Bollobás, B. (1984), The evolution of random graphs., *Trans. Amer. Math. Soc.*, *1*, 257–274. (Cited page 34.)

Bondy, J., et U. Murty (1976), *Graph theory with applications.*, Elsevier Science Publishing. (Cited page 18.)

Bouffard, D., et L. Boegman (2013), A diapycnal diffusivity model for stratified environmental flows., *Dynam. Atmos. Oceans*, *61-62*, 14–34. (Cited pages 62, 69, 109, and 120.)

Bougeault, P., et P. Lacarrere (1989), Parameterization of orography-induced turbulence in a meso-beta scale model., *Mon. Weather Rev.*, *117*, 1872–1890. (Cited pages 99 and 100.)

Bourrin, F., P. Friend, C. Amos, E. Manca, A. P. C. Ulses and, X. D. de Madron, et C. Thompson (2011), Sediment dispersal from a typical Mediterranean flood: The Têt River, Gulf of Lion, *Cont. Shelf Res.*, *28*, 1895–1910. (Cited page 62.)

Boussinesq, J. (1872), Mémoires présentés par divers savants à l'Académie des Sciences : Essai sur la théorie des eaux courantes, rapport sur un mémoire de M. Boussinesq., *Institut National de France*, *23*. (Cited pages 6 and 57.)

Bowman, A., et A. Azzalini (1997), *Applied Smoothing Techniques for Data Analysis.*, New York: Oxford University Press Inc. (Cited page 72.)

Brandes, U. (2006), A faster algorithm for betweenness centrality., *J. Math. Sociol.*, *13*, 163–177. (Cited page 20.)

Bretherton, F. (1969), Momentum transport by gravity waves., *Quart. J. Roy. Meteor. Soc.*, *95*, 213–243. (Cited page 125.)

Britter, R. (1974), An experiment on turbulence in a density stratified fluid, *PhD thesis, Monash University, Victoria, Australia.* (Cited page 120.)

Burchard, H., et K. Bolding (2001), Comparative analysis of four second-moment turbulence closure models for the oceanic mixed layer., *J. Phys. Oceanogr.*, *31*, 1943–1968. (Cited pages 5, 56, 65, 66, 93, 98, and 159.)

Burchard, H., K. Bolding, T. Rippeth, A. Stips, J. Simpson, et J. Sundermann (2002), Microstructure of turbulence in the northern north sea: a comparative study of observations and model simulations., *J. Sea Res.*, *47*, 223–238. (Cited pages 62, 63, 65, 72, 89, and 162.)

Burchard, H., P. Craig, J. Gemmrich, H. van Haren, P.-P. Mathieu, H. M. Meier, W. M. N. Smith, H. Prandke, E. S. T.P. Rippeth and, W. Smyth, D. Welsh, et H. Wijesekera (2008), Observational and numerical modeling methods for quantifying coastal ocean turbulence and mixing., *Progr. Oceanogr.*, *76*, 399–442. (Cited pages 8 and 60.)

Burgess, S., K. Nickols, C. Griesemer, L. Barnett, A. Dedrick, et E. S. L. Yamane (2014), Beyond connectivity: how empirical methods can quantify population persistence to improve marine protected area design., *Ecol. Apppl.*, *24*, 257–270. (Cited page 3.)

Butler, M. J., C. Paris, J. Goldstien, H. Matsuda, et R. Cowen (2011), Behavior constrains the dispersal of long lived spiny lobster larvae., *Mar. Ecol.-Prog. Ser.*, *422*, 223–237. (Cited page 4.)

Canuto, V., A. Howard, Y. Cheng, et M. Dubovikov (2001), Ocean turbulence. Part I: One-point closure model-momentum and heat vertical diffusivities., *J. Phys. Oceanogr.*, *31*, 1413–1426. (Cited pages 8, 61, 66, 67, 84, 94, and 160.)

Caswel, H. (2001), *Matrix population models; 2nd edition*, Sinauer Associates, Sunderland, MA, USA. (Cited page 3.)

Charles, F., E. Jordana, j. M. Amouroux, A. Gremare, M. Desmalades, et L. Zudaire (2003), Reproduction, recruitment and larval metamorphosis in the serpulid polychaete *Ditrupa Arietina* (O.F. Müller)., *Estuar. Coast. Shelf S.*, *57*, 435–443. (Cited page 30.)

Chen, H., M. Hondzo, et A. R. Rao (2002), Segmentation of temperature microstructure., *J. Geophys. Res.*, *12*, 4–1 4–13. (Cited page 110.)

Conan, P., et C. Millot (1992), Variability of the Northern Current off Marseille, western Mediterranean Sea, from February to June 1992., *Oceanol. Acta*, *2*, 193–205. (Cited page 62.)

Corrsin, S. (1951), On the spectrum of isotropic temperature fluctuations in an isotropic turbulence., *J. Appl. Phys.*, *22*, 469–473. (Cited page 5.)

Costa, A., A. Doglioli, K. Guizien, et A. Petrenko (In Preparation), A new interpretation of graph theory measures in evaluating marine metapopulations persistence: The study case of soft-bottom polychaetes in the Gulf of Lion. (Cited pages 24 and 35.)

Cowen, R., K. Lwiza, et S. Sponaugle (2000), Connectivity of marine populations: open or closed?, *Science*, *287*, 857–859. (Cited pages 4 and 161.)

Craig, P., et M. Banner (1994), Modelling wave-enhanced turbulence in the ocean surface layer., *J. Phys. Oceanogr.*, *24*, 2546–2559. (Cited pages 8, 61, 67, 93, and 97.)

Cuypers, Y., P. Bouruet-Aubertot, C. Marec, et J.-L. Fuda (2012), Characterization of turbulence and validation of fine-scale parametrization in the Mediterranean Sea during BOUM experiment., *Biogeosciences*, *9*, 3131–3149. (Cited pages 62, 71, 101, and 111.)

Davidson, P. (2004), *Turbulence. An introduction for scientists and engineers.*, Oxford University Press, Oxford, UK. (Cited pages xv, 5, 6, and 58.)

Davis, R. (1991), Observing the general circulation with floats., *Deep Sea Res. Part A*, *38*, S531–S571. (Cited page 137.)

Djikstra, E. (1959), A note on two problems in connection with graphs., *Numerische Mathematik*, *1*, 269–271. (Cited page 20.)

Doglioli, A., M. Magaldi, L. Vezzulli, et S. Tucci (2004), Development of a numerical model to study the dispersion of wastes coming from a marine fish farm in the Ligurian Sea (Western Mediterranean)., *Aquaculture*, *231*, 215–235. (Cited page 17.)

Dufau-Julliand, C., P. Marsaleix, A. Petrenko, et I. Dekeyser (2004), Three-dimensional modeling of the Gulf of Lion's hydrodynamics (northwest Mediterranean) during January 1999 (MOOGLI3 Experiment) and late winter 1999: Western Mediterranean Intermediate Water's (WIW's) formation and its cascading over the shelf break., *J. Geophys. Res.*, *190*. (Cited page 63.)

Duraiappah, A., et N. Shahid (2005), Millennium ecosystem assessment.ecosystems and human well-being: Biodiversity synthesis, *World Resources Institute, Washington, DC. USA*. (Cited page 1.)

Eriksen, C. (1985), Implications of ocean bottom reflection for internal wave spectra and mixing., *J. Phys. Oceanogr.*, *15*, 1145–1156. (Cited page 125.)

Estournel, C., et D. Guedalia (1987), A new parameterization of eddy diffusivities for nocturnal boundary layer modelling., *Bound. Lay. Meteorol.*, *39*, 191–203. (Cited pages 64 and 98.)

Estournel, C., P. Broche, P. Marsaleix, J.-L. Devenon, F. Auclair, et R. Vehil (2001), The Rhone river plume in unsteady conditions : numerical and experimental results., *Estuar. Coast. Shelf S.*, *53*, 25–38. (Cited pages 8 and 61.)

Estournel, C., X. D. de Madron, P. Marsaleix, F. Auclair, C. Julliand, et R. Vehil (2003), Observation and modeling of the winter coastal oceanic circlation in the Gulf of Lion under wind conditions influenced by the continental orography (FETCH experiment)., *J. Geophys. Res.-Oceans*, *108*, 25–38. (Cited pages 27, 62, and 64.)

Estournel, C., V. Zervakis, P. Marsaleix, A. Papadopoulos, F. Auclair, L. Perovolitis, et E. Tragou (2005), Dense water formation and cascading in the Gulf of Thermalikos (North Aegean), from observaions and modelling., *Cont. Shelf Res.*, *25*, 2366–2386. (Cited page 64.)

Estournel, C., P. Testor, P. Damien, F. D'Ortenzio, P. Marsaleix, P. Conan, F. Kessouri, X. D. de Madron, L. Coppola, J.-M. Lellouche, S. Belamari, L. Mortier, C. Ulses, M. Bouin, et L. Prieur (2016), High resolution modeling of dense water formation in the north-western Mediterranean during winter 2012-2013: Processes and budget., *J. Geophys. Res.*, *121*. (Cited pages 63 and 64.)

Fagiolo, G. (2007), Clustering in complex directed networks., *Phys. Rev. E*, *76*. (Cited pages 34 and 35.)

Farmer, D., et M. Li (1995), Patterns of bubble clouds organized by Langmuir circulation., *J. Phys. Oceanogr.*, *25*, 1426–1440. (Cited page 60.)

Ferrari, R. (2014), Oceanography: What goes down must come up., *Nature*, *513*, 179–180. (Cited page 63.)

Ferrari, R., et C. Wunsh (2009), Ocean circulation kinetic energy-reservoirs, sources and sinks., *Ann. Rev. Fluid Mech.*, *41*, 253–282. (Cited pages 60 and 118.)

Finnigan, T. D., D. Luther, et R. Lukas (2002), Observations of enhanced diapycnal mixing near the Hawaiian Ridge, *J. Phys. Oceanogr.*, *32*, 2988–3002. (Cited page 118.)

Flexas, M. M., X. D. de Madron, M. A. Garcia, M. Canals, et P. Arnau (1997), Flow variability in the Gulf of Lions during the MATER HFF experiment (March-May 1997)., *J. Mar. Sys.*, *33*, 197–214. (Cited page 62.)

Flierl, G., et C. Davis (1993), Biological effects of Gulf Stream meandering, *J. Marine Res.*, *51*, 529–560. (Cited page 61.)

Fraysse, M., I. Pairaud, O. Ross, V. Faure, et C. Pinazo (2014), Intrusion of Rhone river diluted water into the Bay of Marseille: generation processes and impacts on ecosystem functioning., *J. Geophys. Res. Oceans*, *119.* (Cited page 63.)

Frisch, U. (1995), *Turbulence: The Legacy of A. N. Kolmogorov.*, Cambridge University Press, Cambridge, UK. (Cited pages xv and 6.)

Galperin, B., L. Kantha, S. Hassid, et A. Rosati (1988), A quasi-equilibrium turbulent energy model for geophysical flows., *J. Atmos. Sci*, *55*, 45–62. (Cited pages 7, 61, and 64.)

Galpern, P., M. Manseau, et A. Fall (2011), Patch-based graphs of landscape connectivity: A guide to construction, analysis and application for conservation, *Biol. Conserv.*, *144*, 44–55. (Cited page 26.)

Garner, S. (2005), A topographic drag closure built on an analytical base flux., *J. Atmos. Sci.*, *62*, 2302–2315. (Cited page 126.)

Garrett, C. (1979), Comment on 'some evidence for boundary mixing in the deep ocean' by laurence armi., *J. Geophys. Res.*, *84*, 5095. (Cited page 141.)

Garrett, C., et D. Gilbert (1988), *Estimates of vertical mixing by internal waves reflected off a sloping bottom, in Small-Scale Turbulence and Mixing in the Ocean, Proceedings of the* 19^{th} *International Liege Colloquium on Ocean Hydrodynamics.*, 405-424 pp., edited by J.C.J. Nihoul and B.M. Jamart, Elsevier, New York. (Cited page 125.)

Garrett, C., P. MacCready, et P. Rhines (1993), Boundary mixing and arrested Ekman layers: Rotating, stratified flow near a sloping boundary., *Ann. Rev. Fluid Mech.*, *25*, 291–323. (Cited page 141.)

Gaspar, P., Y. Gregoris, et J. Lefevre (1990), A simple eddy kinetic energy model for the simulation of oceanic vertical mixing: Tests at station Papa and long-term upper ocean study site., *J. Geophys. Res.*, *95*, 16,179–16,193. (Cited pages 7, 58, 61, 63, 65, 66, 93, 94, 99, 100, 109, and 160.)

Gatti, J., A. Petrenko, Y. Leredde, J.-L. Devenon, et C. Ulses (2006), The Rhone river dilution zone present in the northeastern shelf of the Gulf of Lion in December 2003., *Continental Shelf Res.*, *26*, 1794–1805. (Cited page 62.)

Ghezzo, M., F. D. Pascalis, G. Umgiesser, P. Zemlys, M. Sigovini, C. Marcos, et A. Perez-Ruzafa (2015), Connectivity in three European coastal lagoons., *Estuar. Coasts*, *38*, 1764–1781. (Cited page 17.)

Gibson, C., et W. Schwarz (1963), The universal equilibrium spectra of turbulent velocity and scalar fields., *J. Fluid Mech.*, *16*, 365–384. (Cited pages 68 and 112.)

Guizien, K., T. Brochier, J.-C. Duchene, B.-S. Koh, et P. Marsailax. (2006), Dispersal of *Owenia fusiformis* larvae by wind-driven currents: turbulence, swimming behavior and mortality in a three-dimensional stochastic model, *Marine Ecology Progress Series*, *311.* (Cited page 30.)

Guizien, K., P. Belharet, P. Marsaleix, et J. Guarini (2012), Using larval dispersal simulations for marine protected area design: application to the Gulf of Lion (NW mediterranean)., *Limnology and Oceanography*, (2), 026,113. (Cited pages 30, 31, and 32.)

Guizien, K., M. Belharet, C. Moritz, et J. Guarini (2014), Vulnerability of marine benthic metapopulations: implications of spatially structured connectivity for conservation practice., *Divers. Distrib.*, *20*(12), 1392–1402. (Cited pages xv, xvi, 2, 3, 10, 21, 23, 24, 26, 28, 29, and 30.)

Hamilton, P., et A. Lugo-Fernandez (2001), Observations of high speed deep currents in the northern Gulf of Mexico., *Geophys. Res. Lett.*, *28*, 2867–2870. (Cited page 125.)

Hamilton, P., et A. Lugo-Fernandez (2009), Topographic rossby waves in the Gulf of Mexico., *Prog. Oceanogr.*, *82*, 1–31. (Cited page 125.)

Han, L. (2014), A two-time-level split-explicit ocean circulation model (MASNUM) and its validation., *Acta Oceanol. Sin.*, *33*, 11–35. (Cited page 66.)

Hanski, I. (1999), Habitat connectivity, habitat continuity, and metapopulations in dynamics landscapes., *Oikos*, *87*, 201–219. (Cited pages 1 and 25.)

Hastings, A., et L. Botsford (2006), Persistence of spatial populations depends on returning home., *Proc. Natl. Acad. Sci. USA*, *103*, 6067–6072. (Cited pages 2, 35, and 36.)

Hauser, D., H. Branger, S. Bouffies-Cloché, S. Despiau, W. M. Drennan, H. Dupuis, P. Durand, X. D. de Madron, C. Estournel, L. Eymard, C. Flamant, H. C. Graber, C. Guérin, K. Kahma, G. Lachaud, J.-M. Lefèvre, J. Pelon, H. Pettersson, B. Piguet, P. Queffeulou, D. Tailliez, J. Tournadre, et A. Weill (2003), The FETCH experiment: an overview., *J. Geophys. Res. Oceans*, *108.* (Cited page 62.)

Hensler, S., et J. Jude (2007), Diel vertical migration of Round Goby larvae in the Great Lakes., *J. Great Lakes Res.*, *33*, 295–302. (Cited page 10.)

Herrmann, M., F. Diaz, C. Estournel, P. Marsaleix, et C. Ulses (2013), Impact of atmospheric and oceanic interannual variability on the Northwestern Mediterranean Sea pelagic planktonic ecosystem and associated carbon cycle., *J. Geophys. Res.*, *118*, 1–22. (Cited page 64.)

Herrmann, M., C. Estournel, F. Adloff, et F. Diaz (2014), Impact of climate change on the northwestern Mediterranean Sea pelagic planktonic ecosystem and associated carbon cycle., *J. Geophys. Res. Oceans*, *119*. (Cited page 62.)

Hobday, A., et K. Hartmann (2006), Near real-time spatial management based on habitat predictions for a longline by catch species., *Fisheries Manag. Ecol.*, *13*, 365–380. (Cited page 162.)

Hogg, A., G. Ivey, et K. Winters (2001), Hydraulics and mixing in controlled exchange flows., *J. Geophys. Res.*, *106*, 959–972. (Cited page 71.)

Horson, G. (1946), Reproduction and larval development of danish marine bottom invertebrates, with special reference to the planktonic larvae in the sound (Oresund)., *Medr. Komm. Danm. Fisk. og Havunders. S. Plankton*, *IV*, 1–523. (Cited page 25.)

Hossain, M. S. (1980), *Mathematische Modellierung von turbulenten Auftriebsstromungen. Ph.D. dissertation*, 145 pp., University of Karlsruhe, Karlsruhe, Germany. (Cited page 66.)

Houde, E. (1989), Comparative growth, mortality, and energetics of marine fish larvae: temperature and implied latitudinal effects., *US Fish Bull.*, *87*, 471–495. (Cited pages 4 and 161.)

Hu, Z., A. Doglioli, A. Petrenko, P. Marsaleix, et I. Dekeyser (2009), Numerical simulations of eddies in the Gulf of Lion., *Ocean Modell.*, *28/4*, 203–208. (Cited page 63.)

Hu, Z., A. Petrenko, A. Doglioli, et I. Dekeyser (2011), Numerical study of eddy generation in the western part of the Gulf of Lion., *J. Geophys. Res. Oceans*, *116*. (Cited page 62.)

Hwang, W., K. Taehyong, R. Murali, et Z. Aidong (2008), Bridging centrality: Graph mining from element level to group level., *Proceedings KDD 2008*, , 366–344. (Cited pages 47 and 48.)

Ilicak, L., T. Özgökmen, H. Peters, H. Baumert, et M. Iskandarini (2008), Performance of two-equation turbulence closures in three-dimensional simulations of the Red Sea overflow., *Ocean Modell.*, *24*, 122–139. (Cited pages 56, 61, 66, and 159.)

Imberger, J., et B. Boashash (1986), Application of the wigner-ville distribution to temperature gradient microstructure: A new technique to study small-scale variations., *J. Phys. Oceanogr.*, *16*, 1997–2012. (Cited page 110.)

Imberger, J., et G. Ivey (1991), On the nature of turbulence in a stratified fluid. Part I: The energetics of mixing., *J. Phys. Oceanogr.*, *21*, 650–658. (Cited pages 101, 110, and 120.)

Itsweire, E., J. Koseff, D. Briggs, et J. Ferziger (1993), Turbulence in stratified shear flows: Implications for interpreting shear-induced mixing in the ocean, *J. Phys. Oceanogr.*, *23*, 1508–1522. (Cited page 120.)

Jacobi, M., C. André, K. Doos, et P. Jonsson (2012), Identification of subpopulations from connectivity matrices., *Ecography*, *35*, 31–44. (Cited pages 3, 17, and 26.)

Jonsson, P., M. Jacobi, et P.-O. Moksnes (2015), How to select networks of marine protected areas for multiple species with different dispersal strategies., *PLoS ONE*, *22*, 1–13. (Cited page 17.)

Jurado, E., H. van der Woerd, et H. Dijkstra (2012), Microstructure measurements along a quasi-meridional transect in the northeastern Atlantic Ocean., *J. Geophys. Res.*, *117*, C04,016. (Cited pages 62, 71, and 112.)

Kantha, L. (2005), Barotropic tides in the Gulf of Mexico, in circulation in the Gulf of Mexico: Observations and models, *Geophys. Monogr. Ser.*, *161*, 159–163. (Cited page 125.)

Kantha, L., et C. Clayson (1994), An improved mixed layer model for geophysical applications., *J. Geophys. Res.*, *99*, 25,235–25,266. (Cited pages 7, 61, 66, 67, 84, 94, and 99.)

Kantha, L., et C. Clayson (2000), Processes en geophysical fluid flows., *International Geophysics Series. Vol. 67.* (Cited page 60.)

Kelvin, L. S. W. T. (1887), On the propagation of laminar motion through a turbulently moving inviscid liquid., *Phil. Mag.*, *24*, 342–353. (Cited page 5.)

Kersalé, M., Petrenko, A. Doglioli, I. Dekeyser, et F. Nencioli (2013), Physical characteristics and dynamics of the coastal Latex09 Eddy derived from in situ data and numerical modeling., *J. Geophys. Res.*, *118*, 1–11. (Cited page 62.)

Kininmonth, S., M. J. van Oppen, et H. Possingham (2010a), Determining the community structure of the coral *Seriatopora hystirix* from hydrodynamic and genetic networks., *Ecol. Model.*, *221*, 2870–2880. (Cited pages 3, 17, 18, 19, and 38.)

Kininmonth, S., M. De'ath, et H. Possingham (2010b), Graph theoretic topology of the Great but small Barrier Reef world., *Theor. Ecol.*, *3*, 75–88. (Cited pages 17, 18, 19, 22, 24, 26, and 35.)

Kolmogorov, A. (1942), The equation of turbulent motion in an incompressible fluid., *Izv. Akad. Nauk SSSR ser Fiz.*, *6*, 56–58. (Cited pages 5, 58, and 99.)

Kolmogorov, A. (1962), A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number., *J. Fluid Mech.*, *13*, 82–85. (Cited page 5.)

Kool, J., A. Moilansen, et E. Treml (2013), Population connectivity: recent advances and new perspectives., *Landscape Ecol.*, *28*, 165–185. (Cited page 25.)

Kunze, E., et S. L. Smith (2002), The role of smallscale topography in turbulent mixing of the global ocean, *Oceanography*, *17*, 51–60. (Cited page 118.)

Labrune, C., A. Gremare, J.-M. Amoroux, R. Sarda, J. Gil, et S. Taboada (2007), Assessment of soft-bottom polychaete assemblages in the Gulf of Lion (NW Mediterranean) based on a mesoscale survey., *Estuar. Coast. Shelf S.*, *71*, 133–147. (Cited pages 10, 26, and 29.)

Lagabrielle, E., E. Crochelet, M. Andrello, S. Schill, S. Arnaud-Haond, N. Alloncle, et B. Ponge (2014), Connecting MPAs - eight challenges for science and management., *Aquatic Conservation Marine And Freshwater Ecosystems*, *24*, 94–110. (Cited pages 25 and 26.)

Large, W., et S. Yaeger (2004), Diurnal to decadal global forcing for ocean and sea-ice models: The data set and flux climatologies., *NCAR Tech. Note NCAR/TN-4601STR*, doi:10.5065/D6KK98Q6. (Cited page 64.)

Large, W., J. McWilliams, et S. Doney (1994), Oceanic vertical mixing: a review and a model with a non-local boundary layer parametrization., *Rev. Geophys.*, *32*, 363–403. (Cited pages 7 and 61.)

Ledwell, J., E. Montgomery, K. Polzin, L. S. Laurent, R. Schmitt, et J. M. Toole (2000), Evidence for enhanced mixing over rough topography in the abyssal ocean., *Nature*, *403*, 179–182. (Cited page 118.)

Ledwell, J., R. He, Z. Xue, S. DiMarco, L. Spencer, et P. Chapman (2016), Dispersion of a tracer in the deep gulf of mexico, *J. Geophys. Res.*, *121*, 1110–1132. (Cited pages xvi, 9, 11, 119, 123, 124, 131, 132, 138, 139, and 161.)

Lott, F., et M. Miller (1997), A new sub-grid orographic drag parameterization: Its formulation and testing., *Q. J. Roy. Meteor. Soc.*, *123*, 101–127. (Cited page 126.)

Lozovatsky, I., M. Figueroa, E. Roget, H. Fernando, et S. Shapovalov (2005), Observation and scaling of the upper mixed layer in the North Atlantic., *J. Geophys. Res.*, *110*(C05013). (Cited page 71.)

Lubchenco, J., S. Palumbi, S. Gaines, et S. Andelman (1999), Plugging a hole in the ocean: the emerging science of marine reserves., *Ecol. Appl.*, *13*, S3–S7. (Cited pages 1 and 25.)

Luketina, D., et J. Imberger (2000), Determining turbulent kinetic energy dissipation from Batchelor curve fitting., *JAOT*, *18*, 100–113. (Cited page 68.)

Mansui, J., A. Molcard, et Y. Ourmieres (2015), Modelling the transport and accumulation of floating marine debris in the Mediterranean basin., *Marine Poll. Bull.*, *91*, 249–257. (Cited page 17.)

Maraldi, C., J. Chanut, B. Levier, N. Ayoub, P. D. Mey, G. Reffray, F. Lyard, S. Cailleau, M. Drevillon, E. Fanjul, M. Sotillo, P. Marasalaix, the Mercator Research, et D. Team (2013), NEMO on the shelf: assessment of the Iberia-Biscay-Ireland configuration., *Ocean Sci.*, *9*, 745–771. (Cited page 67.)

Marchesiello, P., L. Debreu, et X. Couvelard (2009), Spurious diapycnal mixing in terrain-following coordinate models: The problem and a solution., *Ocean Modell.*, *26*, 156–169. (Cited page 63.)

Marsaleix, P., F. Auclair, et C. Estournel (2006), Considerations on open boundary conditions for regional and coastal ocean models., *Ocean Technol.*, *23*, 1604–1613. (Cited pages 21, 29, and 64.)

Marsaleix, P., F. Auclair, J. Floor, M. Herrmann, C. Estournel, I. Pairaud, et C. Ulses (2008), Energy conservation issues in sigma-coordinate free-surface ocean models., *Ocean Model.*, *20*, 61–89. (Cited pages 8, 56, 62, 63, and 64.)

Marsaleix, P., F. Auclair, et C. Estournel (2009), Low-order pressure gradient schemes in sigma coordinate models: The seamount test revisited., *Ocean Model.*, *30*, 169–177. (Cited page 64.)

Marsaleix, P., F. Auclair, T. Duhaut, C. Estournel, C. Nguyen, et C. Ulses (2012), Alternatives to the Robert-Asselin filter., *Ocean Modell.*, *41*, 53–66. (Cited page 64.)

Mashayek, A., R. Ferrari, M. Nikurashin, et W. Peltier (2015), Influence of enhanced abyssal diapycnal mixing on stratification and the ocean overturning circulation., *J. Phys. Oceanogr.*, *45*, 2580–2597. (Cited pages xx, 118, and 119.)

Mastrandrea, R., J. Fournet, et A. Barrat (2015), Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys, *PLOS ONE*, *10*(9), 1–26, doi:10.1371/journal.pone.0136497. (Cited page 158.)

Maxwell, S., E. Hazen, R. Lewison, D. Dunn, H. B. S. Bograd, D. Briscoe, S. Fossette, A. Hobday, M. Bennett, S. Benson, M. Caldwell, D. Costa, H. Dewar, T. Eguchi, L. Hazen, S. Kohin, T. Sippel, et L. Crowder (2015), Dynamic ocean management: Defining and conceptualizing real-time management of the ocean., *Mar. Policy*, *58*, 45–20. (Cited page 162.)

McRae, B., B. Dickson, T. Keitt, et V. Shah (2008), Using circuit theory to model connectivity in ecology, evolution, and conservation., *Ecology*, *89*(10), 2712–2724. (Cited page 3.)

Mellor, G., et T. Yamada (1982), Development of a turbulence closure model for geophysical fluid problems., *Geophys. Res. Phys.*, *20*, 851–855. (Cited pages 7 and 61.)

Michaud, H., P. Marsaleix, Y. Leredde, C. Estournel, F. Bourrin, F. Lyard, C. Mayet, et F. Ardhuin (2012), Three-dimensional modelling of wave-induced current from the surf zone to the inner shelf., *Ocean Sci.*, *8*, 657–681. (Cited pages 64 and 98.)

Millot, C. (1990), The Gulf of Lions' hydrodynamics., *Cont. Shelf Res.*, *10*, 885–894. (Cited pages 27 and 62.)

Moilanen, A. (2011), On the limitations of graph-theoretic connectivity in spatial ecology and conservation., *Journal of Applied Ecology*, *48*, 1543–1547. (Cited pages 3, 15, 17, 25, 26, and 157.)

Moniz, R., D. Fong, C. Woodson, S. Willis, M. Stacey, et S. Monismith (2012), Scale-dependent dispersion within the stratified interior on the shelf of northern Monterey Bay., *J. Phys. Oceanogr*, *44*, 1049–1064. (Cited pages 101 and 111.)

Moum, J. N., M. C. Gregg, R. C. Lien, et M. E. Carr (1995), Comparison of turbulence kinetic energy dissipation rate estimates from two ocean microstructure profilers., *Journal of Atmospheric and Oceanic Technology*, *12*(2), 346–366. (Cited page 113.)

Munk, W. (1966), Abyssal recipes., *Deep-Sea Research*, *13*, 707–730. (Cited page 118.)

Nanda, S., et D. Kotz (2008), Localized bridging centrality for distributed., *Proceedings of the 2008 IEEE*, pp. 1–6. (Cited page 48.)

Naveira Garabato, A., K. Polzin, B. K. K. Heywood, et M. Visbeck (2004), Widespread intense turbulent mixing in the southern ocean., *Science*, *303*, 210–212. (Cited page 9.)

Nerini, D., et B. Ghattas (2007), Classifying densities with functional regression trees : an application in oceanology, *Comput. Stat. Data An.*, *101*, 4287–4297. (Cited page 72.)

Newman, M. (2006), Modularity and community structure in networks., *Proc. Natl. Acad. Sci. USA*, *103*, 8577–8582. (Cited page 37.)

Newman, M., et M. Girwan (2004), Finding and evaluating community structure in networks., *Phys. Rev. E*, *69*(2), 026,113. (Cited page 37.)

Nicosia, V., G. Mangioni, V. Carchiolo, et M. Malgeri (2009), Extending the definition of modularity to directed graphs with overlapping communities., *J. Stat. Mech-Theory E*, *3*, 1742–5468. (Cited page 38.)

Nikurashin, M., et R. Ferrari (2009), Radiation and dissipation of internal waves generated by geostrophic motions impinging on small-scale topography: Application to the Southern Ocean, *J. Phys. Oceanogr.*, *40*, 2025–2042. (Cited page 118.)

Nikurashin, M., R. Ferrari, N. Grisouard, et K. Polzin (2014), The impact of finite-amplitude bottom topography on internal wave generation in the Southern Ocean., *J. Phys. Oceanogr.*, *32*, 3499–2066. (Cited page 131.)

Obukhof, A. (1949), The temperature field structure in a turbulent flow, *Izv. Akad. N auk SSSR, Ser. Geoqr. Geojiz.*, *13*, 58–69. (Cited page 5.)

O'Connor, M., J. B. ans S.D. Gaines, B. Halpern, S. Lester, B. Kinlan, et J. Weiss (2007), Temperature control of larval dispersal and the implications for marine ecology, evolution, and conservation., *Proc. Natl Acad. Sci. USA*, *104*, 1266–1271. (Cited pages 4 and 161.)

Oey, L.-Y., et H.-C. Lee (2002), Deep eddy energy and topographic Rossby waves in the Gulf of Mexico., *J. Phys. Oceanogr.*, *32*, doi:3499-3527. (Cited page 125.)

Osborn, T. (1980), Estimates of the local rate of vertical diffusion from dissipation measurements., *J. Phys. Oceanogr*, *10*, 83–89. (Cited pages 71 and 119.)

Padron, M., et K. Guizien (2015), Modelling the effect of demographic traits and connectivity on the genetic structuration of marine metapopulations of sedentary benthic invertebrates., *ICES J.Mar. Sci.*, doi:10.1093/icesjms/fsv158. (Cited pages 10, 26, and 161.)

Palmer, T., G. Schutts, et R. Swinbank (1986), Alleviation of systematic westerly wind bias in general circulation and numerical weather prediction models through an orographic gravity wave drag parameterization., *Quart. J. Roy. Meteor. Soc.*, *112*, 2056–2066. (Cited pages 9 and 125.)

Penna, A. D., P. Koubbi, C. Cotte, C. Bon, C.-A. Bost, et F. d'Ovidio (In Press), Lagrangian analysis of multi-satellite data in support of open ocean Marine Protected Area design., *Deep-Sea Res.* (Cited page 162.)

Peters, H., et H. Baumert (2007), Validating a turbulence closure against estuarine microstructure measurements., *Ocean Model.*, *19*, 183–203. (Cited pages 4 and 61.)

Peters, H., M. Gregg, et J. Toole (1988), On the parametrization of equatorial turbulence., *J. Geophys. Res.*, *93*, 19–1218. (Cited page 71.)

Peters, H., H. Baumert, et J. Jacob (2004), *Partially mixed estuaries: the Hudson River.*, 324-333 pp., Cambridge University Press. (Cited page 4.)

Peters, H., H. Baumert, et J. Jacob (2009), Observations of turbulent mixing in a phytoplankton thin layer: Implications for formation, maintenance, and breakdown., *Limnol. Oceanogr.*, *54*, 1353–1368. (Cited page 62.)

Petrenko, A., A. Doglioli, F. Nencioli, M. Kersalé, Z. Hu, et F. d'Ovidio (2017), A review of the LATEX project: mesoscale to submesoscale processes in a coastal environment, *Ocean Dynamics*, pp. 1–21. (Cited page 62.)

Petrenko, A. A. (2003), Variability of circulation features in the Gulf of Lion NW Mediterranean Sea. Importance of inertial currents., *Oceanol. Acta*, *26*, 323–338. (Cited pages 27 and 62.)

Petrenko, A. A., Y. Leredde, et P. Marsaleix (2005), Circulation in a stratified and wind-forced Gulf of Lions, NW Mediterranean Sea: in situ and modeling data)., *Cont. Shelf Res.*, *25*, 7–27. (Cited pages 27 and 62.)

Petrenko, A. A., C. Defau, et C. Estournel (2008), Barotropic eastward currents in the western Gulf of Lion, northwestern Mediterranean Sea, during stratified conditions., *J. Mar. Sys.*, *74*, 406–428. (Cited page 27.)

Pinazo, C., P. Marsaleix, B. Millet, C. Estournel, V. Kondrachoff, et R. Véhil (2001), Phytoplankton variability in summer in the northwestern Mediterranean: Modelling of the wind and freshwater impacts., *J. Coastal Res.*, *17*. (Cited page 62.)

Polzin, K. (2004a), Idealized solutions for the energy balance of the finescale internal wavefield., *J. Phys. Oceanogr.*, *34*, 231–246. (Cited page 134.)

Polzin, K. (2004b), An abyssal recipe., *Ocean Model.*, *30*, 298–309. (Cited page 134.)

Polzin, K. (2014), Statistics of the Richardson number: Mixing models and finestructure., *J. Phys. Oceanogr.*, *26*, 1409–1425. (Cited page 134.)

Polzin, K., J. Toole, J. Ledwell, et R. Schmitt (1997), Spatial variability of turbulent mixing in the abyssal ocean., *Science*, *276*, 93–96. (Cited pages 9, 118, and 134.)

Polzin, K., A. N. Garabato, T. Huussen, B. Sloyan, et S. Waterman (2014a), Finescale parameterizations of turbulent dissipation., *J. Geophys. Res. Oceans*, *119*, 1383–1419. (Cited pages xxii, 124, 128, 129, and 143.)

Polzin, K., A. N. Garabato, E. Abrahamsen, L. Jullion, et M. Meredith (2014b), Boundary mixing in Orkney Passage Outflow., *J. Geophys. Res.*, *119*, doi:10.1002/2014JC010099. (Cited page 161.)

Polzin, K., A. Costa, A. R.-A. A. Thurnherr, J. Ledwell, Z. Wang, et S. DiMarco (Submitted), Inference of spatially inhomogeneous and temporally intermittent boundary mixing along the northern deepwater Gulf of Mexico. (Cited pages 9, 10, and 161.)

Qiu, Z., A. Doglioli, Z. Hu, P. Marsaleix, et F. Carlotti (2010), The influence of hydrodynamic processes on zooplankton transport and distributions in the North Western Mediterranean: Estimates from a Lagrangian model., *Ecol. Model.*, *221*, 2816–2827. (Cited page 62.)

Qiu, Z., A. Doglioli, Y. He, et F. Carlotti (2011), Lagrangian model of zooplankton dispersion: numerical schemes comparisons and parameter sensitivity tests., *Chin. J. Oceanol. Limn.*, *29/2*, 438–445. (Cited page 4.)

Rayfield, B., M.-J. Fortin, et A. Fall (2010), Connectivity for conservation: a framework to classify network measures., *Ecology*, *92*, 847–858. (Cited page 26.)

Reffray, G., F. Fraunie, et P. Marsaleix (2004), Secondary flows induced by wind forcing in the Rhône region of freshwater influence., *Ocean Dynamics*, *54*, 179–196. (Cited page 62.)

Reffray, G., R. Bourdalle-Badie, et C. Calone (2015), Modelling turbulent vertical mixing sensitivity using a 1-D version of NEMO., *Geosci. Model Dev.*, *8*, 69–86. (Cited pages 56, 61, 63, 65, 66, 94, and 159.)

Reynolds, O. (1895), On the dynamical theory of incompressible viscous fluids and the determination of the criterion., *Philosophical Transactions of the Royal Society of London*, *186*, 123–164. (Cited page 5.)

Rhines, P. (1988), Mixing and large-scale ocean dynamics., *Elsevier Oceanography Series*, *46*, 263–284. (Cited page 63.)

Rodi, W. (1980), *Turbulence models and their application in hydraulics.*, 104 pp., International Association for Hydraulic Research, Delft, Netherlands. (Cited page 66.)

Rodi, W. (1987), Examples of calculation methods for flow and mixing in stratified fluids., *J. Geophys. Res.*, *92*, 5305–5328. (Cited pages 7 and 61.)

Rossi, V., E. S. G. ans A.A. Lopez Cristobal, et E. Hernandez-Garcia (2014), Hydrodynamic provinces and oceanic connectivity from a transport network help designing marine reserves., *Geoph. Res. Lett.*, *93*, 2883–2891. (Cited pages 17, 18, 19, and 27.)

Rozenfeld, A., S. Arnaud-Haond, E. Hernandez-Garcia, V. Eguiluz, E. Serrao, et C. Duarte (2008), Network analysis identifies weak and strong links in a metapopulation system., *Proc. Natl. Acad. Sci. USA*, *105*, 18,824–18,829. (Cited pages 17, 25, 26, 32, and 37.)

Rubio, A., B. Barnier, G. Jorda, M. Espino, et P. Marsaleix (2009), Origin and dynamics of mesoscale eddies in the Catalan Sea (NW Mediterranean): Insight from a numerical model study., *J. Geophys. Res.*, *114*. (Cited page 62.)

Ruddick, B., A. Anis, et K. Thompson (2000), Maximum likelihood spectral fitting: The Batchelor spectrum., *J. Atmos. Oceanic Technol.*, *17*, 1541–2012. (Cited pages 62, 68, 101, 111, and 116.)

Sale, P., I. Hanski, et J. Kritzer (2006), *The merging of metapopulation theory and marine ecology: establishing the historical context.*, 26 pp., Academic Press, Burlington. (Cited page 2.)

Sammari, C., C. Millot, et L. Prieur (1995), Aspects of the seasonal and mesoscale variabilities of the Northern Current inferred from the PROLIG-2 and PROS-6 experiments., *Deep-Sea Res.*, *6*, 893–917. (Cited page 62.)

Sanchez, X., E. Roget, J. Planella, et F. Forcat (2011), Small-scale spectrum of a scalar field in water: The Batchelor and Kraichnan models., *J. Phys. Oceanogr*, *41*, 2155–2167. (Cited pages 101 and 111.)

Schmitz, W. (1984), Abyssal eddy kinetic energy in the North Atlantic., *J. Mar. Res.*, *42*, doi:509-536. (Cited page 141.)

Scinocca, J. (2001), Low-level topographic drag in atmospheric flows., pp. 111–118. (Cited page 126.)

Scinocca, J., et N. McFarlane (2000), The parameterization of drag induced by stratified flow over anisotropic orography., *Q. J. Roy. Met. Soc.*, *126*, doi:2353-2393. (Cited pages 9 and 126.)

Seidman, S. (1983), Network structure and minimum degree., *Social Networks*, *5*, 269–287. (Cited page 34.)

Sharples, J., et M. Moore (2001), Internal tide dissipation, and vertical nitrate flux at the shelf edge of NE New Zealand., *J. Geophys. Res.*, *106*, 14,069–14,081. (Cited page 62.)

Sharples, J., M. Coates, et J. Sherwood (2003), Quantifying turbulent mixing and oxygen fluxes in a mediterranean-type, microtidal estuary., *Ocean Dynam.*, *53*, 126–136. (Cited page 62.)

Shih, L., J. Koseff, G. Ivey, et J. Ferziger (2005), Parameterization of turbulent fluxes and scales using homogeneous sheared stably stratified turbulence simulations., *J. Fluid Mech.*, *525*, 193–21. (Cited pages 69 and 120.)

Simpson, J., T. Rippeth, et T. Campbell (2000), The phase-lag of turbulent dissipation in tidal flow., *In: Yanagi, T. (Ed.), Interactions between Estuaries, Coastal Seas and Shelf Seas. Terra Scientific Publishing, Tokyo, Japan*, pp. 57–67. (Cited page 6.)

Skyllingstad, E., W. Smyth, J. Moum, et H. Wijesekera (1998), Upper-ocean turbulence during a westerly wind burst: a comparisnon of Large-Eddy Simulation results and microstructure measurements., *J. Phys. Oceanogr.*, *29*, 5–16. (Cited page 5.)

Smyth, W., J. Moum, et D. Caldwell (2001), The efficiency of mixing in turbulent patches: inferences from direct simulations and microstructure observations., *J. Phys. Oceanogr.*, *31*, 1969–1992. (Cited page 5.)

Stacey, M., et S. Pond (1997), On the Mellor-Yamada turbulence closure scheme: The surface boundary condition for q^2., *J. Phys. Oceanogr.*, *27*, 2081–2086. (Cited pages 8 and 61.)

Steinbuck, J., M. Stacey, et S. Monismith (2009), An evaluation of χ_t estimation techniques: Implications for Batchelor fitting and ε., *J. Atmos. Oceanic Technol.*, *26*, 1652–1662. (Cited pages 68 and 102.)

Steinbuck, J. V., A. Genin, S. G. Monismith, J. R. Koseff, R. Holzman, et R. G. Labiosa (2010), Turbulent mixing in fine-scale phytoplankton layers: Observations and inferences of layer dynamics., *Cont. Shelf Res.*, *30*, 442–455. (Cited page 62.)

Steinbuck, J. V., J. Koseff, A. Genin, M. Stacey, et S. Monismith (2011), Horizontal dispersion of ocean tracers in internal wave shear., *J. Geophys. Res.*, *116*. (Cited page 62.)

Stommel, H. (1958), The abyssal circulation. letter to the editors., *Deep-Sea Res.*, *5*, 80–82. (Cited page 117.)

Sverdrup, H. (1933), On vertical circulation in the ocean due to the action of the wind with application to conditions within the Antarctic Circumpolar Current., *Discov. Rep.*, *7*, 139–170. (Cited page 117.)

Taylor, J., et R. Ferrari (2011), Shutdown of turbulent convection as a new criterion for the onset of spring phytoplankton blooms., *Limnol. Oceanogr.*, *56*, 2293–2307. (Cited page 61.)

Tennekens, H., et J. Lumley (1972), *A first course in turbulence*, The MIT Press, Cambridge, MA, USA. (Cited pages 4 and 57.)

Thomas, C., J. Lambrechts, E. Wolansky, V. Traag, V. Blondel, E. Deleersnijder, et E. Hanert (2014), Numerical modelling and graph theory tools to study ecological connectivity in the Great Barrier Reef., *Ecol. Model.*, *272*, 160–174. (Cited pages 3 and 159.)

Thomas, W., et C. Gibson (1990), Effects of small-scale turbulence on microalgae., *J. Appl. Phycology*, *2*, 71–77. (Cited page 61.)

Thorpe, S. (2005), The turbulent ocean., *Cambridge University Press.* (Cited pages xxi, 60, and 122.)

Toole, J., et R. Schmitt (1987), Small-scale structures in the north-west Atlantic subtropical front., *Nature*, *327*, 47–49. (Cited page 60.)

Treml, E., P. Halpin, D. Urban, et L. Pratson (2008), Modeling population connectivity by ocean currents, a graph theoretic approach for marine conservation., *Lansc. Ecol.*, *23*, 19–36. (Cited pages 17, 25, 26, and 37.)

Trowbridge, J., et S. Lentz (1991), Asymmetric behavior of an oceanic boundary layer above a sloping bottom., *J. Phys. Oceanogr.*, *21*, doi:1171-1185. (Cited page 141.)

Ulses, C. (2005), Dynamique océanique et transport de la matière particulaire dans le Golfe de Lion: Crue, tempête et période hivernale., *PhD thesis, Universitè Paul Sebatier Toulouse.* (Cited page 64.)

Ulses, C., P.-A. Auger, K. Soetaert, P. Marsaleix, F. Diaz, L. Coppola, M. Herrmann, F. Kessouri, et C. Estournel (2016), Budget of organic carbon in the North-Western Mediterranean Open Sea over the period 2004-2008 using 3-D coupled physical-biogeochemical modeling., *J. Geophys. Res.*, *121.* (Cited page 64.)

Urban, D., et T. Keitt (2001), Landscape connectivity: a graph theoretic perspective., *Ecology*, *82*, 1205–1218. (Cited pages 3 and 25.)

Venayagamoorthy, S., et J. Koseff (2006), On the mixing efficiency in stably stratified turbulence, pp. 1–7. (Cited page 119.)

Vishveshwara, S., K. Brinda, et K. Kannan (2002), Protein structure: insights from graph theory., *Journ. Theo. Comput. Chem.*, *1*(1), 1–25. (Cited page 158.)

Warner, J., C. Sherwood, H. Arango, et R. P. Signell (2005), Performance of four turbulence closure models implemented using a generic length scale method., *Ocean Model.*, *8*, 81–113. (Cited pages xxvi, 4, 8, 61, 97, 98, and 99.)

Waterhouse, A., J. MacKinnon, J. Nash, M. Alford, E. Kunze, H. Simmons, K. Polzin, L. S. Laurent, O. Sun, R. Pinkel, L. Talley, C. Whalen, T. Huussen, G. Carter, I. Fer,

S. Waterman, A. N. Garabato, T. Sanford, et C. Lee (2014), Global patterns of diapycnal mixing from measurements of the turbulent dissipation rate., *J. Phys. Oceanogr.*, *44*(7), 1854–1872. (Cited page 9.)

Watson, A., J. L. an M.-J. Messias, B. King, N. Mackay, M. Meredith, B. Mills, et A. N. Garabato (2013), Rapid cross-density ocean mixing at mid depths in Drake Passage., *Nature*, *501*, 408–411. (Cited page 118.)

Wilcox, D. (1988), Reassessment of the scale-determining equation for advanced turbulence models., *AIAA Journal*, *26*, 1299–1310. (Cited pages 7 and 61.)

Wright, S. (1931), Evolution in Mendelian populations., *Genetics*, *16*, 1871–1878. (Cited page 25.)

Zhang, A., M. Ramanathan, W. Hwang, et Y. Y. Cho (2010), Bridging centrality: a concept and formula to identify bridging nodes in scale-free networks., uS Patent 7,808,921. (Cited page 48.)

Zhang, X., S. DiMarco, D. S. IV, M. H. nad A.E. Jochens, et R. D. Hetland (1991a), Near-resonant ocean response to sea breeze on a stratified continental shelf., *J. Phys. Oceanogr.*, *39*, doi:2137-2155. (Cited page 125.)

Zhang, X., D. S. IV, S. DiMarco, et R. Hetland (1991b), A numerical study of sea-breeze-driven ocean Poincare wave propagation and mixing near the critical latitude., *J. Phys. Oceanogr.*, *40*, doi:48-66. (Cited page 125.)