Deep learning for ocean satellite altimetry : specificities and practical implications

Quentin Febvre

COLLEGES SCIENCES
BRETAGNE POUR L'INGENIEUR
LOIRE ET LE NUMERIQUE

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

THÈSE DE DOCTORAT DE

l'École Nationale Supérieure Mines-Télécom Atlantique Bretagne Pays de la Loire
IMT Atlantique

ÉCOLE DOCTORALE N° 648
Sciences pour l'Ingénieur et le numérique
Spécialité : *Signal, image et vision*

Par
Quentin FEBVRE

Deep Learning for ocean satellite altimetry : specificities and practical implications

Thèse présentée et soutenue à IMT Atlantique, Brest, le 5 décembre 2023
Unité de recherche : Lab-STICC
Thèse N° : 2023IMTA0374

Rapporteurs avant soutenance :

Jocelyn Chanussot	Professeur Grenoble Institute of Technology
Marc Bocquet	Professeur, École des Ponts ParisTech

Composition du Jury :

Président :	François Rousseau	Professeur, IMT Atlantique
Examinateurs :	Claire Monteleoni	Professeure, INRIA
	Florence Tupin	Professeure, Telecom Paris
	Jocelyn Chanussot	Professeur Grenoble Institute of Technology
	Marc Bocquet	Professeur, École des Ponts ParisTech
Invité :	Gérald Dibarboure	Directeur technique, CNES
Dir. de thèse :	Ronan Fablet	Professeur, IMT Atlantique
Encadrant :	Julien Le Sommer	Directeur de recherche, CNRS
Co-encadrant :	Clément Ubelmann	Docteur, Datlas

Acknowledgement

The work presented in this manuscript was supported by the French National Research Agency (ANR) Melody and OceaniX and CNES, through projects number ANR-17- CE01-0009-01, ANR-19-CE46-0011 and ANR-19-CHIA-0016); by the French National Space Agency (CNES) through the SWOT Science Team program (projects MIDAS and DIEGO) and the OSTST program (project DUACS-HR); by the French National Centre for Scientific Research (CNRS) through the LEFE-MANU program (project IA-OAC). This project also received funding from the European Union's Horizon Europe research and innovation programme under the grant No 101093293 (EDITO-Model Lab project). This project benefited from HPC and GPU computing resources from Ifremer and GENCI-IDRIS (Grant 2021-101030).

Résumé en français

Motivations

Dans le contexte d'un climat en évolution, le suivi des changements de notre environnement est un aspect critique de notre capacité à réagir et à nous adapter. Les océans sont des systèmes physiques régis par des dynamiques connues mais chaotiques, ce qui rend l'utilisation de données d'observation essentielle pour surveiller leur état. Notre capacité de surveillance dépend donc à la fois des systèmes d'observation déployés et de notre capacité à exploiter les données d'observation.

Depuis plusieurs décennies, les altimètres NADIR satellitaires ont considérablement amélioré nos capacités d'observation en fournissant une couverture mondiale de la hauteur de la surface de la mer (SSH). Cependant, en raison du prélèvement rare et irrégulier des constellations d'altimétrie, les produits opérationnels actuels ne fournissent que des informations limitées sur les phénomènes aux petites échelles. La figure 1 montre les échelles approximatives des processus d'intérêt en altimétrie et la limite des capacités d'observation d'une seule constellation d'altimétrie. Ces processus liés à la dynamique méso et sous-mésoéchelle de la surface de l'océan jouent un rôle important dans la redistribution de la chaleur dans l'océan, ce qui a des implications pour la surveillance du climat.

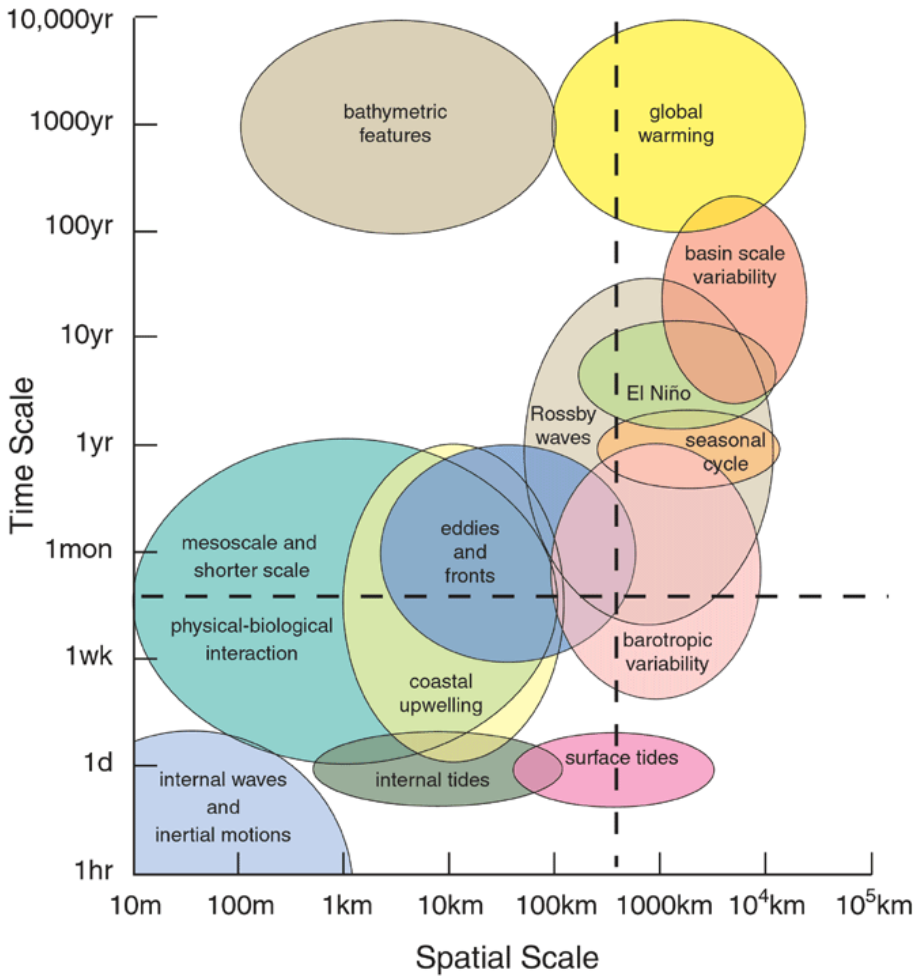


Figure 1 – **Échelles des processus océaniques.** Crédits à Dudley B. Chelton. Illustration de la variété des processus d'intérêt pour l'altimétrie, affichés en fonction de leurs échelles spatiales et temporelles. Les lignes pointillées indiquent les limites d'observation lors de l'utilisation de données altimétriques uniques. (Phénomènes observés dans la section supérieure droite)

Cette thèse s'inscrit dans le contexte de la mission Surface Water Ocean topography (SWOT)[1], qui présente des opportunités pour améliorer nos capacités d'observation des océans. La figure 2 montre un exemple simulé pour illustrer l'impact des observations de la mission SWOT. Le capteur Ka-band Radar Interferometer (KaRIn) fournira les images bidimensionnelles de la topographie de la surface de la mer comme illustré, mais introduira également des défis d'étallonage [2] en raison d'erreurs jamais vues auparavant.

Ce manuscrit se concentre sur le développement de méthodes d'exploitation d'observations de SSH satellitaires afin d'améliorer nos connaissances sur

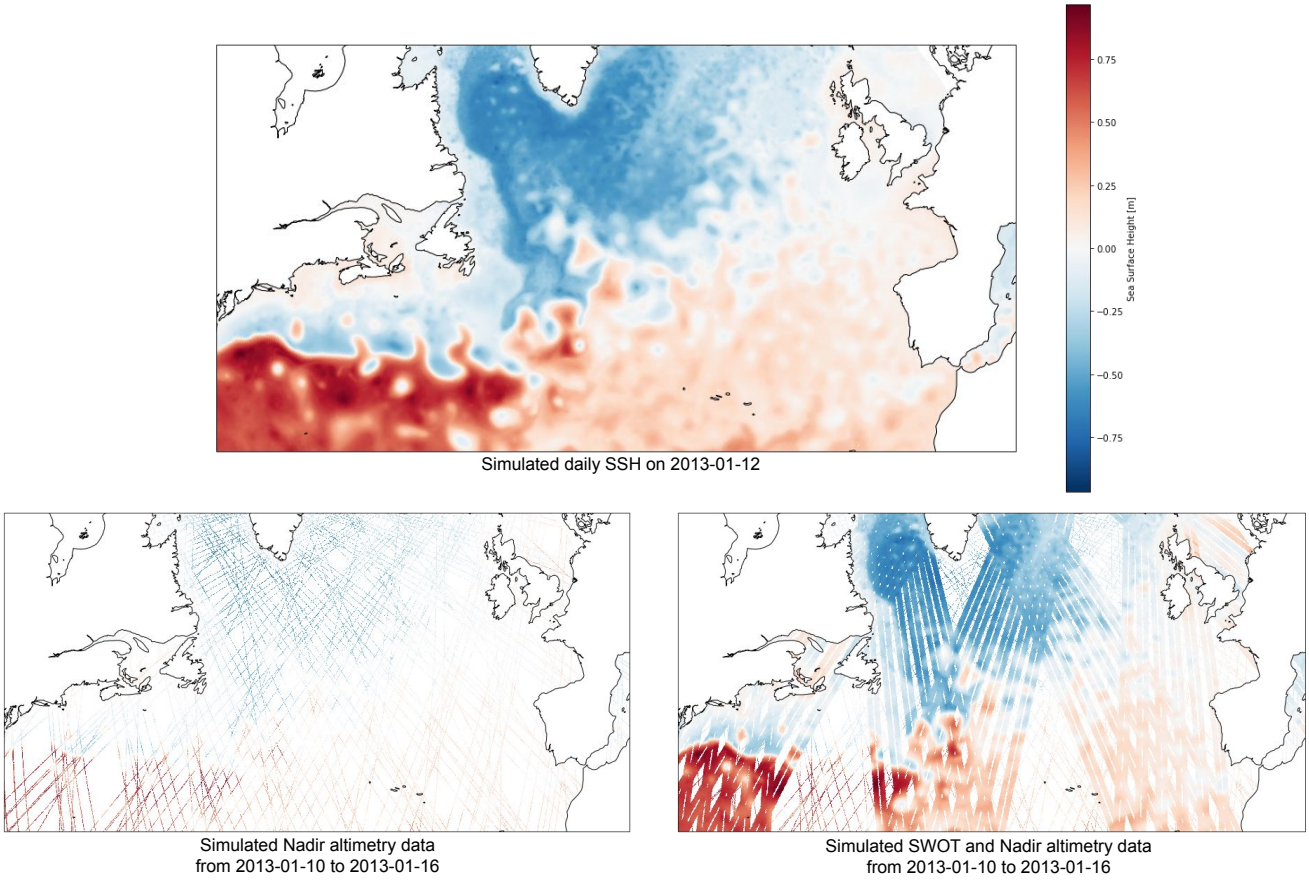


Figure 2 – **Contexte de l'altimetrie.** La figure du haut affiche un exemple de SSH moyen sur l'Atlantique Nord issu de la simulation NATL60. La rangée du bas illustre l'impact observation du satellite SWOT.

la dynamique de la surface de l'océan. Plus précisément, nous cherchons à savoir comment les progrès en apprentissage profond peuvent être bénéfiques à l'analyse de l'altimétrie océanique.

La recherche sur l'apprentissage profond fournit un ensemble d'outils en évolution rapide qui ont été appliqués avec succès à une large gamme de domaines, surpassant les méthodes existantes et réussissant dans des problèmes auparavant non résolus.

Cette thèse vise à étudier le potentiel des outils apportés par le domaine de l'apprentissage profond pour relever les défis de l'observation océanique.

En apprentissage profond, les architectures existantes utilisées pour l'incrustation vidéo[3] et le débruitage d'images[4, 5, 6] résolvent des tâches formellement similaires à la cartographie et à la calibration de l'altimétrie. L'universalité de ces modèles peut potentiellement améliorer les méthodes existantes en

capturant des processus liés à la dynamique océanique et aux observations qui sont difficiles à formuler formellement.

L'application de ces modèles à l'altimétrie introduit cependant des défis liés aux données. Le contexte altimétrique ne permet pas d'accéder à l'état SSH que nous voulons estimer, ce qui pose des questions concernant l'entraînement et l'évaluation du modèle. En effet, au début de ma thèse, le schéma de cartographie neuronale 4dVarNet a surpassé les produits opérationnels pour une utilisation dans la région du Gulf Stream[7]. Cependant, l'architecture neuronale et les résultats ont été entraînés et évalués dans un environnement simulé. Cela pose la question traitée dans ce manuscrit : *Comment les approches d'apprentissage profond peuvent-elles surmonter le manque de référence de SSH pour une utilisation sur des données altimétriques réelles ?* De plus, les architectures de vision par ordinateur (CV) ont été, dans une certaine mesure, validées et adaptées aux images naturelles. La transposition aux données altimétriques et aux champs SSH n'est pas triviale. Le schéma d'échantillonnage clairsemé et irrégulier des altimètres ainsi que la nature turbulente de la SSH peuvent ne pas convenir aux architectures CV classiques.

En outre, un autre facteur qui différencie l'altimétrie des domaines précédemment mentionnés est que des méthodes classiques existent déjà pour résoudre les tâches considérées. Les méthodes existantes utilisent les connaissances physiques disponibles sur la dynamique océanique et les systèmes d'observation. Dans une certaine mesure, cela dépeint un environnement moins favorable pour le potentiel des méthodes DL classiques et soulève un autre point abordé dans cette recherche qui est : *Comment intégrer des connaissances altimétriques spécifiques dans la méthodologie d'apprentissage profond ?*

Ces études démontrent que l'analyse altimétrique dépend fortement de l'expertise du domaine pour les contextes de données et d'évaluation. Le développement et l'évaluation de méthodes telles que 4dVarNet ont reposé sur des efforts antérieurs de standardisation de cadre expérimentaux avec données simulées et observationnelle sous la forme de data-challenge [8, 9]. Ces cas d'utilisation ont fourni un accès aux données pour les praticiens de l'apprentissage automatique ainsi que des mesures pertinentes pour les physiciens océaniques. Cela soulève le troisième et dernier point de ce manuscrit : *Que faut-il pour faciliter l'adoption de l'apprentissage profond dans les sciences de l'observation océanique ?*

Contributions

La première contribution met en lumière l'application réussie de l'apprentissage profond pour la correction des biais de données d'observation simulées par SWOT. Alors que les architectures d'apprentissage profond standard peinaient à différencier les signatures SSH fines des biais à grande amplitude, nous avons montré que les méthodes d'apprentissage profond pouvaient être adaptées aux caractéristiques uniques des données d'altimétrie. Nous avons utilisé les spécifications d'erreurs de la mission SWOT pour créer une architecture personnalisée axée sur la calibration des erreurs corrélées de SWOT. Cette étude est prometteuse, mais la méthode développée a été calibrée et évaluée à l'aide de données simulées, soulevant des questions sur son applicabilité aux observations réelles de SWOT.

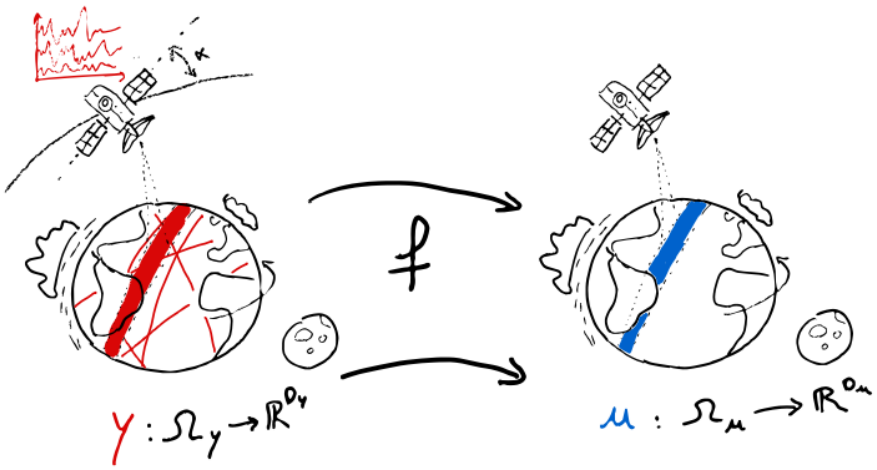


Figure 3 – **Étalonnage SWOT.** La partie gauche illustre le domaine observé en rouge tandis que la partie droite indique le domaine sur lequel nous visons à estimer la SSH.

La deuxième étude s'intéresse à la façon dont les méthodes d'altimétrie basées sur l'apprentissage, une fois calibrées sur des données simulées, peuvent être appliquées à des données réelles. Nous avons évalué les schémas de cartographie 4dVarNet sur des données d'altimétrie réelles après calibration sur des données simulées. Les résultats indiquent des capacités de généralisation élevées même avec des simulations grossières, tandis que des simulations plus précises améliorent les performances de cartographie. Les résultats introduisent des avenues intéressantes dans l'exploration de l'utilisation de la simulation

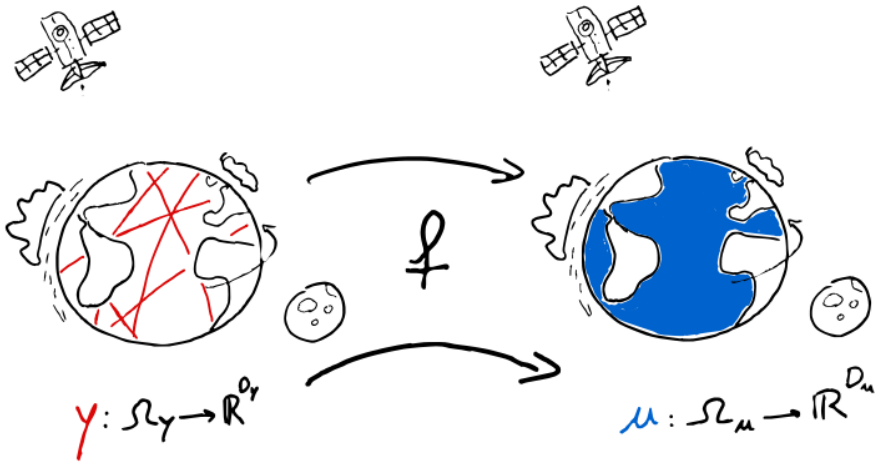


Figure 4 – **Cartographie d'observation altimetrique Nadir.** La partie gauche illustre le domaine observé en rouge tandis que la partie droite indique le domaine sur lequel nous visons à estimer la SSH.

numérique pour entraîner des modèles pour des applications du monde réel.

Les deux premières études mettent en lumière le potentiel de l'application d'approches basées sur l'apprentissage aux défis d'observation en oceanologie. Cependant, elles soulignent également les complexités de l'union de l'expertise en observation, en données de simulation, en techniques d'apprentissage profond et en méthodologies d'évaluation spécifiques au domaine. Cela a stimulé la création de la trousse d'outils spécialisée, Oceanbench, visant à combler le fossé entre les experts en apprentissage profond et les experts en oceanologie. Oceanbench permet aux océanographes de concevoir de manière flexible des configurations d'évaluation à l'aide de données et de métriques. Ces configurations sont accompagnées des outils essentiels pour les praticiens de l'apprentissage profond pour accéder et préparer les données en vue de la formation de leurs modèles. La première iteration présentée dans cet ouvrage porte sur l'interpolation de la hauteur de la surface de la mer, mais elle a été pensée pour être extensible à d'autres défis d'observation de l'océan.

Objectifs et contenu

Une fois le contexte lié à l'altimétrie et l'apprentissage profond introduit dans le **Chapitre 1**, le **Chapitre 2** décrit les principales hypothèses et méth-

odes formulées dans les méthodes actuelles de cartographie et d'étallonage altimétriques. Ce chapitre vise à décrire les modèles et algorithmes d'étalonnage existants disponibles dans l'analyse altimétrique ainsi que les travaux connexes en apprentissage profond. Une description plus détaillée du cadre réseau neuronal basé sur 4DVarNet et ses applications sera présentée car les contributions de cette thèse font un usage intensif de cette architecture.

Le **Chapitre 3** propose une architecture d'apprentissage profond pour l'étalonnage des erreurs corrélées dans les données SWOT. D'un point de vue applicatif, la flexibilité de la méthodologie de l'apprentissage profond ouvre la possibilité de capturer des signaux difficiles à paramétrer explicitement. D'un point de vue méthodologique, cette étude montre comment les architectures d'apprentissage profond peuvent être adaptées aux hypothèses sur l'instrument et ses mesures. Plus précisément, cela se fait en montrant comment les spécificités spectrales des erreurs peuvent être exploitées pour concevoir un schéma de calibration neuronal efficace. Cette étude basée sur des donnée simulée ne traite pas directement des défis posés par le manque de jeu de données de référence, qui sont au centre du chapitre suivant.

Le **Chapitre 4** aborde plus spécifiquement le problème de la disponibilité des données. Il étudie comment les schémas de cartographie neuronaux peuvent être appliqués aux données réelles malgré l'absence de jeu de données de référence. Il s'intéresse plus particulièrement au cadre 4DVarNet qui a été démontré dans un environnement simulé en utilisant des données SSH simulées pour l'entraînement et l'évaluation. Ce chapitre examine les performances sur des données réelles de modèles d'apprentissage profond entraînés sur des données simulées. Il montre comment l'importante connaissance physique de la dynamique océanique peut être exploitée pour pallier le manque de jeu de données de référence en altimétrie par l'utilisation de simulations numériques pour l'entraînement.

Le **Chapitre 5** examine les obstacles à une meilleure synergie entre les communautés de l'altimétrie océanique et de l'apprentissage profond. Les deux domaines sont bien établis avec des connaissances accumulées et des meilleures pratiques. Comme décrit dans ce chapitre, l'apprentissage profond apporte des modèles et des algorithmes puissants. Cependant, les données de calibration et d'évaluation ainsi que les métriques ne peuvent être rationnellement conçues que par un expert du domaine. Nous proposons OceanBench, une interface sous la forme d'un ensemble d'outils logiciels. Oceanbench vise à permettre aux experts du domaine de concevoir facilement des problèmes altimétriques

d'intérêt et à les qualifier avec des métriques pertinentes. Il fournit ensuite aux praticiens de l'apprentissage automatique l'accès aux données nécessaires ainsi qu'à des utilitaires adaptés à la formation et à l'évaluation des méthodes d'apprentissage.

Le **Chapitre 6** discute et conclut sur la recherche présentée dans ce manuscrit. Nous résumons les principaux objectifs et résultats des chapitres précédents ainsi que nous proposons quelques pistes de recherche futures.

BIBLIOGRAPHY

[1] "KaRIn on SWOT: Characteristics of Near-Nadir Ka-Band Interferometric SAR Imagery," https://ieeexplore.ieee.org/document/6553583/.

[2] "Empirical Cross-Calibration of Coherent SWOT Errors Using External References and the Altimetry Constellation," https://ieeexplore.ieee.org/document/6087373/.

[3] D. Kim, S. Woo, J.-Y. Lee, and I. S. Kweon, "Deep Video Inpainting."

[4] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin, "Deep Learning on Image Denoising: An overview," Aug. 2020.

[5] V. Gaya, E. Dalsasso, L. Denis, F. Tupin, B. Pinel-Puysségur, and C. Guérin, "Débruitage multi-modal d'images radar à synthèse d'ouverture par apprentissage profond auto-supervisé."

[6] L. Einig, J. Pety, A. Roueff, P. Vandame, J. Chanussot, M. Gerin, J. H. Orkisz, P. Palud, M. G. Santa-Maria, V. d. S. Magalhaes, I. Bešlić, S. Bardeau, E. Bron, P. Chainais, J. R. Goicoechea, P. Gratier, V. V. Guzmán, A. Hughes, J. Kainulainen, D. Languignon, R. Lallement, F. Levrier, D. C. Lis, H. S. Liszt, J. L. Bourlot, F. L. Petit, K. Öberg, N. Peretto, E. Roueff, A. Sievers, P.-A. Thouvenin, and P. Tremblin, "Deep learning denoising by dimension reduction: Application to the ORION-B line cubes," *Astronomy & Astrophysics*, vol. 677, p. A158, Sep. 2023.

[7] R. Fablet, M. M. Amar, Q. Febvre, M. Beauchamp, and B. Chapron, "END-TO-END PHYSICS-INFORMED REPRESENTATION LEARNING FOR SATELLITE OCEAN REMOTE SENSING DATA: APPLICATIONS TO SATELLITE ALTIMETRY AND SEA SURFACE CURRENTS," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-3-2021, pp. 295–302, Jun. 2021.

[8] M. Ballarotta, E. Cosme, and A. Albert, "Ocean-data-challenges/2020a_SSH_mapping_NATL60: Material for SSH mapping data challenge," Zenodo, Sep. 2020.

[9] M. Ballarotta and F. L. Guillou, "Ocean-data-challenges/2021a_SSH_mapping_OSE: Material for SSH mapping OSE data challenge," Zenodo, Sep. 2021.

Table of Contents

List of acronyms **17**

List of figures **23**

List of tables **26**

1 Introduction **27**
1.1 Broad Context 27
1.2 Key concepts: Graduating a thermometer 30
1.3 The case of ocean altimetry 36
1.4 Deep learning: opportunities and challenges 38
1.5 Thesis objectives and outline 39
Bibliography 40

2 Modeling and solving altimetry problems **43**
2.1 Priors on the Sea Surface Height (SSH) 44
2.2 Solvers: Estimating the state given some observations 46
2.3 Calibrating the method 48
2.4 Evaluation 49
2.5 A closer look on the 4dVarNet 51
Bibliography 53

3 Scale-aware neural calibration for wide swath altimetry observations **57**
3.1 Introduction 57
3.2 Background 60
3.3 Data and Case-study 62
3.4 Proposed Methodology 63
3.5 Experimental results 67
3.6 Conclusion 72
Bibliography 73

4 Training neural mapping schemes for satellite altimetry with simulation data **77**
4.1 Introduction . . . 77
4.2 Background . . . 78
4.3 Method . . . 81
4.4 Results . . . 84
4.5 Discussion . . . 90

5 OceanBench: The Sea Surface Height Edition **101**
5.1 Introduction . . . 101
5.2 Related Work . . . 102
5.3 OceanBench . . . 104
5.4 *Sea Surface Height Edition* . . . 107
5.5 Discussion . . . 121

6 Conclusions and perspectives **137**
6.1 Contributions Summary . . . 137
6.2 Current Limitations and perspectives . . . 138
Bibliography . . . 139

List of publications **140**

Appendix **145**
6.3 Use Case I: Hydra Recipes . . . 145
6.4 Use Case II: XRPatcher . . . 148

List of acronyms

4DVAR	Four Dimensional Variational Data Assimilation
3DVAR	Three Dimensional Variational Data Assimilation
BFN	Back-and-Forth Nudging
CV	Computer Vision
DL	Deep Learning
DUACS	Data Unification and Altimeter Combination System
DYMOST	Dynamic Interpolation Ocean Science Topography
KaRIn	Ka-band Radar Interferometer
KF	Kalman Filter
L4	Level 4
MDT	Mean Dynamic Topography
MIOST	Multiscale Interpolation Ocean Science Topography
ML	Machine Learning
NEMO	Nucleus for European Modelling of the Ocean
NN	Neural Network
NLP	Natural Language Processing
nRMSE	normalized Root Mean Squared Error
OSE	Observing System Experiment
OSSE	Observing System Simulation Experiment
PSD	Power Spectrum Density
QG	Quasi-Geostrophic
RMSE	Root Mean Squared Error
SLA	Sea Level Anomaly
SSH	Sea Surface Height
SST	Sea Surface Temperature
SWOT	Surface Water Ocean Topography

List of Figures

1 **Échelles des processus océaniques.** Crédits à Dudley B. Chelton. Illustration de la variété des processus d'intérêt pour l'altimétrie, affichés en fonction de leurs échelles spatiales et temporelles. Les lignes pointillées indiquent les limites d'observation lors de l'utilisation de données altimétriques uniques. (Phénomènes observés dans la section supérieure droite) 6

2 **Contexte de l'altimetrie.** La figure du haut affiche un exemple de SSH moyen sur l'Atlantique Nord issu de la simulation NATL60. La rangée du bas illustre l'impact observation du satellite SWOT. 7

3 **Étalonnage SWOT.** La partie gauche illustre le domaine observé en rouge tandis que la partie droite indique le domaine sur lequel nous visons à estimer la SSH. 9

4 **Cartographie d'observation altimetrique Nadir.** La partie gauche illustre le domaine observé en rouge tandis que la partie droite indique le domaine sur lequel nous visons à estimer la SSH. 10

1.1 **Scales of Ocean Processes.** Credits to Dudley B. Chelton. Illustration of the variety of processes of interest for altimetry displayed in function of their spatial and temporal scales. The dashed lines indicate the observational limitations when using a single altimeter data. (Observed phenomena in the top-right section) 28

1.2 **Altimetry context.** The top figure display an example of average SSH over the north Atlantic from the NATL60 simulation. The bottom row illustrate the observational impact of the SWOT satellite. 29

1.3 **Thermometer Graduation problem illustration.** Given a simple liquid based thermometer, we aim at finding the matching between height of the liquid within the glass tube and temperature. 30

1.4 **Mapping thermometer level to temperature.** The first step consists in compiling theoretical knowledge to determine a model of the level to temperature relationship. This model define the set of candidate graduations. The second step consists in leveraging data to chose the best candidate graduation through some calibration algorithm. 31

1.5 **Evaluation and errors.** Given some evaluation data and choice of metric, we can compute the errors associated with our graduation. T°_{calib} and T°_{ref} are respectively the temperatures given by our graduation and a reference well graduated thermometer 33

1.6 **Different sources of errors for the thermometer graduation.** From left to right: model errors result from erroneous assumptions about the system. Data errors result from inaccuracies in the calibration data and algorithmic errors result from a failure of the calibration algorithm to select the best candidate from the model. 33

1.7 **SWOT calibration.** The left part illustrate the observed domain in red while the right part indicates the domain on which we aim at estimating the SSH. 37

1.8 **Nadir Altimetry mapping.** The left part illustrate the observed domain in red while the right part indicates the domain on which we aim at estimating the SSH. 38

2.1 **Methodological formulation**. This figure illustrate the organization of the different components described in this chapter . 44

2.2 **Solving altimetry problems**. This figure displays the different methodological components at play when addressing an altimetry challenge. The different approaches for each component are detailed. We highlight which ones are physics or deep learning inspired. 49

2.3 **4dVarNet Method**. 4dvarnet components contextualized within the methodological framework 52

3.1 **Observing System Simulation Experiment Cross-Calibration data:** *Top left:* Sea surface height (SSH) on October 26th 2012 from NATL60 simulation dataset. *Top right:* Calibrated NADIR pseudo-observations sampled using realistic orbits from the SSH, they are used to compute the gridded product for the cross-calibration.*Bottom-left:* NADIR + noise-free-KaRIn pseudo-observations, the 2D sampled SSH is the target of the cross-calibration.*Bottom-right:* NADIR + noisy-KaRIn pseudo-observations, simulated errors added to the swath SSH constitute the uncalibrated input of the cross-calibration problem 58

3.2 **1000km segment of KaRIn observation components in swath geometry:***(a)* Looking at the three signals we see that the large scale instrument errors (middle) are predominant compared to the SSH (top) and geophysical error (bottom). *(b)* Looking at the along-track scales between 10km and 200km, we note that the SSH is dominant w.r.t the error signals. 60

3.3 **Overview of the proposed architecture:** From left to right: The first step interpolates the nadir-based gridded product onto the swath segment. Afterwards, both the nadir-based gridded product and KaRIn observation undergo the scale-space decomposition scheme outlined in 3.4. The scale components are stacked as channels and processed through the neural network. The blue color of the "Split Conv" indicates that each side of the swath is processed independently by the convolution layer whereas the orange coloring of the "Swath Mix" layer tells that the whole data is processed jointly (more details in 3.4). The final convolution computes a correction to be added to the gridded product for computing the calibrated KaRIn data 64

3.4 **Explained variance of scale components before and after re-scaling:** Each bar indicates how much each scale component of the uncalibrated KaRIn contributes to the total variance of the signal, we can see that before re-scaling (blue) there is four orders of magnitude between largest scale and the others. The learnt re-scaling allows for scale component to be spread within a single order of magnitude (orange), which is more suited to the downstream neural architecture. 66

3.5 **Observation and reconstruction error for the SSH at different spatial scales:** The figure shows the relative error w.r.t to the SSH at different along-track scales for the inputs (Uncalibrated KaRIn in orange and nadir based interpolation in blue) and output (calibrated KaRIn in green) of our method. The x axis indicates the standard deviation of the Gaussian blur that was used to remove the high scale components of the different signals. We can see the expected trend of the interpolation error that is concentrated at fine scales. The uncalibrated KaRIn error on the other hand is lower than the interpolation only in the 10km-100km range. We see the calibrated output of our method achieves lower error across all scales. 69

3.6 **Impact of the nadir-based gridded product on the CalCNN output:** The figure shows the RMSE and the RMSE of the $||\nabla_{ssh}||$ of the calibrated observation (stars) and their associated nadir-based gridded products (squares). The improvement brought by the CalCNN is illustrated by the arrows. This improvement can be interpreted as the relevant information extracted from the uncalibrated KaRIn by the CalCNN. Note that the biggest relative improvement concerns the DUACS gridded product (blue) which doesn't uses the SWOT's nadir altimeter. 71

4.1 **Overview of the experimental setup**. On the left side we display the simulation-based training strategy based on an ocean simulation which will be used for 1) generating synthetic observation and 2) computing the training objective of the neural mapping scheme. On the right side we show the evaluation principle of splitting the available satellite observations to evaluate the method on data that were not used for the inference. 81

4.2 **Kinetic energy and relative vorticity on January 6^{th} of training and reconstruction data**. The first two columns (a) and (b) show the training data while columns (c) and (d) show the associated 4DVarNet reconstruction of the 2017 year. Columns ((a) and (c)) display the geostrophic kinetic energy while ((b) and (d)) display the relative vorticity normalized by the Coriolis parameter. Each row corresponds to the dataset: ORCA025 (I), GLORYS12-f (II), GLORYS12-r (III), NATL60 (IV), eNATL60-t (V) and eNATL60-0 (VI) 86

4.3 **Space-time spectral densities of the training datasets (first row) and of their associated reconstruction (second row)**. Darker blue in the lower left corner indicates higher energy at larger wavelength and periods. The different SSH fields exhibit different energy cascades when moving to finer temporal (upward) or spatial (rightward) scales. 87

4.4 **Spectral analysis of the training and reconstructed SSH datasets**. We display the PSD of the training dataset (left plot), reconstructed SSH field (center plot) as well as the associated PSD score (right plot) 88

4.5 **Spectral impact of model reanalysis**. We display the PSD of the training dataset (left plot), reconstructed SSH field (center plot) as well as the associated PSD score (right plot) 89

5.1 **Evaluation of the SSH field reconstructions for the OSSE NADIR experiment**.. Subfigure (a) showcases the normalized root mean squared error (nRMSE), (b) showcases the isotropic power spectrum decomposition (PSD), (c) showcases isotropic PSD scores. The bottom row showcases the space-time PSD for the NEMO simulation (subfigure (d)) and the PSD scores for three reconstruction models: (e) the MIOST model, (f) the BFN-QG model, and (g) the 4DVarNet model. 115

5.2 27^{th} **October, 2012 from the NEMO simulation for the OSSE experiment** outlined in section 5.4. The top row showcases the aggregated NADIR altimetry tracks and the aggregated SWOT altimetry tracks (12 hours before and 12 hours after) as well as the SST from the NEMO simulation. Each subsequent row showcases the following physical variables found in appendix 5.4: (a) Sea Level Anomaly, (b) Kinetic Energy, (c) Relative Vorticity, and (d) Strain. Each column in the subsequent rows showcase the following reconstructed field from the NEMO simulation found in columrn (a): (b) MIOST, (c) BFN-QG, and (d) 4DVarNet. 116

5.3 **Reconstructed quantities by the 4dVarNet method for each of the four tasks**. Each row showcases the following physical variables found in section 5.4: (a) Sea Surface Height, (b) Kinetic Energy, (c) Relative Vorticity, and (d) Strain. Each column showcase the reconstructed from the tasks (a) OSSE using only Nadir tracks: (b) OSSE using Nadir tracks and SWOT swath, (c) Multimodal using Nadir tracks and sea surface temperature, and (d) Reconstruction using real nadir altimetry tracks. 117

5.4 **Power spectrum and associated scores of the 4dVarNet method for the four experiments**. The row display in order: (1) the isotropic PSD, (2) the spatial PSD score (using the isotropic PSD for the first three rows and along track PSD for the last row), (3) the space-time PSD, (4) The spacetime PSD score available only in OSSE task. 119

List of Tables

2.1 **Comparison of SSH models across various methods**. The columns indicate the type of state representation x, the method of obtaining estimated SSH $\hat{u}$, and the prior distribution $p(x)$. Methods annotated with $^{(*)}$ consider the state of a single time step at a time. 46

2.2 **Comparison of state estimation strategies for existing altimetry methods**. Methods annotated with $^{(*)}$ use a sequential resolution of successive time steps. 48

3.1 **Residual error of the benchmarked calibration frameworks** 67

3.2 **Ablation results** 68

3.3 **Impact of network size** 70

3.4 **Calibration metrics in function of the scale decomposition** 72

4.1 **Summary table of the different synthetic SSH fields used for training**. The last column indicate whether the Dynamic Atmospheric Correction was applied on the synthetic SSH. It justify the presence of both eNATL60-0 and NATL60 to isolate the impacts of resolution and tide. 82

4.2 **SSH reconstruction performance of the benchmarked methods** (a) 4DVarNet from this study trained on eNATL60-0 (b) Archambault et al. (2023), (c and d) ConvLstm-SST and ConvLstm from Martin et al. (2023), (e) DYMOST from Ballarotta et al. (2020), (f) MIOST from Ubelmann et al. (2021), (g) BFN-QG from Guillou et al. (2021), (h) DUACS from Taburet et al. (2019), (i) GLORYS12 from Lellouche et al. (2021. The columns indicate from left to right: whether athe mapping schemes rely only on SSH data or also exploit additional data such as gap free SST products; if the method uses deep learning architectures; the data used to calibrate (or train) the mapping scheme; the numerical model of the ocean used for the mapping if any (QG stands for quasi-geostrophic); μ and λ_x are the metrics as described in Section 4.3 85

4.3 **Performance of 4DVarNet mapping schemes trained on different simulated datasets**. The first column shows the source of the training dataset as described in Table 4.1; the subsequent columns indicate the reconstruction metrics described in Section 4.3. Note that the NATL60 could not be evaluated on the OSSE setup since the evaluation data were used for validation during the training stage. 87

5.1 This table gives an extended overview of the datasets provided to complete the data challenges listed in 5.4. The OSSE SST and SSH are outputs from come from the free run NEMO model. The OSSE NADIR and SWOT are pseudo-observations generated from the NEMO simulation. We provide the original simulated satellite tracks as well as a gridded version at the same resolution as the simulation. 109

5.2 This table showcases all of the summary statistics for some methods for each of the data challenges listed in section 5.4. The summary statistics shown are the normalized RMSE and the effective resolution in the spectral domain. The spectral metrics for the effective resolution that were outlined in section 5.4 are: i) λ_a is the spatial score for the alongtrack PSD score, ii) λ_r is the spatial score for the isotropic PSD, iii) λ_x is the spatial score for space-time PSD score, and iv) λ_t is the temporal score for the space-time PSD score. 120

Chapter 1

Introduction

1.1 Broad Context

In the context of an evolving climate, monitoring the changes in our environment is a critical aspect of our ability to react and adapt. The oceans are physical systems ruled by known but chaotic dynamics which makes the use of observational data essential to monitor their state. Our monitoring ability thus depends on the observing systems deployed as well as our ability to exploit the observation data.

In recent decades, satellite NADIR altimeters have greatly improved our observational capabilities by providing a global coverage of the Sea Surface height (SSH). However due to the scarce and irregular sampling of altimeter constellations, current operational products provide limited insights into fine-scale phenomena[1]. Figure 1.1 shows the approximate scales of the processes of interest in altimetry and the limit of observational capabilities of a single altimeter constellation. This thesis is situated within the context of the Surface Water Ocean topography (SWOT)[2] mission, which presents opportunities for enhancing our observational capabilities of the oceans. Figure 1.2 shows an simulated example to illustrate the observational impact of the SWOT mission. The Ka-band Radar Interferometer (KaRIn) sensor will provide the depicted two dimensional images of the ocean surface topography but will also introduce calibration challenges[3] due to previously unseen errors.

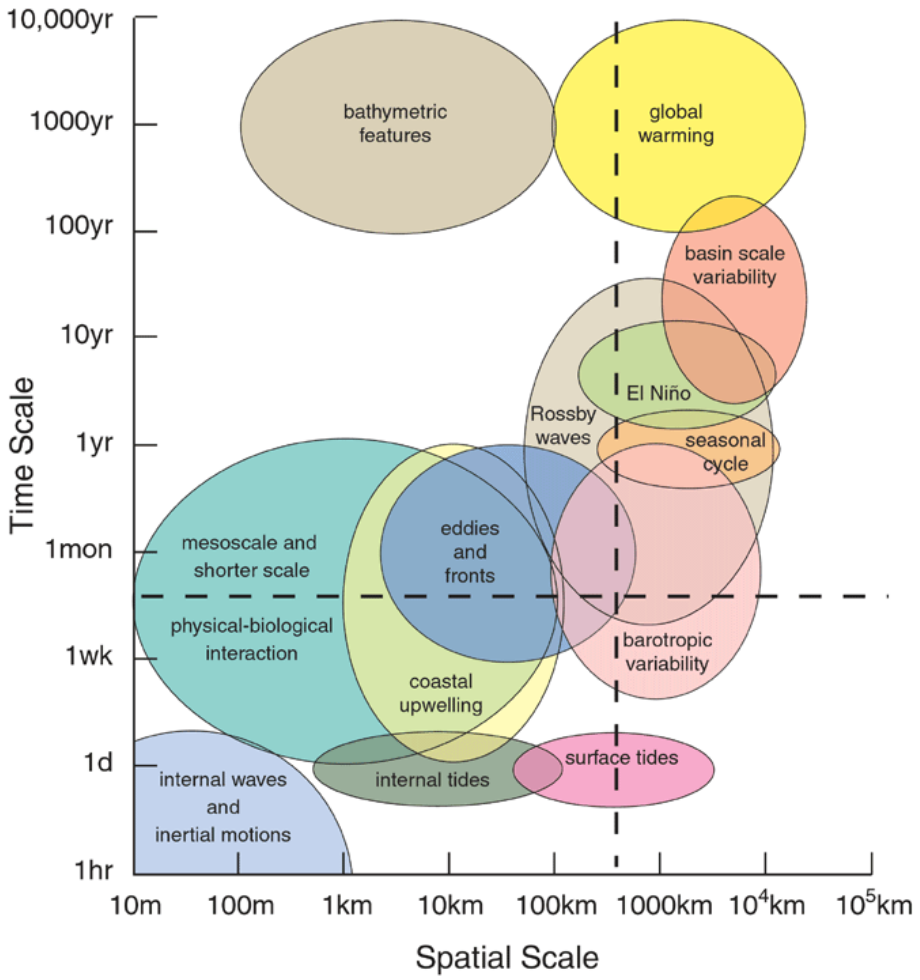


Figure 1.1 – **Scales of Ocean Processes.** Credits to Dudley B. Chelton. Illustration of the variety of processes of interest for altimetry displayed in function of their spatial and temporal scales. The dashed lines indicate the observational limitations when using a single altimeter data. (Observed phenomena in the top-right section)

The research in this thesis focuses on the development of methods to exploit satellite SSH observations for improving our knowledge of ocean surface dynamics. More specifically we're asking how advances in deep learning can be beneficial to ocean altimetry analysis. Deep learning research provides a rapidly evolving set of tools that have been successfully applied to a wide range of domain, surpassing existing methods and succeeding in previously unsolved problems.

In order to study the potentials of deep learning for tackling ocean observation problems, we'll first introduce the necessary methodological components

involved when addressing an observation problem by walking through the illustrative example of a thermometer graduation procedure in section 1.2. We will explicit the similarities between this example and the altimetry challenges studied in this thesis.1.3 This simplified problem will help illustrate and contextualize the complementary roles of data and domain knowledge when addressing this class of problem. In section 1.4, we will then describe how the tools brought by the deep learning field fit in this methodological framework and consider the opportunities and challenges that arise when applying them to ocean altimetry analysis.

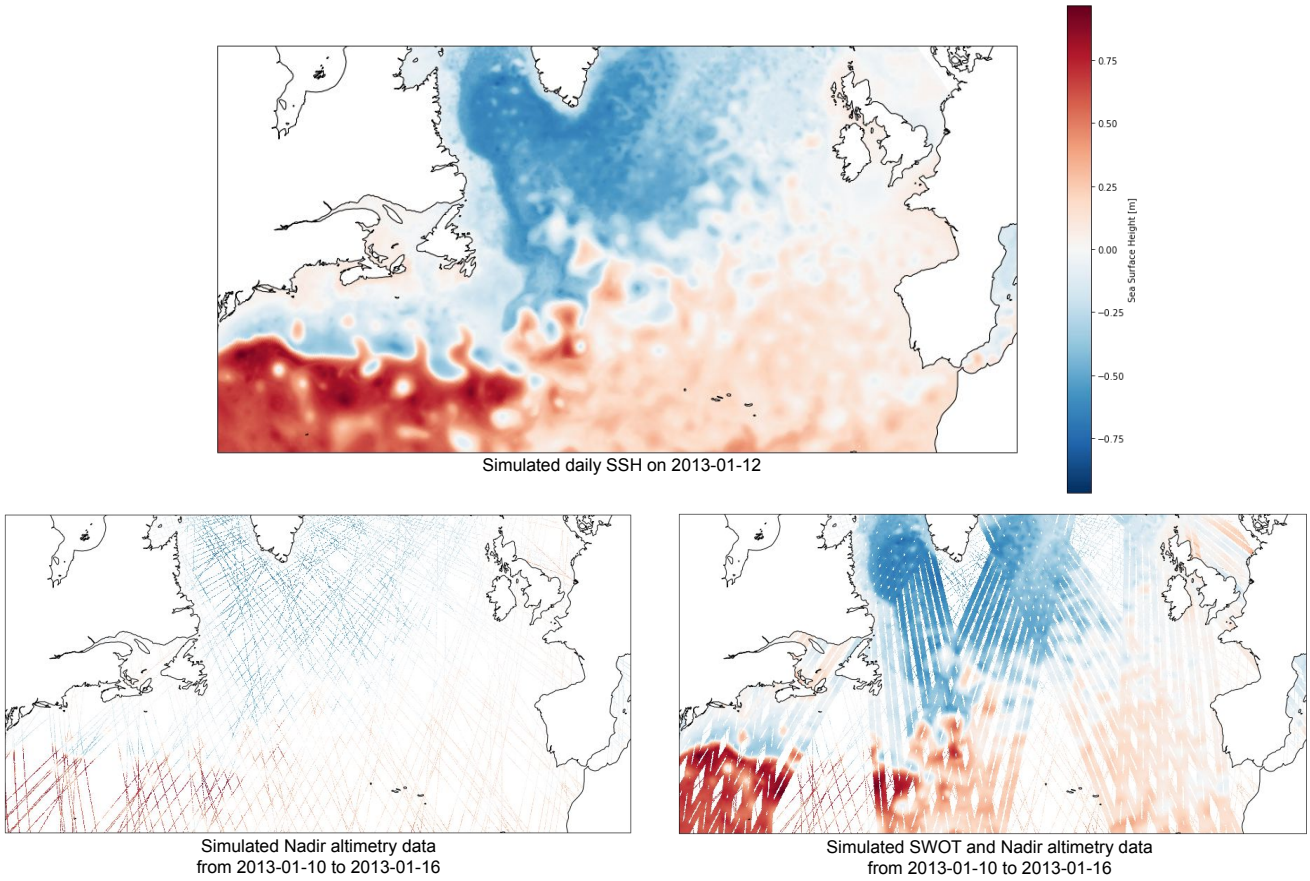


Figure 1.2 – **Altimetry context.** The top figure display an example of average SSH over the north Atlantic from the NATL60 simulation. The bottom row illustrate the observational impact of the SWOT satellite.

Finally we'll outline the structure of this manuscript. We'll present the altimetry use-cases and the deep learning methodological aspect considered in the third and fourth chapter. Finally, we'll introduce the scientific contribution of the OceanBench project which aims at facilitating collaboration between

the ocean altimetry and deep learning communities.

1.2 Key concepts: Graduating a thermometer

Estimating temperature from observations

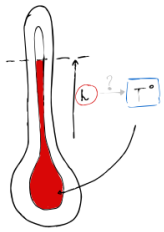


Figure 1.3 – **Thermometer Graduation problem illustration.** Given a simple liquid based thermometer, we aim at finding the matching between height of the liquid within the glass tube and temperature.

Let us consider a standard liquid based thermometer that consists of a liquid in a glass tube. When interested in knowing the temperature, we observe the level of a thermometer. In order to do so, someone had to graduate the thermometer. This seemingly simple action can be detailed in a two-step process, which involves the construction of a theoretical model and its calibration using real-world data.

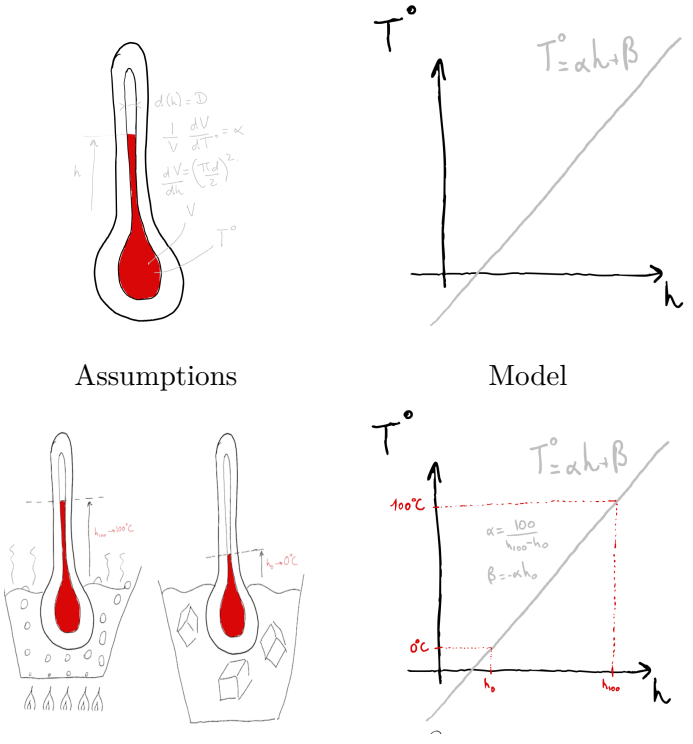
The first step involves compiling theories and assumptions to construct a model linking the observed level and the actual temperature. In this instance, based on our knowledge of fluid dilation in response to temperature, assuming the diameter of the tube is constant with height, we can propose the model that the level is linearly related with the temperature. This model introduces two parameters: the slope and offset of our linear model that need to be ascertained.

The second step involves determining these parameters. This step requires some calibration data as inputs. They are traditionally obtained by immersing the thermometer in icing and boiling water to acquire the levels corresponding to 0°C and 100°C. Using those data points, a linear system can then be used to solve for the parameters. Which finally gives use our level-to-temperature relationship.

The model we chose can have more or less independent parameters that need to be calibrated depending on the assumptions that were made. Interestingly, this introduce a relationship between the assumptions made and the amount

of data required for calibration. For instance, a model with fewer assumptions demands more data. If we were to abandon the assumption of the thermometer tube's constant diameter, we would need to incorporate a parameterization of the tube diameter in our model. This addition creates more parameters and consequently demands additional data for calibration. Conversely, having access to more data can allow us to work with fewer assumptions. Suppose we possess a well-calibrated thermometer that can provide unlimited data points. In that case, we could reduce our assumptions to a minimum and rely heavily on empirical evidence, marking each thermometer graduation using data directly from our well-calibrated thermometer.

With these carefully calibrated graduations now etched onto our thermometer, we can use the liquid level as a convenient stand-in for the temperature. However, an important question remains: How can we assess the accuracy of our newly graduated thermometer?



Assumptions Model

Data Calibration

Figure 1.4 – **Mapping thermometer level to temperature.** The first step consists in compiling theoretical knowledge to determine a model of the level to temperature relationship. This model define the set of candidate graduations. The second step consists in leveraging data to chose the best candidate graduation through some calibration algorithm.

Evaluation

Without evaluation the use of our calibrated instrument would solely rely in the faith given to our mapping above. However one may prefer quantifying the thermometers quality through metrics. In our case the most intuitive metric for characterizing our thermometer's quality is be the precision of the graduations. Each tick of our thermometer has a precision value, different aggregations of the individual values also constitute different metrics (bias versus variance for example). In order to properly evaluate our calibrated instrument, we need to test it in conditions corresponding to its intended use (testing it domestic thermometer 5 kilometers underwater would not give a relevant evaluation).

To do so, let's explicit some assumptions made on what we expect from our thermometer. For example that it needs to "be accurate to the half of degree", "have response time under 10 minutes", "work between -30°C and 200°C" "work at a reasonable atmospheric pressure" etc...

Then we need data to measure the precision of our thermometer in a way that is representative of how we want our thermometer to behave. Using a trustworthy reference like a third-party well-graduated thermometer, we could compare the measurements of the reference with the one given by our solution. An example evaluation procedure could be to confront the measurements of the two instruments at different temperatures such as: in a freezer, in a fridge, at ambient room temperature and in an oven.

Using the procedure above, we can compute our precision metrics and assess if the quality of our thermometer is acceptable. This exercise, raises some critical points about evaluation. The process relies on two components that require a deep understanding of the thermometer's intended use: a suitable choice of metrics and representative data. If the metrics do not align with the intended use of the thermometer, the evaluation will be flawed. Similarly, if the data are not representative of the thermometer's intended use, the evaluation will also be flawed. Furthermore, the reliability of the reference thermometer is pivotal. If the reference thermometer is not well-graduated, the best of metrics will not be able to correctly evaluate our thermometer.

It's also crucial to differentiate between calibration data, which is used to determine the graduation, and evaluation data, which is used to assess the graduation's quality. A well-functioning thermometer should provide accurate temperature readings even for levels it wasn't calibrated on. Thus, evaluation data should differ from calibration data. If we only measure precision at 0°C and 100°C, a thermometer that perfectly fits the calibration data would receive

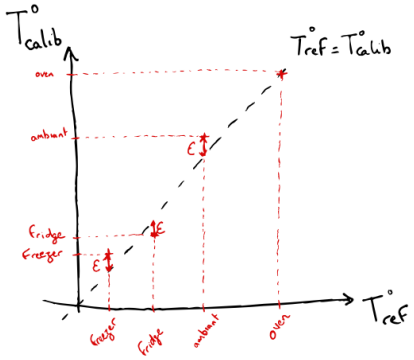
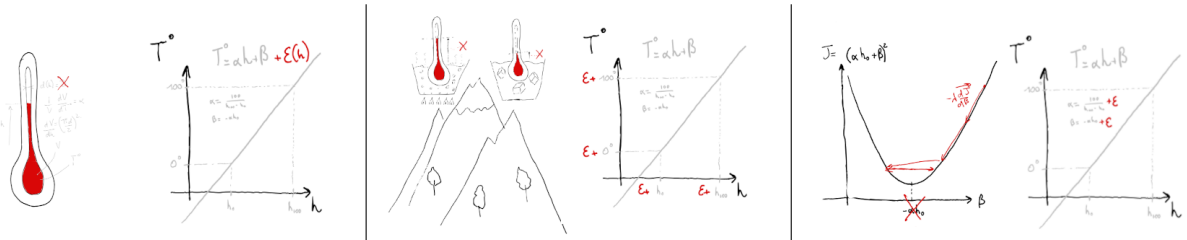


Figure 1.5 – **Evaluation and errors.** Given some evaluation data and choice of metric, we can compute the errors associated with our graduation. T°_{calib} and T°_{ref} are respectively the temperatures given by our graduation and a reference well graduated thermometer



Model | Data | Algorithm

Figure 1.6 – **Different sources of errors for the thermometer graduation.** From left to right: model errors result from erroneous assumptions about the system. Data errors result from inaccuracies in the calibration data and algorithmic errors result from a failure of the calibration algorithm to select the best candidate from the model.

the highest metric, whatever the other graduation indicates.

Sources of errors

Given an evaluation procedure, the errors are the differences to the reference and can be attributed to three sources: the model, the data and the calibration algorithm. The model is a source of error if the assumptions made were inaccurate. For example if the diameter of the tube is not constant with height the linear relationship between level and temperature is not verified and will induce errors when interpreting the level.

Even with perfect assumptions, noisy data can introduce errors in the calibration. If we interpreted our 0°C and 100°C in icing and boiling water at

the top of a mountain with lower atmospheric pressure, we will have calibrated our parameters with erroneous measurements and the subsequent graduation of our thermometer will be inaccurate.

Finally even with perfect assumptions and perfect data, the calibration algorithm used to find the solution's parameters can be a source of errors if it fails to find the optimal parameters. For example if we solve for the parameters with a gradient descent method, using a step size too big will prevent finding the exact parameters which will also results errors in the subsequent measurements.

In order to develop a graduation procedure, we need to take those sources of error into account. The graduation procedure choice will not only depend on the level-temperature relationship but on the whole relationship between calibration data to the final graduated thermometer. We therefore need to incorporate in our reasoning how the calibration data was acquired, what is the best model to map the level to the temperature, and what is the best algorithm to find the optimal parameters of the model.

This example allows us to formulate a generic methodological framework.

Methodological framework

In the process of finding a level-temperature relationship, we chose a model, a calibration algorithm, and had access to calibration data. Additionally, to evaluate our solution, we defined a metric and had access to evaluation data. Interestingly, these components can be specified at a higher level for finding and evaluating the graduation procedure itself, essentially creating a meta-level or "second order" problem.

The **Model**, in this second order scenario, combines different assumptions to determine the parameters of potential graduation procedures.

The **Calibration Algorithm** is used to select the best graduation procedure. This could be as straightforward as testing different combinations and choosing the most effective one, or it could involve complex numerical optimization procedures to determine higher-level parameters.

The **Calibration Data**, at the second order, consists of graduation tasks with a method to assess the performance of candidate procedures. This allows the algorithm to select the best solution.

The **Evaluation Metric** should reflect the intended use of the graduation procedure, including the range of thermometers we plan to use this procedure for. A useful metric might be the precision of all the thermometers we aim to graduate using the proposed solution.

The **Evaluation Data** should be representative of the variety of intended uses. This means it should contain graduation tasks for a range of thermometers of interest. Additionally, we need a reference for these tasks to measure the precision of our solution.

By leveraging these five components, we can select the best graduation procedure, quantify its quality using the evaluation data, and use it to graduate new thermometers with confidence in the resulting graduated instrument. This parallels the problems of "Finding the level-temperature mapping" (which we refer to as the first order problem) and "Finding the graduation procedure" (the second order problem) and offers insights on where general purpose methodological tools can find applications.

Note that second order metrics can extend beyond the scope of the first order problem. These metrics could encompass aspects such as robustness to noise or the computational complexity of the graduation procedure. This means our evaluation of a graduation procedure not only includes how well it measures temperature, but also how well it handles uncertainties or computational burdens.

A second order solution takes first order calibration data as inputs, which contain observations of a specific thermometer and their corresponding temperatures, and outputs a first order solution: a tailored graduation for the thermometer represented in the data.

The second order problem also involves making decisions on parameters to select the best solution, which can take various forms. For instance, these parameters can be discrete choices between different assumptions, like whether to consider the thermometer's tube diameter as constant or not. The parameters can also denote choices between different first order algorithms like choosing a direct linear system inversion or a iterative optimization procedure. Lastly, these second order parameters can be constants in the level-temperature mappings or parameters of an optimization procedure, like step size. This shows that the parameters in the second order problem have a broad range of applicability, affecting both the details of the graduation procedure and how the procedure is chosen.

Finally, a critical note is that the data used to evaluate a solution at the second order level should still be separate from the calibration data. This principle holds true for the same reasons it applies to the first order problem - using distinct data sets helps to ensure that our solutions generalize well beyond the specific scenarios they were trained on.

1.3 The case of ocean altimetry

Introducing space and time

Our previous example implicitly solved the estimation problem of the liquid temperature within the thermometer at a single point in space and time. However, we can extend the problem formulation to estimate a quantity over a spatial and temporal domain. Given some observations y defined on a spatio-temporal domain Ω_y we want to estimate a quantity of interest u on a domain Ω_u. We are therefore looking for a mapping f that estimate u from y The process of determining f can be detailed in two steps, first determining the set $\mathcal{F}$ such that $f \in \mathcal{F}$ by making some assumptions on f. Then determining the calibration algorithm c that will use the calibration data $\mathcal{D}$ to select f from $\mathcal{F}$ The evaluation of the solution rests on the choice of metrics m and evaluation data $\mathcal{E}$

Solving this general problem requires considering the sampling pattern of the observations with respect to the target estimation domain. Incomplete coverage will require the **model** to account for temporal and spatial relationship between y values and u as well as the spatio-temporal structure u. These additional assumptions will require suitable **calibration data and algorithm** and **evaluation data and metrics** to be calibrated and evaluated.

Satellite altimetry

The estimation of the sea surface height (SSH) given satellite altimetry data enters nicely in the methodological presented above.

As illustrated in figure 1.7, the calibration problem considered in this thesis consists in estimating the SSH measured by the KaRIN instrument by removing correlated error signals, using calibrated nadir observations. The target estimation domain here is fully observed, however, some observations contains errors originating from the instrument acquisition process. The **model** for this problem includes assumptions about the processes that produce the errors. Assumptions about the spatio-temporal structure of the SSH are also required here to relate the SSH on the KaRIN instrument to surrounding NADIR altimeter measurements. The mapping problem below isolate this challenge more specifically.

The altimetry mapping problem depicted in figure 1.8 focuses on the spatial and temporal interpolation of NADIR altimetry data. Considering the observation as direct measurements of SSH, we aim at estimating daily maps

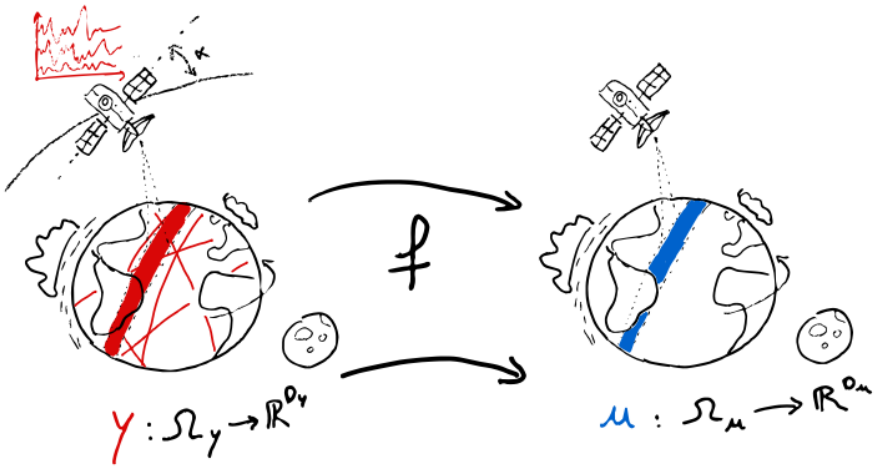


Figure 1.7 – **SWOT calibration.** The left part illustrate the observed domain in red while the right part indicates the domain on which we aim at estimating the SSH.

over a delimited domain. The **model** for such a problem needs to take into account the dynamical structure of the ocean surface topography.

This manuscript, therefore, aims to explore the application of deep learning to these two observation problems. The first is estimating SSH from noisy SWOT observations, and the second is inferring the complete SSH field from partial measurements which we detail in the following two sections. We give an overview in the next chapter of different existing **models** and **calibration algorithm** that have been developed for tackling these challenges.

In order to solve and evaluate solutions to these problems **calibration and evaluation data** are necessary. However, the SSH we aim to estimate is unknown on the target domain. Two separate experimental setups are used to address this issue. Observing System Experiments (OSE) [4] constitute a framework for working directly with observations for calibrating and evaluating new methods. For altimetry mapping for example, some satellite observations may be reserved for the interpolation process while others are employed to calibrate and evaluate the resulting map. Observing System Simulation Experiments (OSSE)[5] use ocean models as well as simulated observing systems to create an controlled environment where simulated ocean quantities are known. For SWOT calibration, this includes simulating processes of the satellite movement such as roll oscillation that are sources of error signals[3]. This peculiar data

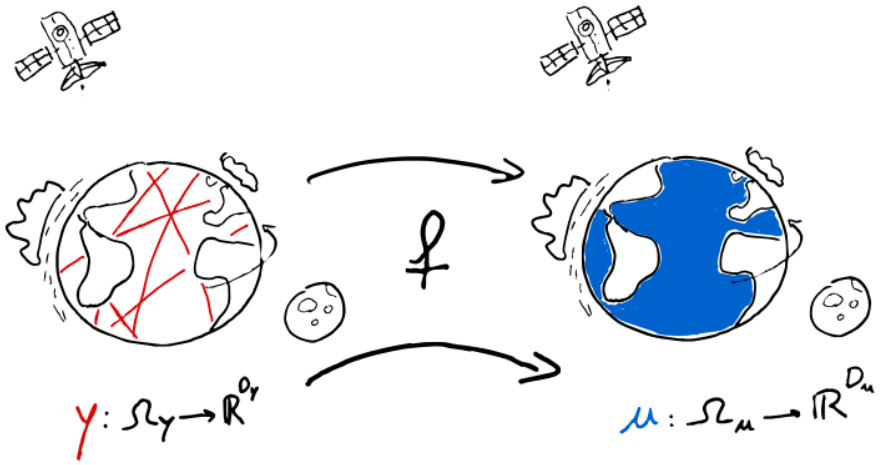


Figure 1.8 – **Nadir Altimetry mapping.** The left part illustrate the observed domain in red while the right part indicates the domain on which we aim at estimating the SSH.

context is a critical factor when developing data-centric methods such as deep learning.

1.4 Deep learning: opportunities and challenges

Success stories in computer vision (CV) and natural language processing (NLP)

In regard to the framework described above, deep learning brings forth **models**, such as neural networks, that are predicated on very weak assumptions. Their strength lies in the fact that, given sufficient parameters, they can approximate any function[6]. This leads to deep learning models defining vast parameter-space, consequently requiring substantial datasets and sophisticated optimization procedures to identify a good solution. These optimization procedures are akin to the **calibration algorithms**.

Deep learning **models** and **calibration algorithms** have advanced in tandem over the last decades. Innovations in model architectures such as ResNet[7], batch normalization[8], and in optimization procedures like Stochastic Gradient Descent (SGD)[9], Adam[10], and various learning rate schedules have consistently improved the calibration large models, therefore enabling the use of even larger neural networks.

However, the fact that deep learning models can in theory approximate any function introduces a peculiar consideration which is that fitting exactly the calibration data gives you no guarantee on how the model will behave on unseen data. Addressing this problem have motivated many innovations in regularization, architectures, initialization schemes and data augmentation techniques. It has also standardized the practice of splitting the **calibration data** in two sets: **training** and **validation**. The training set is used by the optimization procedure to search for the parameters whereas the validation set is used to assess the generalization on "unseen" data.

1.5 Thesis objectives and outline

The following chapters of this thesis are organized as follows:

Chapter 2 describes the main assumptions and methods that are formulated in current methods for altimetry mapping and calibration. This chapter aims at describing the existing **models** and **calibration algorithms** available in altimetry analysis as well as the related work in deep learning. A more detailed description of the neural network-based 4DVarNet framework and its applications will be presented since the contributions of this thesis make extensive use of this architecture.

Chapter 3 propose a deep learning architecture for the calibration of correlated errors in SWOT data. From an applicative standpoint, the flexibility of deep learning methodology opens the potential for capturing signals that are tricky to explicitly parameterize. From a methodological perspective, this study shows how deep learning architectures can be tailored with assumptions on the instrument and its measurements. More specifically, this is done by showing how the spectral specifities of the errors can be leveraged to design a efficient neural calibration scheme. This OSSE study do not directly address the challenges brought by the lack of ground-truthed dataset which are the focus of the following chapter.

Chapter 4 tackles more specifically the data availability problem. It studies how neural mapping schemes can be applied to real data despite the lack of reference dataset. It looks more specifically at the 4DVarNet framework which has been demonstrated in a simulated setup[11] using simulated SSH for training and evaluation. This chapter looks at the performance on real data of deep learning models trained on simulated data. It shows how the extensive physical knowledge of the ocean dynamics can be leveraged to palliate the lack

of ground-truthed dataset in altimetry through the use of numerical simulations for training.

Chapter 5 considers the obstacles to better synergies between the ocean altimetry and deep learning communities. Both fields are well established with accumulated knowledge, and best practices. As described in this chapter, deep learning brings powerful models and algorithms. However the calibration and evaluation data as well as the metrics can only be sensibly designed by an domain expert. We propose OceanBench, an interface in the form of a software suite of tools. Oceanbench aims at empowering domain experts to easily design altimetry problems of interests and qualifying them with relevant metrics. It then provides machine learning practitioners access to the necessary data as well as suited utilities for training and evaluating learning based methods.

Chapter 6 discusses and concludes on the research presented in this manuscript. We summarize the main objectives and results in previous chapters as well as proposing some future avenues of research.

BIBLIOGRAPHY

[1] M. Ballarotta, C. Ubelmann, M.-I. Pujol, G. Taburet, F. Fournier, J.-F. Legeais, Y. Faugère, A. Delepoulle, D. Chelton, G. Dibarboure, and N. Picot, "On the resolutions of ocean altimetry maps," *Ocean Science*, vol. 15, no. 4, pp. 1091–1109, Aug. 2019.

[2] "KaRIn on SWOT: Characteristics of Near-Nadir Ka-Band Interferometric SAR Imagery," https://ieeexplore.ieee.org/document/6553583/.

[3] "Empirical Cross-Calibration of Coherent SWOT Errors Using External References and the Altimetry Constellation," https://ieeexplore.ieee.org/document/6087373/.

[4] M. Hamon, E. Greiner, P.-Y. Le Traon, and E. Remy, "Impact of Multiple Altimeter Data and Mean Dynamic Topography in a Global Analysis and Forecasting System," *Journal of Atmospheric and Oceanic Technology*, vol. 36, no. 7, pp. 1255–1266, Jul. 2019.

[5] S. Verrier, P.-Y. Le Traon, and E. Remy, "Assessing the impact of multiple altimeter missions and Argo in a global eddy-permitting data assimilation system," *Ocean Science*, vol. 13, no. 6, pp. 1077–1092, Dec. 2017.

[6] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, Jan. 1989.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[8] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *Proceedings of the 32nd International Conference on Machine Learning*. PMLR, Jun. 2015, pp. 448–456.

[9] "Large-Scale Machine Learning with Stochastic Gradient Descent Léon Bottou," in *Statistical Learning and Data Science*, 0th ed., M. G. Summa,

L. Bottou, B. Goldfarb, F. Murtagh, C. Pardoux, and M. Touati, Eds. Chapman and Hall/CRC, Dec. 2011, pp. 33–42.

[10] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Jan. 2017.

[11] R. Fablet, M. M. Amar, Q. Febvre, M. Beauchamp, and B. Chapron, "END-TO-END PHYSICS-INFORMED REPRESENTATION LEARNING FOR SATELLITE OCEAN REMOTE SENSING DATA: APPLICATIONS TO SATELLITE ALTIMETRY AND SEA SURFACE CURRENTS," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-3-2021, pp. 295–302, Jun. 2021.

Chapter 2

Modeling and solving altimetry problems

This chapter discusses various existing methodologies for modeling and solving tasks related to ocean altimetry.

Our attention is primarily directed towards techniques that are relevant to the altimetry challenges explored in this study. Most existing methods are geared towards altimetry mapping, as calibration of the Surface Water and Ocean Topography (SWOT) mission is a relatively nascent area with fewer established techniques. Nonetheless, both types of problems aim to estimate the sea surface height u on a domain Ω_u given observations y on a domain Ω_y.

Therefore, the methods of interest $f \in \mathcal{F}$ are such that $f(y) = \hat{u}$ with $\hat{u}$ an estimation of u. The high level problem can be stated as finding best candidate $\hat{f} \in \mathcal{F}$ using a procedure c that relies on data $\mathcal{D}$. Solving this problem rely on the specification of $\mathcal{F}$. We find that the following decomposition is useful to better differentiate between the different approaches: $f = f_{x \to u} \circ f_{y \to x}$. The decomposition mainly introduces the choices of intermediate quantities x that characterize the state of an SSH field and a prior distribution $p(x)$. This implies the definition of $f_{x \to u}$ that provides the SSH estimate given a state x. Secondly, this decomposition implies the choice of a state estimation procedure $f_{y \to x}$ that will determine the state values best suited given observations y. Those components fit nicely in the first and second order formulation introduced in the previous chapter as depicted in Figure 2.1.

This chapter describes existing approaches for formulating the different components. Possible characterizations of the state $x \sim p(x)$ are detailed in Section 2.1 while Section 2.2 depicts existing approaches for formulating $f_{y \to x}$. Sections 2.3 and 2.4 focus respectively on the calibration c and evaluation considerations. The final section 2.5 delves into the specific choices of the 4DVarNet framework, which serves as the backbone for the research presented in this thesis.

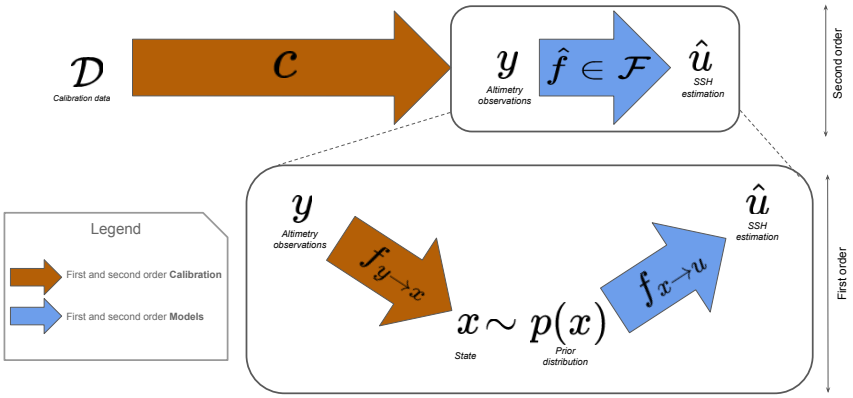


Figure 2.1 – **Methodological formulation**. This figure illustrate the organization of the different components described in this chapter

2.1 Priors on the Sea Surface Height (SSH)

The first distinguishing feature among various methods is the choice of characterization of SSH fields. The characterization can be decomposed in two parts. The first involves the choice of representation x for SSH fields u. This representation essentially outlines the space of all possible SSH states. The second aims to characterize the probability distribution $p(x)$ of x, describing which states are more likely a priori.

State representation

The choice of state representation defines the quantities x that characterizes the estimated SSH field $\hat{u}$. The choice of x implies the choice of the relationship $f_{y \to x}$ between the state values x and the estimated SSH field $\hat{u}$.

Looking at existing methods, the SSH field can be characterized through values sampled on a grid or mesh of the domain. These values can directly quantify the SSH as represented by methods such as the DUACS optimal interpolation [1] (OI), quasi-geostrophic back and forth nudging (BFN-QG) [2], Kalman filtering in GLORYS12 reanalysis [3] or Dynamical Interpolation [4, 5]. Other methods choose more complex or indirect descriptions through large and small scale components like in the 4DVarNet[6] or even projected values on another basis like MIOST[7, 8] which uses a wavelet basis. The choices like the mesh, basis or SSH decomposition used is a way to use prior knowledge to

dimension the state space. A grid representations of the SSH can be propagated to the whole domain Ω_u using interpolation schemes which impose additional choices.

Other methods, like the strong-constraint four dimensional variational data assimilation (s4DVAR)[9], consider only SSH fields that are solutions to differential equations characterizing the system. The state x then takes the form of initial temporal conditions that are propagated in time using a dynamical model numerical integration scheme. Note that parameters of the model and integration scheme can be part of the state and estimated alongside the initial conditions.

Deep learning has introduced new ways of representing the SSH. Notably, Neural fields[10] consists in describing the SSH with the parameters of a coordinate-based neural network. The neural network can then be used to output the SSH value for any given coordinate of the domain. Neural networks can also be used in tandem with previous concepts. For example they can parameterize a latent space[11] and model the basis change. They can also be used to parameterize the error of the dynamical model in s4DVAR[12].

Additionally the state can also contain ancillary quantities that are linked to the SSH. In the GLORYS12[3] reanalysis, the data assimilation scheme considers the state of the ocean beyond the SSH. In the case of the SWOT calibration, estimating the SSH is equivalent to estimating the error signals. Operational methods[13] approach the problem as defining a state representation of the different error signals.

Prior on the state space

The representation of SSH x we chose defined the space of all possible states. Additional assumptions can be made to specify which states are more likely than others. This is the prior distribution of the states $p(x)$. Some approaches like existing work on Neural fields[10] do not define an explicit prior distribution over the state space, which implies a uniform distribution. Others like OI and s4DVAR define $p(x)$ with a background state x_b and an error covariance matrix $\mathbf{B}$ with respect to the background. Under Gaussian assumptions, the prior distribution becomes

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{B}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - x_b)^T \mathbf{B}^{-1} (\mathbf{x} - x_b)\right) \tag{2.1}$$

Instead of considering the difference between the state and a background,

Method	State formulation	SSH estimation	Prior formulation
OI	SSH Grid	Grid Interpolation	Background error
KF $^{(*)}$, 3DVAR $^{(*)}$	SSH Grid at time t	Grid Interpolation	Trajectory error
s4DVar	Initial conditions	Trajectory computation	Initial background error
w4DVar	SSH Grid	Grid Interpolation	Trajectory error
MIOST	Wavelet components	Basis change	Background error
4DVarNet	Scale components	Sum + Grid interpolation	Energy-based model
Neural field	NN parameters	NN inference	None (Uniform)

Table 2.1 – **Comparison of SSH models across various methods**. The columns indicate the type of state representation x, the method of obtaining estimated SSH $\hat{u}$, and the prior distribution $p(x)$. Methods annotated with $^{(*)}$ consider the state of a single time step at a time.

other methods define the prior likelihood of a state based on its distance to the trajectory of a dynamical model. Kalman filters, 3DVAR or weak-constraint four dimensional variational data assimilation (w4DVAR)[9] are such methods. Under Gaussian assumptions $p(x)$ takes a similar formulation as Equation 2.1 but with x_b replaced by a state-dependent quantity.

Deep learning also introduces tools for probabilistic and energy-based modeling[14, 15] that can be used to characterize $p(x)$. For example, 4DVarNet employs a neural network (NN) ϕ to define the following quantity over the state space $E(x) = \|x - \phi(x)\|$. Interpreting this as an energy function, a Gibbs distribution can be defined over the state space, yielding the prior distribution:

$$p(\mathbf{x}) = \frac{\exp(-\frac{E(\mathbf{x})}{T})}{Z} \tag{2.2}$$

with $Z = \int \exp(-E(\mathbf{x})/T)\, d\mathbf{x}$ and T the temperature parameter.

We summarize in table 2.1 the different state characterization approaches.

2.2 Solvers: Estimating the state given some observations

Once all prior assumptions about the SSH field have been made (i.e. the first order **model**), the methods differ by the choice of **calibration procedure** $f_y \rightarrow x$ which estimate the state $\hat{x}$ given observations y.

This problem is classically framed as an inverse problem by defining an observation operator H that describe the relation from state to observations ($f_{y \rightarrow x}$ then becomes the inverse H). For altimetry observations that measure

the quantity of interest u, the H operator can be decomposed as $H(x) = f_{x\to u}(x)_{\Omega_y} + \epsilon$ with ϵ an error term and $f_{x\to u}(x)_{\Omega_y}$ the estimated SSH field for a given state x over the observed domain Ω_y. The field of data assimilation in geoscience propose a variety of methods to solve inverse problems.

Using a Bayesian inference formulation, the estimate of the posterior state is done by assuming Gaussian distributions for observation errors and prior states. For Optimal Interpolation, this can be mathematically expressed as:

$$f_{y\to x}(y) = x_b + \mathbf{K}(y - \mathbf{H}x_b) \tag{2.3}$$

with H a linear observation operator, x_b the background state and $\mathbf{K}$ is the Kalman gain (which depends on the linear observation operator $\mathbf{H}$, $\mathbf{B}$ and the covariance matrix of ϵ). Kalman filters use a similar expression which is applied sequentially for each observation time step. The procedure $f_{x\to u}$ then becomes a sequence of sub-procedure $f^1_{x\to u}t1, ..., f^{t_n}_{x\to u}$ applied to observations $y_{t1}, ..., y_{t_n}$ to estimate the states $x_{t1}, ..., x_{t_n}$

$$f_{y\to x}t + 1(y_{t+1}) = x_t + \mathbf{K_{t+1}}(y_{t+1} - \mathbf{H_{t+1}}\mathbf{M_{t:t+1}}x_t) \tag{2.4}$$

with $\mathbf{M_{t:t+1}}$ a linear forecast operator and $\mathbf{K_{t+1}}$ the Kalman gain similarly computed from the error and prediction covariance matrices.

Another formulation to solve inverse problems is used by variational methods. The estimation is framed as the minimization of a variational cost $f_{y\to x}(y) = \arg\min_x J(x, y)$ that includes a observation term $J_{obs}(x, y)$ and a regularization term $J_{reg}(x)$ related to the prior distribution. The formulation is:

$$f_{y\to x}(y) = \arg\min_x \left[J_{\text{reg}}(x) + J_{\text{obs}}(x, y)\right] \tag{2.5}$$

The minimization procedure is classically performed using an iterative gradient based algorithm[9]. The gradients can be computed using the adjoint method[3] or automatic differentiation libraries[16]. This approach is used in variational data assimilation such as 3DVAR, strong and weak 4DVAR, 4DVarNet and can be used to solve the OI problem. Note that similarly to Kalman filters, the 3DVAR method also perform a sequential resolution of the estimation. The calibration of neural fields is also framed as a the minimization of a training objective. However in existing work [10], since no prior distribution is defined over the states, the minimization objective contains only the observation term.

Deep learning offers alternative strategies outside of the inverse problem

Methods	Procedure f
3DVAR$^{(*)}$, w4DVAR, OI, 4DVarNet	Variational data assimilation
OI, Kalman filters$^{(*)}$	Bayesian inference
Direct inversion, 4dVarNet	NN inference
Neural fields	Neural network training

Table 2.2 – **Comparison of state estimation strategies for existing altimetry methods**. Methods annotated with $^{(*)}$ use a sequential resolution of successive time steps.

formulation coined as direct inversion. In such formulations a neural network directly models the function $f_{y \to x}$. Example of this approach using classical computer vision models have been applied to altimetry mapping in [17] and we apply this method in Chapter 3 for SWOT calibration applications.

The different calibration procedures of different methods are summarized in the table 2.2.

2.3 Calibrating the method

The choice of prior formulation and estimation procedure provide a method with free parameters (i.e. second order **model**) that need to be calibrated. Calibrating the method involves the fine-tuning of several parameters and model factors. These include the background field, covariance matrices for observation and background errors, as well as specifics for variational optimization or numerical integration schemes. In the case of deep learning approaches like direct inversion, the parameters of the neural network also become factors requiring calibration.

The calibration process relies on datasets, typically comprising numerical model outputs and pre-existing observations. These datasets serve for estimating the aforementioned factors and for validating the performance of the state estimation procedure. A widely adopted approach for this calibration is cross-validation. In this technique, a subset of available data is utilized to configure the model, which is then tested on the remaining data to evaluate its performance. The exploration of possible configurations may range from manual adjustments to automated parameter tuning algorithms, depending on the complexity of the method being calibrated.

For methods incorporating neural architectures, advanced optimization algorithms specific to deep learning are often employed. These algorithms efficiently tune the weights and biases of the neural network, aligning them for

better performance in the SSH estimation task. This is particularly advantageous for direct inversion approaches, which depend entirely on the data for calibration and do not require explicit prior models.

The lack of annotated data also motivated the design of training strategies that rely less on a target reference and more focused on capturing the structure of the available data. In this regard unsupervised[18] and self-supervised learning[19] approaches have seen an increase in popularity in earth observation.

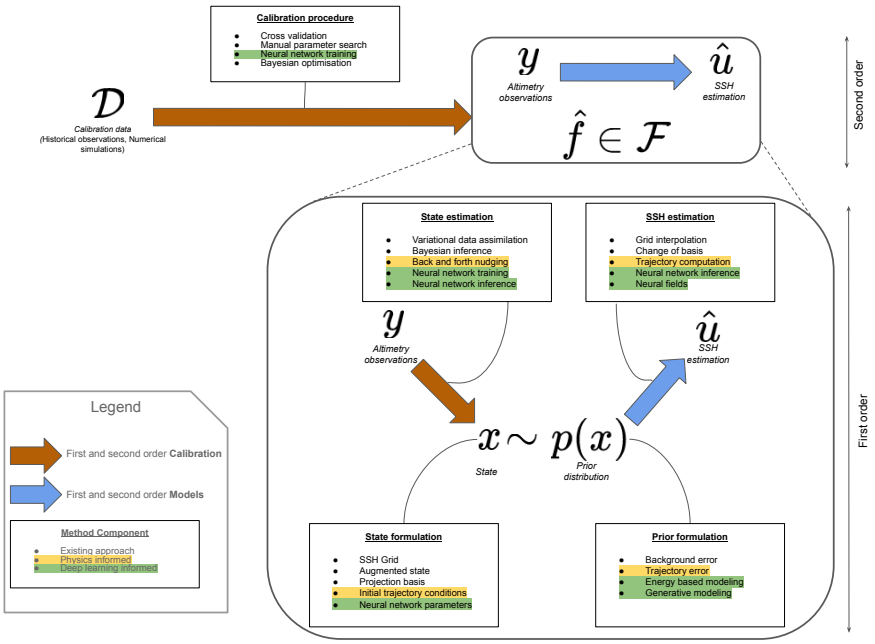


Figure 2.2 – **Solving altimetry problems**. This figure displays the different methodological components at play when addressing an altimetry challenge. The different approaches for each component are detailed. We highlight which ones are physics or deep learning inspired.

2.4 Evaluation

So far, we have described the choices made in developing the methods. Yet, a critical choice remains: selecting the criteria to assess the best method for a given problem. To address this, we consider various evaluation data and metrics.

Evaluation data can stem from two main sources, each with its unique risks and benefits. Firstly, the estimated SSH can be compared directly with

observation data. Termed as the Observing System Experiment (OSE), its primary advantage lies in its close resemblance to real-world scenarios. However, observational data presents its own challenges. Calibrated nadir altimeters are filtered, unable to capture smaller processes - the same processes anticipated in the new SWOT observations. This poses significant challenges when using nadir altimeter data to assess SWOT calibration. It's worth noting that while data from other observations, such as sea surface temperature or in situ drifter data, can be utilized, they present the challenge of discerning their relationship with the estimated SSH. When mapping altimetry, evaluation using nadir tracks means assessing the SSH over only a portion of the estimated domain. The sampling pattern of nadir altimeters also restricts evaluation. Their sporadic acquisition over time hinders the ability to compare the temporal evolution of the estimation with a reference. Furthermore, the one-dimensional nature of this acquisition precludes the evaluation of pertinent geophysical metrics like the vorticity field.

The second source of evaluation data originates from numerical simulations. Dynamical systems, spanning various complexity levels, can be modeled to produce a 'ground truth' SSH field. Similarly, observing systems can be simulated to produce pseudo-observations. Both types of simulations introduce errors when juxtaposed with real-world scenarios. These errors must be carefully accounted for when interpreting evaluations conducted in such settings. However, this evaluation approach offers considerable flexibility due to the availability of a ground truth.

In this thesis, both scenarios are employed. Chapter 3 examines the SWOT calibration in an OSSE context, while Chapter 4 probes how the evaluation of neural mapping schemes varies between OSSE and OSE contexts.

Several metrics are deliberated upon in this thesis. The primary one is the root mean squared error (RMSE) of the estimated SSH compared to the reference evaluation data. Although this metric provides an easily interpretable value in centimeters, it overlooks some aspects. Since the SSH encompasses processes spanning diverse scales and amplitudes, the normalized RMSE (nRMSE) is also employed to better gauge errors relative to the SSH's amplitude. Additionally, oceanic processes of varying amplitudes possess distinctive spatial and temporal scales. The more energetic processes exhibit broader spatial and extended temporal scales. It's imperative for domain experts to ascertain if these processes are aptly depicted in the SSH estimation. Consequently, spectral-based metrics are integral to this thesis to characterize at which scales

the errors become important with respect to the SSH signal.

In closing, a substantial part of assessing an SSH estimation and deciding on pertinent metrics hinges on qualitatively evaluating the field and its temporal evolution, as well as derived geophysical metrics like geostrophic currents and vorticity. Such qualitative evaluations necessitate profound geophysical expertise. Chapter 5 is rooted in these considerations, elucidating and suggesting tools to design appropriate experimental and evaluation frameworks for the adept development and assessment of altimetry mapping methods.

2.5 A closer look on the 4dVarNet

Method overview

The 4dVarNet is a prominent framework frequently employed throughout this thesis which is composed of the following key points. The SSH field is characterized by values on a regular spatial and temporal grid x. The SSH estimate $\hat{u}$ is then obtained using an interpolation scheme $f_{x\to u}$ between the grid points. The prior on the state space is formulated using a neural network ϕ as $\|x-\phi(x)\|_{l2}$. The 4dVarNet framework solves an inverse problem formulation using a variational formulation through the minimization of the cost $J(x,y)$ as stated in Equation 2.6

$$J(x,y) = \|f_{x\to u}(x)(\Omega_y) - y(\Omega_y)\|_{l2} + \|x-\phi(x)\|_{l2} \tag{2.6}$$

The minimization procedure consists in an iterative gradient-based procedure involving a neural network ψ which is described in Equation 2.7.

$$\begin{aligned} f_{y\to x}(y) &= \arg\min_x \left[J(x,y)\right] \\ &= x^N = x^{N-1} - \psi(\nabla_x J(x^{N-1}, y)) \end{aligned} \tag{2.7}$$

The calibration of the neural network parameters of ϕ and ψ are performed through the end-to-end training scheme minimizing a training objective $\mathcal{L}$ on an OSSE dataset $\mathcal{D}$ as put in Equation 2.8.

$$\hat{f} = \arg\min_f \sum_{(y,u)\in\mathcal{D}} \mathcal{L}(f(y), u) \tag{2.8}$$

Existing work

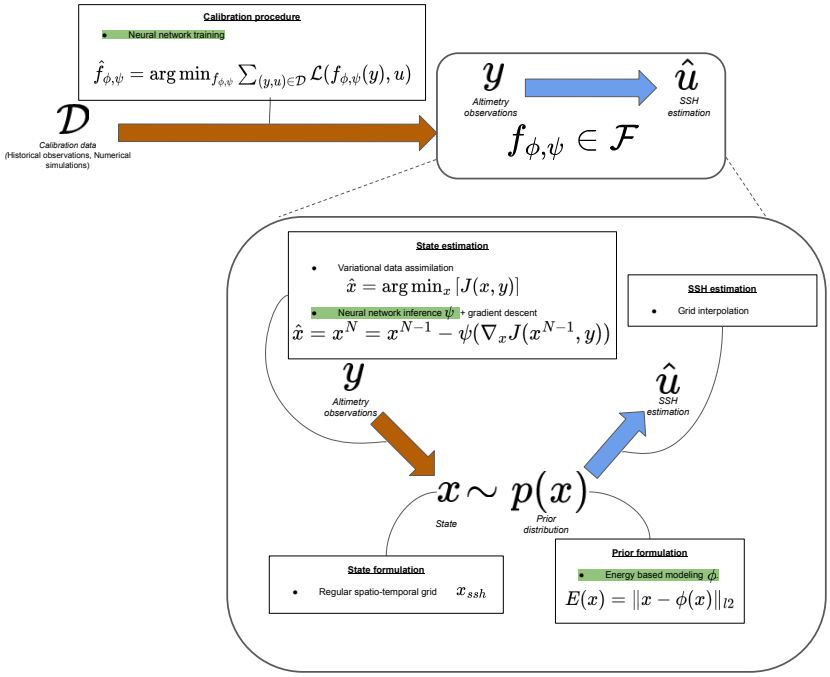


Figure 2.3 – **4dVarNet Method**. 4dvarnet components contextualized within the methodological framework

In this section, we describe in more detail the different variations of the 4DVarNet in existing work. Introduced as a promising approach for mapping altimetry data, 4DVarNet demonstrated robust performance in a study involving simulated SSH based on NATL60 simulation data[16]. Various regions [6] and altimetry configurations were considered in these studies, including setups with 4 nadir altimeters both with and without SWOT observations. Earlier versions considered OI-based products as additional observational data for the inversion [16].

The neural network ϕ employed for the prior formulation consists in bilinear blocks as introduced in [20]. Earlier work [16] used a multiscale architecture while a simpler single scale is used in Chapters 4 and 5. Outside of altimetry, other ϕ have been experimented on Lorenz systems including using the true system's dynamics[21]

A Gaussian assumption for the observation errors is commonly adopted, articulated through a quadratic norm in the observation cost. In a study that incorporate Sea Surface Temperature (SST) observations y_{sst}[22], convolutional filters H_{c1}, H_{c2} are used to formulate an additional observation cost that relates

the state to SST $\|H_{c1}(x) - H_{c2}(y_{sst})\|$.

The neural network ψ is inspired by meta-learning studies [23] that employ a Long-Short-Term-Memory (LSTM) to compute state updates[16]. Earlier works use a fixed-point algorithms which iteratively impose state values at observed grid points and apply a forward pass of the neural network ϕ [24].

Apart from the cross validation and exploration of different architectures and configurations, the training of the neural network's parameters, both for the solver and the prior, is achieved through the Adam optimization algorithm [25]. The aim is to minimize both the mean squared error in SSH reconstruction and its gradients. In order to constrain the estimated states to have low prior costs, a supplementary term is added to ensure effective weightage in the neural prior. This results in the training objective $\mathcal{L}$ described in Equation 2.9.

$$\mathcal{L}(\hat{u}, u) = \alpha_1\|\hat{u} - u\| + \alpha_2\|\nabla\hat{u} - \nabla u\| + \alpha_3\|\hat{u} - \phi(\hat{u})\| \tag{2.9}$$

4DVarNet has emerged as a versatile and effective architecture for handling altimetry data, with various advancements and optimizations over time. By combining neural networks with traditional variational techniques, it opens up promising avenues for state-of-the-art state estimation in oceanographic applications.

BIBLIOGRAPHY

[1] G. Taburet, A. Sanchez-Roman, M. Ballarotta, M.-I. Pujol, J.-F. Legeais, F. Fournier, Y. Faugere, and G. Dibarboure, "DUACS DT2018: 25 years of reprocessed sea level altimetry products," *Ocean Science*, vol. 15, no. 5, pp. 1207–1224, Sep. 2019.

[2] F. L. Guillou, S. Metref, E. Cosme, C. Ubelmann, M. Ballarotta, J. L. Sommer, and J. Verron, "Mapping Altimetry in the Forthcoming SWOT Era by Back-and-Forth Nudging a One-Layer Quasigeostrophic Model," *Journal of Atmospheric and Oceanic Technology*, vol. 38, no. 4, pp. 697–710, Apr. 2021.

[3] J.-M. Lellouche, E. Greiner, R. Bourdallé-Badie, G. Garric, A. Melet, M. Drévillon, C. Bricaud, M. Hamon, O. Le Galloudec, C. Regnier, T. Candela, C.-E. Testut, F. Gasparin, G. Ruggiero, M. Benkiran, Y. Drillet, and P.-Y. Le Traon, "The Copernicus Global 1/12° Oceanic and Sea Ice GLORYS12 Reanalysis," *Frontiers In Earth Science*, vol. 9, Jul. 2021.

[4] C. Ubelmann, P. Klein, and L.-L. Fu, "Dynamic Interpolation of Sea Surface Height and Potential Applications for Future High-Resolution Altimetry Mapping," *Journal of Atmospheric and Oceanic Technology*, vol. 32, no. 1, pp. 177–184, Jan. 2015.

[5] M. Ballarotta, C. Ubelmann, M. Rogé, F. Fournier, Y. Faugère, G. Dibarboure, R. Morrow, and N. Picot, "Dynamic Mapping of Along-Track Ocean Altimetry: Performance from Real Observations," *Journal of Atmospheric and Oceanic Technology*, vol. 37, no. 9, pp. 1593–1601, Sep. 2020.

[6] M. Beauchamp, Q. Febvre, H. Georgenthum, and R. Fablet, "4DVarNet-SSH: End-to-end learning of variational interpolation schemes for nadir and wide-swath satellite altimetry," *Geoscientific Model Development*, vol. 16, no. 8, pp. 2119–2147, Apr. 2023.

[7] C. Ubelmann, G. Dibarboure, L. Gaultier, A. Ponte, F. Ardhuin, M. Ballarotta, and Y. Faugère, "Reconstructing Ocean Surface Current Combining Altimetry and Future Spaceborne Doppler Data," *Journal of Geophysical Research: Oceans*, vol. 126, no. 3, p. e2020JC016560, 2021.

[8] C. Ubelmann, L. Carrere, C. Durand, G. Dibarboure, Y. Faugère, M. Ballarotta, F. Briol, and F. Lyard, "Simultaneous estimation of ocean mesoscale and coherent internal tide sea surface height signatures from the global altimetry record," *Ocean Science*, vol. 18, no. 2, pp. 469–481, Apr. 2022.

[9] A. Carrassi, M. Bocquet, L. Bertino, and G. Evensen, "Data assimilation in the geosciences: An overview of methods, issues, and perspectives," *WIREs Climate Change*, vol. 9, no. 5, p. e535, 2018.

[10] J. E. Johnson, R. Lguensat, R. Fablet, E. Cosme, and J. L. Sommer, "Neural Fields for Fast and Scalable Interpolation of Geophysical Ocean Variables," Nov. 2022.

[11] S. Benaïchouche, C. L. Goff, B. Boussidi, F. Rousseau, and R. Fablet, "Learnable Variational Models for the Reconstruction of Sea Surface Currents Using Ais Data Streams: A Case Study on the Sicily Channel," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, Jul. 2022, pp. 6821–6824.

[12] A. Farchi, M. Chrust, M. Bocquet, P. Laloyaux, and M. Bonavita, "Online Model Error Correction With Neural Networks in the Incremental 4D-Var Framework," *Journal of Advances in Modeling Earth Systems*, vol. 15, no. 9, p. e2022MS003474, 2023.

[13] G. Dibarboure, C. Ubelmann, B. Flamant, F. Briol, E. Peral, G. Bracher, O. Vergara, Y. Faugère, F. Soulat, and N. Picot, "Data-Driven Calibration Algorithm and Pre-Launch Performance Simulations for the SWOT Mission," *Remote Sensing*, vol. 14, no. 23, p. 6070, Jan. 2022.

[14] Y. LeCun, S. Chopra, R. Hadsell, M. Ranzato, and F. Huang, "A tutorial on energy-based learning," *Predicting structured data*, vol. 1, no. 0, 2006.

[15] "A Practical Guide to Training Restricted Boltzmann Machines | SpringerLink," https://link.springer.com/chapter/10.1007/978-3-642-35289-8_32.

[16] R. Fablet, M. M. Amar, Q. Febvre, M. Beauchamp, and B. Chapron, "END-TO-END PHYSICS-INFORMED REPRESENTATION LEARNING FOR SATELLITE OCEAN REMOTE SENSING DATA: APPLICATIONS TO SATELLITE ALTIMETRY AND SEA SURFACE CURRENTS," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-3-2021, pp. 295–302, Jun. 2021.

[17] G. E. Manucharyan, L. Siegelman, and P. Klein, "A Deep Learning Approach to Spatiotemporal Sea Surface Height Interpolation and Estimation of Deep Currents in Geostrophic Ocean Turbulence," *Journal of Advances in Modeling Earth Systems*, vol. 13, no. 1, p. e2019MS001965, 2021.

[18] B. Hosseiny, M. Mahdianpari, M. Hemati, A. Radman, F. Mohammadimanesh, and J. Chanussot, "BEYOND SUPERVISED LEARNING IN REMOTE SENSING: A SYSTEMATIC REVIEW OF DEEP LEARNING APPROACHES," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, pp. 1–22, 2023.

[19] N. Harilal, B.-M. Hodge, A. Subramanian, and C. Monteleoni, "STint: Self-supervised Temporal Interpolation for Geospatial Data," Aug. 2023.

[20] R. Fablet, S. Ouala, and C. Herzet, "Bilinear Residual Neural Network for the Identification and Forecasting of Geophysical Dynamics," in *2018 26th European Signal Processing Conference (EUSIPCO)*, Sep. 2018, pp. 1477–1481.

[21] R. Fablet, B. Chapron, L. Drumetz, E. Mémin, O. Pannekoucke, and F. Rousseau, "Learning Variational Data Assimilation Models and Solvers," *Journal of Advances in Modeling Earth Systems*, vol. 13, no. 10, p. e2021MS002572, 2021.

[22] R. Fablet, Q. Febvre, and B. Chapron, "Multimodal 4DVarNets for the Reconstruction of Sea Surface Dynamics From SST-SSH Synergies," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023.

[23] M. Andrychowicz, M. Denil, S. Gómez, M. W. Hoffman, D. Pfau, and T. Schaul, "Learning to learn by gradient descent by gradient descent."

[24] M. Beauchamp, r. fablet, C. Ubelmann, M. Ballarotta, and B. Chapron, "Data-driven and learning-based interpolations of along-track Nadir and wide-swath SWOT altimetry observations," in *Proceedings of the 10th International Conference on Climate Informatics*, ser. CI2020. New York, NY, USA: Association for Computing Machinery, Jan. 2021, pp. 22–29.

[25] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Jan. 2017.

Chapter 3

Scale-aware neural calibration for wide swath altimetry observations

This chapter is based on a journal publication with the same name ongoing review. The preprint is available here [1]

3.1 Introduction

Nadir altimeter satellites provide invaluable direct measurements of the sea surface height (SSH) to monitor sea surface dynamics. They have played a key role in better understanding ocean circulation and improving climate monitoring. Altimeter-derived SSH data are also of key interest for offshore activities, marine pollution monitoring or maritime traffic routing among others.

However due to the sparse and irregular sampling associated with nadir altimeter constellations, a wide range of ocean processes from the mesoscale to the submesoscale range remains unresolved, typically for horizontal scales below 150 kilometers and time scales below 10 days. The recently launched SWOT mission, with its Ka-band radar interferometer (KaRIn) sensor, provides for the first time higher-resolution and two-dimensional snapshots of the SSH. Once this data is adequately processed, it will likely strongly impact our ability to observe and study upper ocean dynamics [2].

As reported in Figure 3.1, KaRIn data will be affected by instrument and geophysical errors [3] and their exploitation requires to develop robust calibration schemes. We illustrate in Fig.3.2 the two main error sources: instrument errors, especially roll errors, are expected to cause the dominant large-scale signal in both across-swath and along-swath directions; and geophysical errors,

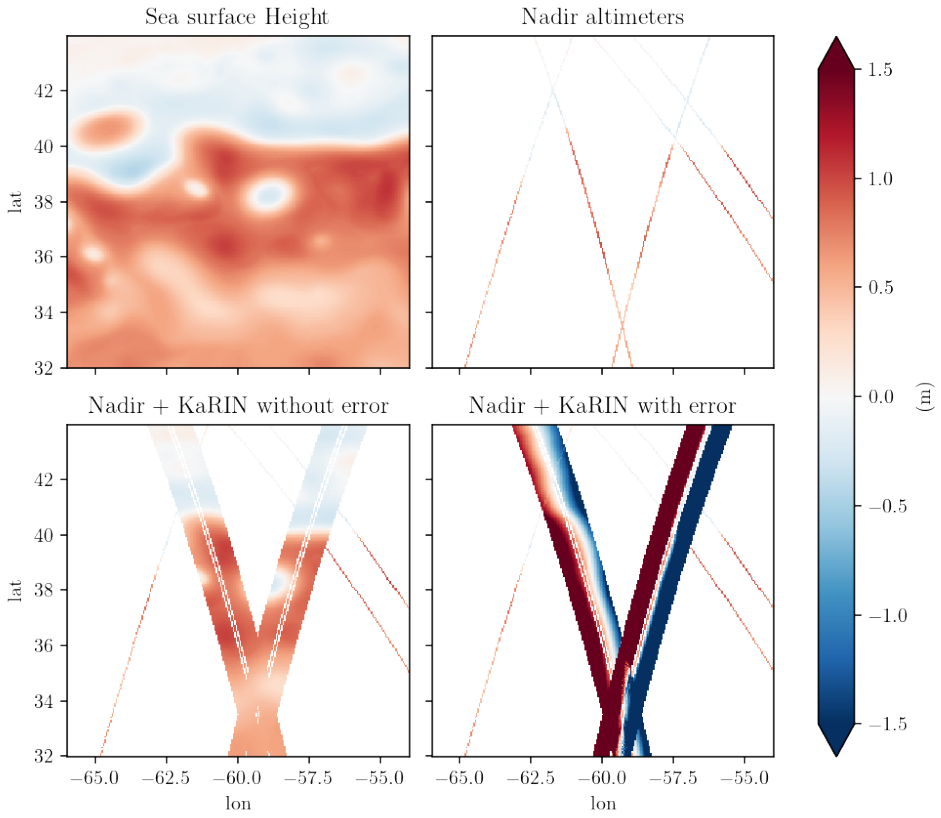


Figure 3.1 – **Observing System Simulation Experiment Cross-Calibration data:** *Top left:* Sea surface height (SSH) on October 26th 2012 from NATL60 simulation dataset. *Top right:* Calibrated NADIR pseudo-observations sampled using realistic orbits from the SSH, they are used to compute the gridded product for the cross-calibration. *Bottom-left:* NADIR + noise-free-KaRIn pseudo-observations, the 2D sampled SSH is the target of the cross-calibration. *Bottom-right:* NADIR + noisy-KaRIn pseudo-observations, simulated errors added to the swath SSH constitute the uncalibrated input of the cross-calibration problem

in particular due to wet-troposphere-induced delays [1]. The amplitude of these errors typically range from a few centimeters to a few meters in simulation, when the variability of the SSH for scales below 150km typically amounts to centimeters (See Fig. 3.6). This makes SWOT calibration a particularly challenging task in terms of signal-to-noise ratio. State-of-the calibration schemes [4] rely on explicit spectral priors to address the calibration problem. The

1. We refer the reader to Section 3.3 for the description of these error signals in raw KaRIn observations.

underlying hypotheses that one can linearly separate the SSH and the different error components may however impede the performance of such calibration methods. This chapter aims to assess the potential of deep learning models in addressing such altimetry challenges. This scenario serves as a compelling use case for evaluating the potential of deep learning in altimetry data analysis since it involves considerations related to both the ocean system and the observing system.

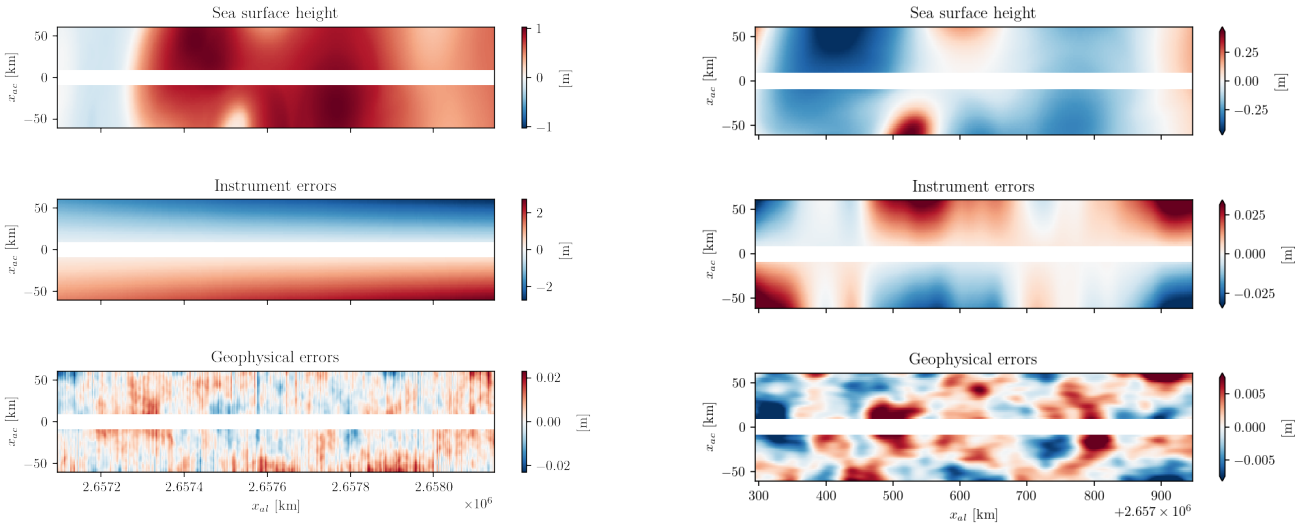
The deep learning solution presented leverages existing mapping methods to provide an initial estimate of SSH on the SWOT swath. It also incorporates a neural network inspired by computer vision architectures to refine this initial estimate using SWOT observation data. Our study demonstrates that incorporating knowledge about error signals when designing the deep learning model is crucial for improving the initial estimate.

This study does not directly address the challenges posed by the lack of precise knowledge regarding SSH and error signals. Instead, we utilize data from state-of-the-art ocean simulations and SWOT error models to calibrate and assess the effectiveness of our method. This approach creates an idealized setup for calibration and evaluation purposes, allowing us to focus on selecting a deep learning architecture suited for the altimetry task. However, it also raises questions about the transferability of deep learning techniques developed in simulated environments to real altimetry data which are posed in the next chapter.

The contributions of this chapter are as follows:

— We state the cross-calibration of KaRIn altimetry observations as a learning problem using both raw KaRIn altimetry data and a gridded altimetry product as inputs to the neural network.
— Our neural network architecture applies a scale-space decomposition scheme in the geometry of the KaRIn swath to improve the separability of the SSH and of the errors.
— Numerical experiments using an Observing System Simulation Experiment (OSSE) demonstrate the relevance of the proposed approach and highlight the impact of the quality of the gridded altimetry product to retrieve finer-scale patterns in the calibrated KaRIn observations.

This chapter is organized as follows. Section 3.2 provides some background on related work. We introduce the considered data and case-study in Section 3.3. Section 3.4 presents our method and we report numerical experiments in Section 3.5. Section 3.6 discusses further our main contributions.



(a) f (b) $\mathcal{G}_{200km}(f) - \mathcal{G}_{10km}(f)$

Figure 3.2 – **1000km segment of KaRIn observation components in swath geometry:** *(a)* Looking at the three signals we see that the large scale instrument errors (middle) are predominant compared to the SSH (top) and geophysical error (bottom). *(b)* Looking at the along-track scales between 10km and 200km, we note that the SSH is dominant w.r.t the error signals.

3.2 Background

Satellite altimeters

In this chapter, we address the cross-calibration of KaRIn observations, meaning that the proposed calibration scheme relies on external calibrated data. More specifically, we consider a constellation of 4 nadir satellite altimeters. We recall that nadir altimeters can provide calibrated measurements of the SSH for medium to large scales along 1D profiles corresponding to satellites' orbiting paths. Over the last decades the constellation counted typically from 4 to 7 satellites.

By contrast, according to the mission's error budget specification [2] the KaRIn instrument samples a two-dimensional swath of approximately 120km-wide with a 2km×2km pixel resolution everywhere over the ocean.

In Figure 3.1, we report simulated altimetry observations for both nadir altimeters and KaRIn along with the reference SSH issued from a numerical simulation dataset (see the Section 3.3 for details). As an illustration of the

complexity of calibration problem, the error signals completely occlude the SSH signal in the uncalibrated KaRIn observation. Figure 3.2 illustrates further this point in the swath geometry. When focusing on along-track scales between 10km and 200km, the SSH signal becomes the main signal (Fig. 3.2). This supports both to consider a scale-space decomposition and to investigate a cross-calibration approach with the exploitation of nadir-altimeter-derived altimetry products, which typically resolve horizontal scales above 100 km.

Interpolation of satellite-derived altimetry data

As mentioned previously, flying nadir altimeter constellations naturally advocate for considering the resulting interpolated SSH products as auxiliary data of interest to address the calibration of KaRIn observations.

Regarding operational SSH products, we may distinguish the optimally-interpolated altimetry-derived product (DUACS) [6] and reanalysis products using ocean general circulation models to assimilate various observation datasets, including satellite altimetry and satellite-derived sea surface temperature data [7]. Both types of products typically retrieve SSH dynamics on a global scale for horizontal scales above 150km and 10 days.

Recently, a renewed interest has emerged in interpolation methods for ocean remote sensing data [8][9]. Especially, deep learning schemes have emerged as appealing approaches to make the most of available observation datasets. Recent benchmarking experiments [10] point out significant potential gains compared with the above-mentioned operational products.

Here, we aim at investigating the extent to which the quality of L4 nadir-altimetry-derived SSH products may impact the calibration of KaRIn observations.

Scale-space theory

The scale-space theory provides a mathematically-sound framework to decompose 2D signals at different spatial scales [5]. Gaussian scale-space methods are among the most widely used. They rely on applying Gaussian blur transformations with different standard deviations. This approach has been widely used in low-level image processing tasks [11, 12]. Recent studies have used the scale-space theory in deep learning architectures [13, 14]. These neural networks better deal with multi-scale patterns in the data. Here, we draw inspiration from the scale-space framework to address the KaRIn calibration problem. We design a scale-aware decomposition scheme as part of our learning

approach with a view to better accounting for the different characteristic scales of the signals in play.

Deep Learning for earth observation

Convolutional neural networks are among the state-of-the-art neural architectures for image processing applications, including image classification[15, 16], image in-painting[17], object detection[18] and more. They have also led to breakthroughs in remote sensing problems such as SAR image segmentation[19, 20], altimetry data interpolation [21] and even sensor calibration [22]. The problem of multi-scale processing in neural networks has traditionally been tackled through the use of pooling layers in architectures such as the UNet [23]. As shown in the reported numerical experiments, these architectures do not reach a state-of-the-art performance for our KaRIn calibration problem. This advocates for the design of neural architectures better accounting for the key features of KaRIn observations.

3.3 Data and Case-study

In this chapter, we run an Observing System Simulation Experiment (OSSE), meaning that we rely on simulated data to apply and evaluate the proposed neural approach. In this section, we present the different datasets considered in this OSSE.

NATL60

The simulation of the sea surface height field is taken from the NATL60 [24] run of the NEMO ocean model. This simulation spans one year and covers the North Atlantic basin with a 1/60° spatial resolution. We more specifically use the data from a 12°×12° domain over the Gulfstream ranging from the longitudes -66° to -54° and latitudes 32° to 44°.

Nadir observations

In order to generate realistic nadir-altimeter pseudo-observations, we consider the real orbits of the years 2012 and 2013 of the four missions Topex-Poseidon, Jason 1, Geosat Follow-On, Envisat, as well as the 21-day cycle phase SWOT orbit from the SWOT simulator [3] project. The sampling of the nadir-altimeter pseudo-observations relies on the interpolation of the hourly SSH

fields of the NATL60 run at the orbit coordinates. We consider nearest-neighbor interpolation in time and a bilinear interpolation in space.

KaRIn observations

The SWOT simulator also generates the swath coordinates on each side of the SWOT nadir. The swath spans from 10km to 60km off nadir with a 2km by 2km resolution. The SSH is then sampled on those coordinates the same way as the nadir observations. Additionally, we also use the SWOT simulator in its "baseline" configuration to generate observation errors. Our simulation includes the systematic instrument errors with the roll, phase, timing and baseline dilation signals. Those signals have time-varying constant, linear or quadratic shape in the across track dimension. We also consider the geophysical error with the wet troposphere residual error as implemented in the simulator. We refer the reader to [3] for a detailed presentation of the SWOT simulator.

Gridded Altimetry Products

As explained in section 3.2, we make use of interpolated SSH products based on nadir altimetry data as inputs for our cross-calibration method. We consider two interpolation schemes in our study:

— the operational state-of-the-art based on optimal interpolation as implemented in the DUACS product [6].

— a state-of-the-art neural interpolation scheme, referred to as 4DVarNet [21]. This method is based on a trainable adaptation of the 4DVAR [25] variational data assimilation method, and out-performs concurrent approaches in the considered OSSE setup [10]. We consider two 4DVarNet interpolation configurations, one using only nadir altimetry data [26], one using jointly nadir altimetry and sea surface temperature data [27]. We also include the latter as it significantly improves the reconstruction of the SSH at finer scales.

3.4 Proposed Methodology

This section presents the proposed methodology for the cross-calibration of raw KaRIn observations. We design trainable neural architectures that take as inputs the uncalibrated KaRIn observations and the nadir-altimeter-derived gridded SSH products interpolated on the KaRIn swath. We train these architectures in a supervised manner on the reconstruction of the SSH on the

Figure 3.3 – **Overview of the proposed architecture:** From left to right: The first step interpolates the nadir-based gridded product onto the swath segment. Afterwards, both the nadir-based gridded product and KaRIn observation undergo the scale-space decomposition scheme outlined in 3.4. The scale components are stacked as channels and processed through the neural network. The blue color of the "Split Conv" indicates that each side of the swath is processed independently by the convolution layer whereas the orange coloring of the "Swath Mix" layer tells that the whole data is processed jointly (more details in 3.4). The final convolution computes a correction to be added to the gridded product for computing the calibrated KaRIn data

KaRIn swath. We first present an overview of the proposed neural architectures (Section 3.4). We then detail two specific components, namely the scale-space decomposition block (Section 3.4) and the swath-mixing layers (Section 3.4).

Proposed neural architecture

The overall architecture considered is shown in figure 3.3. The scale-space decomposition block first decomposed independently the input L4 SSH products and KaRIn observations into N_s-scale tensors, which we concatenate as the channel dimension. This results into a tensor of shape $(2N_s, N_{al}, N_{ac})$ where N_{al} and N_{ac} are respectively the along track and across track sizes of the input swath section. The scale-space decomposition step is described in section 3.4 A linear 2D convolution layer follows. The data is then processed by a series of residual convolutional blocks composed of a convolution layer, a ReLU

non-linearity [28], a skip connection and a swath-mixing layer as described in 3.4. A last linear convolution layer outputs a residual field, which we sum with the input gridded L4 SSH product to produce the calibrated KaRIn observation. The interested reader can refer to our implementation [2].

Scale-space decomposition

We exploit a Gaussian scale-space to compute a scale-space decomposition of the fields provided as inputs to our neural architecture. For given scales σ_1 and σ_2, we extract the associated signal as the difference between filtered versions of the input signal using two Gaussian filters with standard deviation σ_1 and σ_2. We consider one-dimensional filters for the along-track direction. Formally, denoting $\mathcal{G}_\sigma$ the 1-dimensional Gaussian blur operator with standard deviation σ in the along track dimension, the considered scale-space decomposition of a signal f given a sequence of increasing scales $[\sigma_1, \sigma_1, ..., \sigma_S]$ computes the following $S+1$ components: $[\mathcal{G}_{\sigma_1}(f), \mathcal{G}_{\sigma_2}(f) - \mathcal{G}_{\sigma_1}(f), ..., \mathcal{G}_{\sigma_S}(f) - \mathcal{G}_{\sigma_{S-1}}(f), f - \mathcal{G}_{\sigma_S}]$ These different components are then considered as channels for the convolutionnal networks. In our experiments, we consider 20 scales in the along-track direction evenly spaced from 8km to 160km. We discuss in section 3.5 how sensitive the proposed method is to the parameterization of the decomposition. To account for scale-dependent energy levels in the computed scale-space representation (see fig. 3.4), we introduce a batch normalization layer [29]. It re-scales each component to centered and unit-variance variables. We illustrate in Fig.3.4 the impact of the batch normalization step on the relative variance of the signal of each scale of the decomposition.

One may regard the proposed scale-scale decomposition as a convolutional block. Learning such a decomposition from data would however require very large convolutional filters, which does not seem realistic, or a deeper architecture with pooling layers that would require very efficient optimization given the quantity of data available.

Swath mixer block

As observed in Figure 3.2, the swath observed from the KaRIn sensor is not contiguous in the across-track dimension. The observation errors are however clearly correlated between the two sides of the swath. To exploit these

2. `https://github.com/CIA-Oceanix/4dvarnet-core/releases/tag/tgrs-calcnn-2023`

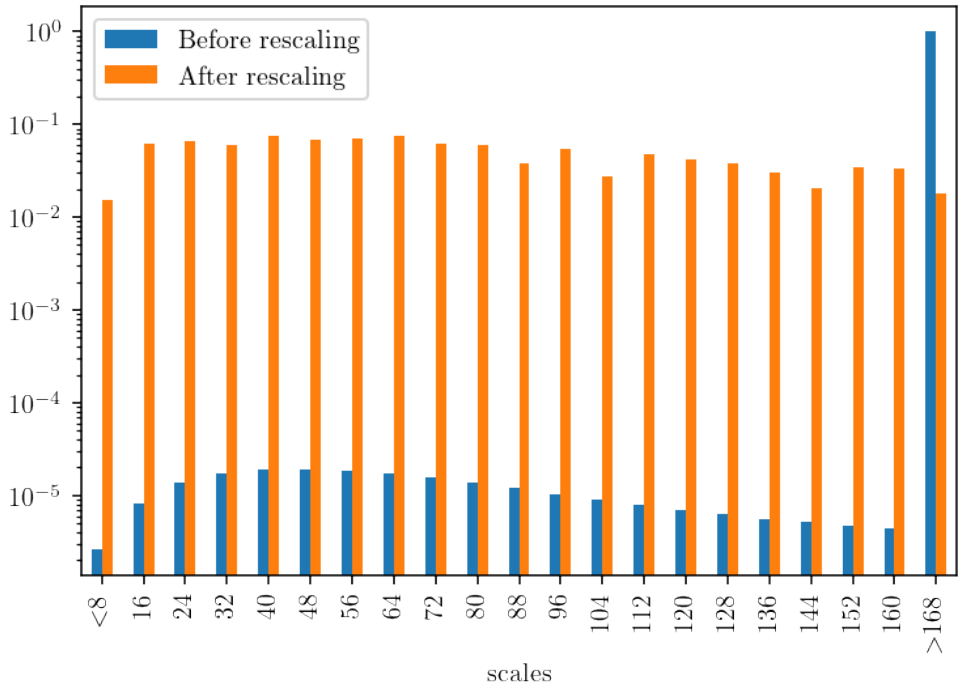


Figure 3.4 – **Explained variance of scale components before and after re-scaling:** Each bar indicates how much each scale component of the uncalibrated KaRIn contributes to the total variance of the signal, we can see that before re-scaling (blue) there is four orders of magnitude between largest scale and the others. The learnt re-scaling allows for scale component to be spread within a single order of magnitude (orange), which is more suited to the downstream neural architecture.

correlations in our architecture, we design a swath-mixer block with two specific layers.

To avoid convolution kernels to mix information from the two sides of the swath which could result in some unwanted side effects, each side is processed separately by each convolution layer noted "Split Conv" in Fig. 3.3. Additionally, each convolution layer input is padded so that the height and width of the input remain unchanged throughout the network.

Besides, to combine relevant features from the two sides of the swath, we introduce a layer denoted as "Swath-mix" in Fig. 3.3. It implements a convolution layer after transposing the across-track dimension as a channel dimension. This idea of a mixer layer has been used in architectures such as the MLP-Mixer [30], in which it has been shown to help with the expressiveness of neural networks.

	RMSE (m)	RMSE $\|\|\nabla_{ssh}\|\|$
CalCNN	1.39e-02	6.46e-03
UNet	2.34e-02	1.07e-02
4DVarNet-5nad	2.17e-02	9.57e-03

Table 3.1 – **Residual error of the benchmarked calibration frameworks**

We analyse in section 3.5 the contribution of the mixing layer.

3.5 Experimental results

Setup

The results of this section have been computed using the one year ocean simulation NATL60, over the 12°x12° domain over the Gulfstream. The model evaluation is done on forty days in the inner 10°x10° region. The training of the mapping and calibration models are done on the remaining days. The experimental setup used is the same as in [10] The base configuration for our architecture uses three convolutional blocks with 128 channels as presented in Figure 3.3. The supervised training loss is a weighted mean of the mean square errors for the reconstruction of the SSH, its gradient and its Laplacian. The default scale-space decomposition used is made of twenty 8 kilometers band. The calibration model is trained for 250 epochs with a annealing triangular cyclical learning rate [31].

Benchmarking experiments

We summarize our benchmarking experiments in Table 3.1. We compare our approach, referred to as CalCNN, with a standard UNet [23] architecture. The latter uses as inputs the gridded altimetry product and the uncalibrated KaRIn observation stacked together as a 2D field with 2 channels. We consider the same training configuration for this UNet model as for the CalCNN. As baseline, we also consider the reconstruction performance for the KaRIn SSH issued from the 4DVarNet method using nadir-altimeter-only data. We evaluate all methods according to the following two metrics, the root mean squared error (RMSE) of the SSH field, and the RMSE of the amplitude of the gradients of the SSH field. Whereas the UNet fails to produce a better estimate than the nadir-only interpolation baseline, our CalCNN improves the estimation of the

xp	RMSE (m)	RMSE $\|\|\nabla_{ssh}\|\|$
CalCNN	1.39e-02	6.46e-03
CalCNN w/o skip connection	2.17e-02	9.57e-03
CalCNN w/o gridded product	1.70e-01	2.47e-02
CalCNN w/o scale decomposition	2.17e-02	9.58e-03
CalCNN w/o mixing layer	1.94e-02	9.60e-03

Table 3.2 – **Ablation results**

SSH and its gradient by over 35% and brings the residual error below 1.4cm (Table 3.1).

In Figure 3.5, we further decompose the calibration error of the CalCNN w.r.t. the spatial scale using 1-dimensional Gaussian blurs as introduced in 3.4. We draw a comparison with the 4DVarNet interpolation baseline and observation errors. The CalCNN reaches a lower error than both KaRIn observations and the interpolation baseline across all scales. At larger scales the error gets closer to the latter as instrument errors dominate the large-scale components of KaRIn observations. Interestingly, at scales lower than 10km, we still retrieve some improvement even though the observation error is quite high. This can be explained by the fact that the high frequency errors on the KaRIn observations is easily separable from the underlying SSH signal. Between 10-100km, our method successfully exploits the lower observation errors to improve the interpolation baseline.

Ablation Study

In this section we analyse further the contribution of the different components of our neural architecture. In Table 3.2, we report the performance metrics of the considered ablation study with the following models: a model without skip connections, one without a gridded product as input, one without the scale-space decomposition scheme (Sec. 3.4) and one without the swath-mixer layers (Sec.3.4. Overall, these four models lead to a significantly lower performance. The largest impact comes from the ablation of the nadir-altimetry-only gridded product which provides large-scale information about the SSH. It leads to a loss which amount to an order of magnitude in the calibration errors. Moreover, we can see that without the skip connections or scale decomposition, we fail to improve on the L4 gridded product. Finally, we can note that we still get a 10% reduction of the RMSE w.r.t the L4 product without the mixing

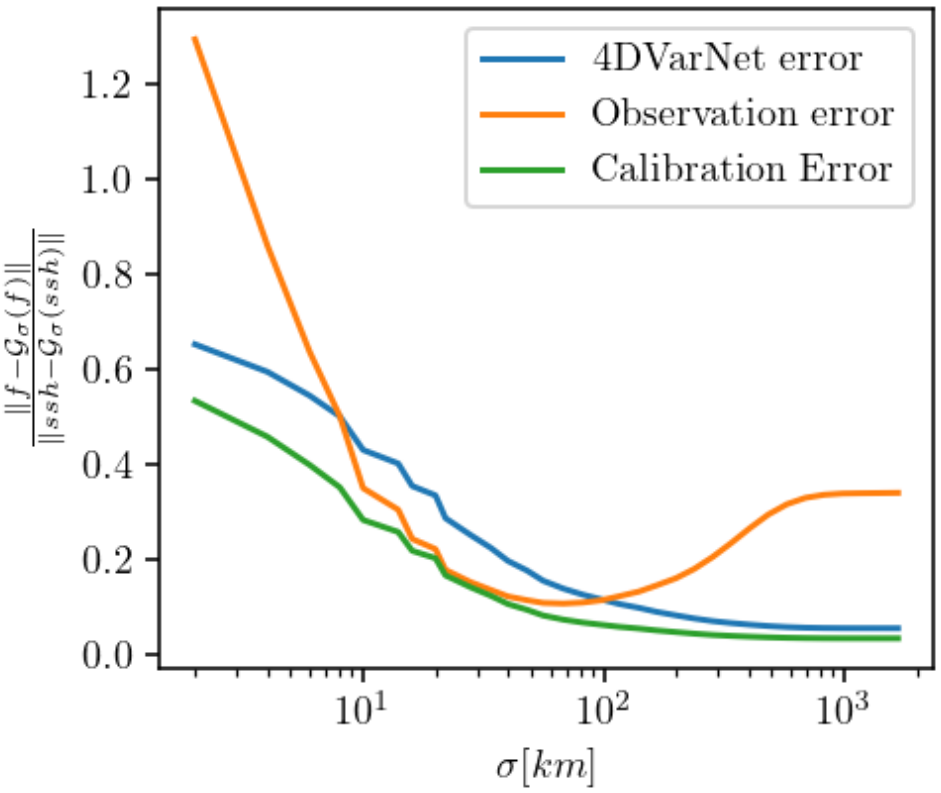


Figure 3.5 – **Observation and reconstruction error for the SSH at different spatial scales:** The figure shows the relative error w.r.t to the SSH at different along-track scales for the inputs (Uncalibrated KaRIn in orange and nadir based interpolation in blue) and output (calibrated KaRIn in green) of our method. The x axis indicates the standard deviation of the Gaussian blur that was used to remove the high scale components of the different signals. We can see the expected trend of the interpolation error that is concentrated at fine scales. The uncalibrated KaRIn error on the other hand is lower than the interpolation only in the 10km-100km range. We see the calibrated output of our method achieves lower error across all scales.

layer, however sharing the information between each side of the swath improves this gain three fold.

In Table 3.3, we show the sensitivity to the size of the network for the same training configuration. We compare the base architecture 3x128 (3 convolution blocks with 128 channels) with a linear operator, as well as a smaller network 1x32 and a bigger one 5x512. The linear version fails to extract geophysical information from the uncalibrated information. This further points out how challenging the considered calibration task is. Interestingly, our architecture

xp	RMSE (m)	RMSE $\|\|\nabla_{ssh}\|\|$
128x3 (Ref)	1.39e-02	6.46e-03
Linear	2.13e-02	1.02e-02
32x1	1.44e-02	6.22e-03
512x5	1.49e-02	7.19e-03

Table 3.3 – **Impact of network size**

leads to a similar performance for different complexity levels. The smaller and larger architectures leads to a slight increase in the residual error but the smaller model shows a slight improvement in the gradient reconstruction and spatial resolution. Overall, these results support the robustness of the proposed learning-based approaches and the conclusions we raise in section 3.5 are not very sensitive to the hyper-parameters of our network architecture.

Gridded product sensitivity

We analyze further how the quality of nadir-altimetry-only gridded product affects the calibration performance. In Figure 3.6, we display the improvement in the RMSE of the SSH on the swath and of the gradients of the SSH obtained by our CalCNN for the three gridded products introduced in Sec. 3.2.

For all three interpolated products, the proposed calibration method improves the reconstruction of the SSH for the KaRIn swath from the joint analysis of the interpolation product and raw KaRIn observations. We report the larger improvement for DUACS product. This relates to the spectral overlap between the SSH information of the uncalibrated KaRIn and SWOT's NADIR. The associated calibration performance remains however significantly worse than that of the two 4DVarNet products, which may relate to the worse interpolation performance of DUACS product [9, 10]. When comparing the impact of the two 4DVarNet products, the results are more nuanced. The 4DVarNet-SST product leads to better metrics. The difference of RMSE is greatly reduced after calibraFKation whereas the gap in RMSE of the gradients is conserved. This could be interpreted as the gain of RMSE we get from using the SST can be obtained from the uncalibrated KaRIn. However some of the gradients we reconstruct through the SST are not easily extracted from the observations. Overall this shows interesting relations between the redundant information in the uncalibrated KaRIn and the interpolated products.

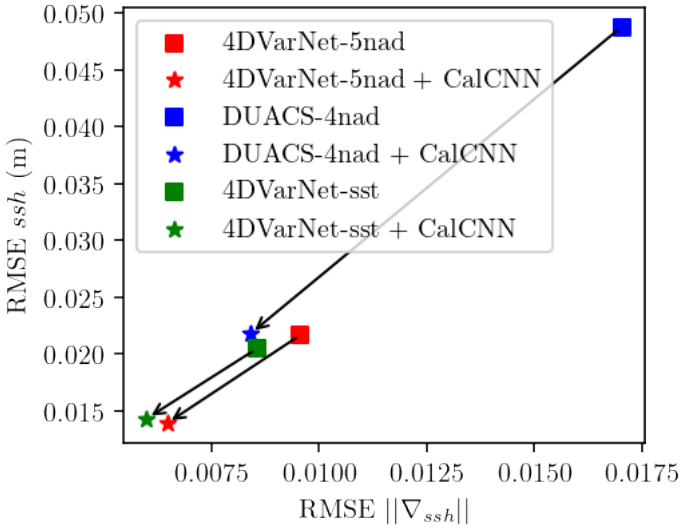


Figure 3.6 – **Impact of the nadir-based gridded product on the CalCNN output:** The figure shows the RMSE and the RMSE of the $||\nabla_{ssh}||$ of the calibrated observation (stars) and their associated nadir-based gridded products (squares). The improvement brought by the CalCNN is illustrated by the arrows. This improvement can be interpreted as the relevant information extracted from the uncalibrated KaRIn by the CalCNN. Note that the biggest relative improvement concerns the DUACS gridded product (blue) which doesn't uses the SWOT's nadir altimeter.

N_{band}	δ_{band}	RMSE (m)	RMSE $\|\|\nabla_{ssh}\|\|$
20	8	1.39e-02	6.46e-03
40	4	1.44e-02	6.64e-03
10	16	1.48e-02	6.75e-03
5	32	1.41e-02	6.65e-03
10	8	1.56e-02	6.81e-03
40	8	1.54e-02	6.88e-03

Table 3.4 – **Calibration metrics in function of the scale decomposition**

Sensitivity to the scale-space decomposition

In Table 3.4, we display the calibration metrics for different scale-space decompositions. We vary the number of scales considered and the spacing between two consecutive scales. When considering the same scale range from 8km to 160km, we retrieve the best performance with 20 scales. But, even with only 5 scales evenly separated by 32 km, the performance decreases only by 3%. By contrast, when considering a scale separation of 8km but varying the number of scales, we note a more significant drop of performance (about 10% in the residual RMSE). This suggests a greater sensitivity to the span of the scale-space decomposition than to the number and spacing of the components. However we still achieve less than 1.6cm residual error for any of the considered variations which is still a competitive calibration outcome.

3.6 Conclusion

We have proposed in this chapter a neural calibration approach which combines a scale-space decomposition of KaRIn observations and a convolutional architecture. This approach proves to be robust with a residual error below 1.5cm which can be compared with the 2cm residual error of the expected operational approaches performance although demonstrated globally using a different ocean simulation [4]. While we can reach a satisfactory calibration performance using the operational nadir altimetry mapping product, our experiments highlight the potential benefit of ongoing effort on neural SSH interpolation schemes to further improve the retrieval of finer-scale features from KaRIn observations. This naturally advocates for future work exploring jointly calibration and mapping problems for nadir and wide-swath altime-

ters, possibly combining our deep learning approach and variational mapping formulations introduced in [32].

This chapter confirmed the potential of deep learning models as alternatives for altimetry data analysis. To further validate this potential in real-world scenarios, we aim to evaluate deep learning methodologies using actual data. At the time of this study, SWOT data was not yet available. Nevertheless, we can assess the mapping of altimetry tracks using real data which is the core of the next chapter.

BIBLIOGRAPHY

[1] Q. Febvre, et al., "Scale-aware neural calibration for wide swath altimetry observations," arXiv:2302.04497, Sep 2023.

[2] E. Peral et al., "Swot mission performance and error budget," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, Jul 2018, p. 8625–8628.

[3] C. Ubelmann, et al., "SWOT Simulator documentation," .

[4] G. Dibarboure, et al., "Data-driven calibration algorithm and pre-launch performance simulations for the swot mission," *Remote Sensing*, vol. 14, no. 2323, pp. 6070, Jan 2022.

[5] A. Witkin, "Scale-space filtering: A new approach to multi-scale description," in *ICASSP '84. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Mar 1984, vol. 9, p. 150–153.

[6] G. Taburet, et al., "DUACS DT2018: 25 years of reprocessed sea level altimetry products," *Ocean Science*, vol. 15, no. 5, pp. 1207–1224, 2019.

[7] J-M. Lellouche, et al., "The copernicus global 1/12° oceanic and sea ice glorys12 reanalysis," *Frontiers in Earth Science*, vol. 9, pp. 585, 2021.

[8] M. Beauchamp, et al., "Intercomparison of Data-Driven and Learning-Based Interpolations of Along-Track Nadir and Wide-Swath SWOT Altimetry Observations," *Remote Sensing*, vol. 12, no. 22, 2020.

[9] R. Fablet, et al., "END-TO-END PHYSICS-INFORMED REPRESENTATION LEARNING FOR SATELLITE OCEAN REMOTE SENSING DATA: APPLICATIONS TO SATELLITE ALTIMETRY AND SEA SURFACE CURRENTS," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-3-2021, pp. 295–302, 2021.

[10] M. Ballarota, et al., "ocean-data-challenges/2020a_ssh_mapping_natl60: Material for ssh mapping data challenge," Sep 2020.

[11] T. Lindeberg, "Edge detection and ridge detection with automatic scale selection," in *IEEE CVPR*. IEEE, 1996, pp. 465–470.

[12] T. Lindeberg, "Image matching using generalized scale-space interest points," *Journal of Mathematical Imaging and Vision*, vol. 52, no. 1, pp. 3–36, May 2015.

[13] S. L. Pintea, et al., "Resolution learning in deep convolutional networks using scale-space theory," *IEEE Transactions on Image Processing*, vol. 30, pp. 8342–8353, Jan 2021.

[14] D. Worrall et al., "Deep scale-spaces: Equivariance over scale," in *Advances in Neural Information Processing Systems*. 2019, vol. 32, Curran Associates, Inc.

[15] Y. Lecun, et al., "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov 1998.

[16] K. He, et al., "Deep residual learning for image recognition," 2016, p. 770–778.

[17] G. Liu, et al., "Image inpainting for irregular holes using partial convolutions," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 85–100.

[18] J. Redmon, et al., "You only look once: Unified, real-time object detection," 2016, p. 779–788.

[19] A. Colin, et al., "Segmentation of sentinel-1 sar images over the ocean, preliminary methods and assessments," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, Jul 2021, p. 4067–4070.

[20] A. Colin, et al., "Semantic segmentation of metoceanic processes using sar observations and deep learning," *Remote Sensing*, vol. 14, no. 44, pp. 851, Jan 2022.

[21] R. Fablet, et al., "Joint Interpolation and Representation Learning for Irregularly Sampled Satellite-Derived Geophysical Fields," *Frontiers in Applied Mathematics and Statistics*, vol. 7, 2021.

[22] X. Li, et al., "A Convolutional Neural Network-Based Relative Radiometric Calibration Method," *IEEE Transactions on Geoscience and Remote*

Sensing, vol. 60, pp. 1–11, 2022, Conference Name: IEEE Transactions on Geoscience and Remote Sensing.

[23] O. Ronneberger, et al., "U-net: Convolutional networks for biomedical image segmentation," , no. arXiv:1505.04597, May 2015, arXiv:1505.04597 [cs].

[24] A. Ajayi, et al., "Spatial and Temporal Variability of the North Atlantic Eddy Field From Two Kilometric-Resolution Ocean Models," *Journal of Geophysical Research: Oceans*, 125, 5. e2019JC015827, 2020.

[25] A. Carrassi, et al., "Data assimilation in the geosciences: An overview of methods, issues, and perspectives," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 9, no. 5, pp. e535, Sept. 2018, Publisher: Wiley.

[26] M. Beauchamp, et al., "4dvarnet-ssh: end-to-end learning of variational interpolation schemes for nadir and wide-swath satellite altimetry," *Geoscientific Model Development Discussions*, vol. 2022, pp. 1–37, 2022.

[27] R. Fablet, et al., "Multimodal 4dvarnets for the reconstruction of sea surface dynamics from sst-ssh synergies," , arXiv:2207.01372, Jul 2022.

[28] V. Nair et al., "Rectified linear units improve restricted boltzmann machines," Jul 2019.

[29] S. Ioffe et al., "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning*. Jun 2015, p. 448–456, PMLR.

[30] I. O. Tolstikhin, et al., "Mlp-mixer: An all-mlp architecture for vision," in *Advances in Neural Information Processing Systems*. 2021, vol. 34, p. 24261–24272, Curran Associates, Inc.

[31] L. N. Smith, "Cyclical learning rates for training neural networks," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Mar 2017, p. 464–472.

[32] Q. Febvre, et al., "Joint calibration and mapping of satellite altimetry data using trainable variational models," in *IEEE ICASSP 2022 - 2022*, May 2022, p. 1536–1540.

Chapter 4

Training neural mapping schemes for satellite altimetry with simulation data

This chapter is based on an journal publication with the same name ongoing review. The preprint is available here [1]

4.1 Introduction

The retrieval of mesoscale-to-submesoscale sea surface dynamics for horizontal scales smaller than 150 km is a challenge for operational systems based on optimal interpolation [2] and data assimilation [3] schemes. This has motivated a wealth of research to develop novel mapping schemes [4, 5, 6].

In this context, data-driven and learning-based approaches [7, 8, 9, 10, 11] appear as appealing alternatives to make the most of the available observation and simulation datasets. Especially, Observing System Simulation Experiments (OSSE) have stressed the potential of neural schemes trained through supervised learning for the mapping of satellite-derived altimetry data [10, 12]. Their applicability to real datasets has yet to be assessed and recent studies have rather explored learning strategies from real gappy multi-year altimetry datasets [11]. The scarce and irregular sampling of the nadir measurements presents a challenge for training deep neural networks directly on observation data.Despite promising results, schemes trained with unsupervised strategies do not reach the relative improvement of the operational processing suggested by OSSE-based studies.

Among the existing methods, the 4dVarNet neural mapping scheme has demonstrated state-of-the-art performance when evaluated on simulated data.

Consequently, it serves as a compelling case study for assessing the transferability of neural schemes from simulated to real data. This chapter explores a specific strategy for applying these neural schemes to real data. Namely, we go beyond using OSSEs as benchmarking-only testbeds. We explore their use for the training of neural mapping schemes and address the space-time interpolation of real satellite altimetry observations. Through numerical experiments on a Gulf Stream case-study with a 5-nadir altimeter constellation, our main contributions are three-fold.

— We demonstrate the relevance of the simulation-based learning of neural mapping schemes and their generalization performance for real nadir altimetry data.

— We benchmark the proposed approach with state-of-the-art operational products as well as neural schemes trained from real altimetry datasets.

— We also assess how the characteristics of the training datasets, especially in terms of resolved ocean processes, drives the mapping performance.

The content of this chapter is organized as follows. Section 4.2 offers background information on related work, Section 4.3 presents our method, Section 4.4 reports our numerical experiments, and Section 4.5 elaborates on our main contributions.

4.2 Background

Gridded satellite altimetry products

The ability to produce gridded maps from scattered along-track nadir altimeter measurements of sea surface height is key to the exploitation of altimeter data in operational services and science studies [13]. As detailed below, we can distinguish three categories of approaches to produce such maps: reanalysis products [3] using data assimilation schemes, observation-based products [2] and learning-based approaches [10].

Reanalysis products such as the GLORYS12 reanalysis [3] leverage the full expressiveness of state-of-the-art ocean models. They aim at retrieving ocean state trajectories close to observed quantities through data assimilation methods including among others Kalman filters and variational schemes [14]. Such reanalyses usually exploit satellite-derived and in situ data sources. For instance, GLORYS12 reanalysis assimilates satellite altimetry data, but also satellite-derived observations of the sea surface temperature, sea-ice concentration as well as in situ ARGO data [15].

The second category involves observation-based products. In contrast to reanalyses, they only rely on altimetry data and address a space-time interpolation problem. They usually rely on simplifying assumptions on sea surface dynamics. In this category, optimal-interpolation-based product DUACS (Data Unification and Altimeter Combination System) [2] exploits a covariance-based prior, while recent studies involve quasi-geostrophic dynamics to guide the interpolation scheme [6, 4].

Data-driven and learning-based approaches form a third category of SSH mapping schemes. Similarly to observation-based methods, they are framed as interpolation schemes. Especially deep learning schemes have gained some attention. Recent studies have explored different neural architectures both for real and OSSE altimetry datasets [16, 17, 11]. These studies investigate both different training strategies as well as different neural architectures from off-the-shelf computer vision ones such as convolutional LSTMs and UNets [18] to data-assimilation-inspired ones [17, 19].

Ocean Modeling and OSSE

Advances in modeling and simulating ocean physics have largely contributed to a better understanding of the processes involved in the earth system and to the development of operational oceanography [20, 21]. High-resolution simulations used in Observing System Simulation Experiments (OSSE) also provide a great test-bed for the design and evaluation of new of ocean observation systems [22]. The availability of numerical model outputs enables the computation of interpretable metrics directly on the quantities of interest. This avoids challenges met when working solely with observation data that may be incomplete, noisy or indirectly related to the desired quantity. For example, in the case of the recently launched SWOT mission, OSSEs combined ocean and instrument simulations to address calibration issues and interpolation performance for SWOT altimetry data [23]. Such OSSEs have also promoted novel developments for the interpolation of satellite altimetry such as the BFN-QG and 4DVarNet schemes [6, 12].

In OSSE settings, we can train learning-based mapping schemes in a supervised manner using model outputs as the "ground truth" during the training phase. Nonetheless, these training methods cannot be straightforwardly applied to Observing System Experiments (OSEs) due to a lack of comprehensive groundtruthed observation datasets. Applied machine learning practitioners often grapple with insufficient amount of labelled data during the training of

supervised learning schemes, as the collection of large annotated datasets for a specific task can be costly or unattainable. Proposed solutions includes the exploitation of large existing datasets (such as ImageNet [24]) to train general purpose models [25]. Another approach involves the generation of synthetic datasets to facilitate the creation of groundtruthed samples [26, 27]. OSSEs, which combine ocean model outputs and observing system simulators [28], can deliver such large synthetic groundtruthed datasets. We propose to investigate how OSSE-based training strategies apply to the analysis of real satellite altimetry datasets. Recent results of SSH super-resolution model trained on simulation datasets and evaluated on real ones [29] support the relevance of such strategies.

Physics-aware deep-learning

In the last decades, DL advances combined with the rise in computational resources and amount of data have shown the power of extracting knowledge from data in domains ranging from computer vision to language processing [30]. Yet, despite to the universality of DL architectures [31], a central challenge persists in learning from data: the generalization performance beyond the distribution of the training data. To tackle this problem, the literature includes a variety of strategies such as data augmentation [32] and regularization techniques, including dropout layers [33] and weight decay schemes [34]. This is of critical importance for physical systems, where models trained on past data will be challenged when the system evolves and reaches dynamics absent from the training data. We can see evidence of this shortcoming in the instability challenges faced by neural closures for climate models [35].

There have been a variety of approaches to harness physical priors within learning schemes to address this issue. Some injects trainable components in classical integration schemes of physical models [36], others leverage physical priors within their learning setups which can been used in the training objective [37, 38], as well as in the architecture [39, 40]. However most of these works have focused on relatively simple physical models and it remains challenging to combine current state-of-the-art ocean models with such methods. Obstacles include the complexity and cost of running the physical models, the differences in programming tools and the computing infrastructures used in each domain, as well as the availability of automatic differentiation tools for state-of-the-art ocean models.

The proposed simulation-based training strategy offers another way to

benefit from the advances in high-resolution ocean modeling in the design of deep neural models for ocean reanalysis problems.

4.3 Method

Overview

We designate our approach as "simulation-based", it consists in leveraging ocean models and simulations of observing systems to design supervised training environments. In this section, we describe the proposed method for assessing the potential of simulation-based neural schemes for the mapping real altimetry tracks. We describe the architecture considered in our study, as well as the different datasets used for training purposes. We also detail our simulation-based training setup and the proposed evaluation framework on real altimetry.

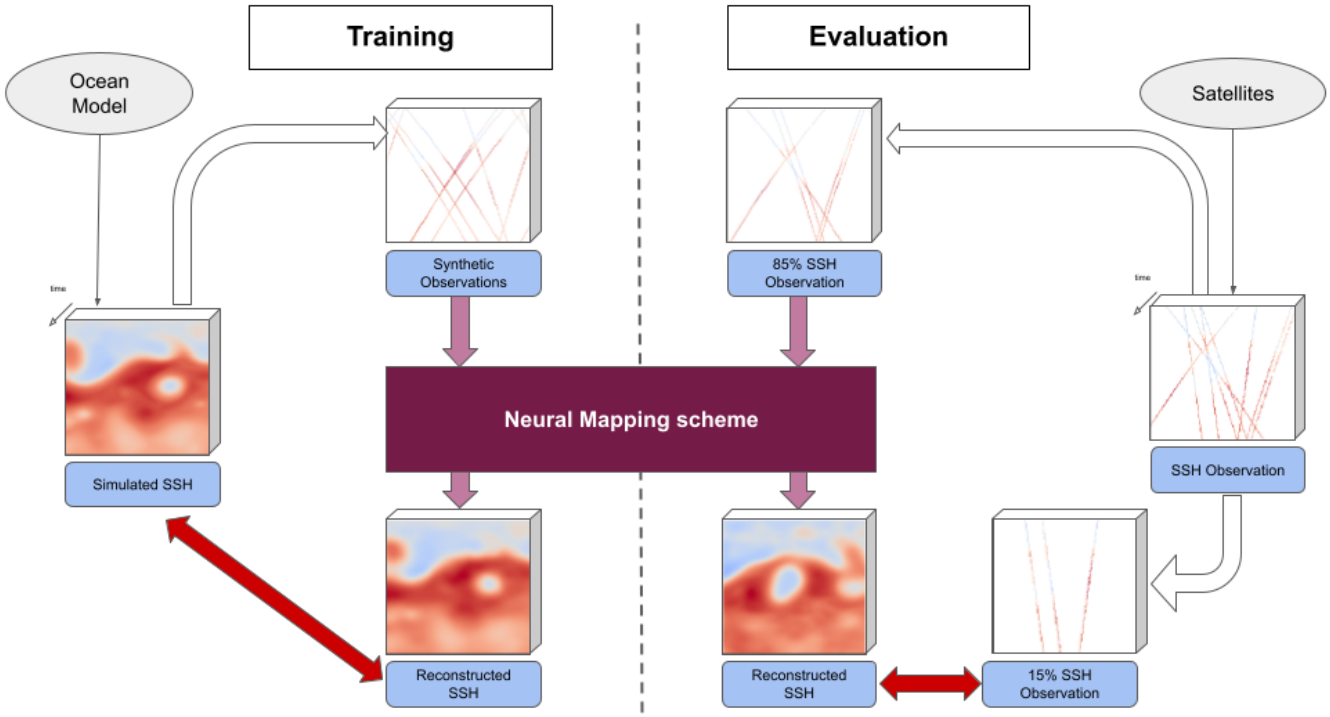


Figure 4.1 – **Overview of the experimental setup**. On the left side we display the simulation-based training strategy based on an ocean simulation which will be used for 1) generating synthetic observation and 2) computing the training objective of the neural mapping scheme. On the right side we show the evaluation principle of splitting the available satellite observations to evaluate the method on data that were not used for the inference.

Neural mapping scheme

The neural mapping scheme considered for this study is the 4DVarNet framework[10]. We choose this scheme due to the performance shown in the OSSE setup. Previous studies [12] report significant better performance than the DUACS product [2] in the targeted Gulf stream region. 4DVarNet relies on a variational data assimilation formulation. The reconstruction results from the minimization of a variational cost. This cost encapsulates a data fidelity term and a regularization term. It exploits a prior on the space-time dynamics through a convolutional neural network inspired from [41], and an iterative gradient-based minimization based on a recurrent neural network as introduced for meta-learning purposes [42]. The overall architecture and components are similar to those presented in existing work [12]. We adapt some implementation details based on cross-validation experiments to improve the performance and reduce the training time. We refer the reader to the code for more details [43].

SSH Data

		Resolution	Reanalysis	Tide	DAC
NATL60	[21]	1/60°	No	No	No
eNATL60-t	[44]	1/60°	No	Yes	Yes
eNATL60-0	[44]	1/60°	No	No	Yes
GLORYS12-r	[3]	1/12°	Yes	No	No
GLORYS12-f	[3]	1/12°	No	No	No
ORCA025	[20]	1/4°	No	No	No

Table 4.1 – **Summary table of the different synthetic SSH fields used for training**. The last column indicate whether the Dynamic Atmospheric Correction was applied on the synthetic SSH. It justify the presence of both eNATL60-0 and NATL60 to isolate the impacts of resolution and tide.

We use numerical simulations of ocean general circulation models (OGCM) to build our reference SSH datasets. Such simulations involve a multitude of decisions that affect the resulting simulated SSH. Here we consider NEMO (Nucleus for European Modelling of the Ocean) [45] which is among the state-of-the art OGCM in operational oceanography [21] as well as in climate studies [46]. The selected SSH datasets reported in Table 4.1 focus on three main aspects: the added-value of high-resolution eddy-rich simulations, the impact of reanalysis datasets and the relevance of tide-resolving simulations.

In order to evaluate the impact of eddy-rich simulations, we consider

NATL60, GLORYS12-f and ORCA025 free runs, respectively with a horizontal grid resolution of 1/60°, 1/12°, and 1/4°. Finer grids allow for more processes to be simulated. We therefore expect higher-resolution simulations to exhibit structures closer to the real ocean and the associated trained deep learning model to perform better. Regarding the impact of reanalysis data, we compare numerical experiments with the GLORYS12-r reanalysis and the associated free run GLORYS12-f. This reanalysis dataset relies on the assimilation of temperature, sea level and sea ice concentration observations. Besides, the recent eNATL60 twin simulations eNATL60-t and eNATL60-0 allow us to evaluate the impact of tide-resolving simulations. We summarize in Table 4.1 the characteristics of the different datasets.

OSSE-based training setup

We sketch the proposed OSSE-based training setup on the left side of the Figure 4.1. In order to fairly evaluate the datasets' quality as a training resource, we standardize the training procedure. We regrid all simulations to the same resolution (1/20°) and we use daily-averaged SSH fields as training targets. We generate noise-free pseudo-observations by sampling values of the daily-averaged fields corresponding to realistic orbits of a 5 altimeter-constellation. We train all models from a one-year dataset in a Gulfstream domain from (66°W, 32°N) to (54°W, 44°N) in which we keep the same two months for validation. The hyper-parameters of the model and training procedure such as the number of epoch, learning rate scheduler are the same for all the experiments. The detailed configuration can be found by the reader in the available implementation. As training objective, we combine the mean square errors for the SSH fields and the amplitude of the gradients as well as a regularization loss for the prior model.

OSE-based evaluation setup

As sketched on the right side of the Figure 4.1, the evaluation setup relies on real altimetry data from the constellation of 6 satellites from 2017 (SARAL/Altika, Jason 2, Jason 3, Sentinel 3A, Haiyang-2A and Cryosat-2). We apply the standardized setup presented in a data-challenge `https://github.com/ocean-data-challenges/2021a_SSH_mapping_OSE`. We use the data from the first five satellites as inputs for the mapping and the last one (Cryosat-2) for computing the performance metrics. We compute these metrics in the along-track geometry. The evaluation domain spans from (65°W, 33°N)

to (55°W, 43°N) and the evaluation period from January 1^{st} to December 31^{st} 2017. Given η_{c2} and $\hat{\eta}$ the measured SSH and the reconstructed SSH respectively, we compute the following two metrics:

— μ_{ssh} is a score based on the normalized root mean squared (nRMSE) error computed as $1 - \frac{RMS(\hat{\eta} - \eta_{c2})}{RMS(\eta_{c2})}$

— λ_x is the wavelength at which the power spectrum density (PSD) score $1 - \frac{PSD(\hat{\eta} - \eta_{c2})}{PSD(\eta_{c2})}$ crosses the 0.5 threshold, which characterize the scales resolved by the reconstruction (the error below that wavelength makes up for more than half of the total signal)

In Table 4.3, we also consider the root mean square error (RMSE) as well as the nRMSE score of the sea level anomaly μ_{sla} obtained by subtracting the mean dynamic topography to the SSH. Lastly, we assess the performance degradation resulting from the transition from simulated to real data by quantifying the improvement relative to DUACS in the resolved scale λ_x on our OSE setup as well as on the OSSE benchmarking setup proposed in related studies [6]. This benchmarking setup relies on NATL60-CJM165 OSSE dataset. We refer the reader to `https://github.com/ocean-data-challenges/2020a_SSH_mapping_NATL60` for a detailed description of this experimental setup.

4.4 Results

This section details our numerical experiments for the considered real altimetry case-study for a Gulf Stream region as described in Section 4.3. We first report the benchmarking experiments to assess the performance of the proposed learning-based strategy with respect to (w.r.t.) state-of-the-art mapping schemes. We then analyse how the characteristics of the training datasets drive the mapping performance.

Benchmarking against the state of the art

We report in Table 4.2 the performance metrics of state-of-the-art approaches including both operational observation products [2, 5], deep-learning-based schemes trained on observation data [16, 11] as well as methods using explicitly a model-based prior on sea surface dynamics [6, 4, 3]. We compare those methods with a 4DVarNet trained on eNATL60-0 OSSE dataset. The latter outperforms all other methods on the two metrics considered (22% improvement in RMSE w.r.t. the DUACS product and 33% improvement in the

resolved scale). We report a significantly worse performance for GLORYS12 reanalysis. This illustrates the challenge of combining large ocean general circulation models and observation data for the mapping of the SSH.

The last column indicates that the 4DVarNet scheme leads to the best mapping scores for both the OSE and OSSE setups. For the latter, the reported improvement of 47% is twice greater than the second best at 22%. The performance of the 4DVarNet drops by 11% when considering the former. By contrast, other methods do not show such differences between the OSE and OSSE case-studies. This suggests that the finer-scale structures that are well reconstructed in the OSSE setup are not as beneficial in the OSE setup. While one could question the representativeness of the OSSE datasets for the fine-scale patterns in the true ocean, real nadir altimetry data may also involve multiple processes which could impede the reconstruction and evaluation of horizontal scales below 100km.

	SSH Only	Deep Learning	Calibrated on data from	Physical Model	rmse (cm)	μ_{ssh} ()	λ_x (km)	$1-\frac{\lambda_x}{\lambda_{ref}}$ (% ose, osse)
(a) **4DVarNet**	Yes	Yes	Simulation	–	**5.9**	**0.91**	**100**	**33**, **47**
(b) MUSTI	No	Yes	Satellite	–	6.3	0.90	112	26, 22
(c) ConvLstm-SST	No	Yes	Satellite	–	6.7	0.90	108	28, –
(d) ConvLstm	Yes	Yes	Satellite	–	7.2	0.89	113	25, –
(e) DYMOST	Yes	No	Satellite	QG	6.7	0.90	131	13, 11
(f) MIOST	Yes	No	Satellite	–	6.8	0.90	135	11, 10
(g) BFN-QG	Yes	No	Satellite	QG	7.6	0.89	122	19, 21
(h) DUACS	Yes	No	Satellite	–	7.7	0.88	151	0, 0
(i) GLORYS12	No	No	Satellite	NEMO	15.1	0.77	241	-60, –

Table 4.2 – **SSH reconstruction performance of the benchmarked methods** (a) 4DVarNet from this study trained on eNATL60-0 (b) Archambault et al. (2023), (c and d) ConvLstm-SST and ConvLstm from Martin et al. (2023), (e) DYMOST from Ballarotta et al. (2020), (f) MIOST from Ubelmann et al. (2021), (g) BFN-QG from Guillou et al. (2021), (h) DUACS from Taburet et al. (2019), (i) GLORYS12 from Lellouche et al. (2021. The columns indicate from left to right: whether athe mapping schemes rely only on SSH data or also exploit additional data such as gap free SST products; if the method uses deep learning architectures; the data used to calibrate (or train) the mapping scheme; the numerical model of the ocean used for the mapping if any (QG stands for quasi-geostrophic); μ and λ_x are the metrics as described in Section 4.3

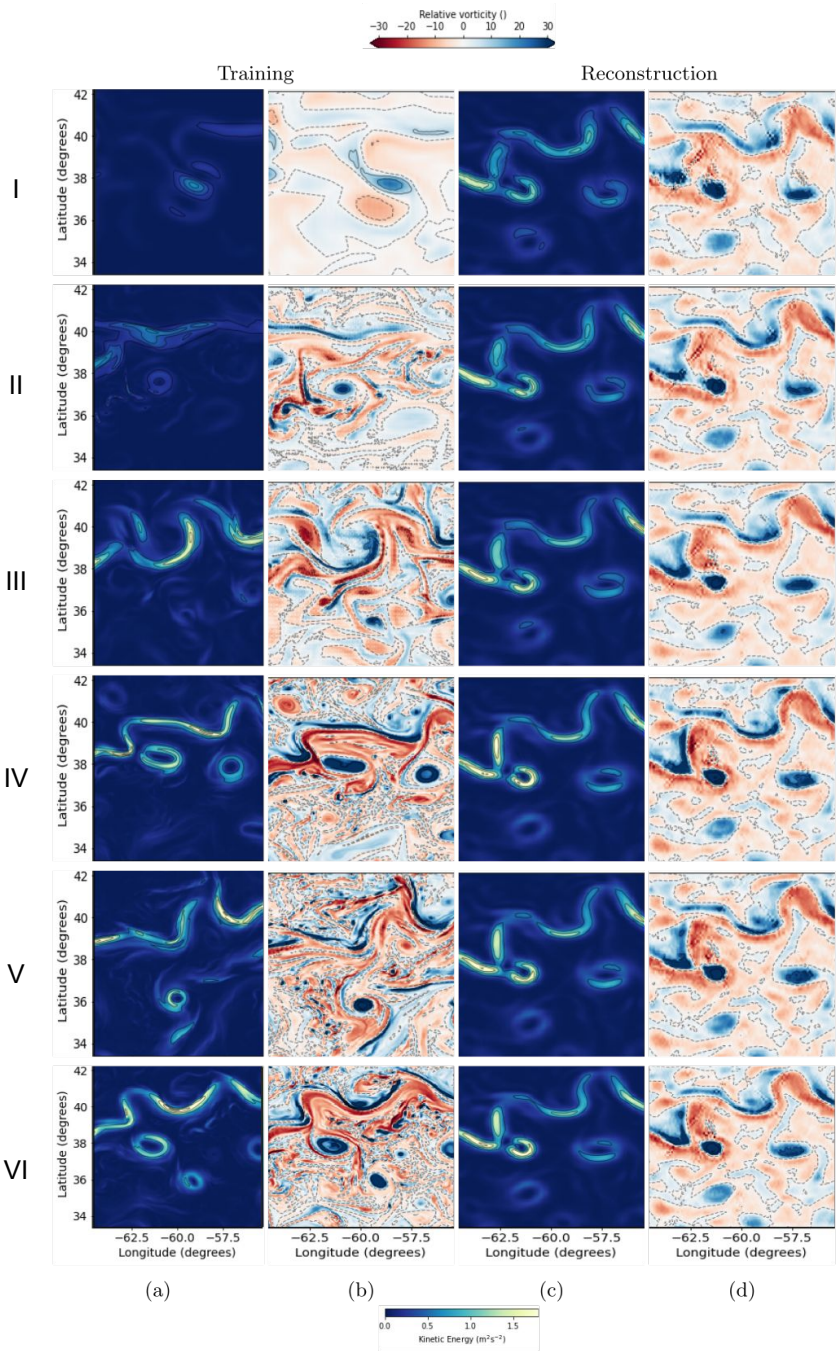


Figure 4.2 – **Kinetic energy and relative vorticity on January 6th of training and reconstruction data**. The first two columns (a) and (b) show the training data while columns (c) and (d) show the associated 4DVarNet reconstruction of the 2017 year. Columns ((a) and (c)) display the geostrophic kinetic energy while ((b) and (d)) display the relative vorticity normalized by the Coriolis parameter. Each row corresponds to the dataset: ORCA025 (I), GLORYS12-f (II), GLORYS12-r (III), NATL60 (IV), eNATL60-t (V) and eNATL60-0 (VI)

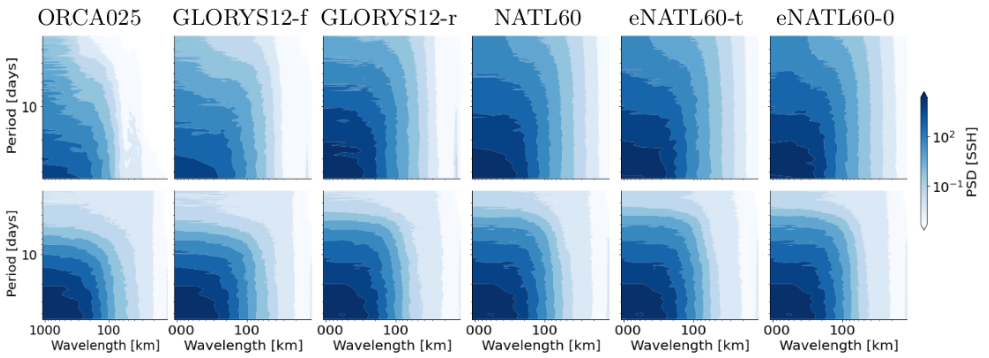


Figure 4.3 – **Space-time spectral densities of the training datasets (first row) and of their associated reconstruction (second row)**. Darker blue in the lower left corner indicates higher energy at larger wavelength and periods. The different SSH fields exhibit different energy cascades when moving to finer temporal (upward) or spatial (rightward) scales.

Training Data	RMSE (cm)	μ_{ssh}	μ_{sla}	λ_x (km)	$1-\frac{\lambda_x}{\lambda_{ref}}$ (% ose, osse)
NATL60	**5.9**	**0.91**	**0.80**	**98**	(**35**, –)
eNATL60-t	**5.9**	**0.91**	**0.80**	100	(33, **48**)
eNATL60-0	**5.9**	**0.91**	**0.80**	100	(33, 47)
GLORYS12-r	6.3	0.90	0.78	106	(30, 28)
GLORYS12-f	6.7	0.90	0.77	119	(21, 23)
ORCA025	7.1	0.89	0.76	126	(17, 17)

Table 4.3 – **Performance of 4DVarNet mapping schemes trained on different simulated datasets**. The first column shows the source of the training dataset as described in Table 4.1; the subsequent columns indicate the reconstruction metrics described in Section 4.3. Note that the NATL60 could not be evaluated on the OSSE setup since the evaluation data were used for validation during the training stage.

Eddy-present datasets versus eddy-rich ones

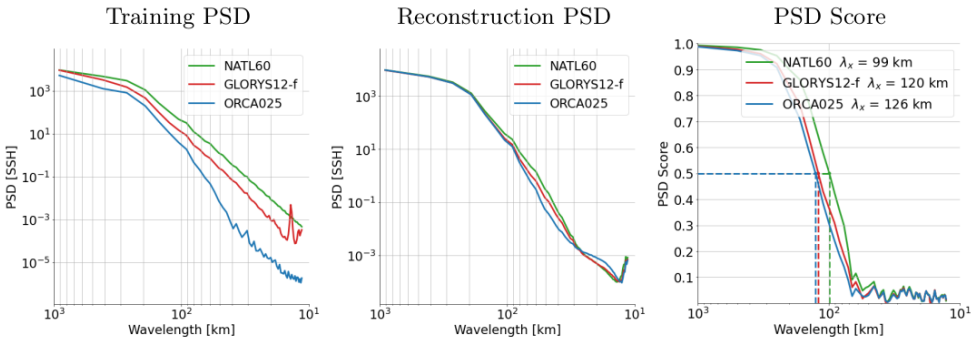


Figure 4.4 – **Spectral analysis of the training and reconstructed SSH datasets**. We display the PSD of the training dataset (left plot), reconstructed SSH field (center plot) as well as the associated PSD score (right plot)

We analyse here in more detail the impact of the spatial resolution of the training dataset onto the reconstruction performance. In Table 4.3, as expected, the higher resolution grid in the ocean run simulation leads to a better mapping with a 22% improvement in λ_x and a 17% improvement in the RMSE score between the experiments with the coarsest (ORCA025) and finest (NATL60) resolutions. We also observe qualitative differences in the relative vorticity fields in Figure 4.2. Residual artifacts due to the altimetry tracks appear (60°W, 39°N) for the two lower-resolution training datasets. They are greatly diminished when considering the NATL60 dataset. Despite these differences, the reconstructed vorticity and kinetic energy fields in Figure 4.2 look very similar for the different 4DVarNet schemes, whatever the training datasets. By contrast, the vorticity and kinetic energy fields in the training datasets clearly depict fewer fine-scale structures and weaker gradients for the lower-resolution simulation datasets, namely ORCA025 and GLORYS12-f. These results support the generalization skills of 4DVarNet schemes to map real altimetry tracks despite being trained on SSH sensibly different from the reconstruction.

We draw similar conclusions from the analysis of the spectral densities shown in Figure 4.4. The differences in the energy distribution of the training data significantly reduce in the reconstructions. 4DVarNet schemes trained from higher-resolution datasets however result in more faithful reconstruction at all scales. The patterns observed for the temporal PSD are slightly different in Figure 4.3. We do not observe the same homogenization as for the spatial PSD. Lower-resolution training datasets involve a significant drop of an order

of magnitude for periods greater than 10 days and wavelength greater than 200km.

Forced simulation datasets versus reanalysis ones

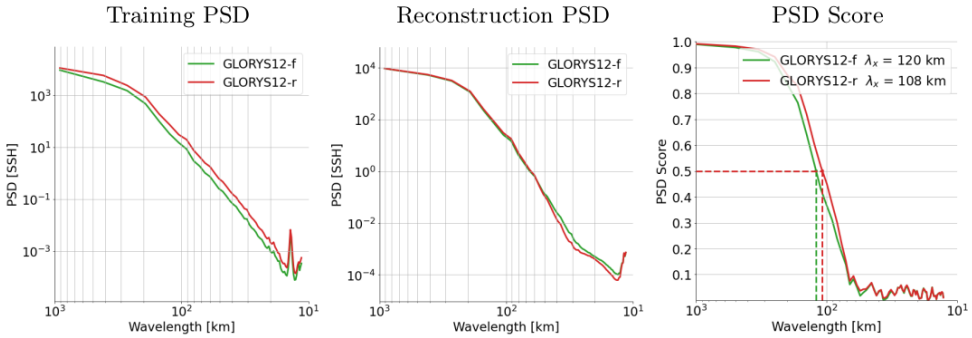


Figure 4.5 – **Spectral impact of model reanalysis**. We display the PSD of the training dataset (left plot), reconstructed SSH field (center plot) as well as the associated PSD score (right plot)

Looking in more specifically at the effect of ocean reanalysis between the two experiments GLORYS12-f and GLORYS12-r. We can first note the impact of observation data assimilation in Figure 4.3 where we see how the power spectrum of the reanalysis is significantly raised compared to the free run. The spectrum is closer to ones of the higher resolution simulations. Visually we also clearly see stronger gradients in the kinetic energy in Figure 4.2.

We can observe a similar behavior as in Section 4.4 in Figure 4.5 with the gap of in spectral density being diminished between the training and reconstruction data, and the PSD score indicating a lower energy of the error at all scales for the reanalysis-based experiment.

Quantitatively in Table 4.1 we see an improvement of 11% in both the RMSE and the scale resolved, besides training on a reanalysis increase the relative gain w.r.t. DUACS significantly more on real data (+9%) than on simulated data (+5%) as we can see in the right most column. This suggests that the reanalysis dataset conveys information on real world observations which improves the generalization performance.

Tide-free datasets versus tide-resolving ones

We assess here the impact of tide-resolving simulation used as training data. We use the twin eNATL60 runs eNATL60-t and eNATL60-0. Contrary

to other runs, those simulations contain barometric and wind forcing, we therefore remove the Dynamic Atmospheric Correction [47] from the SSH fields. Additionally since the barotropic tide signals are removed from real altimetry tracks prior to interpolation, we also remove the signal from the training data by subtracting the spatial mean over the training domain for each hourly snapshot before calculating the daily averages.

Given those processing steps, the two training datasets exhibit very similar wavenumber spectra as shown in Figures 4.3. We also find that training on those two datasets produce little differences in the reconstructions both quantitatively (see Table 4.3) and qualitatively (Fig. 4.2). The resulting performance is comparable to that of the NATL60 experiment.

We identify two hypotheses for explaining why tide-resolving simulation do not lead to better mapping schemes:

— The preprocessing applied on the training field remove the main tide signals. We therefore effectively measure the impact of tide modeling on other ocean processes that may be less significant;

— The evaluation procedure applied on altimetry tracks on which the barotropic tide has been filtered may not be interpretable enough to measure the reconstruction of residual tide signals. New instruments like the KaRIN deployed in the SWOT mission may provide new ways to better quantify those effects.

These findings provide motivation for carefully considering the purpose of the learning-based model when making decisions about the training data. In our case, explicitly modeling tide processes that are removed from the observations in the evaluation setup added overheads in the computational cost of running the simulation as well as in the preprocessing of the training data. Additionally given the considered evaluation data and metrics, we were not able to quantify any significant differences between the two trained mapping schemes.

4.5 Discussion

We have shown in this chapter that training machine learning models on simulations datasets leads good performance on real altimetry data mapping and outperforms current state of the art approaches. The model trained on NATL60 reduces the RMSE by 18% compared neural schemes trained on observation data and improves the scales resolved by 33% compared to the DUACS operational product. Even the coarsest simulation considered ORCA025

provides competitive results with current operational methods. We have shown that using a more realistic SSH fields using reanalysis or higher resolution simulations increases the performances of the trained model. This is an exciting result that shows the potential for training operational products from ocean simulations and how advances in ocean modeling in operational oceanography can be beneficial. The results shown here are limited to the interpolation problem on a regional domain but the robustness of the performance shown are encouraging for further developing these results using a larger domain.

This study has been greatly facilitated by the standardized tasks and evaluation setups proposed in data-challenges `https://ocean-data-challenges.github.io/`. Data-challenges are used to specify a targeted problem of interest to domain experts through datasets and relevant evaluation metrics. This preliminary work have been instrumental in constituting the comprehensive benchmark and combining methods from different teams and institution around the world. Additionally, it also constitutes a strong basis for a trans-disciplinary collaboration between the ocean and machine learning research communities.

The results presented in this study introduce a use of ocean simulations for developing altimetry products. This opens new ways for ocean physicist, modelers and operational oceanographers to collaborate. In order to assess the range of these new synergies, it would be interesting to explore if the approach proposed here of training neural schemes using simulation data would generalize to other tasks such as forecast or sensor calibration and to other quantities like surface temperature, currents, salinity or biochemical tracers.

The next chapter paves the way to facilitate the exploration of new application of learning based methods to ocean science questions.

BIBLIOGRAPHY

[1] Q. Febvre, J. L. Sommer, C. Ubelmann, and R. Fablet, "Training neural mapping schemes for satellite altimetry with simulation data," Sep. 2023.

[2] G. Taburet, A. Sanchez-Roman, M. Ballarotta, M.-I. Pujol, J.-F. Legeais, F. Fournier, Y. Faugere, and G. Dibarboure, "DUACS DT2018: 25 years of reprocessed sea level altimetry products," *Ocean Science*, vol. 15, no. 5, pp. 1207–1224, Sep. 2019.

[3] J. Lellouche, E. Greiner, R. Bourdallé-Badie, G. Garric, A. Melet, M. Drévillon, C. Bricaud, M. Hamon, L. G. Olivier, C. Regnier, T. Candela, C. Testut, F. Gasparin, G. Ruggiero, M. Benkiran, Y. Drillet, and P. Le Traon, "The Copernicus Global 1/12° Oceanic and Sea Ice GLORYS12 Reanalysis," *Frontiers in Earth Science*, vol. 9, 2021.

[4] M. Ballarotta, C. Ubelmann, M. Rogé, F. Fournier, Y. Faugère, G. Dibarboure, R. Morrow, and N. Picot, "Dynamic Mapping of Along-Track Ocean Altimetry: Performance from Real Observations," *Journal of Atmospheric and Oceanic Technology*, vol. 37, no. 9, pp. 1593–1601, Sep. 2020.

[5] C. Ubelmann, G. Dibarboure, L. Gaultier, A. Ponte, F. Ardhuin, M. Ballarotta, and Y. Faugère, "Reconstructing Ocean Surface Current Combining Altimetry and Future Spaceborne Doppler Data," *Journal of Geophysical Research: Oceans*, vol. 126, no. 3, p. e2020JC016560, 2021.

[6] F. L. Guillou, S. Metref, E. Cosme, C. Ubelmann, M. Ballarotta, J. L. Sommer, and J. Verron, "Mapping Altimetry in the Forthcoming SWOT Era by Back-and-Forth Nudging a One-Layer Quasigeostrophic Model," *Journal of Atmospheric and Oceanic Technology*, vol. 38, no. 4, pp. 697–710, Apr. 2021.

[7] A. Alvera Azcarate, A. Barth, M. Rixen, and J.-M. Beckers, "Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: Application to the Adriatic Sea surface temperature," *Ocean Modelling*, vol. 9, no. 4, 2005.

[8] A. Barth, A. Alvera-Azcárate, C. Troupin, and J.-M. Beckers, "DINCAE 2.0: Multivariate convolutional neural network with error estimates to reconstruct sea surface temperature satellite and altimetry observations," *Geoscientific Model Development*, vol. 15, no. 5, pp. 2183–2196, Mar. 2022.

[9] R. Lguensat, P. Tandeo, P. Ailliot, M. Pulido, and R. Fablet, "The Analog Data Assimilation," *Monthly Weather Review*, vol. 145, no. 10, pp. 4093–4107, Oct. 2017.

[10] R. Fablet, M. M. Amar, Q. Febvre, M. Beauchamp, and B. Chapron, "END-TO-END PHYSICS-INFORMED REPRESENTATION LEARNING FOR SATELLITE OCEAN REMOTE SENSING DATA: APPLICATIONS TO SATELLITE ALTIMETRY AND SEA SURFACE CURRENTS," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-3-2021, pp. 295–302, Jun. 2021.

[11] S. A. Martin, G. E. Manucharyan, and P. Klein, "Synthesizing Sea Surface Temperature and Satellite Altimetry Observations Using Deep Learning Improves the Accuracy and Resolution of Gridded Sea Surface Height Anomalies," *Journal of Advances in Modeling Earth Systems*, vol. 15, no. 5, p. e2022MS003589, 2023.

[12] M. Beauchamp, Q. Febvre, H. Georgenthum, and R. Fablet, "4DVarNet-SSH: End-to-end learning of variational interpolation schemes for nadir and wide-swath satellite altimetry," *Geoscientific Model Development*, vol. 16, no. 8, pp. 2119–2147, Apr. 2023.

[13] S. Abdalla, A. Abdeh Kolahchi, M. Ablain, S. Adusumilli, S. Aich Bhowmick, E. Alou-Font, L. Amarouche, O. B. Andersen, H. Antich, L. Aouf, B. Arbic, T. Armitage, S. Arnault, C. Artana, G. Aulicino, N. Ayoub, S. Badulin, S. Baker, C. Banks, L. Bao, S. Barbetta, B. Barceló-Llull, F. Barlier, S. Basu, P. Bauer-Gottwein, M. Becker, B. Beckley, N. Bellefond, T. Belonenko, M. Benkiran, T. Benkouider, R. Bennartz, J. Benveniste, N. Bercher, M. Berge-Nguyen, J. Bettencourt, F. Blarel, A. Blazquez, D. Blumstein, P. Bonnefond, F. Borde, J. Bouffard, F. Boy, J.-P. Boy, C. Brachet, P. Brasseur, A. Braun, L. Brocca, D. Brockley, L. Brodeau, S. Brown, S. Bruinsma, A. Bulczak, S. Buzzard, M. Cahill, S. Calmant, M. Calzas, S. Camici, M. Cancet, H. Capdeville, C. C. Carabajal, L. Carrere, A. Cazenave, E. P. Chassignet, P. Chauhan, S. Cherchali, T. Chereskin, C. Cheymol, D. Ciani, P. Cipollini, F. Cirillo,

E. Cosme, S. Coss, Y. Cotroneo, D. Cotton, A. Couhert, S. Coutin-Faye, J.-F. Crétaux, F. Cyr, F. d'Ovidio, J. Darrozes, C. David, N. Dayoub, D. De Staerke, X. Deng, S. Desai, J.-D. Desjonqueres, D. Dettmering, A. Di Bella, L. Díaz-Barroso, G. Dibarboure, H. B. Dieng, S. Dinardo, H. Dobslaw, G. Dodet, A. Doglioli, A. Domeneghetti, D. Donahue, S. Dong, C. Donlon, J. Dorandeu, C. Drezen, M. Drinkwater, Y. Du Penhoat, B. Dushaw, A. Egido, S. Erofeeva, P. Escudier, S. Esselborn, P. Exertier, R. Fablet, C. Falco, S. L. Farrell, Y. Faugere, P. Femenias, L. Fenoglio, J. Fernandes, J. G. Fernández, P. Ferrage, R. Ferrari, L. Fichen, P. Filippucci, S. Flampouris, S. Fleury, M. Fornari, R. Forsberg, F. Frappart, M.-l. Frery, P. Garcia, A. Garcia-Mondejar, J. Gaudelli, L. Gaultier, A. Getirana, F. Gibert, A. Gil, L. Gilbert, S. Gille, L. Giulicchi, J. Gómez-Enri, L. Gómez-Navarro, C. Gommenginger, L. Gourdeau, D. Griffin, A. Groh, A. Guerin, R. Guerrero, T. Guinle, P. Gupta, B. D. Gutknecht, M. Hamon, G. Han, D. Hauser, V. Helm, S. Hendricks, F. Hernandez, A. Hogg, M. Horwath, M. Idžanović, P. Janssen, E. Jeansou, Y. Jia, Y. Jia, L. Jiang, J. A. Johannessen, M. Kamachi, S. Karimova, K. Kelly, S. Y. Kim, R. King, C. M. M. Kittel, P. Klein, A. Klos, P. Knudsen, R. Koenig, A. Kostianoy, A. Kouraev, R. Kumar, S. Labroue, L. S. Lago, J. Lambin, L. Lasson, O. Laurain, R. Laxenaire, C. Lázaro, S. Le Gac, J. Le Sommer, P.-Y. Le Traon, S. Lebedev, F. Léger, B. Legresy, F. Lemoine, L. Lenain, E. Leuliette, M. Levy, J. Lillibridge, J. Liu, W. Llovel, F. Lyard, C. Macintosh, E. Makhoul Varona, C. Manfredi, F. Marin, E. Mason, C. Massari, C. Mavrocordatos, N. Maximenko, M. McMillan, T. Medina, A. Melet, M. Meloni, S. Mertikas, S. Metref, B. Meyssignac, J.-F. Minster, T. Moreau, D. Moreira, Y. Morel, R. Morrow, J. Moyard, S. Mulet, M. Naeije, R. S. Nerem, H. Ngodock, K. Nielsen, J. E. Ø. Nilsen, F. Niño, C. Nogueira Loddo, C. Noûs, E. Obligis, I. Otosaka, M. Otten, B. Oztunali Ozbahceci, R. P. Raj, R. Paiva, G. Paniagua, F. Paolo, A. Paris, A. Pascual, M. Passaro, S. Paul, T. Pavelsky, C. Pearson, T. Penduff, F. Peng, F. Perosanz, N. Picot, F. Piras, V. Poggiali, É. Poirier, S. Ponce de León, S. Prants, C. Prigent, C. Provost, M.-I. Pujol, B. Qiu, Y. Quilfen, A. Rami, R. K. Raney, M. Raynal, E. Remy, F. Rémy, M. Restano, A. Richardson, D. Richardson, R. Ricker, M. Ricko, E. Rinne, S. K. Rose, V. Rosmorduc, S. Rudenko, S. Ruiz, B. J. Ryan, C. Salaün, A. Sanchez-Roman, L. Sandberg Sørensen, D. Sandwell, M. Saraceno, M. Scagliola, P. Schaeffer, M. G. Scharffenberg, R. Scharroo,

A. Schiller, R. Schneider, C. Schwatke, A. Scozzari, E. Ser-giacomi, F. Seyler, R. Shah, R. Sharma, A. Shaw, A. Shepherd, J. Shriver, C. K. Shum, W. Simons, S. B. Simonsen, T. Slater, W. Smith, S. Soares, M. Sokolovskiy, L. Soudarin, C. Spatar, S. Speich, M. Srinivasan, M. Srokosz, E. Stanev, J. Staneva, N. Steunou, J. Stroeve, B. Su, Y. B. Sulistioadi, D. Swain, A. Sylvestre-baron, N. Taburet, R. Tailleux, K. Takayama, B. Tapley, A. Tarpanelli, G. Tavernier, L. Testut, P. K. Thakur, P. Thibaut, L. Thompson, J. Tintoré, C. Tison, C. Tourain, J. Tournadre, B. Townsend, N. Tran, S. Trilles, M. Tsamados, K.-H. Tseng, C. Ubelmann, B. Uebbing, O. Vergara, J. Verron, T. Vieira, S. Vignudelli, N. Vinogradova Shiffer, P. Visser, F. Vivier, D. Volkov, K. von Schuckmann, V. Vuglinskii, P. Vuilleumier, B. Walter, J. Wang, C. Wang, C. Watson, J. Wilkin, J. Willis, H. Wilson, P. Woodworth, K. Yang, F. Yao, R. Zaharia, E. Zakharova, E. D. Zaron, Y. Zhang, Z. Zhao, V. Zinchenko, and V. Zlotnicki, "Altimetry for the future: Building on 25 years of progress," *Advances in Space Research*, vol. 68, no. 2, pp. 319–363, Jul. 2021.

[14] A. Carrassi, M. Bocquet, L. Bertino, and G. Evensen, "Data assimilation in the geosciences: An overview of methods, issues, and perspectives," *WIREs Climate Change*, vol. 9, no. 5, p. e535, 2018.

[15] A. P. S. Wong, S. E. Wijffels, S. C. Riser, S. Pouliquen, S. Hosoda, D. Roemmich, J. Gilson, G. C. Johnson, K. Martini, D. J. Murphy, M. Scanderbeg, T. V. S. U. Bhaskar, J. J. H. Buck, F. Merceur, T. Carval, G. Maze, C. Cabanes, X. André, N. Poffa, I. Yashayaev, P. M. Barker, S. Guinehut, M. Belbéoch, M. Ignaszewski, M. O. Baringer, C. Schmid, J. M. Lyman, K. E. McTaggart, S. G. Purkey, N. Zilberman, M. B. Alkire, D. Swift, W. B. Owens, S. R. Jayne, C. Hersh, P. Robbins, D. West-Mack, F. Bahr, S. Yoshida, P. J. H. Sutton, R. Cancouët, C. Coatanoan, D. Dobbler, A. G. Juan, J. Gourrion, N. Kolodziejczyk, V. Bernard, B. Bourlès, H. Claustre, F. D'Ortenzio, S. Le Reste, P.-Y. Le Traon, J.-P. Rannou, C. Saout-Grit, S. Speich, V. Thierry, N. Verbrugge, I. M. Angel-Benavides, B. Klein, G. Notarstefano, P.-M. Poulain, P. Vélez-Belchí, T. Suga, K. Ando, N. Iwasaska, T. Kobayashi, S. Masuda, E. Oka, K. Sato, T. Nakamura, K. Sato, Y. Takatsuki, T. Yoshida, R. Cowley, J. L. Lovell, P. R. Oke, E. M. van Wijk, F. Carse, M. Donnelly, W. J. Gould, K. Gowers, B. A. King, S. G. Loch, M. Mowat, J. Turton, E. P. Rama Rao,

M. Ravichandran, H. J. Freeland, I. Gaboury, D. Gilbert, B. J. W. Greenan, M. Ouellet, T. Ross, A. Tran, M. Dong, Z. Liu, J. Xu, K. Kang, H. Jo, S.-D. Kim, and H.-M. Park, "Argo Data 1999–2019: Two Million Temperature-Salinity Profiles and Subsurface Velocity Observations From a Global Array of Profiling Floats," *Frontiers in Marine Science*, vol. 7, 2020.

[16] T. Archambault, A. Filoche, A. Charantonnis, and D. Béréziat, "Multimodal Unsupervised Spatio-Temporal Interpolation of satellite ocean altimetry maps," in *VISAPP*, Feb. 2023.

[17] M. Beauchamp, r. fablet, C. Ubelmann, M. Ballarotta, and B. Chapron, "Data-driven and learning-based interpolations of along-track Nadir and wide-swath SWOT altimetry observations," in *Proceedings of the 10th International Conference on Climate Informatics*, ser. CI2020. New York, NY, USA: Association for Computing Machinery, Jan. 2021, pp. 22–29.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, ser. Lecture Notes in Computer Science, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[19] R. Fablet, B. Chapron, L. Drumetz, E. Mémin, O. Pannekoucke, and F. Rousseau, "Learning Variational Data Assimilation Models and Solvers," *Journal of Advances in Modeling Earth Systems*, vol. 13, no. 10, p. e2021MS002572, 2021.

[20] B. Barnier, M. Gurvan, P. Thierry, M. Jean-Marc, T. Anne-Marie, L. S. Julien, B. Aike, B. Arne, B. Claus, D. Joachim, D. Corine, D. Edmée, G. Sergei, R. Elizabeth, T. Claude, T. Sébastien, M. Mathew, M. Julie, and D. C. Beverly, "Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy-permitting resolution," *Ocean Dynamics*, vol. 56, no. 5, pp. 543–567, Dec. 2006.

[21] A. Ajayi, J. Le Sommer, E. Chassignet, J.-M. Molines, X. Xu, A. Albert, and E. Cosme, "Spatial and Temporal Variability of the North Atlantic Eddy Field From Two Kilometric-Resolution Ocean Models," *Journal of Geophysical Research: Oceans*, vol. 125, no. 5, p. e2019JC015827, 2020.

[22] M. Benkiran, G. Ruggiero, E. Greiner, P.-Y. Le Traon, E. Rémy, J. M. Lellouche, R. Bourdallé-Badie, Y. Drillet, and B. Tchonang, "Assessing

the Impact of the Assimilation of SWOT Observations in a Global High-Resolution Analysis and Forecasting System Part 1: Methods," *Frontiers in Marine Science*, vol. 8, 2021.

[23] G. Dibarboure, C. Ubelmann, B. Flamant, F. Briol, E. Peral, G. Bracher, O. Vergara, Y. Faugère, F. Soulat, and N. Picot, "Data-Driven Calibration Algorithm and Pre-Launch Performance Simulations for the SWOT Mission," *Remote Sensing*, vol. 14, no. 23, p. 6070, Jan. 2022.

[24] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2009, pp. 248–255.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[26] C. A. Gomez Gonzalez, O. Wertz, O. Absil, V. Christiaens, D. Defrère, D. Mawet, J. Milli, P.-A. Absil, M. Van Droogenbroeck, F. Cantalloube, P. M. Hinz, A. J. Skemer, M. Karlsson, and J. Surdej, "VIP: Vortex Image Processing Package for High-contrast Direct Imaging," *The Astronomical Journal*, vol. 154, p. 7, Jul. 2017.

[27] A. Dosovitskiy, P. Fischer, E. Ilg, P. Häusser, C. Hazirbas, V. Golkov, P. van der Smagt, D. Cremers, and T. Brox, "FlowNet: Learning Optical Flow with Convolutional Networks," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec. 2015, pp. 2758–2766.

[28] S.-A. Boukabara, K. Ide, Y. Zhou, N. Shahroudi, R. N. Hoffman, K. Garrett, V. K. Kumar, T. Zhu, and R. Atlas, "Community Global Observing System Simulation Experiment (OSSE) Package (CGOP): Assessment and Validation of the OSSE System Using an OSSE–OSE Intercomparison of Summary Assessment Metrics," *Journal of Atmospheric and Oceanic Technology*, vol. 35, no. 10, pp. 2061–2078, Oct. 2018.

[29] B. Buongiorno Nardelli, D. Cavaliere, E. Charles, and D. Ciani, "Super-Resolving Ocean Dynamics from Space with Computer Vision Algorithms," *Remote Sensing*, vol. 14, no. 5, p. 1159, Jan. 2022.

[30] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[31] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, Jan. 1989.

[32] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, Jul. 2019.

[33] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014.

[34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, vol. 25. Curran Associates, Inc., 2012.

[35] N. D. Brenowitz, T. Beucler, M. Pritchard, and C. S. Bretherton, "Interpreting and Stabilizing Machine-Learning Parametrizations of Convection," *Journal of the Atmospheric Sciences*, vol. 77, no. 12, pp. 4357–4375, Dec. 2020.

[36] Y. Yin, V. L. Guen, J. Dona, E. de Bézenac, I. Ayed, N. Thome, and P. Gallinari, "Augmenting physical models with deep networks for complex dynamics forecasting*," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2021, no. 12, p. 124012, Dec. 2021.

[37] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686–707, Feb. 2019.

[38] S. Greydanus, M. Dzamba, and J. Yosinski, "Hamiltonian Neural Networks," in *Advances in Neural Information Processing Systems*, vol. 32. Curran Associates, Inc., 2019.

[39] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, "Fourier neural operator for parametric partial differential equations," 2020.

[40] R. Wang, K. Kashinath, M. Mustafa, A. Albert, and R. Yu, "Towards physics-informed deep learning for turbulent flow prediction," *Proceed-*

ings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020.

[41] R. Fablet, S. Ouala, and C. Herzet, "Bilinear Residual Neural Network for the Identification and Forecasting of Geophysical Dynamics," in *2018 26th European Signal Processing Conference (EUSIPCO)*, Sep. 2018, pp. 1477–1481.

[42] M. Andrychowicz, M. Denil, S. Gómez, M. W. Hoffman, D. Pfau, and T. Schaul, "Learning to learn by gradient descent by gradient descent," Nov. 2016.

[43] Febvre, "Code and Data release Simulation-based 4dVarNets for satellite altimetry," Jun. 2023.

[44] L. Brodeau, J. L. Sommer, and A. Albert, "Ocean-next/eNATL60: Material describing the set-up and the assessment of NEMO-eNATL60 simulations," Zenodo, Sep. 2020.

[45] M. Gurvan, R. Bourdallé-Badie, J. Chanut, E. Clementi, A. Coward, C. Ethé, D. Iovino, D. Lea, C. Lévy, T. Lovato, N. Martin, S. Masson, S. Mocavero, C. Rousset, D. Storkey, S. Müeller, G. Nurser, M. Bell, G. Samson, P. Mathiot, F. Mele, and A. Moulin, "NEMO ocean engine," Mar. 2022.

[46] A. Voldoire, E. Sanchez-Gomez, D. Salas y Mélia, B. Decharme, C. Cassou, S. Sénési, S. Valcke, I. Beau, A. Alias, M. Chevallier, M. Déqué, J. Deshayes, H. Douville, E. Fernandez, G. Madec, E. Maisonnave, M.-P. Moine, S. Planton, D. Saint-Martin, S. Szopa, S. Tyteca, R. Alkama, S. Belamari, A. Braun, L. Coquart, and F. Chauvin, "The CNRM-CM5.1 global climate model: Description and basic evaluation," *Climate Dynamics*, vol. 40, no. 9, pp. 2091–2121, May 2013.

[47] L. Carrere, Y. Faugère, and M. Ablain, "Major improvement of altimetry sea level estimations using pressure-derived corrections based on ERA-Interim atmospheric reanalysis," *Ocean Science*, vol. 12, no. 3, pp. 825–842, Jun. 2016.

Chapter 5

OceanBench: The Sea Surface Height Edition

This chapter is based on an accepted international conference publication with the same name. The preprint is available here [1]

5.1 Introduction

The ocean is vital to the Earth's system [2]. It plays a significant role in climate regulation regarding carbon [3] and heat uptake [4]. It is also a primary driver of human activities (e.g., maritime traffic and world trade, marine resources and services) [5, 6]. Satellite remote sensing is one of the most effective ways of measuring essential sea surface quantities [7] such as sea surface height (SSH) [8], sea surface temperature (SST) [9], and ocean color [10]. While these variables characterize only a tiny portion of the ocean ecosystem, they present a gateway to many other derived physical quantities [6].

Although we can access observable sea surface quantities, they can be noisy and irregularly sampled like the altimetry data previously considered in this thesis [8]. This makes the characterization of ocean processes highly challenging for operational products and downstream tasks that depend on relevant gap-free variables. As presented in previous chapters, deep learning schemes [11, 12, 13] have become appealing solutions to benefit from existing large-scale observation and simulation datasets and reach significant breakthroughs in the monitoring of upper ocean dynamics from scarcely and irregularly sampled observations. To ensure that these methods provide genuine value, evaluation criteria and metrics must be defined with domain expertise by ocean experts. The quality of SSH estimations, for instance, depends on factors such as geographical region, season, physical plausibility of derived quantities. The choice of using observational or simulated data for metric computation also yields different assessments

Furthermore, the heterogeneity and characteristics of the observation data present major challenges for effectively applying these methods beyond idealized case studies. A data source can have different variables, geometries, and noise levels, resulting in many domain-specific preprocessing procedures that can vastly change the solution outcome. Accessibility to the data and the relevant processing steps can significantly lower the entry barriers for aspiring machine learning practitioners.

These considerations provide the motivation for **OceanBench**, a framework for co-designing machine-learning-driven high-level experiments from ocean observations. It consists of an end-to-end framework for piping data from its raw form to an ML-ready state and from model outputs to interpretable quantities. We regard `OceanBench` as a key facilitator for the uptake of MLOPs tools and research [14, 15] for ocean-related datasets and case studies. This first edition provides datasets and ML-ready benchmarking pipelines for SSH interpolation problems, an essential topic for the space oceanography community, related to ML communities dealing with issues like in-painting [16], denoising [17, 18], and super-resolution [19]. We expect `OceanBench` to facilitate new challenges to the applied machine learning community and contribute to meaningful ocean-relevant breakthroughs. The remainder of the chapter is organized as follows: in Section 2, we outline some related work that was inspirational for this work; in Section 3, we formally outline `OceanBench` by highlighting the target audience, code structure, and problem scope; in Section 4, we provide some insight into different tasks related to SSH interpolation where `OceanBench` could provide some helpful utility; and in Section 5 we outline current limitations of the project and give some concluding remarks.

5.2 Related Work

Machine learning applied to geosciences is becoming increasingly popular, but there are few examples of transparent pipelines involving observation data. After a thorough literature review, we have divided the field into three camps of ML applications that pertain to this work: 1) toy simulation datasets, 2) reanalysis datasets, and 3) observation datasets. We outline the literature for each of the three categories below.

Toy Simulation Data. One set of benchmarks focuses on learning surrogate models for well-defined but chaotic dynamical systems in the form of ordinary differential equations (ODEs) and partial differential equations

(PDEs) and there are freely available code bases which implement different ODEs/PDEs [20, 21, 22, 23, 24, 25, 26, 27]. This is a great testing ground for simple toy problems that better mimic the structures we see in real-world observations. Working with simulated data is excellent because it is logistically simple and allows users to test their ideas on toy problems without increasing the complexity when dealing with real-world data. However, these are ultimately simple physical models that often do not reflect the authentic structures we see in real-world, observed data.

Reanalysis Data. This is assimilated data of real observations and model simulations. There are a few major platforms that host ocean reanalysis data like the Copernicus Marine Data Store [28, 29, 30, 31], the Climate Data Store [32], the BRAN2020 Model [33], and the NOAA platform [34]. However, to our knowledge, there is no standard ML-specific ocean-related tasks to accompany the data. On the atmospheric side, platforms like `WeatherBench` [35], `ClimateBench` [36], `ENS10` [37] were designed to assess short-term and medium-term forecasting using ML techniques with recent success of ML [38, 39] The clarity of the challenges set by the benchmark suites has inspired the idea of `OceanBench`, where we directly focus on problems dealing with ocean observation data.

Observation Data. These observation datasets (typically sparse) stem from satellite observations that measure surface variables or in-situ measurements that measure quantities within the water column. Some major platforms to host data include the Marine Data Store [40, 41], the Climate Data Store [42, 43, 44], ARGO [45], and the SOCAT platform [46]. However, it is more difficult to assess the efficacy of operational ML methods that have been trained only on observation data and, to our knowledge, there is no coherent ML benchmarking system for ocean state estimation. There has been significant effort by the *Ocean-Data-Challenge* Group [1] which provides an extensive suite of datasets and metrics for SSH interpolation. Their efforts heavily inspired our work, and we hope that `OceanBench` can build upon their work by adding cohesion and facilitating the ease of use for ML research and providing a high-level framework for providing ML-related data products.

1. Ocean Data Challenge group: Freely associated scientist for oceanographic algorithm and product improvements (ocean-data-challenges.github.io)

5.3 OceanBench

Why OceanBench?

There is a high barrier to entry in working with ocean observations for researchers in applied machine learning as there are many processing steps for both the observation data and the domain-specific evaluation procedures. `OceanBench` aims to lower the barrier to entry cost for ML researchers to make meaningful progress in the field of state prediction. We distribute a standardized, transparent, and flexible procedure for defining data and evaluation pipelines for data-intensive geoscience applications. Proposed examples and case studies provide a plug-and-play framework to benchmark novel ML schemes w.r.t. state-of-the-art, domain-specific ML baselines. In addition, we adopt a pedagogical abstraction that allows users to customize and extend the pipelines for their specific tasks. To our knowledge, no framework embeds processing steps for earth observation data in a manner compatible with MLOps abstractions and standards regarding reproducibility and evaluation procedures. Ultimately, we aim to facilitate the uptake of ML schemes to address ocean observation challenges and to bring new challenges to the ML community to extend additional ML tools and methods for irregularly-sampled and partially-observed high-dimensional space-time dynamics. The abstractions proposed here apply beyond ocean sciences and SSH interpolation to other geosciences with similar tasks that intersect with machine learning.

Code Structure

`OceanBench` is lightweight in terms of the core functionality. We keep the code base simple and focus more on how the user can combine each piece. We adopt a strict functional style because it is easier to maintain and combine sequential transformations. There are five features we would like to highlight about `OceanBench`: 1) Data availability and version control, 2) an agnostic suite of geoprocessing tools for `xarray` datasets that were aggregated from different sources, 3) Hydra integration to pipe sequential transformations, 4) a flexible multi-dimensional array generator from `xarray` datasets that are compatible with common deep learning (DL) frameworks, and 5) a JupyterBook [47] that offers library tutorials and demonstrates use-cases. In the following section, we highlight these components in more detail.

Data Availability. The most important aspect is the public availability of the datasets. We aggregate all pre-curated datasets from other sources, e.g.

the *Ocean-Data-Challenge* [48, 49], and organize them to be publicly available from a single source [2]. We also offer a few derived datasets which can be used for demonstrations and evaluation. Data is never static in a pipeline setting, as one can have many derived datasets which stem from numerous preprocessing choices. In fact, in research, we often work with derived datasets that have already been through some preliminary preprocessing methods. To facilitate the ever-changing nature of data, we use the Data Version Control (`DVC`) tool [50], which offers a git-like version control of the datasets.

Geoprocessing Tools. The core `OceanBench` library offers a suite of functions specific to processing geo-centric data. While a few particular functionalities vary from domain to domain, many operations are standard, e.g., data variable selections, filtering/smoothing, regridding, coordinate transformations, and standardization. We almost work exclusively with the `xarray` [51] framework because it is a coordinate-aware, flexible data structure. In addition, the geoscience community has an extensive suite of specialized packages that operate in the `xarray` framework to accomplish many different tasks. Almost all `OceanBench` toolsets are exclusively within the `xarray` framework to maintain compatibility with a large suite of tools already available from the community.

Hydra Integration. As discussed above, many specific packages accomplish many different tasks. However, what needs to be added is the flexibility to mix and match these operations as the users see fit. `Hydra` [52] provides a configurable way to aggregate and *pipe* many sequential operations together. It also maintains readability, robustness, and flexibility through the use of `.yaml` files which explicitly highlights the function used, the function parameters chosen, and the sequence of operations performed. In the ML software stack, `Hydra` is often used to manage the model, optimizer, and loss configurations which helps the user experiment with different options. We apply this same concept in preprocessing, geoprocessing, and evaluation steps, often more important than the model configuration in geoscience-related tasks.

`XRPatcher` [3]. Every machine learning pipeline will inevitably require moving data from the geo-specific data structure to a multi-dimensional array easily digestible for ML models. A rather underrated, yet critical, feature of ML frameworks such as `PyTorch` [53] (`Lightning` [54]) and `TensorFlow` [55] (`Keras` [56]) is the abstraction of the dataset, dataloader, datamodules, and data pipelines. In applied ML in geosciences, the data pipelines are often more

2. Available at: oceanbench-data-registry.github.com
3. Available at: github.com/jejjohnson/xrpatcher

important than the actual model [57]. The user can control the *patch*-size and the *stride*-step, which can generate arbitrary coordinate-aware items directly from the `xarray` data structure. In addition, `XRPatcher` provides a way to reconstruct the fields from an arbitrary patch configuration. This robust reconstruction step is convenient to extend the ML inference step where one can reconstruct entire fields of arbitrary dimensions beyond the training configuration, e.g., to account for the border effects within the field (see appendix 6.4) or to reconstruct quantities in specific regions or globally.

JupyterBook. Building a set of tools is relatively straightforward; however, ensuring that it sees a broader adoption across a multi-disciplinary community is much more challenging. We invested heavily in showing use cases that appeal to different users with the `JupyterBook` platform [47]. Code with context is imperative for domain and ML experts as we need to explain and justify each component and give many examples of how they can be used in other situations. Thus, we have paid special attention to providing an extensive suite of tutorials, and we also highlight use cases for how one can effectively use the tools.

Problem Scope

There are many problems that are of great interest the ocean community [58] but we limit the scope to state estimation problems [59]. Under this scope, there are research questions that are relevant to operational centers which are responsible for generating the vast majority of global ocean state maps [28, 30, 29, 31] that are subsequently used for many downstream tasks [6]. For example: how can we effectively use heterogeneous observations to predict the ocean state on the sea surface [60, 61, 62, 63, 64, 9]; how can we incorporate prior physics knowledge into our predictions of ocean state trajectories [60, 58, 6]; and how can we use the current ocean state at time T to predict the future ocean state at time $T + \tau$ [65, 35, 66]. In the same vain, there are more research questions that are of interest to the academic modeling community. For example: is simulated or reanalysis data more effective for learning ML emulators that replace expensive ocean models [67, 68]; what metrics are more effective for assessing our ability to mimic ocean dynamics [69, 70]; and how much model error can we characterize when learning from observations [71, 72].

We have cited many potential applications of how ML can be applied to tackle the state estimation problem. However, to our knowledge there is no publicly available, standardized benchmark system that is caters to ML-research standards. We believe that, irrespective of the questions posed above and the

data we access, there are many logistical similarities for each of the problem formulations where we can start to set standards for a subset of tasks like interpolation or forecasting. On the front-end, we need a way to select regions, periods, variables, and a valid train-test split (see sec. 6.3). On the back-end, we need a way to transform the predictions into more meaningful variables with appropriate metrics for validation (see sec. 6.3 and 6.3). `OceanBench` was designed to be an agnostic tool that is extensible to the types of datasets, processing techniques and metrics needed for working with a specific class of Ocean-related datasets. We strongly feel that a suite like this is the first step in designing task-specific benchmarks within the ocean community that is compatible with ML standards. In the remainder of the chapter, we will demonstrate how `OceanBench` can be configured for the sea surface height interpolation use-case.

5.4 *Sea Surface Height Edition*

The `OceanBench` project is currently at a first iteration dedicated to SSH interpolation. The previous chapter highlighted the potential of learning-based methodologies for this task. Integrating and extending the corresponding experimental setups in `OceanBench` is a natural first step. The next part of this section details datasets, metrics and evaluation quantities made available by the `OceanBench` platform. Finally we introduce four data challenges implemented in this SSH Edition with examples of metrics and maps for different methods.

Data

The availability of multi-year simulation and observation datasets naturally advocates for the design of synthetic (or twin) experiments, referred to as observing system simulation experiments (OSSE), and of real-world experiments, referred to as observing system experiments (OSE). We outline these two experimental setups below followed by Table 5.1 which describe in detail the data made available by `OceanBench`.

Observing System Simulation Experiments (OSSE). A staple and groundtruthed experimental setup uses a reference simulation dataset to simulate the conditions we can expect from actual satellite observations. This setup allows researchers and operational centers to create a fully-fledged pipeline that mirrors the real-world experimental setting. An ocean model simulation is deployed over a specified spatial domain and period, and a satellite observation

simulator is deployed to simulate satellite observations over the same domain and period. This OSSE setup has primarily been considered for performance evaluation, as one can assess a reconstruction performance over the entire space-time domain. It also provides the basis for the implementation of classic supervised learning strategies [13, 12, 11]. The domain expert can vary the experimental conditions depending on the research question. For example, one could specify a region based on the expected dynamical regime [49] or add a certain noise level to the observation tracks based on the satellite specifications. The biggest downside to OSSE experiments is that we train models exclusively with ocean simulations which could produce models that fail to generalize to the actual ocean state. Furthermore, the simulations are often quite expensive, which prevents the community from having high spatial resolution over very long periods, which would be essential to capture as many dynamical regimes as possible.

Observing System Experiments (OSE). As more observations have become available over the past few decades, we can also design experiments using real data. This involves aggregating as many observations from real ocean altimetry satellites as possible with some specific independent subset left out for evaluation purposes. A major downside to OSE experiments is that the sparsity and spatial coverage of the observations narrow the possible scope of performance metrics and make it very challenging to learn directly from observation datasets. The current standard altimetry data are high resolution but cover a tiny area. As such, it can only inform fine-scale SSH patterns in the along-track satellite direction and cannot explicitly reveal two-dimensional patterns. Despite these drawbacks, it provides a quantitative evaluation of the generalizability of the ML methods concerning the true ocean state.

	OSSE SSH	OSSE SSH NADIR		OSSE SSH SWOT		OSSE SST	OSE SSH NADIR
Data Structure	Gridded	AlongTrack	Gridded	AlongTrack	Gridded	Gridded	AlongTrack
Source	NEMO [73]	NEMO [73]	NEMO [73]	NEMO [73]	NEMO [73]	NEMO [73]	Altimetry [40]
Region	GulfStream	GulfStream	GulfStream	GulfStream	GulfStream	GulfStream	GulfStream
Domain Size [°]	$10 \times 10°$	$10 \times 10°$	$10 \times 10°$	$10 \times 10°$	$10 \times 10°$	$10 \times 10°$	$10 \times 10°$
Domain Size [km]	1100×1100	1100×1100	1100×1100	1100×1100	1100×1100	1100×1100	1100×1100
Longitude Extent	$[-65°, -55°]$	$[-65°, -55°]$	$[-65°, -55°]$	$[-65°, -55°]$	$[-65°, -55°]$	$[-65°, -55°]$	$[-65°, -55°]$
Latitude Extent	$[33°, 43°]$	$[33°, 43°]$	$[33°, 43°]$	$[33°, 43°]$	$[33°, 43°]$	$[33°, 43°]$	$[33°, 43°]$
Resolution [°]	$0.05° \times 0.05°$	N/A	$0.05° \times 0.05°$	N/A	$0.05° \times 0.05°$	$0.05° \times 0.05°$	N/A
Resolution [km]	5.5×5.5	6	5.5×5.5	6	5.5×5.5	5.5×5.5	7
Grid Size	200×200	N/A	200×200	N/A	200×200	200×200	N/A
Num. Datapoints	∼14.6M	∼205K	∼14.6M	∼955K	∼14.6M	∼14.6M	∼1.79M
Period Start	2012-10-01	2012-10-01	2012-10-01	2012-10-01	2012-10-01	2012-10-01	2016-12-01
Period End	2013-09-30	2013-09-30	2013-09-30	2013-09-30	2013-09-30	2013-09-30	2018-01-31
Frequency	Daily	1 Hz	Daily	1 Hz	Daily	Daily	1 Hz
Period Length	365 Days	365 Days	365 Days	365 Days	365 Days	365 Days	427 Days
Evaluation Start	2012-10-22	2012-10-22	2012-10-22	2012-10-22	2012-10-22	2012-10-22	2017-01-01
Evaluation End	2012-12-02	2012-12-02	2012-12-02	2012-12-02	2012-12-02	2012-12-02	2017-12-31
Evaluation Length	45 Days	45 Days	45 Days	45 Days	45 Days	45 Days	365 Days

Table 5.1 – This table gives an extended overview of the datasets provided to complete the data challenges listed in 5.4. The OSSE SST and SSH are outputs from come from the free run NEMO model. The OSSE NADIR and SWOT are pseudo-observations generated from the NEMO simulation. We provide the original simulated satellite tracks as well as a gridded version at the same resolution as the simulation.

Metrics

There are many metrics that are standard within the ML community but unconvincing for many parts the geoscience community. Specifically, many of these standard scores do not capture the important optimization criteria in the scientific machine learning tasks. However, there is not consensus within domain-specific communities about the perfect metric which captures every aspect we are interested. Therefore, we should have a variety of scores from different perspectives to really assess the pros and cons of each method we wish to evaluate thoroughly. Below, we outline two sets of scores we use within this framework: skill scores and spectral scores.

Skill Scores

We classify one set of metrics as *skill scores*. These are globally averaged metrics which tend to operate within the real space. Some examples include the root mean squared error (RMSE), the normalized root mean squared (nRMSE) error, and the nRMSE score. The RMSE metric can also be calculated w.r.t. the spatial domain, temporal domain or both. For example, figure 5.1 showcases the nRMSE score calculated only on the spatial domain and visualized for each time step.

$$\text{RMSE}: \qquad \text{RMSE}(\eta, \hat{\eta}) = ||\eta - \hat{\eta}||_2 \tag{5.1}$$

$$\text{nRMSE}: \qquad \text{nRMSE}(\eta, \hat{\eta}) = \frac{\text{RMSE}(\eta, \hat{\eta})}{||\eta||_2} \tag{5.2}$$

$$\text{nRMSE}_{\text{score}}: \qquad \text{nRMSE}_{\text{score}}(\eta, \hat{\eta}) = 1 - \text{nRMSE}(\eta, \hat{\eta}) \tag{5.3}$$

However, we are not limited to just the standard MSE metrics. We can easily incorporate more higher-order statistics like the Centered Kernel Alignment (CKA) [74] or information theory metrics like mutual information (MI) [75, 76]. In addition, we could also utilize the same metrics in the frequency domain as is done in [21].

Spectral Scores

Another class of scores that we use in `OceanBench` are the *spectral scores*. These scores are calculated within the spectral space via the wavenumber power spectral density (PSD). This provides a spatial-scale-dependent metric which is useful for identifying the largest and smallest scales that were resolved by the reconstruction map. In general, we use these to measure the expected energy at different spatiotemporal scales and we can also construct custom score functions which gives us a summary statistic for how well we reconstructed

certain scales.

$$\text{PSD}: \qquad \text{PSD}(\eta) = \sum_{k_{min}}^{k_{max}} \|\mathcal{F}(\eta)\|^2 \tag{5.4}$$

$$\text{PSD}_{score}: \qquad \text{PSD}_{score}(\eta, \hat{\eta}) = 1 - \frac{\text{PSD}(\eta - \hat{\eta})}{\text{PSD}(\eta)} \tag{5.5}$$

where $\mathcal{F}$ is the Fast Fourier Transformation (FFT). In our application, there are various ways to construct the PSD which depend on the FFT transformation. We denote the *space-time PSD* as $\lambda_{\mathbf{x}}$ which does the 2D FFT in the longitude and time direction, then takes the average over the latitude. We denote the *space-time PSD* as $\lambda_{\mathbf{t}}$ which does the 2D FFT in the longitude and latitude direction, then takes the average over the time. We denote the *isotropic PSD* as λ_r which assumes a radial relationship in the spatial domain and then averages over the temporal domain. Lastly, we denote the standard PSD score as λ_a which is the 1D FFT over a prescribed distance along the satellite track; this is what is done for the OSE NADIR experiment. We recognize that the FFT configurations are limited due to their global treatment of the spectral domain and we need more specialized metrics to handle the local scales. This opens the door to new metrics that handle such cases such as the Wavelet transformation [77].

Physical Variables

Many machine learning methods are unconstrained so they may provide solutions that are physically inconsistent and visualizing the field is a very easy eye test to assess the validity. We have access to many physical quantities which can be derived from sea surface height. This gives us a way to analyze how effective and trustworthy are our reconstructions.

We are interested in the domain across the earth's surface. let us define the earth's domain by some spatial coordinates, $\mathbf{x} = [\text{longitude}, \text{latitude}]^\top \in \mathbb{R}^{d_s}$, and temporal coordinates, $t = [\text{time}] \in \mathbb{R}^+$, where d_s is the dimensionality of the coordinate vector. we can define some spatial (sub-)domain, $\Omega \subseteq \mathbb{R}^{d_s}$, and a temporal (sub-)domain, $\mathcal{T} \subseteq \mathbb{R}^+$. this domain could be the entire globe for 10 years or a small region within the north atlantic for 1 year.

$$\text{spatial coordinates}: \qquad \mathbf{x} \in \omega \subseteq \mathbb{R}^{d_s} \tag{5.6}$$

$$\text{temporal coordinates}: \qquad t \in \mathcal{T} \subseteq \mathbb{R}^+. \tag{5.7}$$

in this case $d_s = 2$ because we only have a two coordinates, however we can do some coordinate transformations like spherical to cartesian. likewise, we can do some coordinate transformation for the temporal coordinates like cyclic transformations or sinusoidal embeddings [78].

Sea Surface Height is the deviation of the height of the ocean surface from the geoid of the Earth. We can define it as:

$$\text{Sea Surface Height } [m]: \qquad \eta = \boldsymbol{\eta}(\mathbf{x}, t) \qquad \Omega \times \mathcal{T} \to \mathbb{R} \tag{5.8}$$

This quantity is the actual value that is given from the satellite altimeters and is presented in the products for SSH maps [8]. An example can be seen in the first row of figure 5.3.

Sea Surface Anomaly is the anomaly wrt to the spatial mean which is defined by

$$\text{Sea Level Anomaly } [m]: \qquad \bar{\eta} = \boldsymbol{\eta}(\mathbf{x}, t) - \bar{\eta}(t) \qquad \Omega \times \mathcal{T} \to \mathbb{R} \tag{5.9}$$

where $\bar{\eta}(t)$ is the spatial average of the field at each time step. An example can be seen in the first row of figure 5.2.

Another important quantity is the **geostrophic velocities** in the zonal and meridional directions. This is given by

$$\text{Zonal Velocity}[ms^{-2}]: \qquad u = -\frac{g}{f_0}\frac{\partial \eta}{\partial y} \qquad \Omega \times \mathcal{T} \to \mathbb{R} \tag{5.10}$$

$$\text{Meridional Velocity}[ms^{-2}]: \qquad v = \frac{g}{f_0}\frac{\partial \eta}{\partial x} \qquad \Omega \times \mathcal{T} \to \mathbb{R} \tag{5.11}$$

where g is the gravitational constant and f_0 is the mean Coriolis parameter. These quantities are important as they can be an related to the sea surface current. The geostrophic assumption is a very strong assumption however it can still be an important indicator variable. The **kinetic energy** is a way to summarize the (geostrophic) velocities as the total energy of the system. This is given by

$$KE = \frac{1}{2}\left(u^2 + v^2\right) \tag{5.12}$$

An example can be seen in the second row of figure 5.3.

Another very important quantity is the *vorticity* which measures the spin and rotation of a fluid. In geophysical fluid dynamics, we use the **relative vorticity** which is the vorticity observed within at rotating frame. This is

given by

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \tag{5.13}$$

An example can be seen in the third row of figure 5.3.

We can also use the **Enstrophy** to summarize the relative voriticty to measure the total contribution which is given by

$$E = \frac{1}{2}\zeta^2 \tag{5.14}$$

The **Strain** is a measure of deformation of a fluid flow.

$$\sigma = \sqrt{\sigma_n^2 + \sigma_s^2} \tag{5.15}$$

where σ_n is the shear strain (aka the shearing deformation) and σ_s is the normal strain (aka stretching deformation). An example can be seen in the fourth row of figure 5.3.

The **Okubo-Weiss Parameter** is high-order quantity which is a linear combination of the strain and the relative vorticity.

$$\sigma_{ow} = \sigma_n^2 + \sigma_s^2 - \zeta^2 \tag{5.16}$$

This quantity is often used as a threshold for determining the location of Eddies in sea surface height and sea surface current fields [79, 80, 81].

Data Challenges

We rely on existing OSSE and OSE experiments for SSH interpolation designed by domain experts [48, 49] and recast them into `OceanBench` framework to deliver a ML-ready benchmarking suites. The selected data challenges for this first edition address SSH interpolation for a 1000km×1000km Gulfstream region. We describe each of them below.

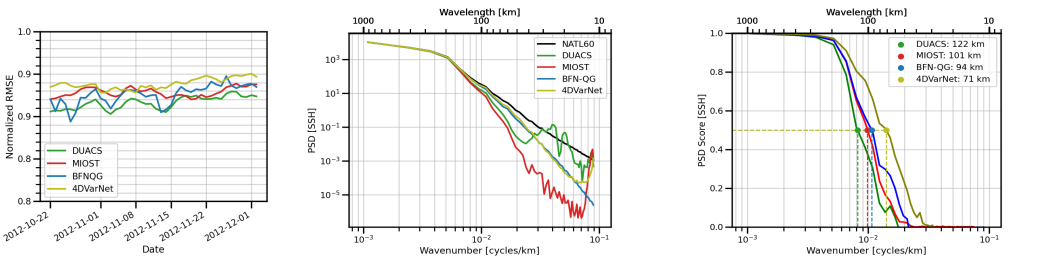
Experiment I (*OSSE NADIR*) addresses SSH interpolation using NADIR altimetry tracks which are very fine, thin ocean satellite observations (see Figure 5.2). It relies on an OSSE using high-resolution (1/60° resolution) ocean simulations generated by the NEMO model over one year with a whole field every day. The reference simulation is the *NATL60* simulation based on the NEMO model [73]. This particular simulation was run over an entire year without any tidal forcing. The simulation provides the outputs of SSH, SST, sea surface salinity (SSS) and the u,v velocities every 1 hour. For the purposes of this data challenge, the spatial domain is over the Gulfstream

with a spatial domain of $[-65°, -55°]$ longitude and $[33°, 43°]$ latitude. The resolution of the original simulation is $1/60°$ resolution with hourly snapshots, and we consider a daily downsampled trajectory at $1/20°$ for the data challenge which results in a 365x200x200 spatio-temporal grid. This simulation resolves finescale dynamical processes ($\sim$15km) which makes it a good test bed for creating an OSSE environment for mapping. The SSH observations include simulations of ocean satellite NADIR tracks. In particular, they are simulations of Topex-Poseidon, Jason 1, Geosat Follow-On, and Envisat. There is no observation error considered within the challenge. We use a the entire period from 2012-10-10 until 2013-09-30. A training period is only from 2013-01-02 to 2013-09-30 where the users can use the reference simulation as well as all available simulated observations. The evaluation period is from 2012-10-22 to 2012-12-02 (i.e. 41 days) which is considered decorrelated from the training period. During the evaluation period, the user cannot use the reference NATL60 simulation but they can use all available simulated observations. There is also a spin-up period allowance from 2012-10-01 where the user can also use all available simulated observations.

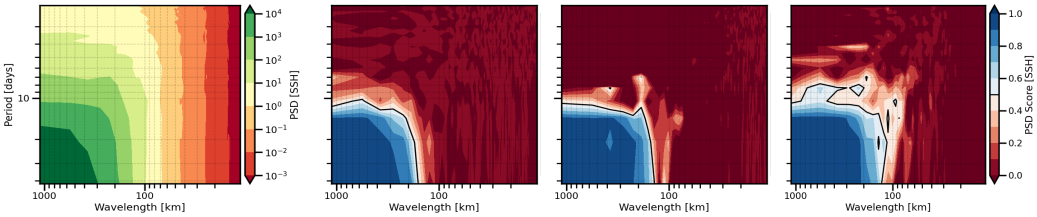
Experiment II (*OSSE SWOT*) addresses SSH interpolation using jointly NADIR and SWOT altimetry data where we complement the **OSSE NADIR** configuration with simulated SWOT observations. SWOT is a new satellite altimetry mission with a much higher spatial coverage but a much lower temporal resolution as illustrated in Figure 5.2. The higher spatial resolution allows us to see structures at a smaller resolution but at the cost of a massive influx of observations (over $\times 100$).

Experiment III (*OSSE SST*) addresses SSH interpolation using altimetry and SST satellite data jointly. We complement the **OSSE SWOT** challenge with simulated SST observations. Satellite-derived SST observations are more abundantly available in natural operational settings than SSH at a finer resolution, and structures have visible similarities [82, 60]. So this challenge allows for methods to take advantage of multi-modal learning [63, 11].

For the OSSE SWOT and OSSE SST experiments, the reference simulation, domain, and evaluation period is the same as the OSSE NADIR experiment. However, the OSSE SWOT includes simulated observations of the novel KaRIN sensor recently deployed during the SWOT mission, the pseudo-observations were generated using the SWOT simulator [82]. This OSSE SST experiment allows the users to utilize the full fields of SST as inputs to help reconstruct the SSH field in conjunction with the NADIR and SWOT SSH observation.



(a) Normalized RMSE (b) Isotropic Power Spectrum (c) Isotropic Power Spectrum Score



(d) NEMO Simulation (e) MIOST (f) BFN-QG (g) 4DVarNet

Figure 5.1 – **Evaluation of the SSH field reconstructions for the OSSE NADIR experiment**.. Subfigure (a) showcases the normalized root mean squared error (nRMSE), (b) showcases the isotropic power spectrum decomposition (PSD), (c) showcases isotropic PSD scores. The bottom row showcases the space-time PSD for the NEMO simulation (subfigure (d)) and the PSD scores for three reconstruction models: (e) the MIOST model, (f) the BFN-QG model, and (g) the 4DVarNet model.

Because the SST comes from the same NATL60 simulation, the geometry characteristics SST and SSH are exactly the same.

Experiment IV (*OSE NADIR*) addresses SSH interpolation for real NADIR altimetry data. In contrast to the three OSSE data challenges, it only looks at actual observations aggregated from the currently available ocean altimetry data from actual satellites. It involves a similar space-time sampling as Experiment (**OSSE NADIR**) to evaluate the generalization of ML methods trained in Experiment I to real altimetry data. The training problem's complexity increases significantly due to the reference dataset's sparsity compared with the **OSSE NADIR** dataset. One may also explore transfer learning or fine-tuning strategies from the available OSSE dataset.

The OSE NADIR experiment only uses real observations aggregated from different altimeters. These SSH observations include observations from the SARAL/Altika, Jason 2, Jason 3, Sentinel 3A, Haiyang-2A and Cryosat-2 altimeters. The Cryosat-2 altimeter is used as the independent evaluation track used to assess the performance of the reconstructed SSH field.

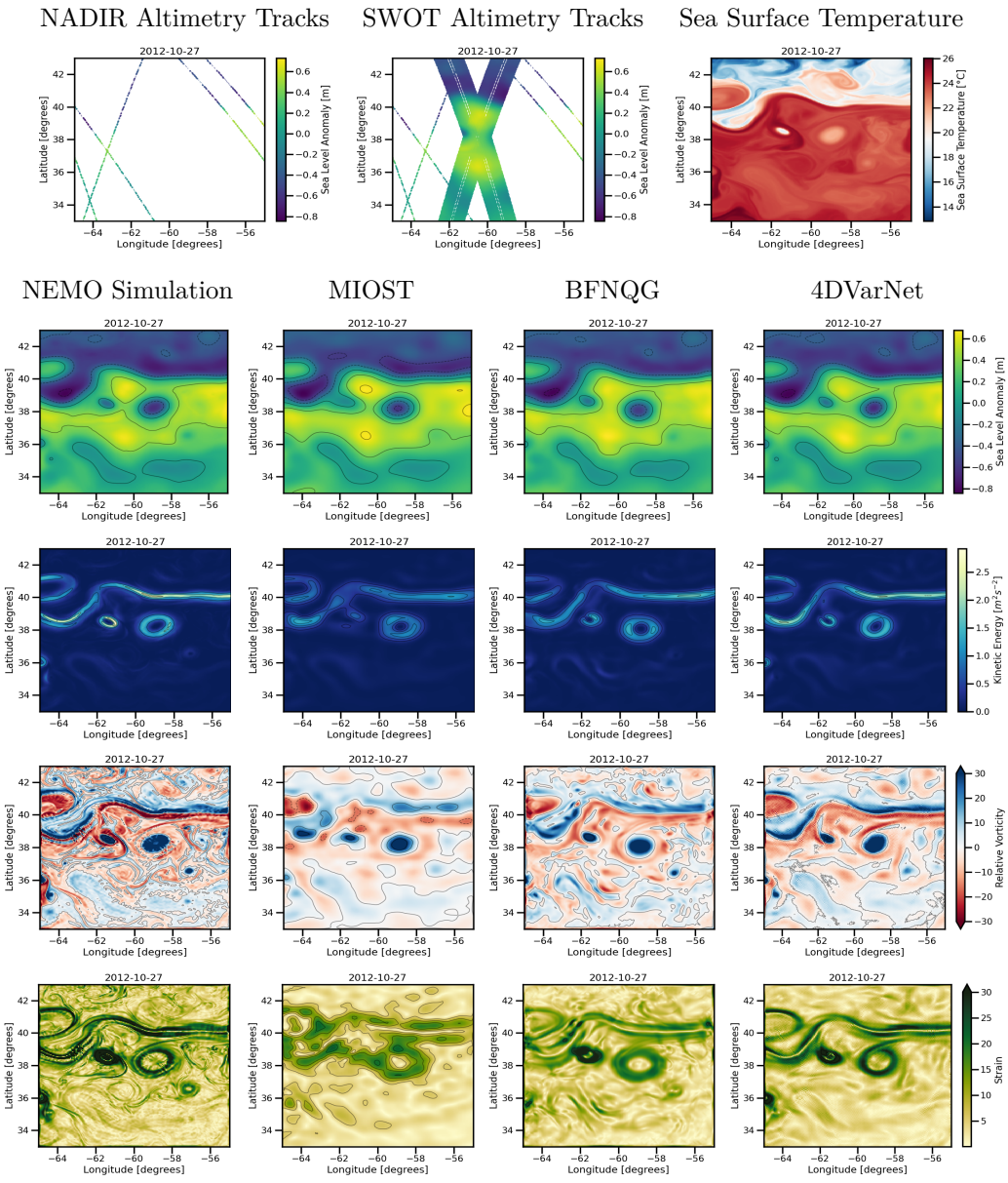
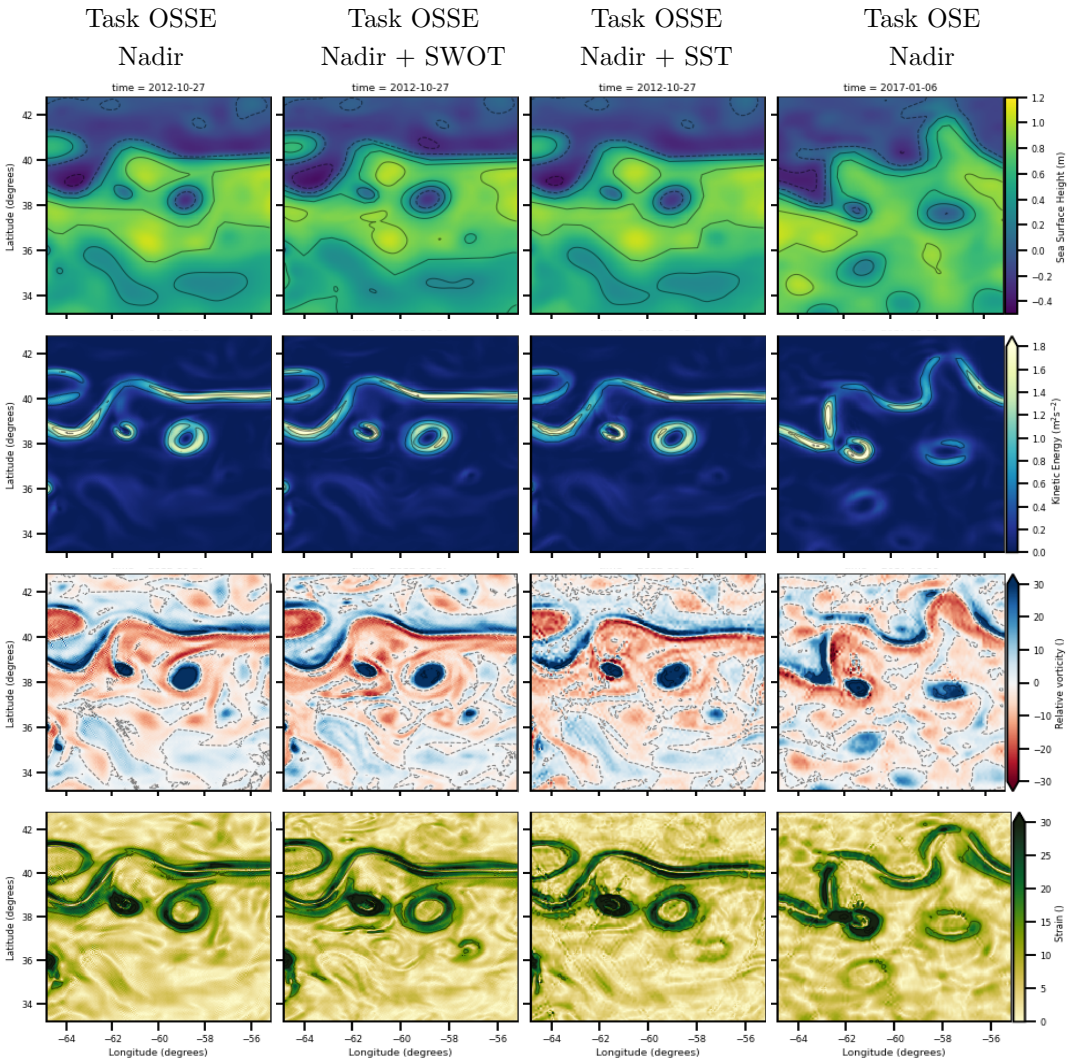


Figure 5.2 – 27^{th} **October, 2012 from the NEMO simulation for the OSSE experiment** outlined in section 5.4. The top row showcases the aggregated NADIR altimetry tracks and the aggregated SWOT altimetry tracks (12 hours before and 12 hours after) as well as the SST from the NEMO simulation. Each subsequent row showcases the following physical variables found in appendix 5.4: (a) Sea Level Anomaly, (b) Kinetic Energy, (c) Relative Vorticity, and (d) Strain. Each column in the subsequent rows showcase the following reconstructed field from the NEMO simulation found in columrn (a): (b) MIOST, (c) BFN-QG, and (d) 4DVarNet.



(a) (b) (c) (d)

Figure 5.3 – **Reconstructed quantities by the 4dVarNet method for each of the four tasks**. Each row showcases the following physical variables found in section 5.4: (a) Sea Surface Height, (b) Kinetic Energy, (c) Relative Vorticity, and (d) Strain. Each column showcase the reconstructed from the tasks (a) OSSE using only Nadir tracks: (b) OSSE using Nadir tracks and SWOT swath, (c) Multimodal using Nadir tracks and sea surface temperature, and (d) Reconstruction using real nadir altimetry tracks.

Results

We use `OceanBench` to generate maps of relevant quantities from the 4DVarNet method [64, 63]. Figure 5.3 showcases some demo maps for some key physical variables outlined in section 5.4. We showcase the 4DVarNet method because it is the SOTA method that was applied to each of the data challenges. We can see that the addition of more information, i.e. NADIR -> SWOT -> SST, results in maps look more similar to the NEMO simulation in the OSSE challenges. It also produces sensible maps for the OSE challenge as well.

`OceanBench` also generated figure 5.4 which shows plots of the PSD and PSD scores of SSH for the different challenges. Again, as we increase the efficacy of the observations via SWOT and allow for more external factors like the SST, we get an improvement in the isotropic and spacetime PSD scores. In addition, we see that the PSD plots for the OSE task look very similar to the OSE challenges.

Lastly, we used `OceanBench` to generate a leaderboard of metrics for a diverse set of algorithms where the maps were available online. Table 5.2 displays all of the key metrics outlined in section 5.4 including the normalized RMSE and various spectral scores which are appropriate for the challenge. We see that as the complexity of the method increases, the metrics improve. In addition, the methods that involve end-to-end learning perform the best overall, i.e. 4DVarNet.

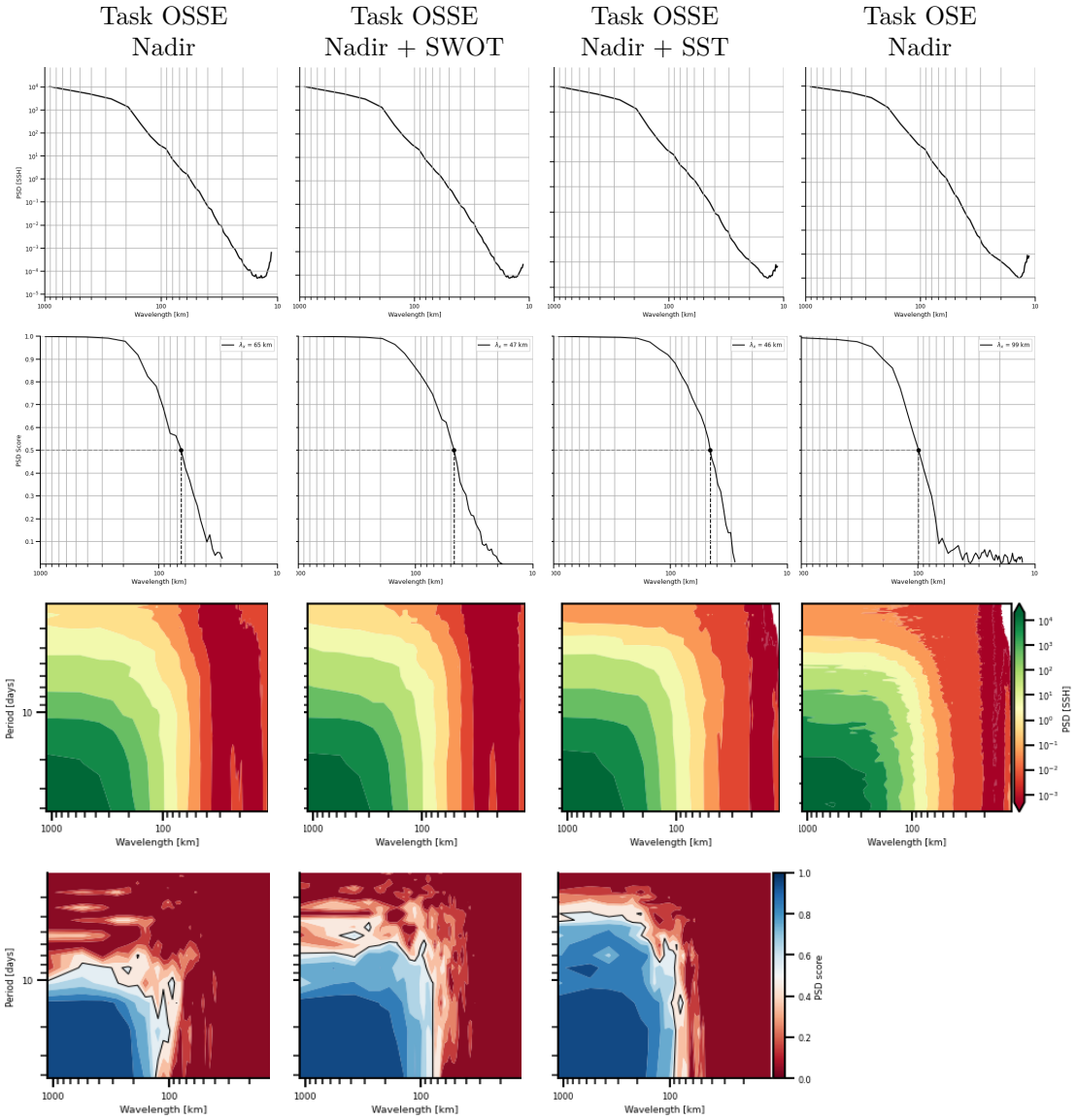


Figure 5.4 – **Power spectrum and associated scores of the 4dVarNet method for the four experiments**. The row display in order: (1) the isotropic PSD, (2) the spatial PSD score (using the isotropic PSD for the first three rows and along track PSD for the last row), (3) the space-time PSD, (4) The spacetime PSD score available only in OSSE task.

Experiment	Algorithm	nRMSE Score	Effective Resolution			
			λ_a [km]	λ_r [km]	$\lambda_{\mathbf{x}}$ [km]	λ_t [days]
OSSE NADIR	OI	0.92	-	123	174	10.8
OSSE NADIR	MIOST	0.93	-	100	157	10.1
OSSE NADIR	BFNQG	0.93	-	88	139	10.4
OSSE NADIR	4DVarNet	**0.94**	-	**65**	**117**	**7.7**
OSSE SWOT	OI	0.92	-	106	139	11.7
OSSE SWOT	MIOST	0.94	-	88	131	10.1
OSSE SWOT	BFNQG	0.94	-	64	118	36.5
OSSE SWOT	4DVarNet	**0.96**	-	**47**	**77**	**5.6**
OSSE SST	Musti	0.95	-	46	138	4.1
OSSE SST	4DVarNet	**0.96**	-	**46**	**87**	**3.7**
OSE NADIR	OI	0.88	151	-	-	-
OSE NADIR	MIOST	0.90	135	-	-	-
OSE NADIR	BFNQG	0.88	122	-	-	-
OSE NADIR	ConvLSTM	0.89	113	-	-	-
OSE NADIR	4DVarNet	**0.91**	**98**	-	-	-

Table 5.2 – This table showcases all of the summary statistics for some methods for each of the data challenges listed in section 5.4. The summary statistics shown are the normalized RMSE and the effective resolution in the spectral domain. The spectral metrics for the effective resolution that were outlined in section 5.4 are: i) λ_a is the spatial score for the alongtrack PSD score, ii) λ_r is the spatial score for the isotropic PSD, iii) λ_x is the spatial score for space-time PSD score, and iv) λ_t is the temporal score for the space-time PSD score.

`OceanBench` Pipelines

For the four data challenges presented in the previous section, we used `OceanBench` pipelines to deliver a ML-ready benchmarking framework. We used the `hydra` and the geoprocessing tools outlined in section 5.3 with specialized routines for regridding the ocean satellite data to a uniformly gridded product and vice versa when necessary. Appendix 6.3 showcases an example of the hydra integration for the preprocessing pipeline. A key feature is the creation of a custom patcher for the appropriate geophysical variables using our `XRPatcher` tool, which is later integrated into custom datasets and dataloaders for the appropriate model architecture, e.g., coordinate-based or grid-based. We provide an example snippet of how this can be done easily in section 6.4. `OceanBench` also features some tools specific to the analysis of SSH. For example, physically-interpretable variables like geostrophic currents and relative vorticity, which

can be derived from first-order and second-order derivatives of the SSH, are essential for assessing the quality of the reconstructions generated by the models. Figure 5.2 showcases some fields of the most common physical variables used in the oceanography literature for the SSH-based analysis of sea surface dynamics.

Regarding the evaluation framework, we include domain-relevant performance metrics beyond the standard ML loss and accuracy functions. They account for the sampling patterns of the evaluation data. Spectral analytics are widely used in geoscience [60], and here, we consider spectral scores computed as the minimum spatial and temporal scales resolved by the reconstruction methods proposed in [60]. For example, figure 5.1 showcases how `OceanBench` generated the isotropic power spectrum and score and the space-time power spectrum decomposition and score. Table 5.2 outlines some standard and domain-specific scores for the experiments outlined in section 5.4.

5.5 Discussion

Framework Limitations

While we have advertised `OceanBench` as a unifying framework that provides standardized processing steps that comply with domain-expert standards, we also highlight some potential limitations that could hinder its adoption for the wider community.

Data Serving. We provide a few datasets but we omit some of the original simulations. We found that the original simulations are terabytes/petabytes of data which becomes infeasible for most modest users (even with adequate CPU resources). This is very big problem and if we want to have a bigger impact, we may need to do more close collaborations with specified platforms like the Marine Data Store [28, 29, 30, 41, 31, 40, 83] or the Climate Data Store [32, 42, 44, 43]. Furthermore, there are many people that will not be able to do a lot of heavy duty research which indirectly favours institutions with adequate resources and marginalizing others. This is also problematic as those communities tend to be the ones who need the most support from the products of such frameworks. We hope that leaving this open-source at least ensure that the knowledge is public.

Framework Dependence. The user has to "buy-into" the `hydra` framework to really take advantage of `OceanBench`. This adds a layer of abstraction and a new tool to learn. However, we designed the project so that high level usage does not require in-depth knowledge of the framework. In addition, we

hope that, despite the complexity of project, users will appreciate the flexibility and extensibility of this framework.

Lack of Metrics. We do not provide the most exhaustive list of metrics available with the ocean community. In fact, we also believe that many of these metrics are often poor and do not effectively assess the goodness of our reconstructions. However, we do provide a platform that will hopefully be useful and easy to implement new and improved metrics. Furthermore, having a wide range of metrics that are trusted across communities may help to improve the overall assessment of the different model performances [84].

Limited ML Scope. The framework does not support nor promote any machine learning methods and we lack any indication of comparing ML training and inference performance. However, we argue that a benchmark framework will allow us to effectively compare whichever ML methods are demonstratively the best which is a necessary preliminary step which offers users more flexibility in the long-run.

Broad Oceans Application Scope. We have targeted a broad ocean-application scope of state estimation. However, there may be more urgent applications such as maritime monitoring, object tracking, and general ocean health. However, we feel that many downstream applications require high-quality maps. In addition, those downstream applications tend to be very complicated and are not always straightforward to apply ML under those instances.

Full Pipeline Transparency. We use a lot of different `xarray`-specific packages which have different design principles, assumptions and implementations. This may give the users an illusion of simplicity and transparency to real-world use. However, there are many underlying assumptions within each of the packages that may occlude a lot of design decisions. Despite this limitation, we believe that being transparent about the processing steps and being consistent with the evaluation procedure will be beneficial for the ML research community.

Scalability. Scaling this to many terabytes or petabytes of data is easily the biggest limitation of the framework. In addition, we have only showcased demonstrations for 2D+T fields which are much less expensive than 3D+T fields.

Deployability. MLOPs has many wheels and it is not easy to integrate into existing systems. We offer no solutions to this. However, we believe that our framework is fully transparent in the assumptions and use cases which will

facilitate some adoption into operational systems where they can further modify it for their use cases (see the evolution of `WeatherBench` and `ClimateBench`).

Visualization Tools. We do not incorporate a high quality visualization tool that allows users to do pre- and post-analysis at a large scale. We do provide some simple visualization steps that are ML-relevant (see the GitHub repo) but it is very limited to ML standards. One solution is to interface our pipeline with the source of many ocean datasets, e.g. Climate Data Store [32] or Marine Data Store [28], then we can offset this task to them where they can offer better quality visualization tools.

Conclusion

The ocean community faces technological and algorithmic challenges to make the most of available observation and simulation datasets. In this context, recent studies evidence the critical role of ML schemes in reaching breakthroughs in our ability to monitor ocean dynamics for various space-time scales and processes. Nevertheless, domain-specific preprocessing steps and evaluation procedures slow down the uptake of ML toward real-world applications. The application considered here is SSH mapping which facilities the production of many crucial derived products that are used in many downstream tasks like subsequent modeling [6], ocean health monitoring [85, 86, 87] and maritime risk assessment [5].

We proposed four challenges towards a ML-ready benchmarking suite for ocean observation challenges. We outlined the inner workings `OceanBench` and demonstrated its usefulness by recreating some preprocessing and analysis pipelines from a few data challenges involving SSH interpolation. We firmly believe that the `OceanBench` platform is instrumental in fostering greater ML method adoption by the ocean community, while also rallying a larger portion of the ML community to tackle the ocean's scientific complexities.

Bibliography

[1] J. E. Johnson, Q. Febvre, A. Gorbunova, S. Metref, M. Ballarotta, J. L. Sommer, and R. Fablet, "OceanBench: The Sea Surface Height Edition," Sep. 2023.

[2] L. Cheng, K. Schuckmann, J. Abraham, K. Trenberth, M. Mann, L. Zanna, M. England, J. Zika, J. Fasullo, Y. Yu, Y. Pan, J. Zhu, E. Newsom, B. Bronselaer, and X. Lin, "Past and future ocean warming," *Nature Reviews Earth and Environment*, pp. 1–19, 10 2022.

[3] T. DeVries, "The ocean carbon cycle," *Annual Review of Environment and Resources*, vol. 47, no. 1, pp. 317–341, 2022. [Online]. Available: https://doi.org/10.1146/annurev-environ-120920-111307

[4] L. Resplandy, R. F. Keeling, Y. Eddebbar, M. Brooks, R. Wang, L. Bopp, M. C. Long, J. P. Dunne, W. Koeve, and A. Oschlies, "Quantification of ocean heat uptake from changes in atmospheric O2 and CO2 composition," *Scientific Reports*, vol. 9, no. 1, p. 20244, Dec. 2019.

[5] K. von Schuckmann, P.-Y. L. Traon, N. Smith, A. Pascual, P. Brasseur, K. Fennel, S. Djavidnia, S. Aaboe, E. A. Fanjul, E. Autret, L. Axell, R. Aznar, M. Benincasa, A. Bentamy, F. Boberg, R. Bourdallé-Badie, B. B. Nardelli, V. E. Brando, C. Bricaud, L.-A. Breivik, R. J. Brewin, A. Capet, A. Ceschin, S. Ciliberti, G. Cossarini, M. de Alfonso, A. de Pascual Collar, J. de Kloe, J. Deshayes, C. Desportes, M. Drévillon, Y. Drillet, R. Droghei, C. Dubois, O. Embury, H. Etienne, C. Fratianni, J. G. Lafuente, M. G. Sotillo, G. Garric, F. Gasparin, R. Gerin, S. Good, J. Gourrion, M. Grégoire, E. Greiner, S. Guinehut, E. Gutknecht, F. Hernandez, O. Hernandez, J. Høyer, L. Jackson, S. Jandt, S. Josey, M. Juza, J. Kennedy, Z. Kokkini, G. Korres, M. Kõuts, P. Lagemaa, T. Lavergne, B. le Cann, J.-F. Legeais, B. Lemieux-Dudon, B. Levier, V. Lien, I. Maljutenko, F. Manzano, M. Marcos, V. Marinova, S. Masina, E. Mauri, M. Mayer, A. Melet, F. Mélin, B. Meyssignac, M. Monier, M. Müller, S. Mulet, C. Naranjo, G. Notarstefano, A. Paulmier, B. P. Gomez, I. P. Gonzalez, E. Peneva, C. Perruche, K. A. Peterson,

N. Pinardi, A. Pisano, S. Pardo, P.-M. Poulain, R. P. Raj, U. Raudsepp, M. Ravdas, R. Reid, M.-H. Rio, S. Salon, A. Samuelsen, M. Sammartino, S. Sammartino, A. B. Sandø, R. Santoleri, S. Sathyendranath, J. She, S. Simoncelli, C. Solidoro, A. Stoffelen, A. Storto, T. Szerkely, S. Tamm, S. Tietsche, J. Tinker, J. Tintore, A. Trindade, D. van Zanten, L. Vandenbulcke, A. Verhoef, N. Verbrugge, L. Viktorsson, K. von Schuckmann, S. L. Wakelin, A. Zacharioudaki, and H. Zuo, "Copernicus marine service ocean state report," *Journal of Operational Oceanography*, vol. 11, no. sup1, pp. S1–S142, 2018. [Online]. Available: https://doi.org/10.1080/1755876X.2018.1489208

[6] M. Sonnewald, R. Lguensat, D. C. Jones, P. D. Dueben, J. Brajard, and V. Balaji, "Bridging observations, theory and numerical simulation of the ocean using machine learning," *Environmental Research Letters*, vol. 16, no. 7, p. 073008, July 2021.

[7] S. Abdalla, A. Abdeh Kolahchi, (...), and V. Zlotnicki, "Altimetry for the future: Building on 25 years of progress," *Advances in Space Research*, vol. 68, no. 2, pp. 319–363, 2021, 25 Years of Progress in Radar Altimetry. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0273117721000594

[8] G. Taburet, A. Sanchez-Roman, M. Ballarotta, M.-I. Pujol, J.-F. Legeais, F. Fournier, Y. Faugere, and G. Dibarboure, "Duacs dt2018: 25 years of reprocessed sea level altimetry products," *Ocean Science*, vol. 15, no. 5, pp. 1207–1224, 2019. [Online]. Available: https://os.copernicus.org/articles/15/1207/2019/

[9] P. Minnett, A. Alvera-Azcárate, T. Chin, G. Corlett, C. Gentemann, I. Karagali, X. Li, A. Marsouin, S. Marullo, E. Maturi, R. Santoleri, S. Saux Picart, M. Steele, and J. Vazquez-Cuervo, "Half a century of satellite remote sensing of sea-surface temperature," *Remote Sensing of Environment*, vol. 233, p. 111366, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425719303852

[10] S. Groom, S. Sathyendranath, Y. Ban, S. Bernard, R. Brewin, V. Brotas, C. Brockmann, P. Chauhan, J.-k. Choi, A. Chuprin, S. Ciavatta, P. Cipollini, C. Donlon, B. Franz, X. He, T. Hirata, T. Jackson, M. Kampel, H. Krasemann, S. Lavender, S. Pardo-Martinez, F. Mélin, T. Platt, R. Santoleri, J. Skakala, B. Schaeffer, M. Smith, F. Steinmetz,

A. Valente, and M. Wang, "Satellite ocean colour: Current status and future perspective," *Frontiers in Marine Science*, vol. 6, 2019. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmars.2019.00485

[11] K. Zhang, L. Huang, Z. Wei, C. An, and X. Lv, "Sea surface height data reconstruction via inter and intra layer features based on dual attention," *Neurocomputing*, vol. 545, p. 126313, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231223004368

[12] S. A. Martin, G. E. Manucharyan, and P. Klein, "Synthesizing sea surface temperature and satellite altimetry observations using deep learning improves the accuracy and resolution of gridded sea surface height anomalies," *Journal of Advances in Modeling Earth Systems*, vol. 15, no. 5, p. e2022MS003589, 2023, e2022MS003589 2022MS003589. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022MS003589

[13] T. Archambault, A. Filoche, A. Charantonnis, and D. Béréziat, "Multimodal Unsupervised Spatio-Temporal Interpolation of satellite ocean altimetry maps," in *VISAPP*, Lisboa, Portugal, Feb. 2023. [Online]. Available: https://hal.sorbonne-universite.fr/hal-03934647

[14] D. Kreuzberger, N. Kühl, and S. Hirschl, "Machine learning operations (mlops): Overview, definition, and architecture," 2022.

[15] G. Symeonidis, E. Nerantzis, A. Kazakis, and G. A. Papakostas, "Mlops - definitions, tools and challenges," *CoRR*, vol. abs/2201.00162, 2022. [Online]. Available: https://arxiv.org/abs/2201.00162

[16] H. Xiang, Q. Zou, M. A. Nawaz, X. Huang, F. Zhang, and H. Yu, "Deep learning for image inpainting: A survey," *Pattern Recognition*, vol. 134, p. 109046, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S003132032200526X

[17] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin, "Deep learning on image denoising: An overview," *Neural networks : the official journal of the International Neural Network Society*, vol. 131, pp. 251–275, 2019.

[18] R. S. Thakur, S. Chatterjee, R. N. Yadav, and L. Gupta, "Image de-noising with machine learning: A review," *IEEE Access*, vol. 9, pp. 93 338–93 363, 2021.

[19] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, pp. 3365–3387, 2019.

[20] W. Gilpin, "Chaos as an interpretable benchmark for forecasting and data-driven modelling," in *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, J. Vanschoren and S. Yeung, Eds., vol. 1. Curran, 2021. [Online]. Available: https://datasets-benchmarks-proceedings.neurips.cc/paper_files/paper/2021/file/ec5decca5ed3d6b8079e2e7e7bacc9f2-Paper-round2.pdf

[21] M. Takamoto, T. Praditia, R. Leiteritz, D. MacKinlay, F. Alesiani, D. Pflüger, and M. Niepert, "Pdebench: An extensive benchmark for scientific machine learning," in *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., vol. 35. Curran Associates, Inc., 2022, pp. 1596–1611. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2022/file/0a9747136d411fb83f0cf81820d44afb-Paper-Datasets_and_Benchmarks.pdf

[22] R. Abernathey, rochanotes, A. Ross, M. Jansen, Z. Li, F. J. Poulin, N. C. Constantinou, A. Sinha, D. Balwada, SalahKouhen, S. Jones, C. B. Rocha, C. L. P. Wolfe, C. Meng, H. van Kemenade, J. Bourbeau, J. Penn, J. Busecke, M. Bueti, and Tobias, "pyqg/pyqg: v0.7.2," May 2022. [Online]. Available: https://doi.org/10.5281/zenodo.6563667

[23] D. Kochkov, J. A. Smith, A. Alieva, Q. Wang, M. P. Brenner, and S. Hoyer, "Machine learning–accelerated computational fluid dynamics," *Proceedings of the National Academy of Sciences*, vol. 118, no. 21, 2021. [Online]. Available: https://www.pnas.org/content/118/21/e2101784118

[24] J. Anderson, T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Avellano, "The data assimilation research testbed: A community facility," *Bulletin of the American Meteorological Society*, vol. 90, no. 9, pp. 1283 – 1296, 2009. [Online]. Available: https://journals.ametsoc.org/view/journals/bams/90/9/2009bams2618_1.xml

[25] UCAR/NCAR/CISL/DAReS, "The data assimilation research testbed," https://keras.io, 2023.

[26] D. Häfner, R. L. Jacobsen, C. Eden, M. R. B. Kristensen, M. Jochum, R. Nuterman, and B. Vinter, "Veros v0.1 – a fast and versatile ocean simulator in pure python," *Geoscientific Model Development*, vol. 11, no. 8, pp. 3299–3312, 2018. [Online]. Available: https://gmd.copernicus.org/articles/11/3299/2018/

[27] A. Ramadhan, G. L. Wagner, C. Hill, J.-M. Campin, V. Churavy, T. Besard, A. Souza, A. Edelman, R. Ferrari, and J. Marshall, "Oceananigans.jl: Fast and friendly geophysical fluid dynamics on gpus," *Journal of Open Source Software*, vol. 5, no. 53, p. 2018, 2020. [Online]. Available: https://doi.org/10.21105/joss.02018

[28] E. C. M. S. I. (CMEMS), "Global ocean physics analysis and forecast."

[29] ——, "Global ocean biogeochemistry analysis and forecast."

[30] ——, "Global ocean ensemble physics reanalysis."

[31] ——, "Global ocean waves reanalysis."

[32] E. C. C. C. S. (CCCS), "Oras5 global ocean reanalysis monthly data from 1958 to present," 2021.

[33] M. A. Chamberlain, P. R. Oke, R. A. S. Fiedler, H. M. Beggs, G. B. Brassington, and P. Divakaran, "Next generation of bluelink ocean reanalysis with multiscale data assimilation: Bran2020," *Earth System Science Data*, vol. 13, no. 12, pp. 5663–5688, 2021. [Online]. Available: https://essd.copernicus.org/articles/13/5663/2021/

[34] D. W. Behringer, M. Ji, and A. Leetmaa, "An improved coupled model for enso prediction and implications for ocean initialization. part i: The ocean data assimilation system," *Monthly Weather Review*, vol. 126, no. 4, pp. 1013 – 1021, 1998. [Online]. Available: https://journals.ametsoc.org/view/journals/mwre/126/4/1520-0493_1998_126_1013_aicmfe_2.0.co_2.xml

[35] S. Rasp, P. D. Dueben, S. Scher, J. A. Weyn, S. Mouatadid, and N. Thuerey, "Weatherbench: A benchmark data set for data-driven weather forecasting," *Journal of Advances in Modeling Earth Systems*, vol. 12, no. 11, p. e2020MS002203, 2020, e2020MS002203 10.1029/2020MS002203. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002203

[36] D. Watson-Parris, Y. Rao, D. Olivié, Ø. Seland, P. Nowack, G. Camps-Valls, P. Stier, S. Bouabid, M. Dewey, E. Fons, J. Gonzalez, P. Harder, K. Jeggle, J. Lenhardt, P. Manshausen, M. Novitasari, L. Ricard, and C. Roesch, "Climatebench v1.0: A benchmark for data-driven climate projections," *Journal of Advances in Modeling Earth Systems*, vol. 14, no. 10, p. e2021MS002954, 2022, e2021MS002954 2021MS002954. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021MS002954

[37] S. Ashkboos, L. Huang, N. Dryden, T. Ben-Nun, P. Dueben, L. Gianinazzi, L. Kummer, and T. Hoefler, "Ens-10: A dataset for post-processing ensemble weather forecasts," in *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., vol. 35. Curran Associates, Inc., 2022, pp. 21 974–21 987. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2022/file/89e44582fd28ddfea1ea4dcb0ebbf4b0-Paper-Datasets_and_Benchmarks.pdf

[38] R. R. Lam, A. Sanchez-Gonzalez, M. Willson, P. Wirnsberger, M. Fortunato, A. Pritzel, S. V. Ravuri, T. Ewalds, F. Alet, Z. Eaton-Rosen, W. Hu, A. Merose, S. Hoyer, G. Holland, J. Stott, O. Vinyals, S. Mohamed, and P. W. Battaglia, "Graphcast: Learning skillful medium-range global weather forecasting," *ArXiv*, vol. abs/2212.12794, 2022.

[39] J. Pathak, S. Subramanian, P. Harrington, S. Raja, A. Chattopadhyay, M. Mardani, T. Kurth, D. Hall, Z. Li, K. Azizzadenesheli, P. Hassanzadeh, K. Kashinath, and A. Anandkumar, "Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators," 2022.

[40] E. C. M. S. I. (CMEMS), "Global ocean alon-track l3 sea surface heights reprocessed (1993-ongoing) tailored for data assimilation."

[41] ——, "Global ocean- in-situ near real time observations of ocean currents."

[42] E. C. C. C. S. (CCCS), "Sea surface temperature daily data from 1981 to present derived from satellite observations."

[43] ——, "Sea surface temperature daily gridded data from 1981 to 2016 derived from a multi-product satellite-based ensemble."

[44] ——, "Ocean colour daily data from 1997 to present derived from satellite observations."

[45] A. P. S. Wong, S. E. Wijffels, S. C. Riser, S. Pouliquen, S. Hosoda, D. Roemmich, J. Gilson, G. C. Johnson, K. Martini, D. J. Murphy, M. Scanderbeg, T. V. S. U. Bhaskar, J. J. H. Buck, F. Merceur, T. Carval, G. Maze, C. Cabanes, X. André, N. Poffa, I. Yashayaev, P. M. Barker, S. Guinehut, M. Belbéoch, M. Ignaszewski, M. O. Baringer, C. Schmid, J. M. Lyman, K. E. McTaggart, S. G. Purkey, N. Zilberman, M. B. Alkire, D. Swift, W. B. Owens, S. R. Jayne, C. Hersh, P. Robbins, D. West-Mack, F. Bahr, S. Yoshida, P. J. H. Sutton, R. Cancouët, C. Coatanoan, D. Dobbler, A. G. Juan, J. Gourrion, N. Kolodziejczyk, V. Bernard, B. Bourlès, H. Claustre, F. D'Ortenzio, S. Le Reste, P.-Y. Le Traon, J.-P. Rannou, C. Saout-Grit, S. Speich, V. Thierry, N. Verbrugge, I. M. Angel-Benavides, B. Klein, G. Notarstefano, P.-M. Poulain, P. Vélez-Belchí, T. Suga, K. Ando, N. Iwasaska, T. Kobayashi, S. Masuda, E. Oka, K. Sato, T. Nakamura, K. Sato, Y. Takatsuki, T. Yoshida, R. Cowley, J. L. Lovell, P. R. Oke, E. M. van Wijk, F. Carse, M. Donnelly, W. J. Gould, K. Gowers, B. A. King, S. G. Loch, M. Mowat, J. Turton, E. P. Rama Rao, M. Ravichandran, H. J. Freeland, I. Gaboury, D. Gilbert, B. J. W. Greenan, M. Ouellet, T. Ross, A. Tran, M. Dong, Z. Liu, J. Xu, K. Kang, H. Jo, S.-D. Kim, and H.-M. Park, "Argo data 1999–2019: Two million temperature-salinity profiles and subsurface velocity observations from a global array of profiling floats," *Frontiers in Marine Science*, vol. 7, 2020. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmars.2020.00700

[46] D. C. E. Bakker, B. Pfeil, C. S. Landa, N. Metzl, K. M. O'Brien, A. Olsen, K. Smith, C. Cosca, S. Harasawa, S. D. Jones, S. Nakaoka, Y. Nojiri, U. Schuster, T. Steinhoff, C. Sweeney, T. Takahashi, B. Tilbrook, C. Wada, R. Wanninkhof, S. R. Alin, C. F. Balestrini, L. Barbero, N. R. Bates, A. A. Bianchi, F. Bonou, J. Boutin, Y. Bozec, E. F. Burger, W.-J. Cai, R. D. Castle, L. Chen, M. Chierici, K. Currie, W. Evans, C. Featherstone, R. A. Feely, A. Fransson, C. Goyet, N. Greenwood, L. Gregor, S. Hankin, N. J. Hardman-Mountford, J. Harlay, J. Hauck, M. Hoppema, M. P. Humphreys, C. W. Hunt, B. Huss, J. S. P. Ibánhez, T. Johannessen, R. Keeling, V. Kitidis, A. Körtzinger, A. Kozyr, E. Krasakopoulou, A. Kuwata, P. Landschützer, S. K. Lauvset, N. Lefèvre,

C. Lo Monaco, A. Manke, J. T. Mathis, L. Merlivat, F. J. Millero, P. M. S. Monteiro, D. R. Munro, A. Murata, T. Newberger, A. M. Omar, T. Ono, K. Paterson, D. Pearce, D. Pierrot, L. L. Robbins, S. Saito, J. Salisbury, R. Schlitzer, B. Schneider, R. Schweitzer, R. Sieger, I. Skjelvan, K. F. Sullivan, S. C. Sutherland, A. J. Sutton, K. Tadokoro, M. Telszewski, M. Tuma, S. M. A. C. van Heuven, D. Vandemark, B. Ward, A. J. Watson, and S. Xu, "A multi-decade record of high-quality fco_2 data in version 3 of the surface ocean co_2 atlas (socat)," *Earth System Science Data*, vol. 8, no. 2, pp. 383–413, 2016. [Online]. Available: https://essd.copernicus.org/articles/8/383/2016/

[47] E. B. Community, "Jupyter book," Feb. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.4539666

[48] M. Ballarotta and F. L. Guillou, "ocean-data-challenges/2021a_SSH_mapping_OSE: Material for SSH mapping OSE data challenge," Sep. 2021. [Online]. Available: https://doi.org/10.5281/zenodo.5511905

[49] M. Ballarotta, E. Cosme, and A. Albert, "ocean-data-challenges/2020a_SSH_mapping_NATL60: Material for SSH mapping data challenge," Sep. 2020, This challenge is part of the BOOST-SWOT project funded by ANR (project number ANR-17-CE01-0009-01) and a contribution to the MIDAS project funded by CNES for the NASA/CNES SWOT Science Team. [Online]. Available: https://doi.org/10.5281/zenodo.4045400

[50] R. Kuprieiev, skshetry, P. Rowlands, D. Petrov, P. Redzyński, C. da Costa-Luis, D. de la Iglesia Castro, A. Schepanovski, Gao, I. Shcheklein, B. Taskaya, J. Orpinel, D. Berenbaum, F. Santos, daniele, R. Lamy, A. Sharma, Z. Kaimuldenov, D. Hodovic, N. Kodenko, A. Grigorev, Earl, N. Dash, G. Vyshnya, maykulkarni, M. Hora, Vera, and S. Mangal, "Dvc: Data version control - git for data & models," May 2023. [Online]. Available: https://doi.org/10.5281/zenodo.7990791

[51] S. Hoyer and J. Hamman, "xarray: N-D labeled arrays and datasets in Python," *Journal of Open Research Software*, vol. 5, no. 1, 2017. [Online]. Available: https://doi.org/10.5334/jors.148

[52] O. Yadan, "Hydra - a framework for elegantly configuring complex

applications," Github, 2019. [Online]. Available: https://github.com/facebookresearch/hydra

[53] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," pp. 8024–8035, 2019. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[54] W. Falcon and The PyTorch Lightning team, "PyTorch Lightning," Mar. 2019. [Online]. Available: https://github.com/Lightning-AI/lightning

[55] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning." in *OSDI*, vol. 16, 2016, pp. 265–283.

[56] F. Chollet *et al.*, "Keras," https://keras.io, 2015.

[57] N. Sambasivan, S. Kapania, H. Highfill, D. Akrong, P. Paritosh, and L. M. Aroyo, ""everyone wants to do the model work, not the data work": Data cascades in high-stakes ai," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: https://doi.org/10.1145/3411764.3445518

[58] S. Cheng, C. Quilodrán-Casas, S. Ouala, A. Farchi, C. Liu, P. Tandeo, R. Fablet, D. Lucor, B. Iooss, J. Brajard *et al.*, "Machine learning with data assimilation and uncertainty quantification for dynamical systems: a review," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 6, pp. 1361–1387, 2023.

[59] A. Carrassi, M. Bocquet, L. Bertino, and G. Evensen, "Data assimilation in the geosciences: An overview of methods, issues, and perspectives," *WIREs Climate Change*, vol. 9, no. 5, p. e535, 2018. [Online]. Available: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcc.535

[60] F. L. Guillou, S. Metref, E. Cosme, C. Ubelmann, M. Ballarotta, J. L. Sommer, and J. Verron, "Mapping altimetry in the forthcoming swot era

by back-and-forth nudging a one-layer quasigeostrophic model," *Journal of Atmospheric and Oceanic Technology*, vol. 38, no. 4, pp. 697 – 710, 2021. [Online]. Available: https://journals.ametsoc.org/view/journals/atot/38/4/JTECH-D-20-0104.1.xml

[61] J. E. Johnson, R. Lguensat, R. Fablet, E. Cosme, and J. L. Sommer, "Neural fields for fast and scalable interpolation of geophysical ocean variables," *ArXiv*, vol. abs/2211.10444, 2022.

[62] C. Ubelmann, G. Dibarboure, L. Gaultier, A. Ponte, F. Ardhuin, M. Ballarotta, and Y. Faugère, "Reconstructing ocean surface current combining altimetry and future spaceborne doppler data," *Journal of Geophysical Research: Oceans*, vol. 126, no. 3, p. e2020JC016560, 2021, e2020JC016560 2020JC016560. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020JC016560

[63] R. Fablet, Q. Febvre, and B. Chapron, "Multimodal 4dvarnets for the reconstruction of sea surface dynamics from sst-ssh synergies," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2022.

[64] M. Beauchamp, Q. Febvre, H. Georgentum, and R. Fablet, "4dvarnet-ssh: end-to-end learning of variational interpolation schemes for nadir and wide-swath satellite altimetry," *Geoscientific Model Development*, 2022.

[65] L. Espeholt, S. Agrawal, C. Sønderby, M. Kumar, J. Heek, C. Bromberg, C. Gazen, R. Carver, M. Andrychowicz, J. Hickey, A. Bell, and N. Kalchbrenner, "Deep learning for twelve hour precipitation forecasts," *Nature Communications*, vol. 13, no. 1, p. 5145, 2022. [Online]. Available: https://doi.org/10.1038/s41467-022-32483-x

[66] R. Berlinghieri, B. L. Trippe, D. R. Burt, R. J. Giordano, K. Srinivasan, T. Özgökmen, J. Xia, and T. Broderick, "Gaussian processes at the helm(holtz): A more fluid model for ocean currents," in *Proceedings of the 40th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, Eds., vol. 202. PMLR, 23–29 Jul 2023, pp. 2113–2163. [Online]. Available: https://proceedings.mlr.press/v202/berlinghieri23a.html

[67] H. Frezat, J. Le Sommer, R. Fablet, G. Balarac, and R. Lguensat, "A posteriori learning for quasi-geostrophic turbulence parametrization," *Journal*

of Advances in Modeling Earth Systems, vol. 14, no. 11, p. e2022MS003124, 2022, e2022MS003124 2022MS003124. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022MS003124

[68] L. Zanna and T. Bolton, "Data-driven equation discovery of ocean mesoscale closures," *Geophysical Research Letters*, vol. 47, no. 17, p. e2020GL088376, 2020.

[69] S. A. Martin, G. E. Manucharyan, and P. Klein, "Synthesizing sea surface temperature and satellite altimetry observations using deep learning improves the accuracy and resolution of gridded sea surface height anomalies," *Journal of Advances in Modeling Earth Systems*, vol. 15, no. 5, p. e2022MS003589, 2023, e2022MS003589 2022MS003589. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022MS003589

[70] H. Frezat, G. Balarac, J. Le Sommer, R. Fablet, and R. Lguensat, "Physical invariance in neural networks for subgrid-scale scalar flux modeling," *Phys. Rev. Fluids*, vol. 6, p. 024607, Feb 2021. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevFluids.6.024607

[71] M. Bonavita and P. Laloyaux, "Estimating model error covariances with artificial neural networks," 2022.

[72] P. Laloyaux, T. Kurth, P. D. Dueben, and D. Hall, "Deep learning to estimate model biases in an operational nwp assimilation system," *Journal of Advances in Modeling Earth Systems*, vol. 14, no. 6, p. e2022MS003016, 2022, e2022MS003016 2022MS003016. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022MS003016

[73] A. Ajayi, J. Le Sommer, E. Chassignet, J.-M. Molines, X. Xu, A. Albert, and E. Cosme, "Spatial and temporal variability of the north atlantic eddy field from two kilometric-resolution ocean models," *Journal of Geophysical Research: Oceans*, vol. 125, no. 5, p. e2019JC015827, 2020, e2019JC015827 10.1029/2019JC015827. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JC015827

[74] S. Kornblith, M. Norouzi, H. Lee, and G. Hinton, "Similarity of neural network representations revisited," in *Proceedings of the 36th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, K. Chaudhuri and R. Salakhutdinov, Eds.,

vol. 97. PMLR, 09–15 Jun 2019, pp. 3519–3529. [Online]. Available: https://proceedings.mlr.press/v97/kornblith19a.html

[75] J. E. Johnson, V. Laparra, M. Piles, and G. Camps-Valls, "Gaussianizing the earth: Multidimensional information measures for earth data analysis," *IEEE Geoscience and Remote Sensing Magazine*, vol. 9, pp. 191–208, 2020.

[76] V. Laparra, J. E. Johnson, G. Camps-Valls, R. Santos-Rodríguez, and J. Malo, "Information theory measures via multidimensional gaussianization," *ArXiv*, vol. abs/2010.03807, 2020.

[77] T. Uchida, Q. Jamet, A. C. Poje, N. Wienders, W. K. Dewar, and B. Deremble, "Wavelet-based wavenumber spectral estimate of eddy kinetic energy: Idealized quasi-geostrophic flow," *Journal of Advances in Modeling Earth Systems*, vol. 15, no. 3, p. e2022MS003399, 2023, e2022MS003399 2022MS003399. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022MS003399

[78] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

[79] A. Okubo, "Horizontal dispersion of floatable particles in the vicinity of velocity singularities such as convergences," *Deep Sea Research and Oceanographic Abstracts*, vol. 17, no. 3, pp. 445–454, 1970. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0011747170900598

[80] J. Weiss, "The dynamics of enstrophy transfer in two-dimensional hydrodynamics," *Physica D: Nonlinear Phenomena*, vol. 48, no. 2, pp. 273–294, 1991. [Online]. Available: https://www.sciencedirect.com/science/article/pii/016727899190088Q

[81] B. K. Shivamoggi, G. J. F. van Heijst, and L. P. J. Kamp, "The okubo–weiss criterion in hydrodynamic flows: geometric aspects and further extension," *Fluid Dynamics Research*, vol. 54, no. 1, p. 015505, jan 2022. [Online]. Available: https://dx.doi.org/10.1088/1873-7005/ac495d

[82] L. Gaultier, C. Ubelmann, and L.-L. Fu, "The challenge of using future swot data for oceanic field reconstruction," *Journal of Atmospheric and Oceanic Technology*, vol. 33, no. 1, pp. 119 – 126, 2016. [Online]. Available: https://journals.ametsoc.org/view/journals/atot/33/1/jtech-d-15-0160_1.xml

[83] E. C. M. S. I. (CMEMS), "Global ocean gridded l 4 sea surface heights and derived variables reprocessed 1993 ongoing."

[84] M. Gauch, F. Kratzert, O. Gilon, H. Gupta, J. Mai, G. Nearing, B. Tolson, S. Hochreiter, and D. Klotz, "In defense of metrics: Metrics sufficiently encode typical human preferences regarding hydrological model performance," *Water Resources Research*, vol. 59, no. 6, p. e2022WR033918, 2023, e2022WR033918 2022WR033918. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022WR033918

[85] D. Tuia, B. Kellenberger, S. Beery, B. R. Costelloe, S. Zuffi, B. Risse, A. Mathis, M. W. Mathis, F. van Langevelde, T. Burghardt, R. Kays, H. Klinck, M. Wikelski, I. D. Couzin, G. van Horn, M. C. Crofoot, C. V. Stewart, and T. Berger-Wolf, "Perspectives in machine learning for wildlife conservation," *Nature Communications*, vol. 13, no. 1, p. 792, 2022. [Online]. Available: https://doi.org/10.1038/s41467-022-27980-y

[86] C. S. Longo, M. Frazier, S. C. Doney, J. E. Rheuban, J. M. Humberstone, and B. S. Halpern, "Using the ocean health index to identify opportunities and challenges to improving southern ocean ecosystem health," *Frontiers in Marine Science*, vol. 4, 2017. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmars.2017.00020

[87] A. Franke, T. Blenckner, C. M. Duarte, K. Ott, L. E. Fleming, A. Antia, T. B. Reusch, C. Bertram, J. Hein, U. Kronfeld-Goharani, J. Dierking, A. Kuhn, C. Sato, E. van Doorn, M. Wall, M. Schartau, R. Karez, L. Crowder, D. Keller, A. Engel, U. Hentschel, and E. Prigge, "Operationalizing ocean health: Toward integrated research on ocean health and recovery to achieve ocean sustainability," *One Earth*, vol. 2, no. 6, pp. 557–565, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590332220302499

Chapter 6

Conclusions and perspectives

We review in this chapter the primary contributions outlined in this manuscript and the future avenues of research they open.

6.1 Contributions Summary

This thesis is part of a broader movement towards developing deep learning methods to address observation challenges in ocean science. It emphasizes altimetry applications, especially in the context of the recent launch of the SWOT mission.

The first contribution highlights the successful application of deep learning for bias correction of simulated SWOT observation data. While standard deep learning architectures struggled to differentiate fine SSH signatures from high amplitude bias, we demonstrated that deep learning methods could be tailored to suit the unique characteristics of altimetry data. We employed SWOT mission's error specifications to craft a custom architecture focused on calibrating SWOT's correlated errors. This study is promising, yet the method developed was calibrated and assessed using simulated data, bringing up questions about its applicability to actual SWOT observations.

The second study delves into how learning-based altimetry methods, once calibrated on simulated data, can be applied to real data. We evaluated the 4dVarNet mapping schemes on real altimetry after calibration on simulated data. The findings indicate strong generalization capabilities even with coarse simulations, while more accurate simulations enhance the mapping performance. The results introduce interesting avenues in exploring the use of numerical simulation for training models for real-world applications.

The initial two studies shed light on the potential of applying learning-based approaches to ocean science's observational challenges. Yet, they also spotlight the complexities in melding expertise in observation, simulation data,

deep learning techniques, and domain-specific evaluation methodologies. This spurred the creation of the specialized toolset, Oceanbench, aiming to narrow the gap between deep learning and ocean science experts. Oceanbench enables ocean scientists to flexibly design evaluation setups using data and metrics. These setups come with the essential tools for deep learning practitioners to access and prepare the data in view of training their models. The first iteration presented in this manuscript focuses on sea surface height interpolation but has been thought to be extensible to other ocean observation challenges.

6.2 Current Limitations and perspectives

Several avenues can be explored to further extend the work presented in this thesis.

Global SSH Estimation. The research presented here pertains to particular region and periods over the Gulfstream which is not representative of the different global regimes. This use-case contains a dynamical regime and a well-studied area which has some importance for specific communities and is small enough to mitigate the problems involving scale. However confirming the robustness of deep learning schemes on the global ocean is a necessary step to validate their potential. On this note, Beauchamp et al. (2023)[1] have made promising strides by applying the 4dVarNet to other regions within the North Atlantic.

Toward operational products. Real altimetry use-cases involve global and/or high-resolution data. This involves dealing with very high-dimensional spatio-temporal global state-space. In practice, the necessity for the scalability of the method is of paramount importance. Transitioning the methods demonstrated in this thesis to functional products would entail considerable scaling challenges. These encompass both scientific aspects, such as dealing with earth earth geometry [2], coastlines and varying ocean regimes, and engineering concerns like handling large datasets for the training and assessment of the models.

Beyond altimetry. This thesis centers on SSH, a surface field that is relatively well-observed in the realm of ocean quantities. Exploring other quantities, observed through different instruments, with different sampling or not directly-observed would introduce many more challenges requiring domain-informed problem specifications that deep learning could be applied to. Therefore the work presented here is still far away from actual reanalysis[3] and

forecasting goals of full state estimation. Achieving more ambitious estimation challenge will require a lot of interdisciplinary work across communities and we hope the work done with Oceanbench can help to that regard.

Deep learning interpretability. While deep learning methods offer promising results, it's understandable to remain cautious. Concerns regarding the interpretability of deep learning models and their robustness compared to physically descriptive systems are valid points of discussion. Two potential paths forward can help address these concerns. First, emphasizing the importance of quantifying the uncertainty [4, 5] associated with model estimations. Such uncertainty quantification (UQ) is crucial when addressing ill-posed inverse problems and can play a significant role in bolstering confidence in the results. Second, exploring the realm of physics-informed deep learning and theory-guided data-science, which marries our physical understanding of the ocean with the adaptive nature of deep learning models. Existing studies have dabbled in approaches involving dynamical systems[6, 7], which, while usually simpler than the ocean, can provide valuable insights for ocean observation applications.

BIBLIOGRAPHY

[1] M. Beauchamp, Q. Febvre, H. Georgenthum, and R. Fablet, "4DVarNet-SSH: End-to-end learning of variational interpolation schemes for nadir and wide-swath satellite altimetry," *Geoscientific Model Development*, vol. 16, no. 8, pp. 2119–2147, Apr. 2023.

[2] B. Bonev, T. Kurth, C. Hundt, J. Pathak, M. Baust, K. Kashinath, and A. Anandkumar, "Spherical fourier neural operators: Learning stable dynamics on the sphere," 2023.

[3] L. Jean-Michel, G. Eric, B.-B. Romain, G. Gilles, M. Angélique, D. Marie, B. Clément, H. Mathieu, L. G. Olivier, R. Charly, C. Tony, T. Charles-Emmanuel, G. Florent, R. Giovanni, B. Mounir, D. Yann, and L. T. Pierre-Yves, "The copernicus global 1/12 ° oceanic and sea ice glorys12 reanalysis," *Frontiers in Earth Science*, vol. 9, 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/feart.2021.698876

[4] A. Carrassi, M. Bocquet, L. Bertino, and G. Evensen, "Data assimilation in the geosciences: An overview of methods, issues, and perspectives," *WIREs Climate Change*, vol. 9, no. 5, p. e535, 2018. [Online]. Available: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcc.535

[5] M. Beauchamp, Q. Febvre, and R. Fablet, "Ensemble-based 4DVarNet uncertainty quantification for the reconstruction of sea surface height dynamics," *Environmental Data Science*, vol. 2, p. e18, Jan. 2023.

[6] A. Karpatne, G. Atluri, J. H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, and V. Kumar, "Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 10, pp. 2318–2331, Oct. 2017.

[7] A. Farchi, M. Chrust, M. Bocquet, P. Laloyaux, and M. Bonavita, "Online Model Error Correction With Neural Networks in the Incremental 4D-Var Framework," *Journal of Advances in Modeling Earth Systems*, vol. 15, no. 9, p. e2022MS003474, 2023.

List of publications

International Journals

Ongoing review as main author

— Febvre, Q., Sommer, J. L., Ubelmann, C., & Fablet, R. (2023, September 19). Training neural mapping schemes for satellite altimetry with simulation data. arXiv. https://doi.org/10.48550/arXiv.2309.14350 (Submitted at Journal of Advances in Modelling Earth Systems)

— Febvre, Q., Ubelmann, C., Sommer, J. L., & Fablet, R. (2023). Scale-aware neural calibration for wide swath altimetry observations. ArXiv, abs/2302.04497. (Submitted at IEEE Transactions on Geoscience and Remote Sensing)

Accepted as contributor

— Beauchamp, M., Febvre, Q., & Fablet, R. (2023). Ensemble-based 4DVar-Net uncertainty quantification for the reconstruction of sea surface height dynamics. Environmental Data Science, 2, e18. https://doi.org/10.1017/eds.2023.19

— Fablet, R., Febvre, Q., & Chapron, B. (2023). Multimodal 4DVarNets for the Reconstruction of Sea Surface Dynamics From SST-SSH Synergies. IEEE Transactions on Geoscience and Remote Sensing, 61, 1–14. https://doi.org/10.1109/TGRS.2023.3268006

— Beauchamp, M., Febvre, Q., Georgenthum, H., & Fablet, R. (2023). 4DVarNet-SSH: end-to-end learning of variational interpolation schemes for nadir and wide-swath satellite altimetry. Geoscientific Model Development, 16(8), 2119–2147. https://doi.org/10.5194/gmd-16-2119-2023

— Beauchamp, M., Febvre, Q., Georgentum, H., & Fablet, R. (2022). 4DVarNet-SSH: end-to-end learning of variational interpolation schemes for nadir and wide-swath satellite altimetry. Geoscientific Model Development.

— Fablet, R., Amar, M. M., Febvre, Q., Beauchamp, M., & Chapron, B. (2021). END-TO-END PHYSICS-INFORMED REPRESENTATION LEARNING FOR SATELLITE OCEAN REMOTE SENSING DATA:

APPLICATIONS TO SATELLITE ALTIMETRY AND SEA SURFACE CURRENTS. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, V-3–2021, 295–302. https://doi.org/10.5194/isprs-annals-V-3-2021-295-2021

International Conferences

Proceedings as first author or co-author

— Johnson, J. E., Febvre, Q., Gorbunova, A., Metref, S., Ballarotta, M., Sommer, J. L., & Fablet, R. (2023, September 27). OceanBench: The Sea Surface Height Edition. arXiv. https://doi.org/10.48550/arXiv.2309.15599 (Accepted at NEURIPS 2023)

— Febvre, Q., Fablet, R., Sommer, J. L., & Ubelmann, C. (2022). Joint calibration and mapping of satellite altimetry data using trainable variational models. In ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1536–1540). https://doi.org/10.1109/ICASSP43922.2022.9746889

Oral presentation (as contributor)

— Febvre, Q., Le Sommer, J., Ubelmann, C., Fablet, R., Beauchamp, M., (2023), Simulation-based learning of neural interpolation methods for the mapping of real altimetry data, Climate Informatics 2023

— Beauchamp, M., Febvre, Q., Fablet, R., (2023) Ensemble-based 4DVar-Net uncertainty quantification for the reconstruction of Sea Surface Height dynamics, Climate Informatics 2023

— Febvre, Q., Le Sommer, J., Ubelmann, C., Fablet, R., Benaichouche, S., (2023), Learning operational altimetry mapping from ocean models, EGU General Assembly 2023

— Febvre, Q., Le Sommer, J., Ubelmann, C., Fablet, R., Joint calibration and mapping of satellite altimetry data using trainable variaitional models, EGU General Assembly 2022

Poster presentation (as contributor)

— Febvre, Q., Le Sommer, J., Ubelmann, C., Fablet, R., (2023), Simulation-based learning of neural interpolation methods for the mapping of real altimetry data, IEEE IGARSS 2023

— Febvre, Q., Ubelmann, C., Le Sommer, J., Fablet, R., (2023), Scale-aware neural calibration for wide swath altimetry observations, IEEE IGARSS 2023

— Dorffer, C., Jourdin, F., Mouillot, D., Devillers, R., Fablet, R., Febvre, Q., (2023) Data-driven reconstruction of sea surface turbidity dynamics with 4dvarnet neural scheme applied to gappy satellite images, IEEE IGARSS 2023

Appendix

OceanBench: The Sea Surface Height Edition - Supplementary Material

6.3 Use Case I: Hydra Recipes

This framework has drastically reduced the overhead for the ML researcher while also enhancing the reprducibility and replicability of the preprocessing steps. In this section we showcase a few examples for how one can use oceanbench in conjunction with hydra to provide recipes for some standard processes.

Task Recipe

In this example, we showcase how we define an interpolation task for the OSE NADIR data challenge. We need to state the list of datasets available and specify which datasets are to be using for training and testings. We also specify the spatial region we would like to train on and the train-test period. There are a few simple changes one could do here to extend this task provided that one has uploaded standardized data that follows our set conventions. For example, for this interpolation task, the test period is a complete subset of the train period but one could imagine a forecasting task whereby the test period is at a completely different time period. Similarly, for this task, the train-test domain is the same but we could easily change the region of interest to see how the models perform in a completely different domain.

```yaml
#@package _global_.task
outputs:
    # name of data challenge
    name: DC2021 OSE Gulfstream
    # list of datasets and locations
    data:
      train: # train data list
        alg: ${....data.outputs.alg}
        h2g: ${....data.outputs.h2g}
        j2g: ${....data.outputs.j2g}
        j2n: ${....data.outputs.j2n}
        j3: ${....data.outputs.j3}
        s3a: ${....data.outputs.s3a}
      test: # test data list
        c2: ${....data.outputs.c2}
    # spatial region specification
    domain: {lat: [33, 43], lon: [-65, -55]}
    # temporal period specification
    splits: {
        test: ['2017-01-01', '2017-12-31'],
        train: ['2016-12-01', '2018-01-31']
    }
```

Listing 1 – This is a `.yaml` which showcases how we can communicate with `Hydra` framework to list a predefined set of specifications for the spatial region and the temporal period. This is an interpolation task for the OSE NADIR data challenge listed in section 5.4.

GeoProcessing Recipe

In this example, we showcase how one can pipe a sequential transformation through the hydra framework. In this example, we open the dataset, validate the coordinates to comply to our standards, select the region of interest, subset the data, regrid the alongtrack data to a uniform grid, and save the data to a netcdf file. See the listing 2 for more information.

Evaluation Recipe - OSSE

In this example, we showcase how one can use hydra to do the evaluation procedure. This is the same evaluation procedure that is used to evaluate the effectiveness of the OSSE NADIR experiment. From code snippet 2, we see that we choose which target function to initialize and we choose the data directory

```yaml
# Target Function to initialize
_target_: "oceanbench._src.dataset.pipe"
# netcdf file to be loaded
inp: "${data_directory}/nadir_tracks.nc"
# sequential transformations to be applied
fns:
    # Load Dataset
    - {_target_: "xarray.open_dataset", _partial_: True}
    # Validate LatLonTime Coordinates
    - {_target_: "oceanbench.validate_latlon", _partial_: True}
    - {_target_: "oceanbench.validate_time", _partial_: True}
    # Select Specific Region (Spatial | Temporal)
    - {_target_: "xarray.Dataset.sel", args: ${domain}, _partial_: True}
    # Take Subset of Data
    - {_target_: "oceanbench.subset", num_samples: 1500, _partial_: True}
    # Regridding (AlongTrack -> Uniform Grid)
    - {
        _target_: "oceanbench.regrid",
        target_grid: ${grid.high_res},
        _partial_: True
      }
    # Save Dataset
    - {
        _target_: "xarray.Dataset.to_netcdf",
        save_name: "demo.nc",
        _partial_: True
      }
```

Listing 2 – This is a `.yaml` which showcases how we can communicate with `Hydra` framework to list a predefined set of transformations to be *piped* through sequential. In this example, we showcase some standard pre-processing strategies to be saved to another netcdf file.

where the `.netcdf` file is located. Then, we pipe some transformations for the `.netcdf` file: 1) validate the spatiotemporal coordinates, 2) we select the evaluation region, 3) we regrid it to the target get, 4) we fill in the nans with a Gauss-Seidel procedure, 5) we rescale the coordinates to be in meters and days, and 6) we perform the isotropic power spectrum transformation to get the effective resolution outlined in section 5.4.

```yaml
# Target Function to initialize
_target_: "oceanbench._src.dataset.pipe"
# netcdf file to be loaded
inp: "${data_directory}/ml_result.nc"
# sequential transformations to be applied
fns:
    # Load Dataset
    - {_target_: "xarray.open_dataset", _partial_: True}
    # Validate LatLonTime Coordinates
    - {_target_: "oceanbench.validate_latlon", _partial_: True}
    - {_target_: "oceanbench.validate_time", _partial_: True}
    # Select Specific Region (Spatial | Temporal)
    - {_target_: "xarray.Dataset.sel", args: ${domain}, _partial_: True}
    # Regridding (Uniform Grid -> Uniform Grid)
    - {_target_: "oceanbench.regrid",
       target_grid: ${grid.reference}, _partial_: True}
    # Fill NANS (around the corners)
    - {_target_: "oceanbench.fill_nans",
       method: "gauss_seidel", _partial_: True}
    # Coordinate Change (degree -> meters, ns -> days)
    - {_target_: "oceanbench.latlon_deg2m", _partial_: True}
    - {_target_: "oceanbench.time_rescale",
       freq: 1, unit: "days", _partial_: True}
    # Calculate Isotropic Power Spectrum
    - {_target_: "oceanbench.power_spectrum_isotropic",
       reference: ${grid.reference}, _partial_: True}
    # Calculate Resolved Spatial Scale
    - {_target_: "oceanbench.resolved_scale", _partial_: True}
    # Save Dataset
    - {_target_: "xarray.Dataset.to_netcdf",
       save_name: "ml_result_psd.nc", _partial_: True}
```

Listing 3 – This is a `.yaml` which showcases how we can communicate with `Hydra` framework to list a predefined set of transformations to be *piped* through sequential. In this example, we showcase some standard pre-processing strategies to be saved to another netcdf file.

6.4 Use Case II: XRPatcher

There are many usecases for the `XRPatcher`. For example, we can do 1D Time chunking, 2D Spatial-Temporal Patches, or 3D Spatial-Temporal Cubes.

```python
import xarray as xr
import torch
import itertools
from oceanbench import XRPatcher
# Easy Integration with PyTorch Datasets (and DataLoaders)
class XRTorchDataset(torch.utils.data.Dataset):
    def __init__(self, batcher: XRPatcher, item_postpro=None):
        self.batcher = batcher
        self.postpro = item_postpro
    def __getitem__(self, idx: int) -> torch.Tensor:
        item = self.batcher[idx].load().values
        if self.postpro:
            item = self.postpro(item)
        return item
    def reconstruct_from_batches(
            self, batches: list(torch.Tensor), **rec_kws
        ) -> xr.Dataset:
        return self.batcher.reconstruct(
            [*itertools.chain(*batches)], **rec_kws
        )
    def __len__(self) -> int:
        return len(self.batcher)
# load demo dataset
data = xr.tutorial.load_dataset("eraint_uvz")
# Instantiate the patching logic for training
patches = dict(longitude=30, latitude=30)
train_patcher = XRPatcher(
    da=data,
    patches=patches,
    strides=patches,        # No Overlap
    check_full_scan=True    # check no extra dimensions
)
# Instantiate the patching logic for testing
patches = dict(longitude=30, latitude=30)
strides = dict(longitude=5, latitude=5)
test_patcher = XRPatcher(
    da=data,
    patches=patches,
    strides=strides,        # Overlap
    check_full_scan=True    # check no extra dimensions
)
# instantiate PyTorch DataSet
train_ds = XRTorchDataset(train_patcher, item_postpro=TrainingItem._make)
test_ds = XRTorchDataset(test_patcher, item_postpro=TrainingItem._make)
# instantiate PyTorch DataLoader
train_dl = torch.utils.data.DataLoader(train_ds, batch_size=4, shuffle=False)
test_dl = torch.utils.data.DataLoader(test_ds, batch_size=4, shuffle=False)
```

Listing 4 – `XRPatcher` integration in Pytorch. We define a PyTorch dataset that handles the `XRPatcher`. We load an arbitrary dataset with `xarray`, then we instantiate the `XRPatcher` with the patching logic, then we instantiate the PyTorch dataset and dataloaders.

COLLEGES SCIENCES
BRETAGNE POUR L'INGENIEUR
LOIRE ET LE NUMERIQUE

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Titre : Apprentissage profond pour l'altimétrie satellitaire océanique : spécificités et implications pratiques.

Mot clés : Apprentissage profond, Altimétrie, SWOT

Résumé : Cette thèse explore comment les avancées en apprentissage profond peuvent aider à l'analyse des mesures satellitaires de la hauteur de surface de la mer (SSH). Les altimètres actuels fournissent des données échantillonnées de manière irrégulière limitant l'observation des processus les plus fins. Repousser cette limite améliorerait nos capacités de surveillance du climat. D'excitantes opportunités ont émergées avec la mission SWOT. Les approches d'apprentissage ont démontré des capacités remarquables dans de nombreux domaines. Cette thèse aborde les considérations spécifiques de l'application de l'apprentissage profond aux données altimétriques en trois parties.

Premièrement, à travers l'étalonnage du capteur KaRIn, nous démontrons comment des connaissances spécifiques du domaine peuvent être intégrées dans les cadres d'apprentissage profond. Deuxièmement, nous abordons la rareté des données de vérité terrain lors de l'apprentissage de méthodes d'interpolation de données altimétriques. Nous illustrons comment les simulations de modèles océaniques et de systèmes d'observation peuvent surmonter ce défi en fournissant des environnements d'entraînement supervisés qui se généralisent aux données réelles.Enfin, notre troisième contribution traite des défis rencontrés pour combler le fossé entre les communautés "océan" et "apprentissage profond". Nous décrivons comment nous avons abordé ces aspects lors du développement du projet OceanBench.

Title: Deep Learning for ocean satellite altimetry : specificities and practical implications

Keywords: Deep Learning, Altimetry, SWOT

Abstract: This thesis explores how advancements in deep learning can aid in the analysis of satellite measurements of sea surface height (SSH). Current altimeters provide data sampled in an irregular manner, limiting the observation of finer processes. Pushing this limit would enhance our climate monitoring capabilities. Exciting opportunities have emerged with the SWOT mission. Learning approaches have shown remarkable capabilities in many areas. This thesis addresses the specific considerations of applying deep learning to altimetry data in three parts.

First, through the calibration of the KaRIn sensor, we demonstrate how specific domain knowledge can be integrated into deep learning frameworks. Second, we address the scarcity of ground truth data when learning altimetry data interpolation methods. We illustrate how ocean model simulations and observation systems can overcome this challenge by providing supervised training environments that generalize to real data. Lastly, our third contribution discusses the challenges faced in bridging the gap between the "ocean" and "deep learning" communities. We describe how we approached these aspects during the development of the OceanBench project.