

Reconstruction of missing data in HF radar observations using the convolutional autoencoder DINCAE

Auteur : Houyoux, Aimée

Promoteur(s) : Barth, Alexander

Faculté : Faculté des Sciences appliquées

Diplôme : Master : ingénieur civil électricien, à finalité spécialisée en "electronic systems and devices"

Année académique : 2020-2021

URI/URL : <http://hdl.handle.net/2268.2/11517>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.



MASTER THESIS

**RECONSTRUCTION OF MISSING DATA IN
HIGH-FREQUENCY RADAR OBSERVATIONS
USING THE CONVOLUTIONAL AUTOENCODER
DINCAE**

Graduation Studies conducted for obtaining the Master's degree in Electrical Engineering by

Aimée HOUYOUS

Guided By - Dr. Alexander Barth

University of Liège, Faculty of Applied Sciences

Academic year 2020-2021

Abstract

[eng] The goal of this work is to evaluate the possibility to reconstruct missing surface current data from high frequency radars (HFR) with a convolutional neural network. HF radars measure surface currents based on Bragg scattering of electromagnetic waves upon gravity waves on the ocean surface. Perturbation of the process can occur due to interferences or presence of metallic objects in the measured field and cause gaps in the collected data.

Data-Interpolating Convolutional Autoencoder (DINCAE) is here applied to total currents measurements in the Ibiza Channel that are and made freely available by the Balearic Islands Coastal Ocean Observing and Forecasting System (SOCIB) since 2012. The structure of the Autoencoder allows the network to be trained on fields with missing data.

Both zonal currents (U) and meridional currents (V) are given as inputs to the network are reconstructed along with the expected standard deviation of the error of reconstruction, following a gaussian curve. Main patterns in the data are captured by the encoder part of the network thanks to convolutional filters and pooling and from there, the full fields are reconstructed by the decoder part.

During training of the network, a part of the data was set aside to perform cross-validation. Masks were added to those data to assess the ability of DINCAE to reconstruct the hidden data. Performances of the reconstruction were computed for different hyperparameter combinations. Presence of drifters in the area in 2014 further gave the possibility to make a second validation with external data.

Since the two validation strategies gave different results from each other, a configuration selected by each validation was further analysed. It came out that cross-validation selected configuration is presumably trapped in a local minimum because of a too high lower limit on the expected standard deviation. In fact this is an issue encountered with many tests that were carried out. Drifter selected configuration is thus more reliable and will lead to better generalization, though some time instances are polluted with artefacts, explaining the poor results with cross-validation (performed on a longer time interval).

Finally, drifter validation makes comparison with other reconstruction methods possible. DINCAE is found to make more accurate reconstructions than Open-Boundary Model Analysis (OMA), which is a very commonly used method, but not as good as Data Interpolation Variational Analysis in n-dimensions (DIVAnd).

[fr] L'objectif de ce travail est d'évaluer la possibilité de reconstruire des données manquantes parmi les données collectées par des radars à haute fréquence relatives aux courants de surface à l'aide d'un réseau de neurones convolutif. Les mesures de courants marins par les radars HF sont faits sur base de la diffusion de Bragg d'ondes électromagnétiques sur les vagues à la surface de l'océan. Durant le processus de collecte des données, des perturbations telles que la présence d'objets métalliques dans le champ du radar ou des interférences peuvent causer la perte d'une partie des données ou une détérioration de leur qualité.

DINCAE (un AutoEncodeur Convolutif pour l'Interpolation de Données) est appliqué aux courants totaux mesurés dans le Canal de Ibiza, rendus disponibles par la SOCIB (Balearic Islands Coastal Ocean Observing and Forecasting System) depuis 2012. La structure de l'autoencodeur permet l'entraînement du réseau sur des champs comprenant des données manquantes.

Les courants zonaux (U) et méridionaux (V) sont donnés en entrée au réseau neuronal pour être reconstruits en sortie, accompagnés de l'écart-type de l'erreur de reconstruction estimée, sous forme d'une gaussienne. Les principaux schémas présents dans les données sont détectés par la partie d'encodage du réseau grâce aux filtres convolutifs et aux couches de pooling. Les champs complets sont ensuite reconstruits à partir de ces schémas par la partie de décodage du réseau.

Pendant l'entraînement du réseau, une partie des données est mise de côté pour servir à la validation croisée. Des masques sont ajoutés à ces données pour évaluer la capacité de DINCAE à reconstruire les données cachées. La présence de drifters dans la zone étudiée en 2014 permet une deuxième validation basée sur une technique de mesure des courants différente. La capacité de reconstruction du réseau a été évalué avec plusieurs combinaisons d'hyperparamètres.

Les classement des combinaisons d'hyperparamètres sont différents pour la validation croisée et pour la

validation par drifter. Une configuration a donc été sélectionnée pour chaque classement et les résultats ont été analysés plus en profondeur. Il en résulte que la configuration sélectionnée par validation croisée est vraisemblablement coincée dans un minimum local car la limite inférieure sur les écarts-types estimés était trop élevée. Beaucoup de configurations testées sont en fait dans le même cas. La configuration sélectionnée par la validation par drifter ne présente pas ce problème, ce qui mène à une meilleure généralisation. Certains instants temporels sont cependant pollués par des artefacts, ce qui explique la performance médiocre de cette configuration par validation croisée.

Finalement, la validation par drifter rend possible la comparaison avec d'autres méthodes de reconstruction des données manquantes. DINCAE fait des reconstructions plus précises que OMA (Open-Boundary Model Analysis), méthode largement utilisée mais pas aussi bonnes que DIVAnd (Data Interpolation Variational Analysis in n-dimensions).