

ECOLE DOCTORALE DES SCIENCES DE L'ENVIRONNEMENT
D'ILE DE France N° 129
Proposition de sujet de thèse pour la rentrée 2020

Nom du Laboratoire d'accueil : LATMOS

N° UMR :8190

Nom du Directeur du laboratoire : François Ravetta

Adresse complète du laboratoire : 11 Bd d'Alembert, 78 280 Guyancourt

Nom de l'Equipe d'accueil et adresse si différente de celle du laboratoire :

Statistiques Processus Atmosphère Cycle de l'Eau (LATMOS/Guyancourt)

Nom du Directeur de thèse **HDR** :Cécile Mallet

Nom du co-directeur de thèse **HDR** : Laurent Barthes

Téléphone : 01 80 28 52 16

Téléphone : 01 80 28 52 26

Mail : cecile.mallet@latmos.ipsl.fr

Mail : laurent.barthes@latmos.ipsl.fr

OU

Nom du co-encadrant **non HDR** : Nicolas Viltard

Téléphone :

Mail : Nicolas.Viltard@latmos.ipsl.fr

Collaborations : groupe de travail AI4climate <https://ai4climate.lip6.fr/>

• Titre de la thèse en Français :

Analyse statistique des chroniques spatiales sur 20 années d'observations TRMM

• Titre de la thèse en Anglais :

Statistical analysis of 20 years of spaceborne remote sensing times series (TRMM)

• Résumé Sujet en Français (1 page maximum) :

Le travail proposé va s'intéresser à l'évolution des propriétés des systèmes nuageux et de la convection au cours des dernières décennies sous la pression du changement climatique. En effet, les changements dans les conditions thermodynamiques de l'atmosphère doivent se traduire par une modification des propriétés micro- et macro-physiques des systèmes précipitants. Si de nombreuses études se sont penchées sur l'évolution des propriétés microphysiques des nuages et leur impact sur le bilan radiatif terrestre, peu de travaux portent sur l'évolution des propriétés macro-physiques.

Depuis maintenant 30 ans, les satellites défilants en orbite basse sont équipés de capteurs hyperfréquences passifs qui mesurent les températures de brillances au sommet de l'atmosphère. Pour le sujet qui nous intéresse, ces mesures sont réalisées dans les canaux fenêtres entre 10 et 150 GHz. A ces fréquences, les deux principaux processus contributeurs sont l'émission par les gouttes de pluie et la diffusion par les hydrométéores glacés.

Dans une première partie de la thèse, l'évolution des statistiques annuelles de températures de Brillance seront complétées par une analyse par événement. Un algorithme de segmentation d'image permettra de sélectionner automatiquement les événements dans les images de températures de Brillance. L'ensemble des 9 canaux de 10 à 85 GHz seront utilisés, ceux-ci apportant chacun des

informations plus ciblées sur les trois principales du cycle de l'eau atmosphérique : précipitations liquides, glacées et vapeur d'eau. A partir des événements ainsi détectés, il est possible de définir un certain nombre de variables permettant de caractériser ces systèmes : taille, distribution des températures de brillance, minimum, maximum, durée de vie, position etc... L'analyse statistique de ces variables permettra l'étude de l'évolution des événements à l'échelle globale dans un premier temps puis éventuellement régionale.

L'originalité du travail repose sur l'utilisation des observations brutes (températures de Brillance) qui contiennent de manière implicite l'information pertinente, mais combinée avec d'autres variables. En général les études réalisées se fondent sur des produits plus ou moins élaborés tels que les produits IMERG ou TMPA (voir par ex. [1]) qui ont l'avantage de fournir une grandeur physique directement interprétable, mais qui d'une part n'exploitent pas forcément toute l'information contenue dans les données brutes et d'autre part contiennent des hypothèses souvent contraignantes. Cette nouvelle approche globale n'est cependant pas sans poser des problèmes d'interprétation mais elle évite ainsi les erreurs introduites lors de la phase d'inversion.

Les variables définies a priori et mentionnées ci-dessus, sont pertinentes pour caractériser les propriétés des systèmes mais ne sont pas nécessairement pertinentes pour capturer les changements subtils qui peuvent intervenir sur une vingtaine d'années. Dans la deuxième partie de la thèse, en nous appuyant sur la longue expérience en traitement des données satellite par machine learning qui existe au sein de l'équipe, il s'agira de tester des approches non-supervisées ou faiblement supervisées. Une approche possible est d'extraire, à partir des températures de brillance, un ensemble de caractéristiques représentatif d'un système précipitant. La différence fondamentale avec la première partie de la thèse est que ces caractéristiques ne sont pas définies a priori, mais sont déduites de façon automatique par apprentissage non supervisé. Pour cela, on envisage d'utiliser des méthodes de type autoencodeur convolutionnel [2] qui permettent de définir un espace de dimension réduit dit espace latent. Les variables définies dans cet espace à partir des observations seront utilisées pour classifier les événements de manière totalement non supervisée. Une autre approche possible est d'utiliser des cartes autoorganisatrices [3] pour réaliser une quantification vectorielle de l'espace des observations et une classification des événements, fondée sur une approche également multidimensionnelle et non linéaire. Les classes obtenues seront dans la mesure du possible interprétées à posteriori à partir de l'analyse des processus physiques qui peuvent en être à l'origine. La labellisation des classes obtenues sera réalisée par une analyse visuelle complétée par des observations complémentaires (données du radar de TRMM). C'est la raison pour laquelle le travail commencera avec sa série des 20 ans de données disponibles du satellite *Tropical Rainfall Measuring Mission*.

Dans une troisième partie, il s'agira de porter ces résultats vers les 30 ans de données disponibles pour la série des *Defense Meteorological Satellite Program*.

Références

- [1] J. Wang, R.F. Adler, G.J. Huffman, and D. Bolvin, 2014: *An Updated TRMM Composite Climatology of Tropical Rainfall and Its Validation*. *J. Climate*, 27, 273–284, <https://doi.org/10.1175/JCLI-D-13-00331.1>
- [2] Y. Wang, H. Yao and S. Zhao: *Auto-encoder based dimensionality reduction*, *Neurocomputing*, 184, pp 232-242, 2016
- [3] T. Kohonen : *Exploration of very large databases by self-organizing maps: Proceedings of International Conference on Neural Networks (ICNN'97)*, 1997

• **Résumé Sujet en Anglais (1 page maximum) :**

The proposed work will focus on the evolution of the properties of cloud systems and convection in recent decades under the pressure of climate change. Indeed, changes in the thermodynamic conditions of the atmosphere must induce a modification of the micro- and macro-physical properties of precipitating systems. While many studies have looked at the evolution of microphysical properties of clouds and their impact on the Earth's radiation budget, little has been done on the evolution of macro-physical properties of the rain systems themselves.

For 30 years, remote sensing satellites in low orbit have been equipped with passive microwave sensors that measure the brightness temperatures at the top of the atmosphere. These measurements are made in the window channels between 10 and 150 GHz. At these frequencies, the two main contributing processes are emission by rain drops and scattering by ice crystals.

First, an assessment of the brightness temperature statistical properties will be performed. An image segmentation algorithm will automatically select the rain events in the brightness temperature images. All 9 channels from 10 to 85 GHz will be used, each providing more targeted information on the three main variables of the atmospheric water cycles: rain drops, ice crystals and water vapor. From the events thus detected, it is possible to define a certain number of variables to characterize these systems: size, distribution of the brightness temperatures, minimum, maximum, lifetime, position, etc. The statistical analysis of these variables will allow us to study the evolution of rain systems on a global scale and eventually at regional scale.

The originality of the work is based on the use of raw observations (brightness temperatures) which implicitly contain the relevant information, but combined with other variables. In general, the studies carried out are based on products such as IMERG or TMPA (see for example [1]). These provide a physical quantity directly interpretable, but on the one hand do not necessarily exploit all the information contained in the raw data and on the other hand often contain many assumptions. This new global approach is not without raising interpretation problems, but it thus avoids the errors introduced during the retrieval phase.

The a priori defined variables mentioned above, are relevant for characterizing the properties of the systems but are not necessarily relevant for capturing the subtle changes that may occur over the twenty-year period. In the second part of the thesis, unsupervised or weakly supervised approaches. One possible approach will be tested to extract, from the brightness temperatures, a set of characteristics representative of a precipitating system. The fundamental difference with the first part of the thesis is that these characteristics are not defined a priori, but are deduced automatically by the learning process. Methods like the convolutional autoencoder type [2] will be used. This method allows us to define a reduced dimensional space called latent space. The variables defined in this space from the observations will be used to classify the events in a totally unsupervised manner. Another possible approach is to use self-organizing maps [3] to perform vector quantification of the space of observations and classification of events, based on an approach that is also multidimensional and non-linear. The classes obtained will, as far as possible, be interpreted from the analysis of the physical processes which may be at their origin. The labeling of the classes obtained will be carried out by a visual analysis supplemented by additional observations (data from the radar of TRMM). This is why the work will start with its series of 20 years of available data from the *Tropical Rainfall Measuring Mission satellite*.

Last, we will test these methods on the 30 years of data available from the series of *Defense Meteorological Satellite Program*.

• Type de financement autre que ED 129, précisez si envisageable ou acquis (CNES, CEA, ADEME etc...) :

• **Encadrement :**

. **Liste des autres doctorants que vous encadrez ou co-encadrez au 1^{er} janvier 2020**

(Nom, Université d'inscription, type de financement, date de soutenance envisagée)

ED129

Sabrina Derouiche Co-tutelle internationale Paris -Saclay / tunis El Manar (année de dérogation en 2020) , **co-direction Zoubeida Bargaoui** LMHE Tunis, financement soutenance prévue en 2021

hors ED129

Pierre Lepetit Université Paris-Saclay , **co-direction Laurent Barthes**, financement FCPLR