Geophysical Research Abstracts
Vol. 21, EGU2019-8842, 2019
EGU General Assembly 2019

Atmospheric boundary layer classification with unsupervised learning from field campaign data

Thomas Rieutord, Alexandre Paci, and Alain Dabas
CNRM, Météo-France/CNRS, Université de Toulouse, Toulouse, France (thomas.rieutord@meteo.fr)

The atmospheric boundary layer is the lowest part of the atmosphere: the layer influenced by the ground within a typical timescale of one hour. But as the ground is subject to large deviations depending on the diurnal cycle, the atmospheric boundary layer has very different characteristics as well. At night the cold ground usually induces a stable stratification, while at day the heat of the sun changes it into a convective stratification. Moreover, other phenomena are also at play, for example local flows over complex terrain or clouds at the top of the boundary layer, which make things even more complex. Therefore, it is common to distinguish several types of boundary layers (among which "stable" and "convective" boundary layers) to better describe what is happening inside.
This work aims to use unsupervised learning methods to perform such classification. In the light of the recent breakthroughs in other disciplines, multiple improvements are expected from these methods: algorithms returning the same classification a human expert, complex cases better managed than with human-tailored decision trees and a better flexibility when instruments or location are changing. The present work only addresses the first of these expectations. It focuses on agglomerative clustering method: from a metric and a linkage strategy, such algorithm returns classes with a nested structure. A thoroughly detailed case study is presented to test the sensitivity of the algorithm to configuration (several metrics and linkage strategies have been tested) and changes in the data (noise, resolution...). Performance of different options are compared with both expert examination and numeric scores (Calinski-Harabaz index and silhouette score). Data come from two instruments likely to be deployed widely: a microwave radiometer (providing temperature profiles) and a ceilometer (providing aerosol backscatter profiles). They were operating during the Passy-2015 field campaign which focuses on wintertime atmospheric dynamic and air quality in an alpine valley. Other algorithms (K-means family, Expectation-maximisation, DBSCAN) will be evoked but not detailed. The case study shows that, despite a tuning phase that must not be neglected, such methods provide relevant and multi-scale information that is worth to be tested on larger datasets. They open a way to define boundary layer types "from the data" rather than "from the theory".