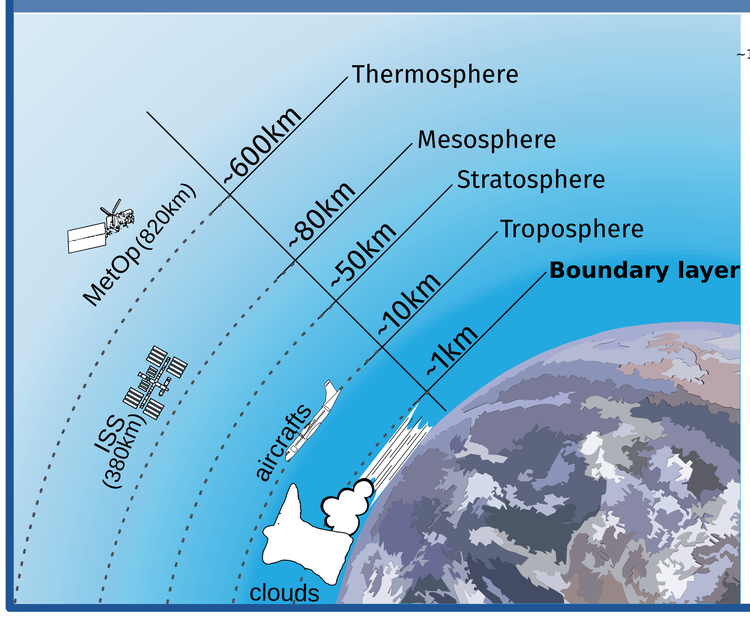
Mixing height derivation from aerosol lidar using machine learning: KABL and ADABL algorithms

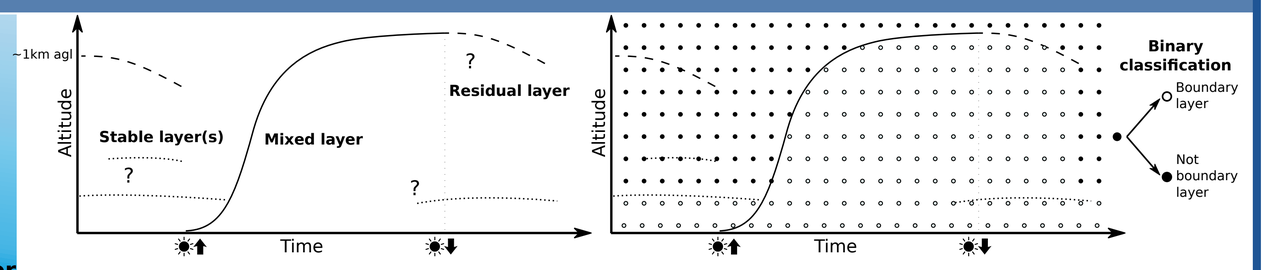
Thomas Rieutord*, Sylvain Aubert
CNRM, Météo-France, CNRS, Université de Toulouse, France
*thomas.rieutord@meteo.fr

Abstract

Atmospheric boundary layer top is a key parameter but its derivation is challenging. This work experiments both supervised and unsupervised learning to derive it. Case studies are encouraging but two-year comparison do not show clear improvement from existing methods. However, they are open-source and have good prospects.

1 Introduction





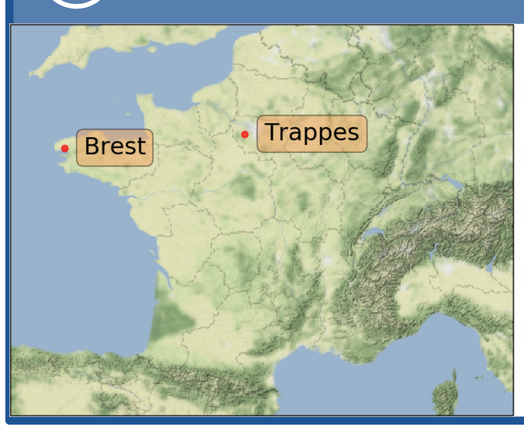
Why study boundary layer?

- Interface between surface and atmosphere, multiples fluxes and forcing driving atmosphere.
- Siege of complex phenomena like turbulence, fog, local circulation.
- Benefit to: air quality, renewable energy, transportation, meteorological forecast...

How do we proceed?

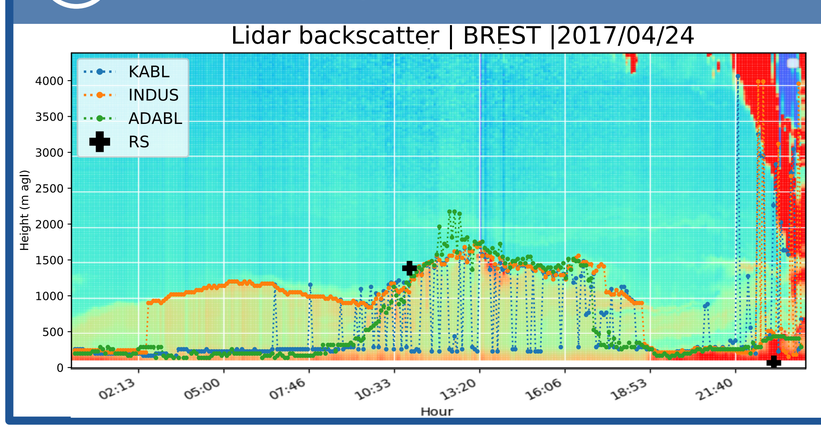
- Lidar backscatter is a proxy of the aerosol content, thus of the atmospheric boundary layer (ABL).
- Unsupervised classification: **K-means for ABL (KABL)**
- Supervised classification: **AdaBoost for ABL (ADABL)**

2 Experimental setup



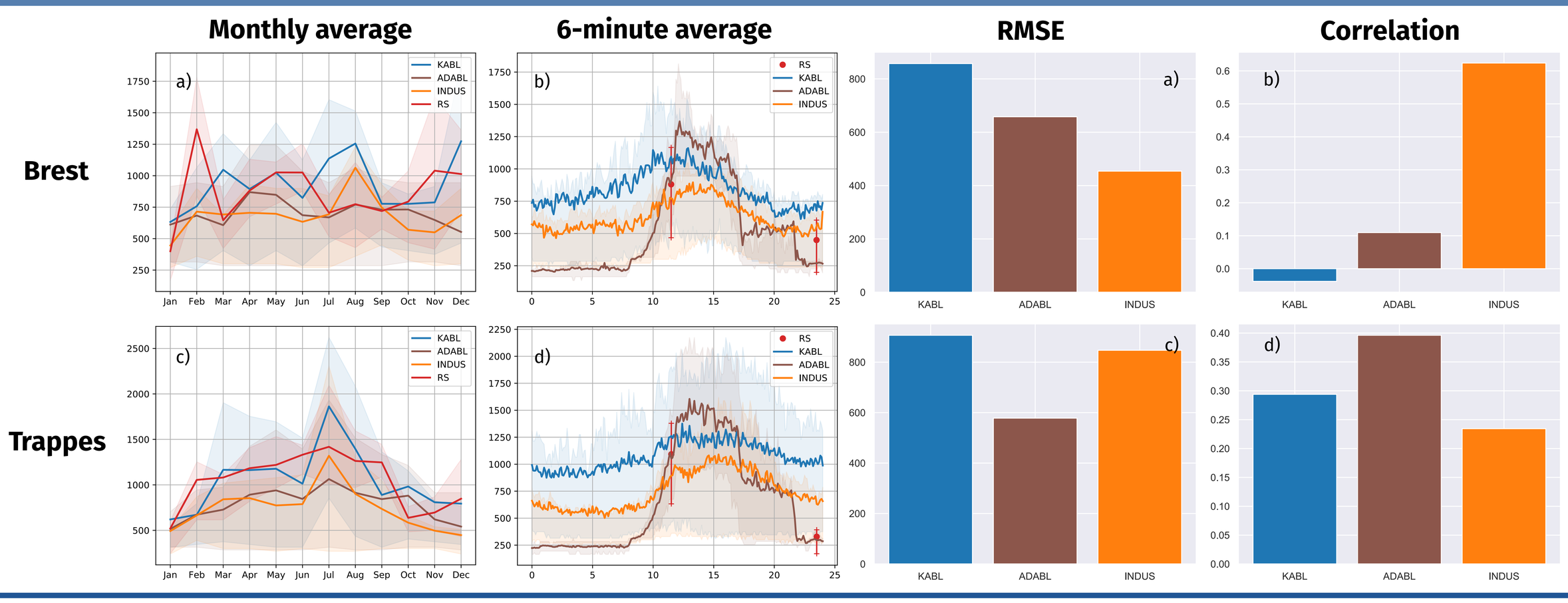
- 2 sites (Brest and Trappes)
- 2 years of data (2017-2018)
- Co-located lidar and RS
- Lidar: MiniMPL (SigmaSpace)
- RS estimation: parcel method

3 Case study

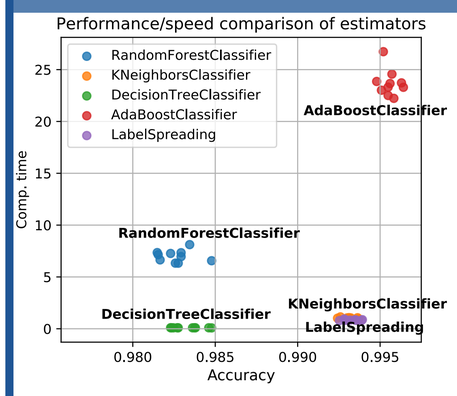


- **Manufacturer's** estimation (INDUS): misses the morning transition, good otherwise.
- **KABL**: most of the time acceptable estimation, but many jumps
- **ADABL**: catches very well morning transition, few odd points around 12:00

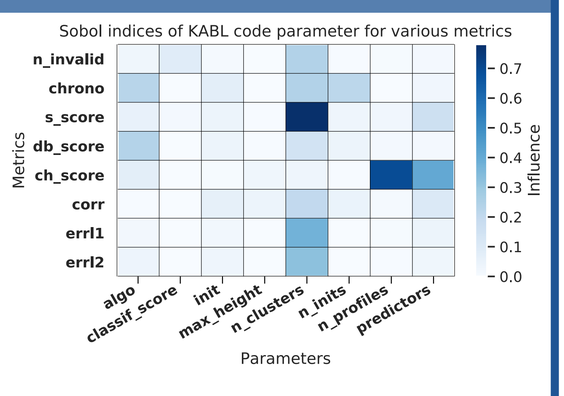
4 Two-year comparison



5 Sensitivity analysis



Among several classifiers tested, AdaBoost was chosen because it has the best accuracy and speed is not critical.



Influence of KABL's parameters over several metrics was estimated. The number of clusters and the predictors are critical.

Conclusions

- This work attempts to derive boundary layer top from lidar backscatter measurements with machine learning. Both supervised (ADABL) and unsupervised (KABL) algorithms have been tried.
- Case study shows acceptable results despite few drawbacks.
- Two-year comparison with RS does not draw clear improvement from manufacturer's algorithm. Results are different on the two sites: KABL and ADABL do not compare to RS at Brest, but they do at Trappes.
- Diurnal cycles are similar for KABL and manufacturer's, ADABL reproduces too much the cycle of the day it has been trained on (overestimated importance of time and altitude predictors)
- Seasonal cycles are acceptable at Trappes, but not at Brest, even for RS estimations

Rieutord et al. (2020), AMT-D

https://github.com/ThomasRieutord/kabl