Sothea Has — Ph.D. in Applied Mathematics

Bâtiment Sophie Germain, office 5032
8 place Aurélie Nemours, 75013 Paris, France
https://hassothea.github.io/
sothea.has@lpsm.paris

SKILL SUMMARY

Postdoc	Apply statistical and machine learning methods in atmospheric science.
Ph.D.	Theoretical machine learning: clustering, aggregation methods, and energy modeling.
Teaching	Coding and practical insights in machine learning, statistics, data analysis and modeling.

EXPERIENCES

2023 — Postdoc
CNRS, LPSM
Build and maintain `gradientcobra` python library.

2022-Present — Postdoc
CNRS, LPSM Université Paris Cité and LMD École Polytechnique
Improving parametrizations in climate modeling using statistical and machine learning.
- Modeling the balloon-observed Gravity Wave Momentum Fluxes (GWMF).
- Extracting important features for reconstruction.
- Interpreting and providing information of the stochastic component of GWMFs.

2018 - 2022 — Ph.D.
LPSM (UMR 8001) - Sorbonne Université
Theoretical study and applications of machine learning methods.
- Energy data modeling using supervised and unsupervised machine learning algorithms.
- Aggregation method for regression problems.
- Aggregation method in high dimension.

2018 - Present — Teaching
UFR Mathematics Université de Paris
Master 1 and Master 2
- Practical class of Data Analysis with `R` and `Rstudio`, Master 1 ISIFAR.
- Practical class of Data Mining with `R` and `Rstudio`, Master 2 ISIFAR.
- Practical class of Exploratory Data Analysis with `R` and `Rstudio`, Master 1 EDA.
- Practical class of Algorithm and Programming with `Python`, Licence 2 MIASHS.
- Practical class of Big Data Technologies with `Python` and `Spark`, Master 1 MATINF.
- Tutorial class of Statistical Inference and Data Modeling, Master 2 M2MO.

2018 — M2 internship
LPSM (UMR 8001) Université de Paris
Predictive models based on clustering with Bregman divergences and local predictions
- Analyzing the sensitivity of K-means clustering with Bregman divergences.
- Constructing local models on different configurations of clusters.

2017 — M1 internship
Laboratory of TELECOM SudParis
Study of optimization problems with marginal simulated annealing algorithm

PUBLICATIONS

2023 — Status
Estimating balloon-observed gravity wave momentum fluxes using ML & input from reanalysis.
In progress, with R. Plougonven, A. Fischer, R. Rani, F. Lott, A. Hertzog, A. Podglajen, M. Corcos.

2023 — Status
Gradient COBRA: A kernel-based consensual aggregation for regression.
Published at Journal of Data Science Statistics and Visualisation, single author.

2022 **A consensual aggregation on randomly projected high-dimensional features of predictions.**
Status *Published in HAL, single author.*

2022 **Machine learning methods applied to the global modeling of event-driven pitch angle diffusion coefficients during high-speed streams.**
Status *Published in Frontiers Physics, with G. Kluth, J.F. Ripoll, A. Fischer, M. Mougeot, and E. Camporeale.*

April 2021 **KFC: A clusterwise supervised learning procedure based on aggregation of distances.**
Status *Published in Journal of Statistical Computation and Simulation, with A. Fischer and M. Mougeot.*

EDUCATION

2022 - Present **CNRS, LPSM - Université Paris Cité & LMD - École Polytechnique, France**
Title **Postdoctoral researcher in atmospheric science**
Research topic Reconstruct Gravity Wave Momentum Flux using statistical and machine learning methods.

2018 - 2022 **Sorbonne University Pierre and Marie Curie - Paris 6, France**
Title **Ph.D. in Applied Mathematics**
Research topic Consensual aggregation and distance measurements for statistical learning. Theoretical contributions and applications to the field of energy.

2018 **University Paris Diderot - Paris 7, France**
Title **Master's degree in Random Modelling and Data Science (M2MO)**
Project Data Science for Company, Massive Data Processing (`R`-programming).
Courses Statistical Learning, Statistical Modeling, Diffusion Statistics, Stochastic Calculus.
Machine Learning (`Python`), Monte Carlo Method (`C++`).

2018 **École Nationale Supérieure d'Informatique pour l'Industrie et l'Enterprise - ENSIIE, France**
Title **Engineering's degree in Applied Mathematics**
Project Time Series, Simulation Methods, Research Project in Finance, Machine Learning.
Courses Stochastic Process, Operation Research, Stochastic Calculus in Finance.
Data Analysis, Numerical Methods for PDE, `C++`.

2015 **Royal University of Phnom Penh - RUPP, Cambodia**
Title **Bachelor's degree in pure mathematics**

LANGUAGES & PROGRAMMING

Languages Khmer (Mother tongue), English (fluent), French (conversational)
Programming **R**: `tidyverse`, `dplyr`, `ggplot`, `plotly`, ...
Python: `Numpy`, `Pandas`, `TensorFlow`, `Scikit-learn`, `PySpark`, ...
Others: SQL, C++, Matlab, Scilab, LaTeX.

PERSONAL INTEREST

Reading Behavioral science and meditation.
Sport Volleyball, basketball and football.
Other interest Music, guitar, and drawing.