Learning from small data in Ecology

Opportunities & Challenges

Laure Berti-Equille

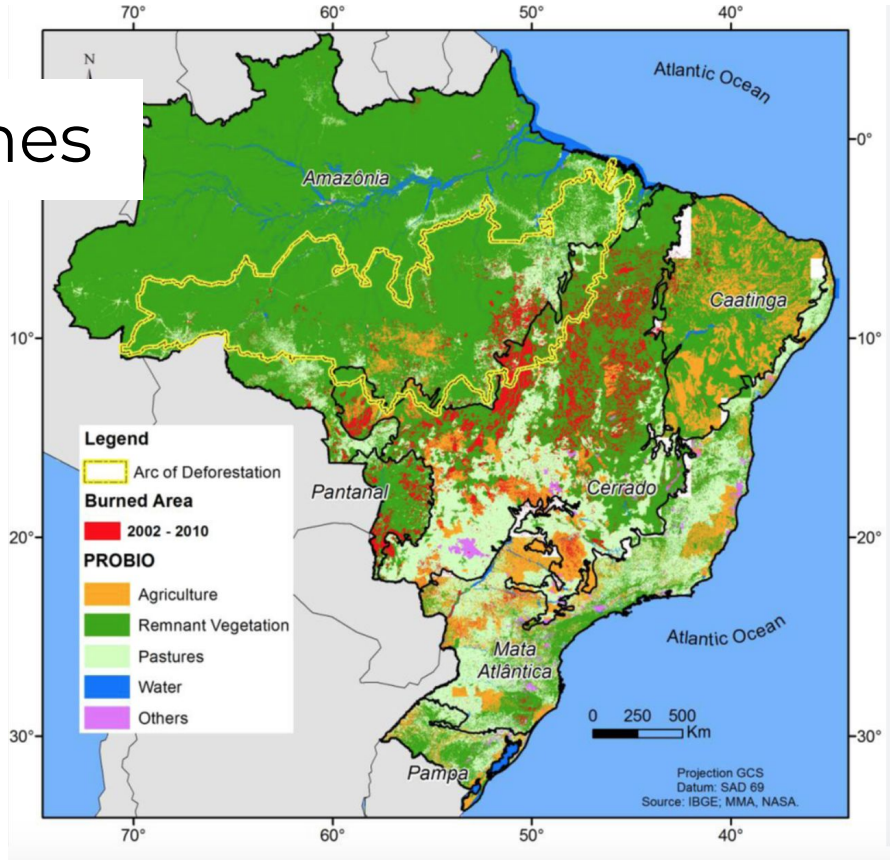
laure.berti@ird.fr

Institut de Recherche pour le Développement
FRANCE

IDEAL
Agroecology & AI Joint Lab

IDEAL Focus: Nordeste Biomes

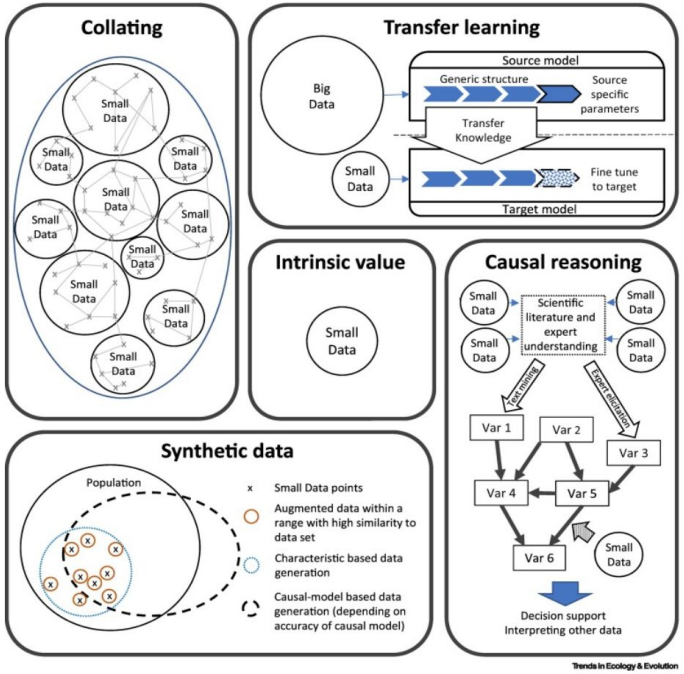
- Significance of Caatinga and Atlantic Forest biomes for sustainable agroforestry, and biodiversity restoration efforts
- Importance of applying machine learning techniques but limited data
- Open science: fair data principles, research collaboration, participatory science, and local community engagement



Source: Moreira de Araújo F, Ferreira LG, Arantes AE. Distribution Patterns of Burned Areas in the Brazilian Biomes: An Analysis Based on Satellite Data for the 2002–2010 Period. *Remote Sensing*. 2012; 4(7):1929-1946. https://doi.org/10.3390/rs4071929

Challenges in Ecology Data: Small but Mighty

- Challenges of field data:
  - Limited sample sizes
  - Sparse data points
  - High variability / low coverage
  - Discrete when continuous is needed
  - Out-of-dateness of observations
- Traditional ML approaches struggle with small data
- Innovative techniques needed



1. Lindsay C. Todman, Alex Bush, Amelia S.C. Hood, 'Small Data' for big insights in ecology. Trends in Ecology & Evolution. Opinion| Volume 38, Issue 7, P615-622, July 2023 https://doi.org/10.1016/j.tree.2023.01.015

2. Miller, D.L. "Challenges and opportunities in small sample environmental data analysis." Environmental Modelling & Software 45 (2013): 45-48.

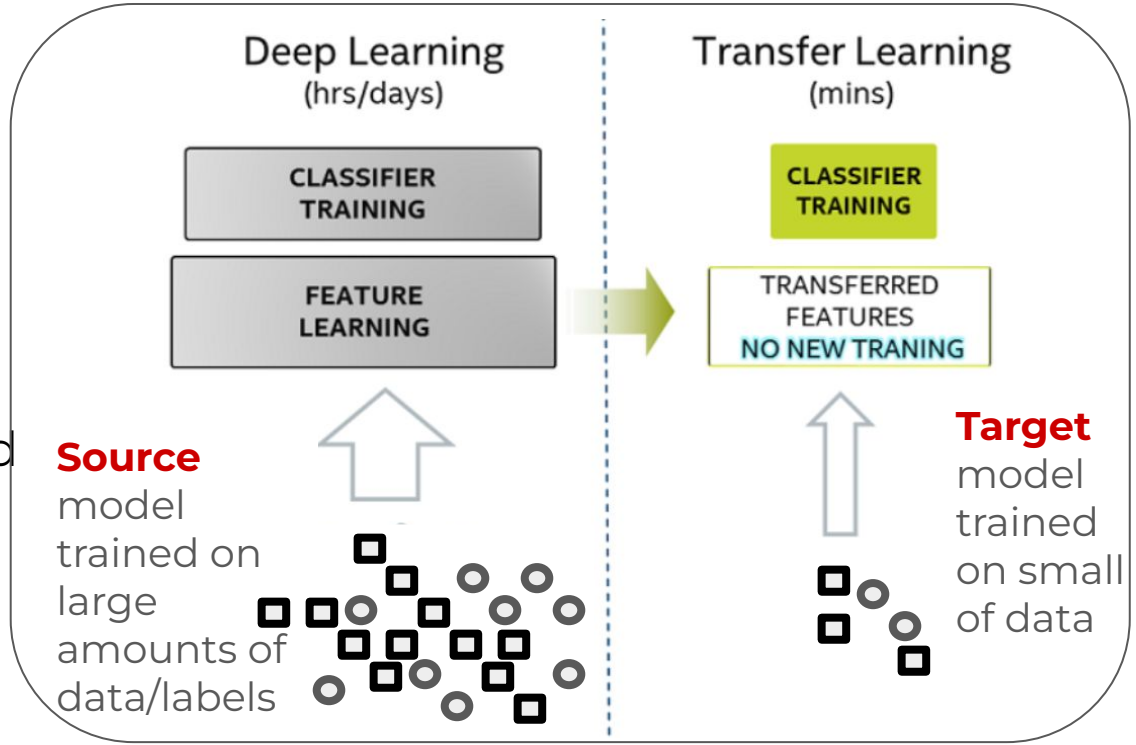
3. Prasad, A.M., et al. "Challenges and Opportunities in the Application of Machine Learning in Ecology." Ecological Modelling 363 (2017): 1-6.

4. Rahman, H., et al. "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review." Computers and Electronics in Agriculture 168 (2020): 105105.

Successful Methods for Small Data

- Transfer learning
- Semi-supervised learning
- Ensemble methods

Domain-specific knowledge informs feature selection and model development.



1. Kämäräinen, J.K., and K. Heikkinen. "Transfer learning in environmental sciences: A review." Ecological Informatics 59 (2020): 101124.
2. Belgiu, M., and L. Drăguţ. "Random forest in remote sensing: A review of applications and future directions." ISPRS Journal of Photogrammetry and Remote Sensing 114 (2016): 24-31.

Example of transfer learning for land cover classification from limited HSI data

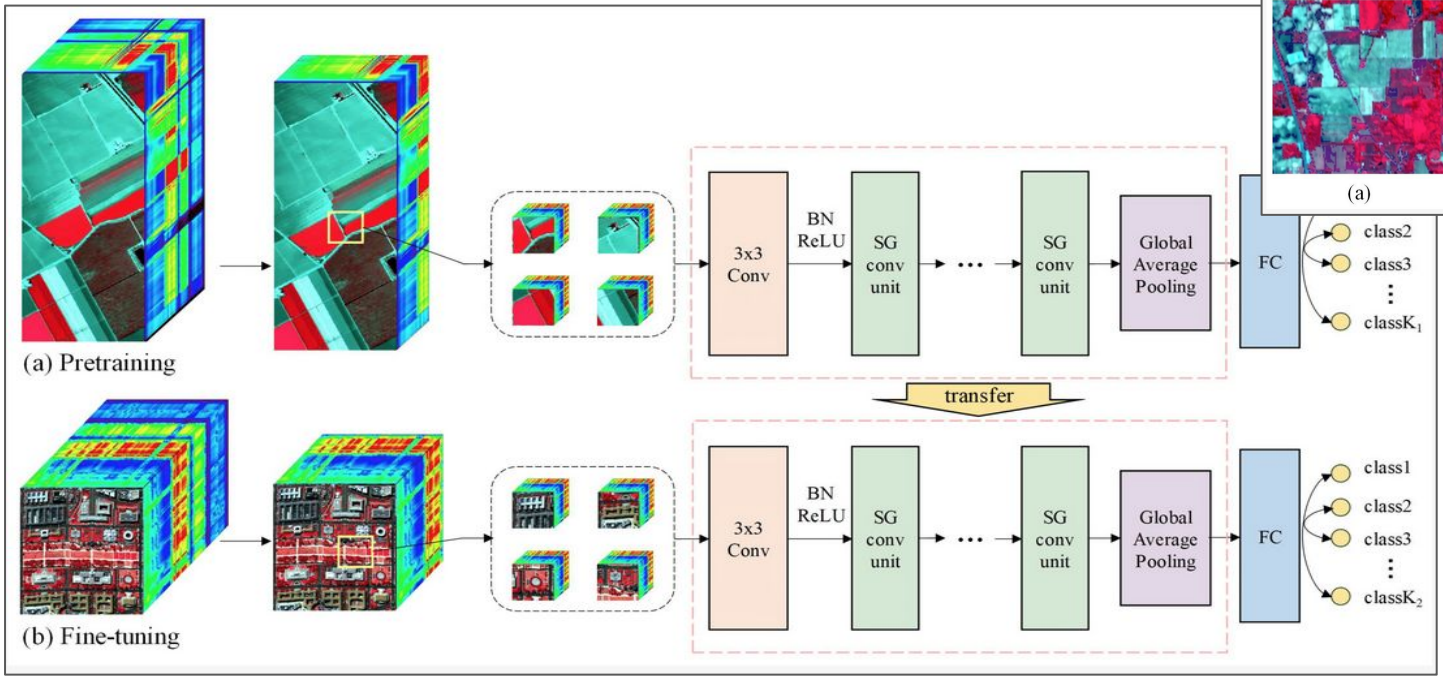
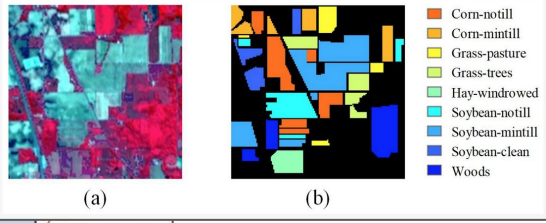


Figure 4. The Indian Pines scene: (**a**) false-color composite image; (b) ground truth.



1. Liu, Y.; Gao, L.; Xiao, C.; Qu, Y.; Zheng, K.; Marinoni, A. Hyperspectral Image Classification Based on a Shuffled Group Convolutional Neural Network with Transfer Learning. *Remote Sens.* **2020**, *12*, 1780. https://doi.org/10.3390/rs12111780
2. Araújo, D.V., et al. "Applying machine learning for land-use and land-cover classification in the Brazilian Amazon." Remote Sensing 12.17 (2020): 2757.

Applications to Nordeste Biomes

Current exploratory research using ML for our study cases

- Micro-deforestation detection
    - Could we transpose the methodology of M2D2 project leaded by Joris to PB, Ceara, RGdN & Pernambuco using & annotating Radar Sentinel 1, Sentinel 2 and PlanetScope images?
- Habitat prediction and abundance maps
    - Inventory of presence/absence data for emblematic species and data collection strategies (e.g., drone?)
- Plant/animal species distribution modeling using Deep Learning
- Design a coviability index (using proxies) and study its evolution

1. Sousa, L., et al. "Machine Learning for Biodiversity Conservation: A Review." Biological Conservation 253 (2021): 108932.
2. Ferreira, F., et al. "Machine learning algorithms applied to the prediction of deforestation in the Brazilian Amazon." Remote Sensing 12.22 (2020): 3808.
3. Sara Beery, Elijah Cole, Joseph Parker, Pietro Perona, and Kevin Winner. 2021. Species Distribution Modeling for Machine Learning Practitioners: A Review. In *ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS) (COMPASS '21), June 28-July 2, 2021, Virtual Event, Australia.* ACM, New York, NY, USA 20 Pages. https://doi.org/10.1145/3460112.3471966