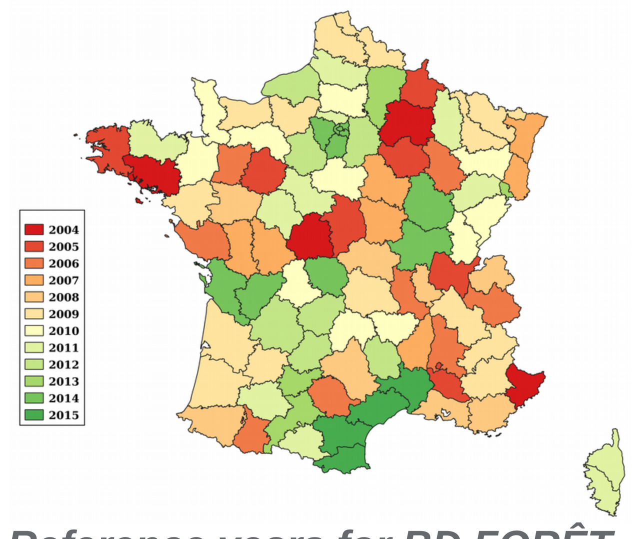
FROM LOCAL TO GLOBAL: A TRANSFER LEARNING-BASED APPROACH FOR MAPPING POPLAR PLANTATIONS AT LARGE SCALE

Yousra Hamrouni[1,2], David Sheeren[1], Éric Paillassa[3], Véronique Chéret[1], Claude Monteil[1]

1. Université de Toulouse, INRAE, UMR DYNAFOR, Castanet-Tolosan, France, 2. Conseil National du Peuplier, Paris, France, 3. Centre National de la propriété Forestière

CONTEXT



Reference years for BD FORÊT

Poplar is one of the fast-growing and wood producing trees which are increasingly considered as an important resource.

In France, accurate and regularly up-to-date maps of poplar plantations are not yet available at the national scale.

The update rate of national forest maps is unsuitable for this species because of its short rotation cycle (15 years on average).

Since the availability of high spatial, spectral and temporal resolution Sentinel time series, new opportunities for monitoring poplar plantations over large areas have come up.

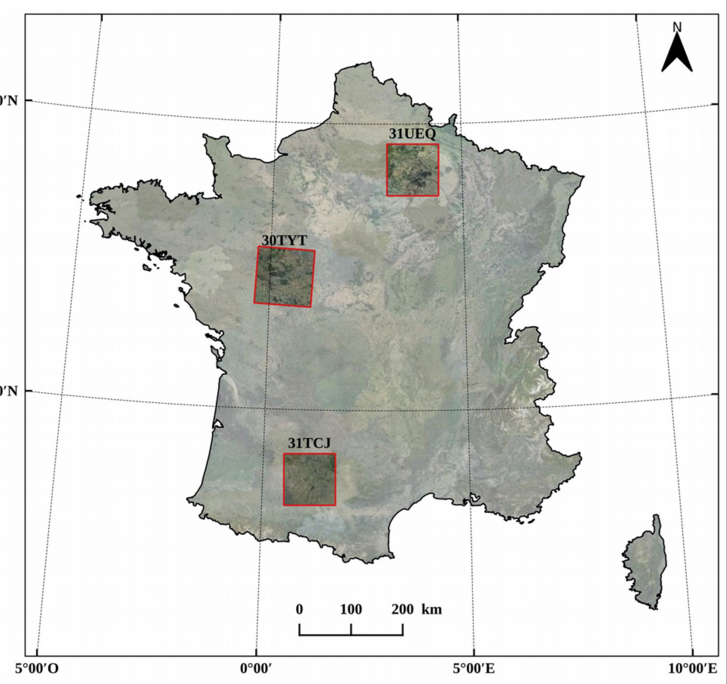
Main objective: large-scale mapping of poplar plantations from Sentinel-2 data

- With all available reference data: supervised classification (***local***)
- With a reduced number of samples & knowledge transfer: use of Active Learning (***global***)

MATERIALS

Study area

The study is focused on three different poplar sites in France with the most contrasting situations in terms of plantations variability (cultivars), management practices and climatic conditions. They are represented by three Sentinel-2 tiles (100 km² area each) in the following figure.



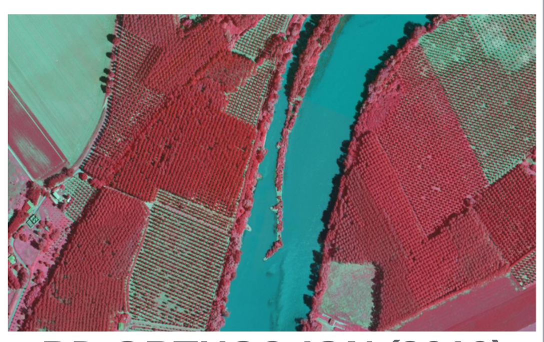
Data

- ***Optical images***

All available Sentinel-2A and Sentinel-2B images over 2017: level 2A products (orthorectified and atmospherically corrected) downloaded from Theia platform.

Tile code	Relative orbit number	No. of available dates in 2017
31UEQ	51	26
30TYT	94	34
31TCJ	51	36

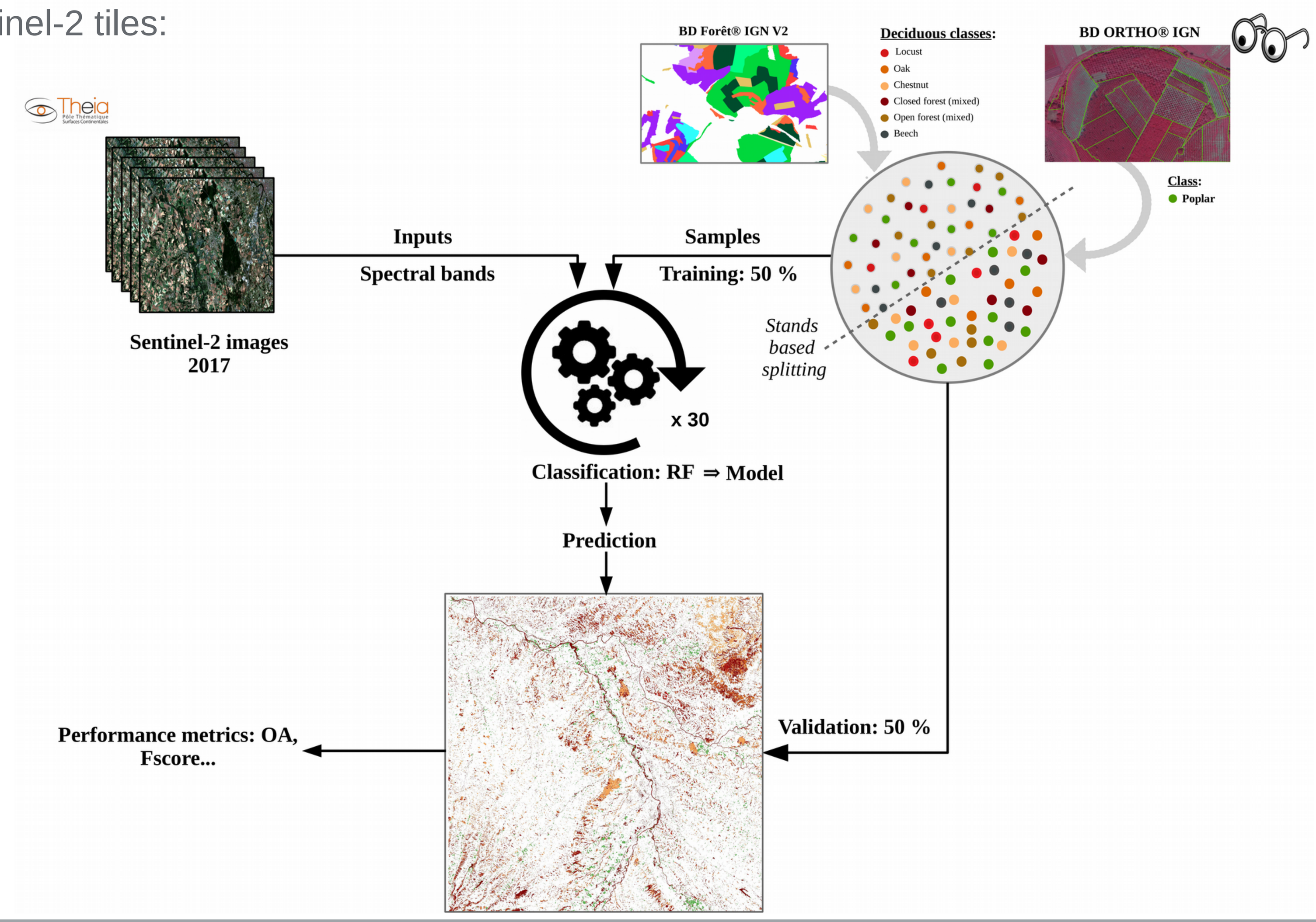
- ***Reference data***
  - *Poplar class*: visual interpretation from Orthophotos **+ validation with Sentinel-2 images.**
  - *Others deciduous classes:* sampling from the national forest database BD Forêt®.



BD ORTHO® IGN (2016)

METHODS: tile by tile classification *(local)*

The following flowchart illustrates the classification process applied to each of the three Sentinel-2 tiles:

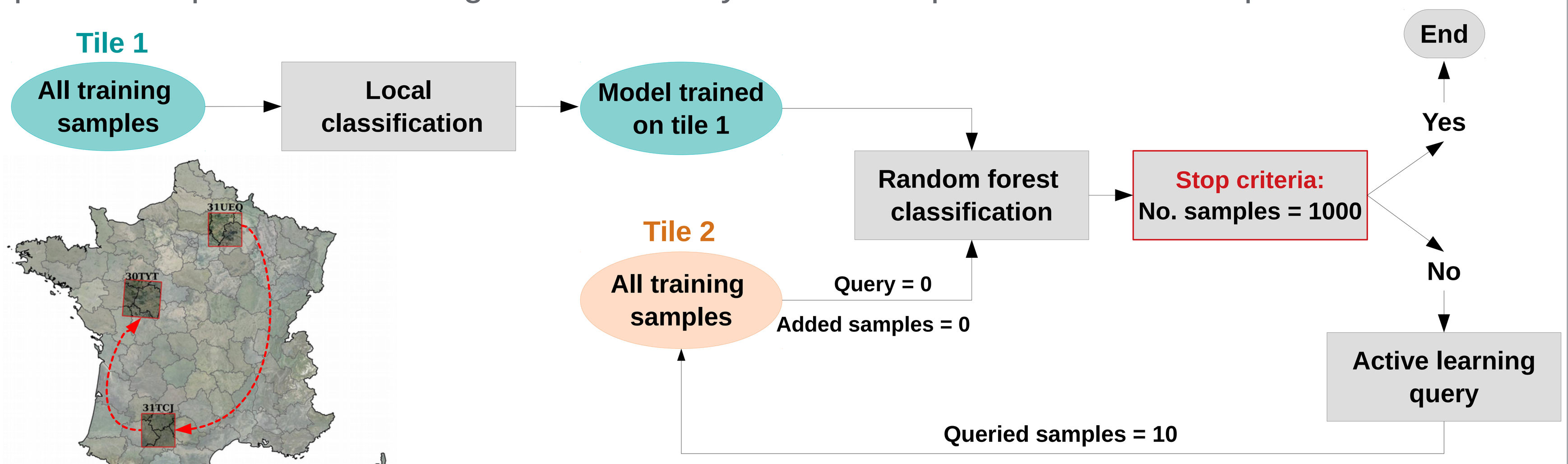


METHODS: tile to tile classification (global)

Active learning (AL) is based on the hypothesis that a machine learning algorithm can achieve greater accuracy with fewer training labels if it is allowed to choose the data from which it learns (Settles, 2010).

Its use is well-motivated since collecting appropriate training samples is an expensive and difficult task especially for large scale.

Starting from a well trained model in a **first** tile, new samples are queried with AL from a **second** tile and iteratively added to the initial model in order to adapt it to both sites. The process is performed among the three study tiles and repeated until the stop criteria are met.



- *The active learning is then compared to a random selection of the same number of samples.*

RESULTS

Local classification

Tile code	Training size per class in pixels	No. classes	Average Overall Accuracy$_{(*30)}$	Average F-score$_{(*30)}$	Average poplar F-score$_{(*30)}$
31UEQ	1250	6	73.7±1.7 %	73.1±2.0 %	89.5±3.3 %
30TYT	2000	6	74.9±1.8 %	75.0±1.9 %	99.3±0.2 %
31TCJ	3850	6	80.0±0.7 %	80.1±0.6 %	97.9±0.8 %

→ Good ability of Sentinel-2 time series to identify poplar plantations among other deciduous species at the tile scale.

→ Additional work of poplars photo-interpretation is required to achieve such performance.

Global classification with Active Learning

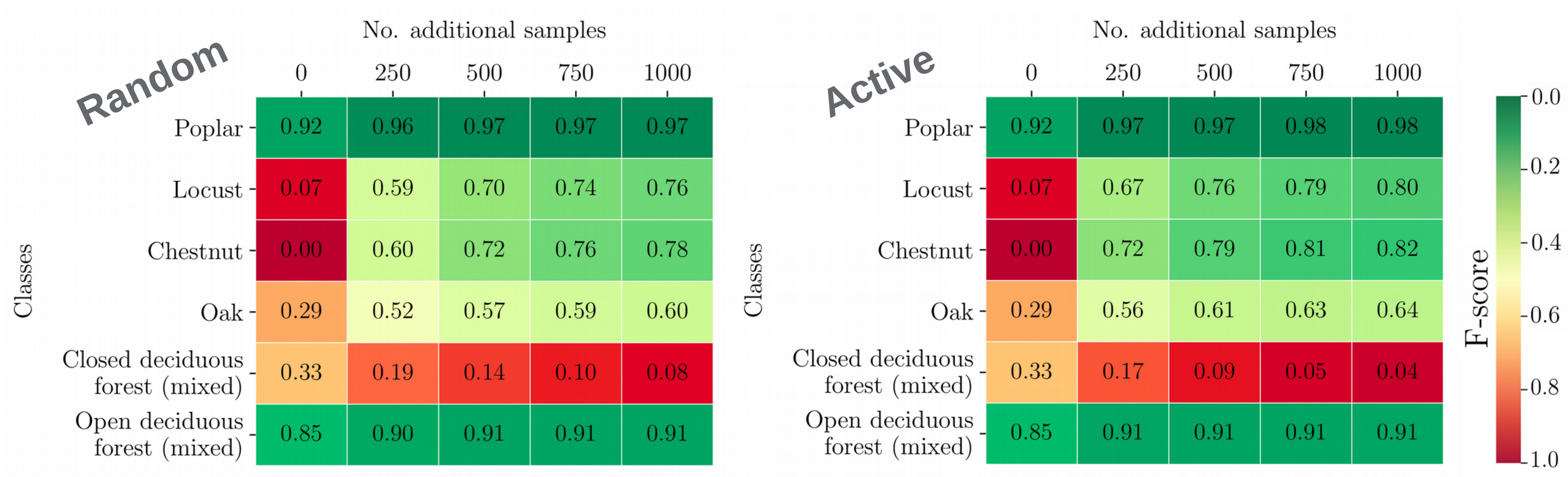
- **Global assessment with Overall Accuracy (OA): all classes**

		Target tiles														
		31UEQ					30TYT					31TCJ				
Source tiles	**OA (%)**	**No. of additional samples**														
		0	250	500	750	1000	0	250	500	750	1000	0	250	500	750	1000
31UEQ	*Random*	-	-	-	-	-	36	50	57	60	61	47	63	69	71	72
	Active	-	-	-	-	-	36	54	61	63	64	47	68	73	74	75
30TYT	*Random*	31	45	52	55	57	-	-	-	-	-	53	69	70	71	72
	Active	31	49	55	58	59	-	-	-	-	-	53	70	72	73	74
31TCJ	*Random*	40	50	58	61	63	52	61	64	64	65	-	-	-	-	-
	Active	40	53	61	64	65	52	63	65	66	66	-	-	-	-	-

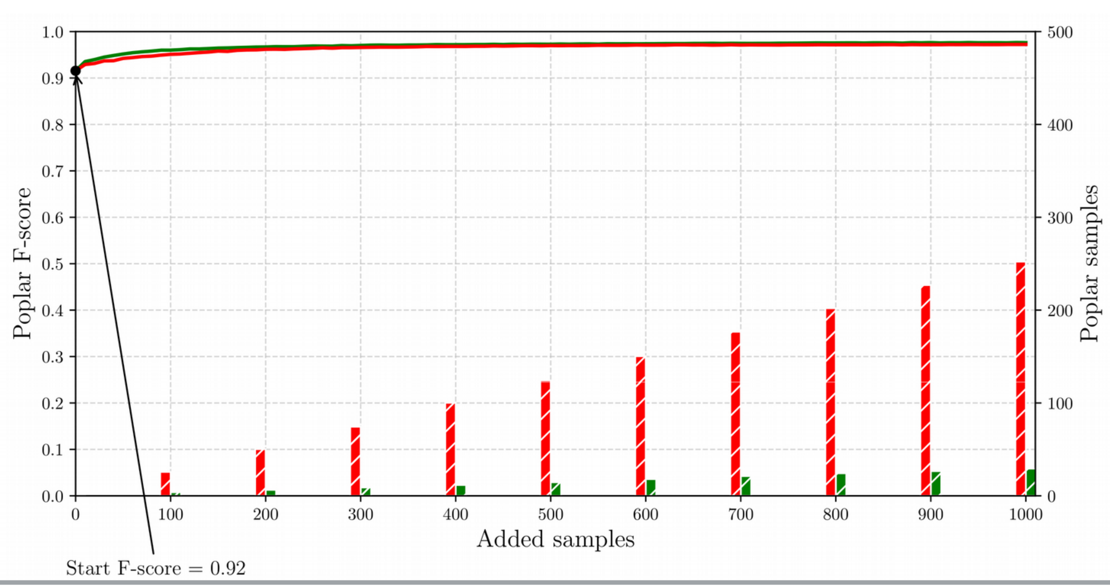
→ In all cases, AL achieved better OA scores with less samples than the random selection.

- **Class specific assessment: use of F-scores**

Example: 31UEQ tile (north) → 31TCJ tile (south)



- **Poplar class:**



→ High performance initial poplar F-score: 92%.

→ For equivalent F-score, random samples outnumber active ones.

CONCLUSIONS

- High potential of Sentinel-2 time series to map poplar plantations at the tile scale.
- Interest of active learning to build easily a global model with a minimum number of samples → approach adapted to a large scale context and is a suitable option where reliable and appropriate training samples are limited, time-consuming or expensive to collect.
- Noise influence on active learning performance: impact of heterogeneous classes (closed deciduous forest), undetected clouds...

Mediterranean and Middle-East Geoscience and Remote Sensing Symposium, 9 – 11 March, 2020 • Tunis, Tunisia