

FAO workshop – Santa Marta (Colombia), July 2024

## Using the deforisk QGIS plugin for making and comparing deforestation risk maps



Ghislain VIEILLEDENT<sup>1</sup> Thomas ARSOUZE<sup>1</sup> FAO team<sup>2</sup>

[1] Cirad UMR AMAP, [2] FAO Rome and Latin America



**cirad** **AIM** **Forests**



UK Government

# Outline

---

- 1 The deforisk QGIS plugin
  - Aim and specificities
  - Website and documentation
  - Installation
- 2 Data preparation
  - Get variables
  - Forest cover change data
  - Spatial explanatory variables
- 3 Models and validation
  - Benchmark model
  - Forestatrisk models
  - Moving window models
  - Validation
- 4 Usage
  - Allocating deforestation
  - Subnational jurisdictions
  - User's data
- 5 Conclusion
  - Workshop agenda
  - Perspectives

# Outline

---

- 1 The deforisk QGIS plugin
  - Aim and specificities
  - Website and documentation
  - Installation
- 2 Data preparation
  - Get variables
  - Forest cover change data
  - Spatial explanatory variables
- 3 Models and validation
  - Benchmark model
  - Forestatrisk models
  - Moving window models
  - Validation
- 4 Usage
  - Allocating deforestation
  - Subnational jurisdictions
  - User's data
- 5 Conclusion
  - Workshop agenda
  - Perspectives

# Aims

---

- Provide a **tool** to create and compare **deforestation risk maps**.
- At the **jurisdictional** level.
- Following **Verra's methodology** for certification.
- **Allocating deforestation** to projects within the jurisdiction.

# Specificities

---

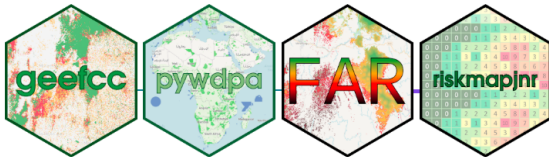
- Open-source and Python based : transparency, reproducibility.
- Computationally efficient :
  - Processing raster by blocks.
  - Running tasks in parallel.
- OS independent : Windows, Linux, MacOS.
- Should run on any computer with average performance.
- Performant alternative statistical models (iCAR).
- Fully documented and translated (English, Spanish, French).
- Help with data preparation.
- Should be (relatively) easy to use.

# Python based

---

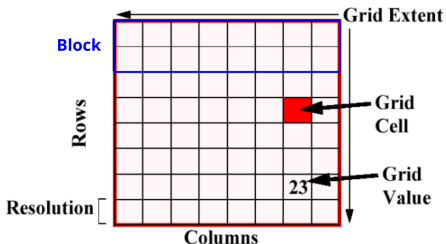
The deforisk plugin relies on four Python packages developed specifically for modelling deforestation :

- `geefcc` : make forest cover change maps from Google Earth Engine (GEE).
- `pywdpa` : downloading protected areas from the World Database on Protected Areas (WDPA).
- `forestatrisk` : model deforestation and predict the spatial deforestation.
- `riskmapjnr` : risk maps following Verra JNR methodologies.



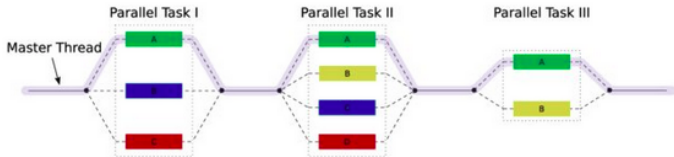
# Processing raster by blocks

- Raster files of forest cover change and explanatory variables might occupy a space of several gigabytes on disk.
- Processing such large rasters in memory can be prohibitively intensive on computers with limited RAM.
- Functions used in the deforisk plugin process large rasters by blocks of pixels representing subsets of the raster data.
- This makes computation efficient, with low memory usage.



# Running tasks in parallel

- State-of-the-art approach to select the best risk map implies repeating tasks (model, periods).
- To save computation time, the deforisk plugin use the QGIS task manager.
- Allows running several analysis in parallel.





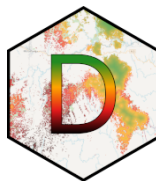
# Website and documentation

---

The website includes all the documentation to use the plugin :

- **Installation page** : How to install the plugin ?
- **Plugin API page** : What is the meaning of each parameter ?
- **Get started page**. How to start using the plugin on a small area of interest ?
- **Articles' page**. How can I use the plugin for specific cases (subnational jurisdictions, user's data) ?
- **References' page** : A page with reference documents including presentations.

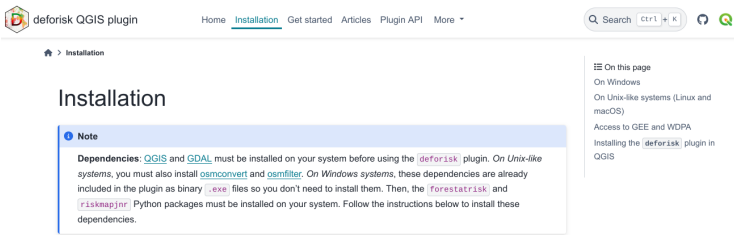
<https://deforisk-qgis-plugin.org>



# Installation



Reduced number of steps for installing the plugin :

- Install QGIS and GDAL on you system (using OSGeo4W on Windows).
- Install the forestatrisk and riskmapjnr Python packages using pip.
- **Download** and install the deforisk plugin from QGIS.
- (Unix-like systems only : install OSM tools).



deforisk QGIS plugin

Home [Installation](#) Get started Articles Plugin API More ▾

Q Search   

🏠 > Installation

## Installation

**Note**

**Dependencies:** [QGIS](#) and [GDAL](#) must be installed on your system before using the [deforisk](#) plugin. On *Unix-like* systems, you must also install [osmconvert](#) and [osmfilter](#). On *Windows* systems, these dependencies are already included in the plugin as binary `.exe` files so you don't need to install them. Then, the [forestatrisk](#) and [riskmapjnr](#) Python packages must be installed on your system. Follow the instructions below to install these dependencies.

☰ On this page

- On Windows
- On Unix-like systems (Linux and macOS)
- Access to GEE and WDPA
- Installing the [deforisk](#) plugin in QGIS

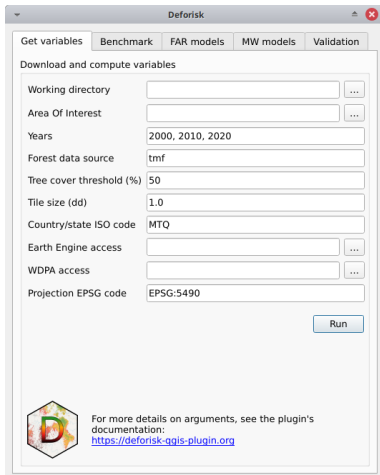
# Outline

---

- 1 The deforisk QGIS plugin
  - Aim and specificities
  - Website and documentation
  - Installation
- 2 Data preparation
  - Get variables
  - Forest cover change data
  - Spatial explanatory variables
- 3 Models and validation
  - Benchmark model
  - Forestatrisk models
  - Moving window models
  - Validation
- 4 Usage
  - Allocating deforestation
  - Subnational jurisdictions
  - User's data
- 5 Conclusion
  - Workshop agenda
  - Perspectives

# Get variables

- Functions to help prepare the data for modelling deforestation.
- Two different sources for **forest cover change** (GFC or TMF).
- Spatial explanatory variables describing **forest accessibility** and **land tenure** (altitude, slope, distance to roads, protected areas, etc.).



# GFC dataset

- Hansen et al. 2013.
- Global dataset encompassing all forest types.
- Tree cover and annual tree cover loss.
- 30m resolution, from 2000 on.
- Data : <https://glad.earthengine.app/view/global-forest-change>

The screenshot shows the Earth Engine Apps interface for the Global Forest Change dataset. At the top left is the GLAD logo and the text "Global Land Analysis & Discovery". A search bar with the placeholder "Search places" is in the top center. On the right, it says "Earth Engine Apps". The main area is a map showing South America and Africa with a forest cover overlay. Below the map are controls for "Layers", "Plan", and "Satellite". On the right side, there is a sidebar with the following text:

**Global Forest Change**  
Published by Hansen, Potapov, Moore, Hancher et al.

**University of Maryland**  
Department of Geographical Sciences

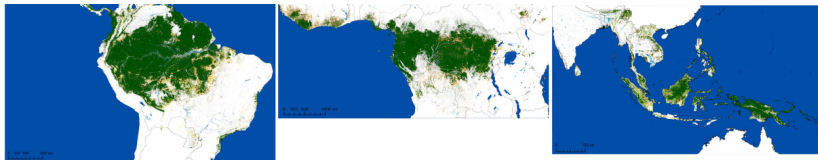
Results from time-series analysis of Landsat images characterizing forest extent and change.

Trees are defined as vegetation taller than 5m in height and are expressed as a percentage per output grid cell as 2000 Percent Tree Cover. 'Forest Cover Loss' is defined as a stand-replacement disturbance, or a change from a forest to non-forest state, during the period 2000-2023. 'Forest Cover Gain' is defined as the inverse of loss, or a non-forest to forest change entirely within the period 2000-2012. 'Forest Loss Year' is a disaggregation of total 'Forest Loss' to annual time scales.

# TMF dataset

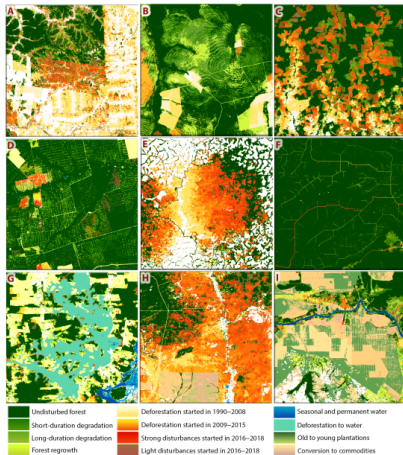
---

- Vancutsem et al. 2021. Tropical Moist Forests (evergreen forest, no dry deciduous forests).
- 30m resolution, from 1990 on.
- Tropical deforestation was underestimated (-33% in 2000–2012, Hansen et al. 2013), especially in Africa.
- Data : <https://forobs.jrc.ec.europa.eu/TMF/>.



# TMF dataset

- Precise enough to visually identify the causes of deforestation (logging, fires, agriculture)



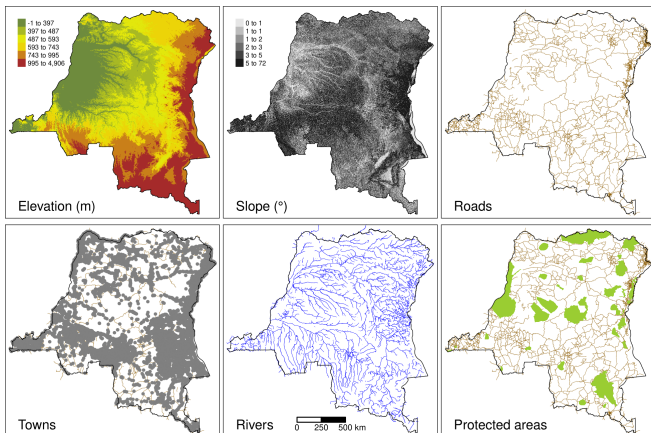
## Spatial variables

The plugin helps computing eight explanatory variables.

Product	Source	Variable derived	Unit	Resolution (m)	Date
Forest maps (2000-2010-2020)	Vancutsem et al. 2021	distance to forest edge	m	30	–
		distance to past deforestation	m	30	–
Digital Elevation Model	SRTM v4.1 CSI-CGIAR	elevation	m	90	–
Highways	OSM-Geofabrik	slope	degree	90	–
		distance to road	m	150	March 2021
Places		distance to town	m	150	March 2021
Waterways		distance to river	m	150	March 2021
Protected areas	WDPA	presence of protected area	–	30	March 2021



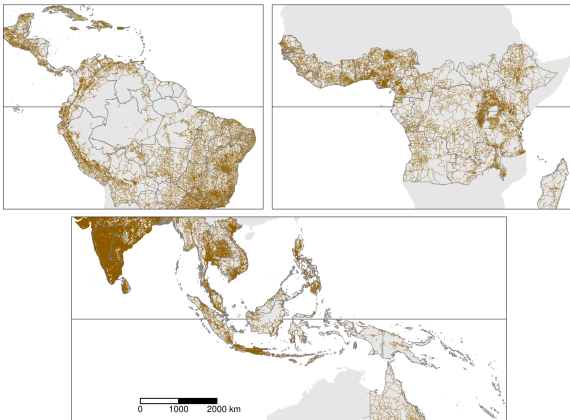
# Spatial variables



## Spatial explanatory variables in DRC

# Roads

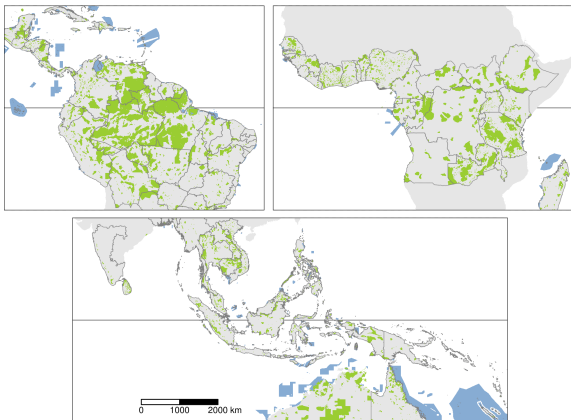
- OpenStreetMap (OSM)
- “motorway”, “trunk”, “primary”, “secondary” and “tertiary” roads
- 3.6 million roads from OSM



# Protected areas

---

- PA status : “Designated”, “Inscribed”, “Established”, or “Proposed”.
- 85,000 protected areas from WDPA.



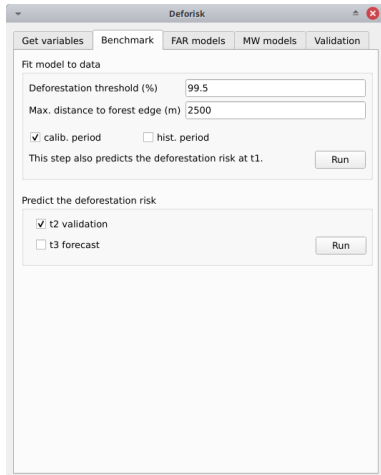
# Outline

---

- 1 The deforisk QGIS plugin
  - Aim and specificities
  - Website and documentation
  - Installation
- 2 Data preparation
  - Get variables
  - Forest cover change data
  - Spatial explanatory variables
- 3 Models and validation
  - Benchmark model
  - Forestatrisk models
  - Moving window models
  - Validation
- 4 Usage
  - Allocating deforestation
  - Subnational jurisdictions
  - User's data
- 5 Conclusion
  - Workshop agenda
  - Perspectives

# Benchmark model

- Benchmark model or reference model.
- A reasonably good deforestation model (better than a null model).
- Assuming a *decrease of deforestation with distance to forest edge* (commonly admitted).
- And a *different model between subjurisdictions* (regional variability).
- See presentation **Cirad and FAO**. 2024. **Jurisdictional risk maps for allocating deforestation.**



# Forestrisk models

- Three statistical models : iCAR, GLM, RF.
- iCAR : Logistic regression with spatial random effects (iCAR process).
- GLM : Generalized Linear Model, simple logistic regression (no random effects).
- Random Forest model : random regression trees.
- Statistical models based on a sample of the observations.

**Deforisk**

Get variables | Benchmark | **FAR models** | MW models | Validation

Sample observations

N# samples: 10000  Adapt sampling

Random seed: 1234

Spatial cell size (km): 2

calib. period  hist. period

Fit models to data

List of variables: C(pa), dist\_edge, altitude, slope, ...

Starting values for beta: -99

Prior Vrho: -1

MCMC: 5000

Variable selection

calib. period  hist. period

Predict the deforestation risk

Spatial cell size interpolation (km): 0.1

iCAR model  GLM  RF model

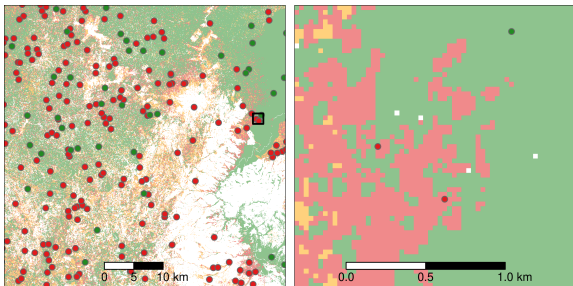
t1 calibration  t2 validation

t1 historical  t3 forecast

## Sampling for FAR models

---

- We consider the forest cover change between  $t$  and  $t + 1$ .
- Stratified sampling between deforested/non-deforested pixels.
- Total number of points proportional to the forest cover (from 20,000 to 100,000 points per study area).



# iCAR model

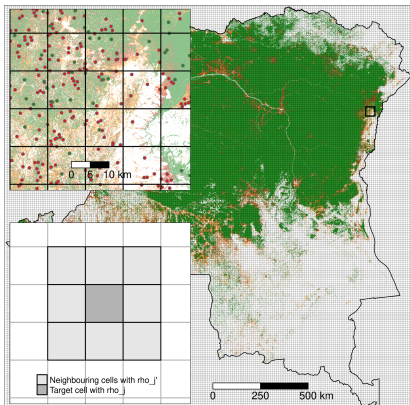
A logistic regression model with iCAR process :

$$y_i \sim \text{Bernoulli}(\theta_i)$$

$$\text{logit}(\theta_i) = \alpha + X_i\beta + \rho_{j(i)}$$

$$\rho_{j(i)} \sim \text{Normal}\left(\sum_{j'} \rho_{j'} / n_j, V_\rho / n_j\right)$$

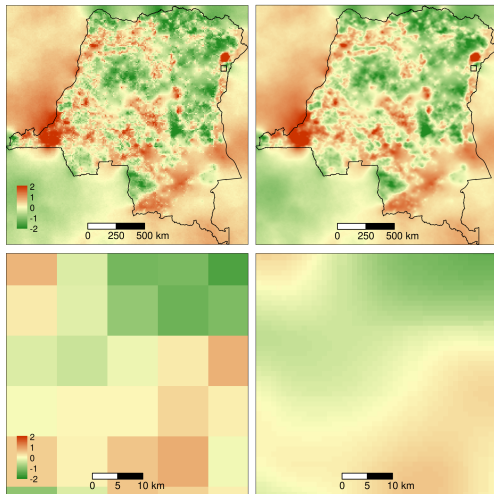
**Random effects  $\rho_{j(i)}$  allows accounting for residual spatial variation not taken into account by model variables  $X_j$ .**



**Square grid of 10km cells over DRC**



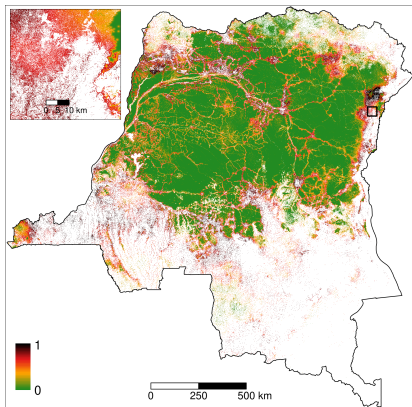
# Spatial random effects



**Interpolation of spatial random effects at 1km in DRC**

# Spatial probability of deforestation

- We use the fitted model to compute the spatial probability of deforestation.
- Probabilities in  $[0, 1]$  are transformed into classes in  $[1, 65535]$ .



**Relative spatial probability of deforestation in DRC**

## GLM model

---

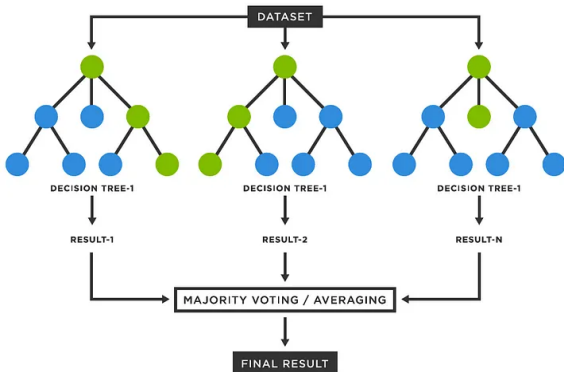
A simple logistic regression model without random effects :

$$y_i \sim \text{Bernoulli}(\theta_i)$$
$$\text{logit}(\theta_i) = \alpha + \mathbf{X}_i\beta$$

Easy to compare with iCAR to see the impact of spatial random effects.

# Random Forest model

- Random Forest is an ensemble machine learning algorithm.
- Combines multiple decision trees to create a more robust and accurate predictive model.



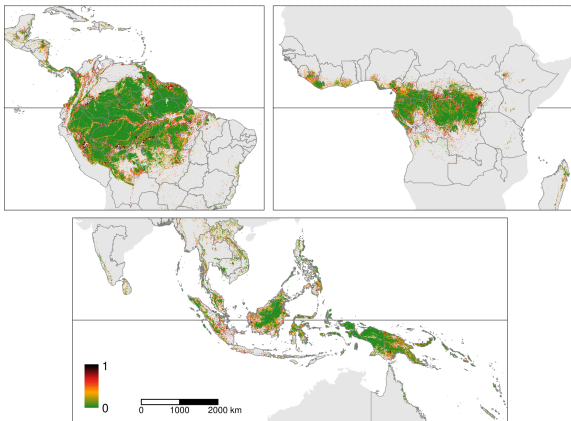
## ForestAtRisk in the tropics

- **i.** Consider tropical moist forest in **92** countries (119 study areas)
- **ii.** Estimate the current deforestation rate and uncertainty in each country
- **iii.** Model the spatial risk of deforestation from environmental factors
- **iv.** Forecast the deforestation assuming a business-as-usual scenario
- **v.** Consequences in terms of carbon emissions



The 119 study areas in the 3 continents

# ForestAtRisk in the tropics



## Pantropical map of the spatial probability of deforestation

Article in review : [10.1101/2022.03.22.485306](https://doi.org/10.1101/2022.03.22.485306)

<https://forestatrisk.cirad.fr/maps.html>

# Moving window models

- Model proposed by previous Verra's methodology.
- Find a distance threshold to define class 1 for the deforestation risk (same thing as for the benchmark model).

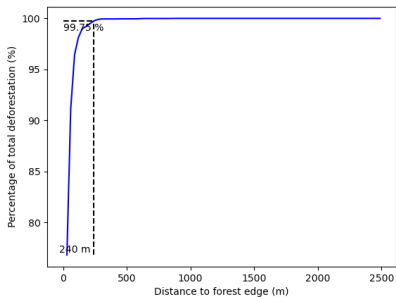


Figure – Cumulative deforestation as a function of the distance to forest edge.

The screenshot shows the 'Deforisk' software interface with the 'MW models' tab selected. The interface is divided into two main sections: 'Fit model to data' and 'Predict the deforestation risk'. In the 'Fit model to data' section, there are three input fields: 'Deforestation threshold (%)' with the value 99.5, 'Max. distance to forest edge (m)' with the value 2500, and 'Window sizes (# pixels)' with the values 11, 21. There are two checkboxes: 'calib. period' (checked) and 'hist. period' (unchecked). A 'Run' button is located to the right. In the 'Predict the deforestation risk' section, there are four checkboxes: 't1 calibration' (checked), 't2 validation' (checked), 't1 historical' (unchecked), and 't3 forecast' (unchecked). A 'Run' button is located to the right.

# Moving window models

- Compute a local risk of deforestation at the pixel level using a moving window.
- The moving window can be of different sizes.
- Deforestation rates in  $[0, 1]$  are converted to  $[2, 65535]$ .

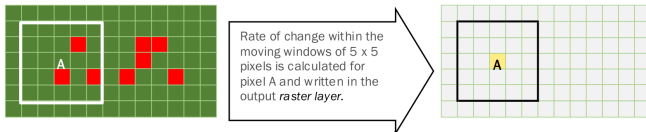
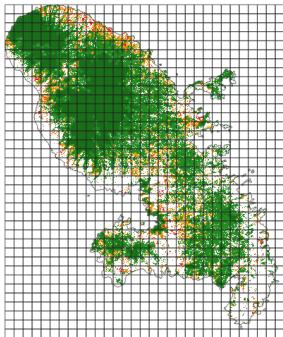


Figure – Moving window.



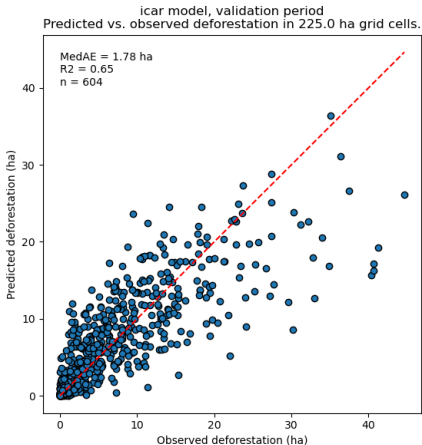
# Validation

- Comparing predicted vs. observed deforestation (in ha) for each cell in a coarse grid.
- For a given period of time.



# Validation

- Performance indices :  $R^2$ , and median of absolute error (MedAE).
- Computed for each model and each period (calibration, validation, historical).



# Outline

---

- 1 The deforisk QGIS plugin
  - Aim and specificities
  - Website and documentation
  - Installation
- 2 Data preparation
  - Get variables
  - Forest cover change data
  - Spatial explanatory variables
- 3 Models and validation
  - Benchmark model
  - Forestatrisk models
  - Moving window models
  - Validation
- 4 Usage
  - Allocating deforestation
  - Subnational jurisdictions
  - User's data
- 5 Conclusion
  - Workshop agenda
  - Perspectives

## Allocating deforestation

For the best model, we obtain at t3 :

- A jurisdictional map with classes of deforestation risk.
- A table with relative deforestation rates for each class.

**Table** – Deforestation rates at t3 for each class of deforestation risk (numbers truncated to three decimal digits).

cat	$n_i$	$d_i$	$\theta_{m,i}$	$\theta_{a,i}$	$T$	$A$	$\delta_i$
1	137575	–	1.000e-06	–	–	0.09	–
2	5425	–	1.625e-05	–	–	0.09	–
3	3523	–	3.151e-05	–	–	0.09	–
4	2458	–	4.677e-05	–	–	0.09	–
5	2078	–	6.203	–	–	0.09	–

## Allocating deforestation

**Table** – Deforestation rates at t3 for each class of deforestation risk (numbers truncated to three decimal digits).

cat	$n_i$	$d_i$	$\theta_{m,i}$	$\theta_{a,i}$	$T$	$A$	$\delta_i$
1	137575	–	1.000e-06	–	–	0.09	–

- Considering a total **deforestation**  $D$  (in ha) for the next  $Y$  **years** at the jurisdictional level.
- **Adjustment factor** is  $\rho = D / (A \sum_i n_i \theta_{m,i})$ , with  $A$  the pixel area in ha.
- **Absolute rate** is  $\theta_{a,i} = \rho \theta_{m,i}$  : so that total predicted deforestation = expected deforestation.
- **Deforestation density** is  $\delta_i = \theta_{a,i} \times A / Y$ . Used to predict the amount of deforestation (in ha/yr) for each forest pixel.

## Allocating deforestation

**Deforestation density** is  $\delta_i$  (in ha/yr) is used to predict the amount of deforestation for each forest pixel.

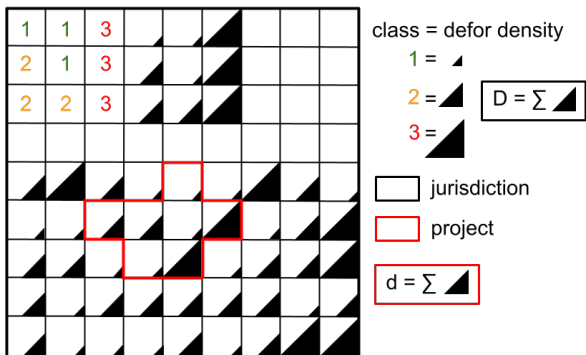


Figure – Allocating deforestation to projects within the jurisdiction.



## User's data

---

- Possibility to use user's data : national forest cover change map, other explanatory variables (e.g. mining concessions).
- Manual steps at the moment.
- Files in the data folder must be replaced with user's data.
- Additional raster variables can be added to the data folder.
- Symbolic links in data\_\* folders must exist.
- More details on the website page [User's data](#).



# Outline

---

- 1 The deforisk QGIS plugin
  - Aim and specificities
  - Website and documentation
  - Installation
- 2 Data preparation
  - Get variables
  - Forest cover change data
  - Spatial explanatory variables
- 3 Models and validation
  - Benchmark model
  - Forestatrisk models
  - Moving window models
  - Validation
- 4 Usage
  - Allocating deforestation
  - Subnational jurisdictions
  - User's data
- 5 Conclusion
  - Workshop agenda
  - Perspectives

# Workshop agenda

---

## Four practical sessions :

- Installing the software and run the Get Started tutorial.
- Chose a small subnational jurisdiction and select the best risk map.
- Derive the best risk map for a large jurisdiction (e.g. country scale).
- Exercises :
  - Change model parameters to see models' behavior (e.g. size of spatial cells for iCAR model).
  - Use country data (e.g. national forest cover change map).
  - Allocate future deforestation to a project.

# Perspectives

---

- Recent plugin (first version in July 2024).
- Improvements are expected :
  - Increase computational speed (for predictions on large areas).
  - Adding more alternative models (MLP).
- Modifications from users' feedback.

... Thank you for attention ...

<https://deforisk-qgis-plugin.org>

> Articles > References > Presentations



**AIM** Forests



UK Government