4th Carbon from Space Workshop

Francesco Minunno

25th - 28th October 2022

Data Assimilation Of Forest Status Using Sentinel 2 Data and a Process-Based Model

Rationale

Large-scale forest monitoring is essential for quantifying greenhouse gas (GHG) exchanges over vast areas.

It is possible to predict ecological processes and quantify GHG exchanges, by means of **process-based models.**

However, **accurate and unbiased information** on ecosystems **current state** are essential to achieve robust estimates.

Nowadays huge amount of **data** is becoming available (e.g., high resolution EO) and there is a need to integrate those information in the modelling frameworks.

Data assimilation (DA) allows to combine model predictions and data from multiple sources, considering the associated uncertainties.

Objectives

Develop two frameworks that allow to assimilate repeated measurements of high resolution (10-30 m) remotely sensed data into a simple process-based forest model.

1. Site fertility class (ST)
2. Forest structural variables (FSV)

Satellite data and field measurements

Areas: three tiles of 100 Km2 across Finland

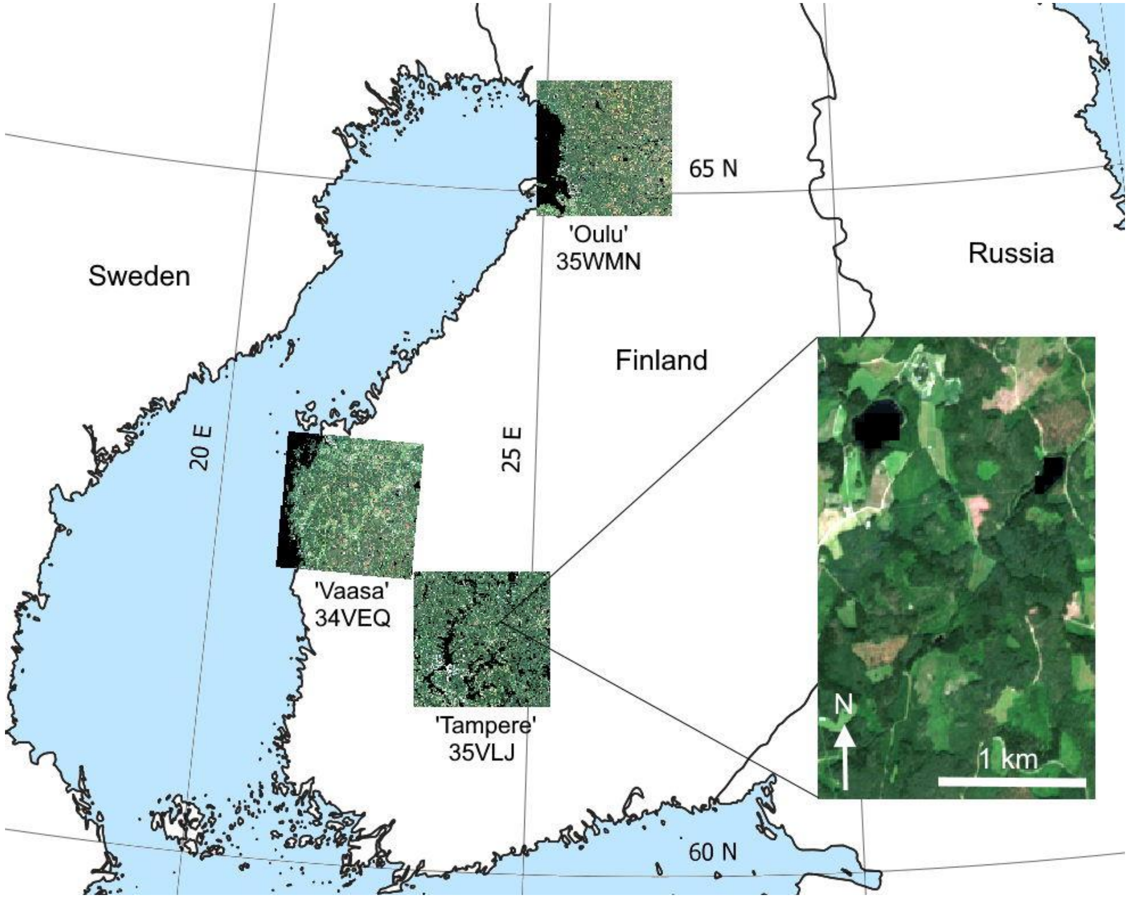
Satellite data collection: sentinel 2 (s2) 2016 and 2019

Resolution: 10x10 m

Field measurements: Finnish Forest Centre campaigns from 2016 and 2019.

Forest state: basal area (B), stand average height (H), diameter at breast height (D), species composition

Managed forests were excluded using a GIS database



Models

PREBAS: simple process-based model

PREBAS emulators: regression models that predict PREBAS outputs.

PREBAS emulators were built and calibrated for each tile using PREBAS outputs from 20000 sampled pixels.

1. FSV and ST estimates based on s2 2016 were used to initialize PREBAS
2. PREBAS output for 2019 was used to calibrate the emulators

Bayesian method for data assimilation

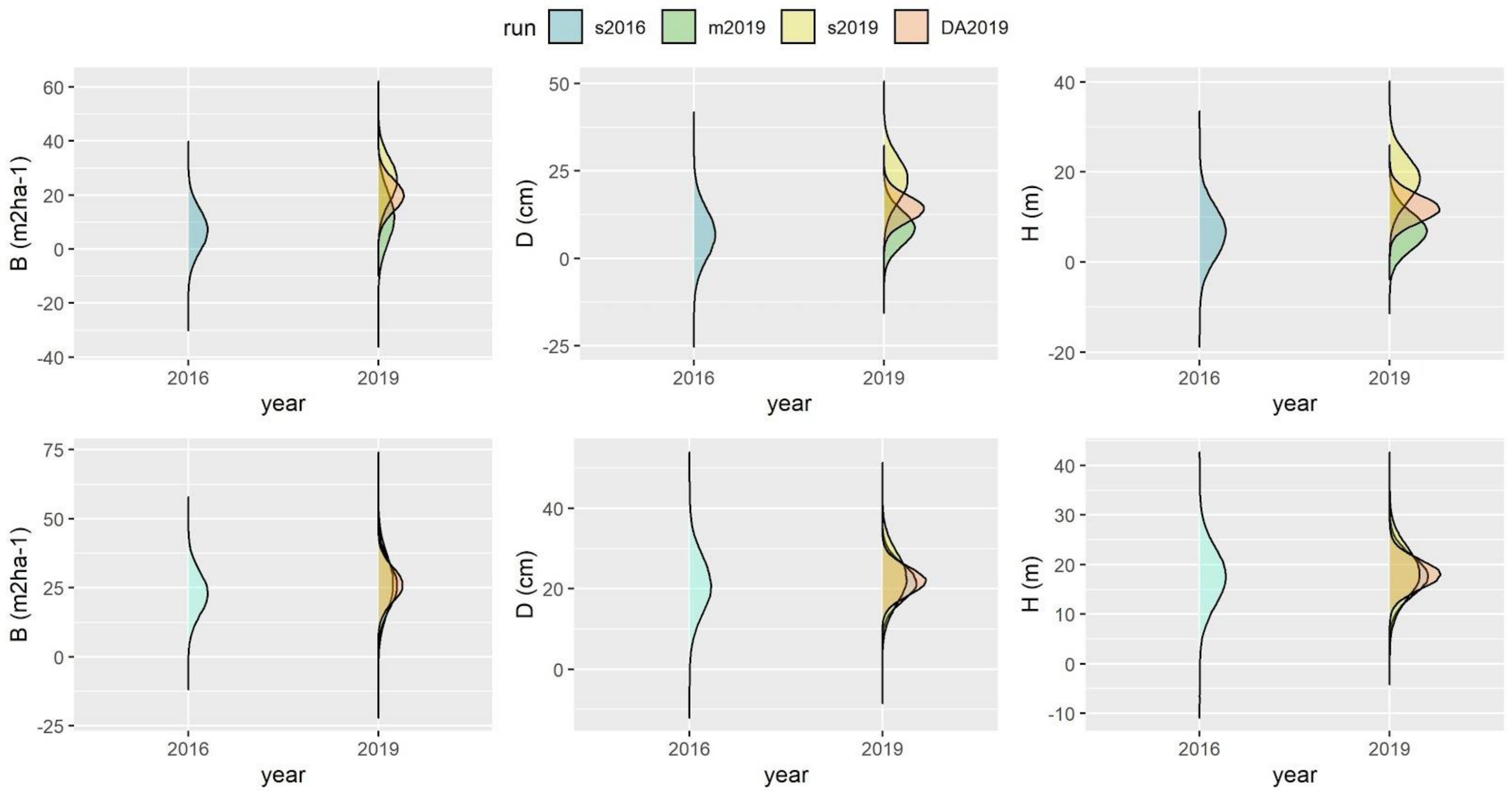
Bayesian method was used for data assimilation in the following steps:

1. The model was initialized with EO based estimates of 2016. Monte Carlo simulations were used to account for the initial state uncertainty.

2. Model forecasts for 2019 (*prior*) were combined with new EO based estimates for 2019 (*likelihood*).

3. New maps were produced for 2019 on the basis of DA results (*posterior*).

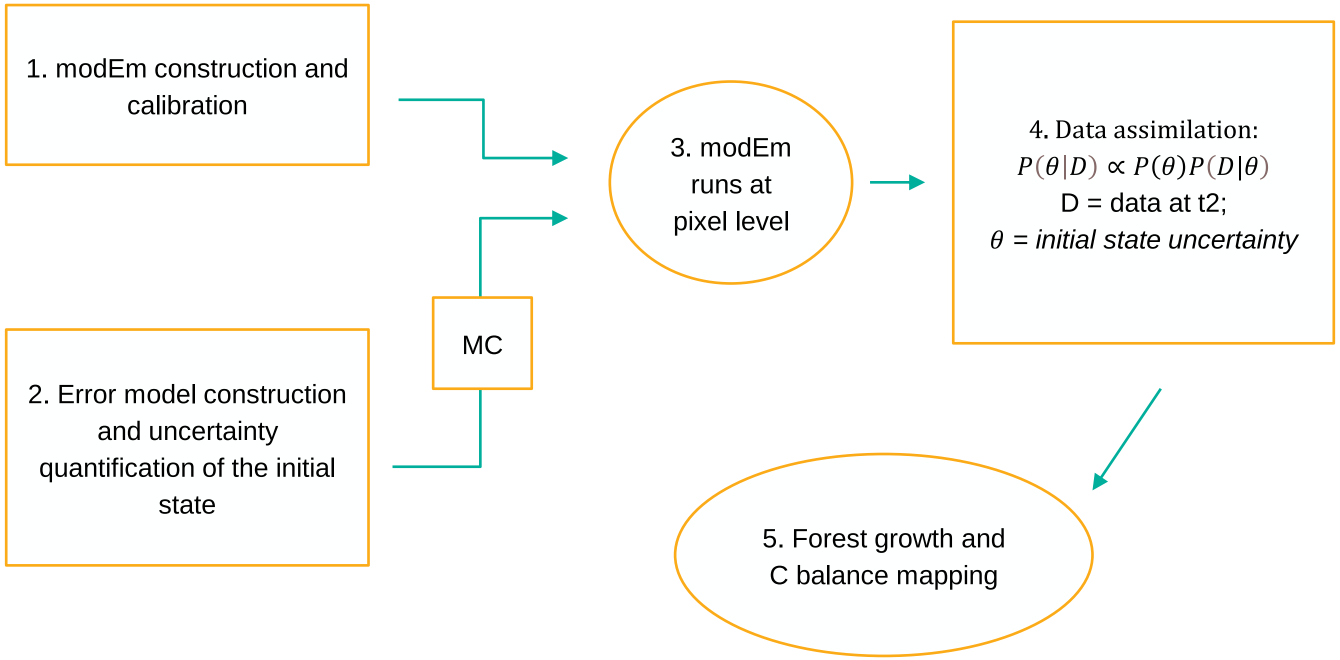
Exanple of DA at pixel level.



Example of data assimilation at pixel level. The blue, green and yellow distributions represent, respectively, the uncertainty of Sentinel 2 satellite based estimates for 2016 (s2016), model predictions for 2019 (m2019) and Sentinel 2 estimates for 2019 (s2019). The distribution of the data assimilation results are reported in red (DA2019). The first row of the figure panels is an example of data assimilation for a pixel where m2019 and s2019 diverged; the second row shows results from a pixel where m2019 and s2019 estimates were consistent

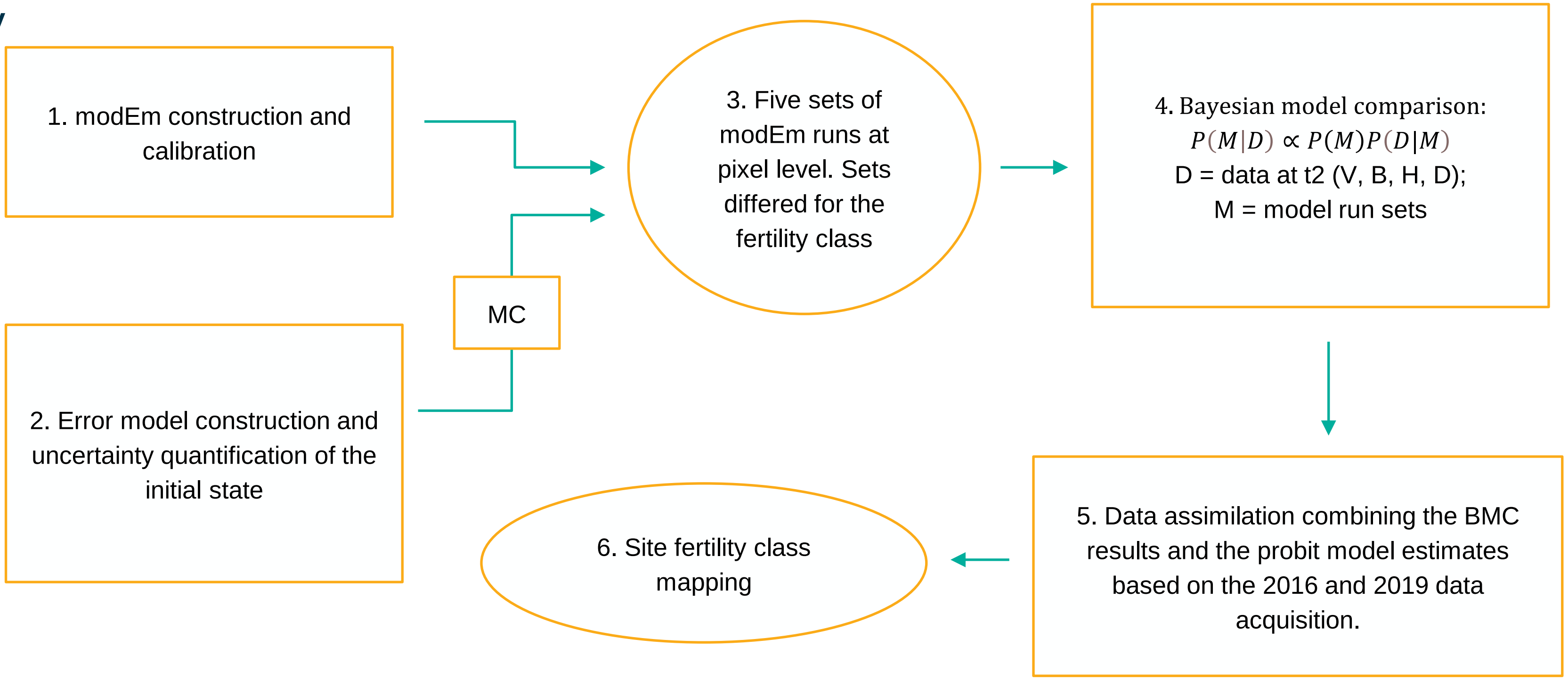
Data assimilation for forest structural variables

1. Emulator calibration;
2. Monte Carlo simulations for the uncertainty quantification of initial state;
3. Emulator runs;
4. Data assimilation;
5. Map production.

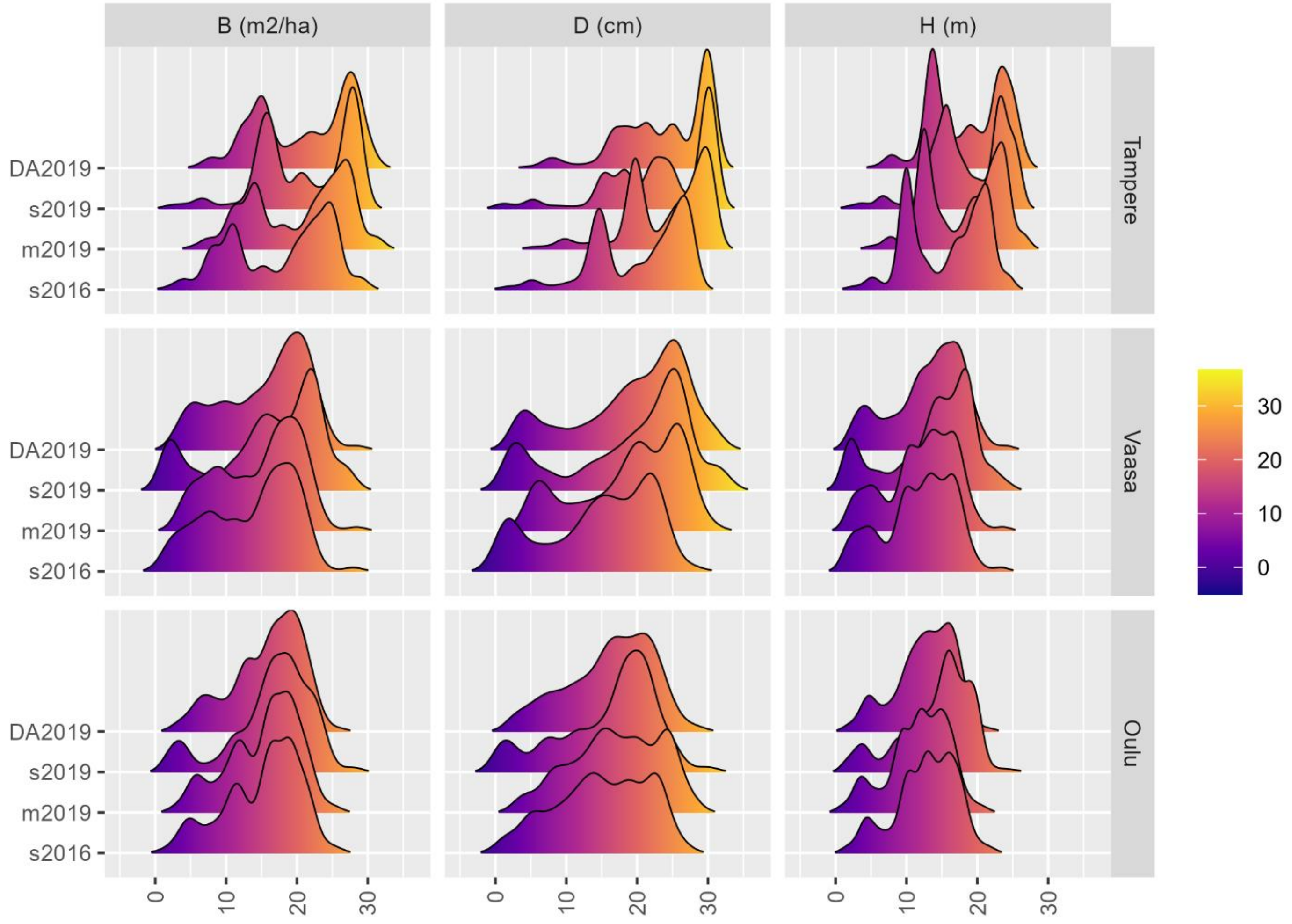


Data assimilation for site fertility

1. Emulator calibration;
2. MC simulations for the uncertainty quantification of initial state;
3. 5 sets of emulator runs;
4. Bayesian model comparison;
5. Data assimilation;
6. Map production.

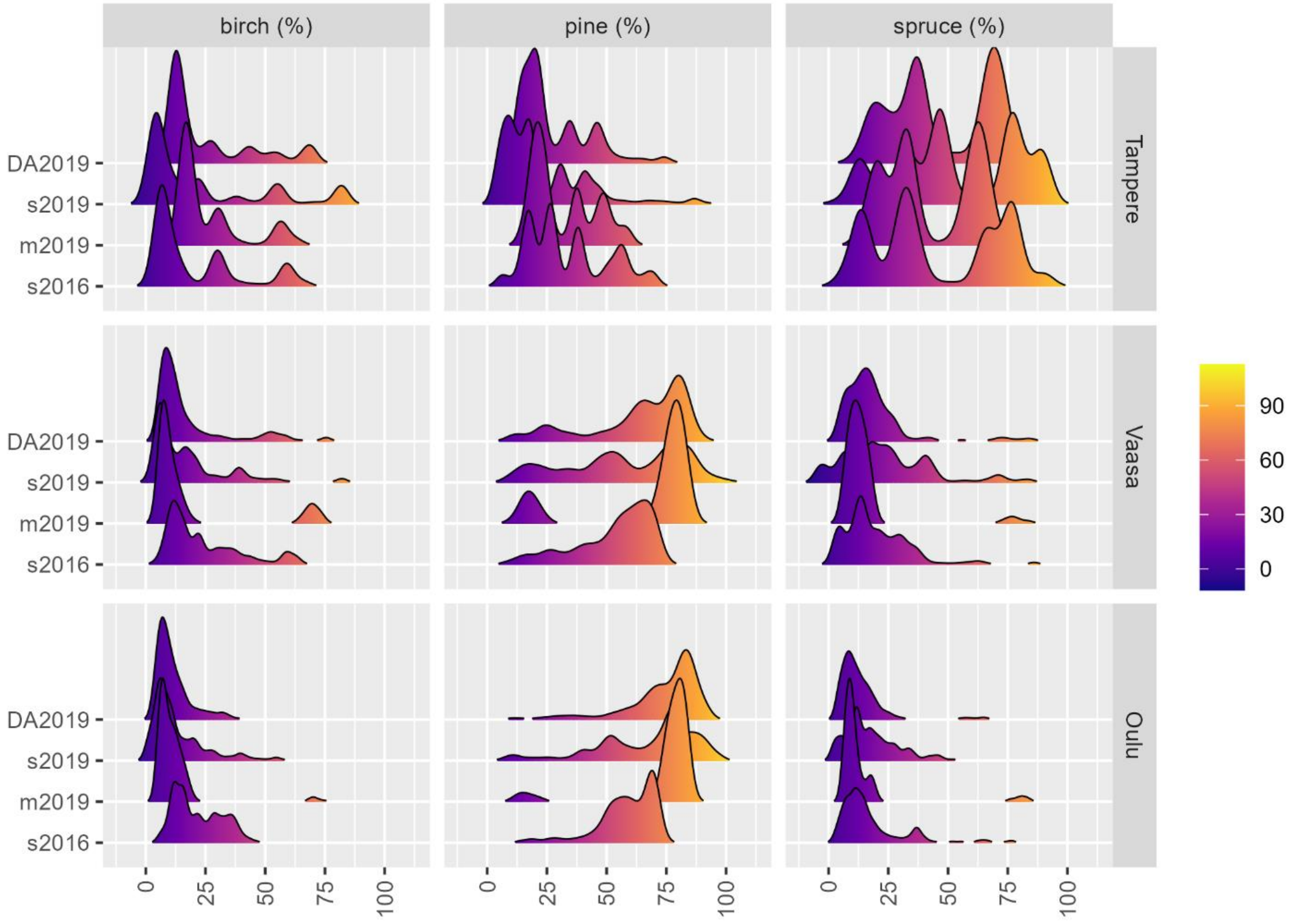


Results



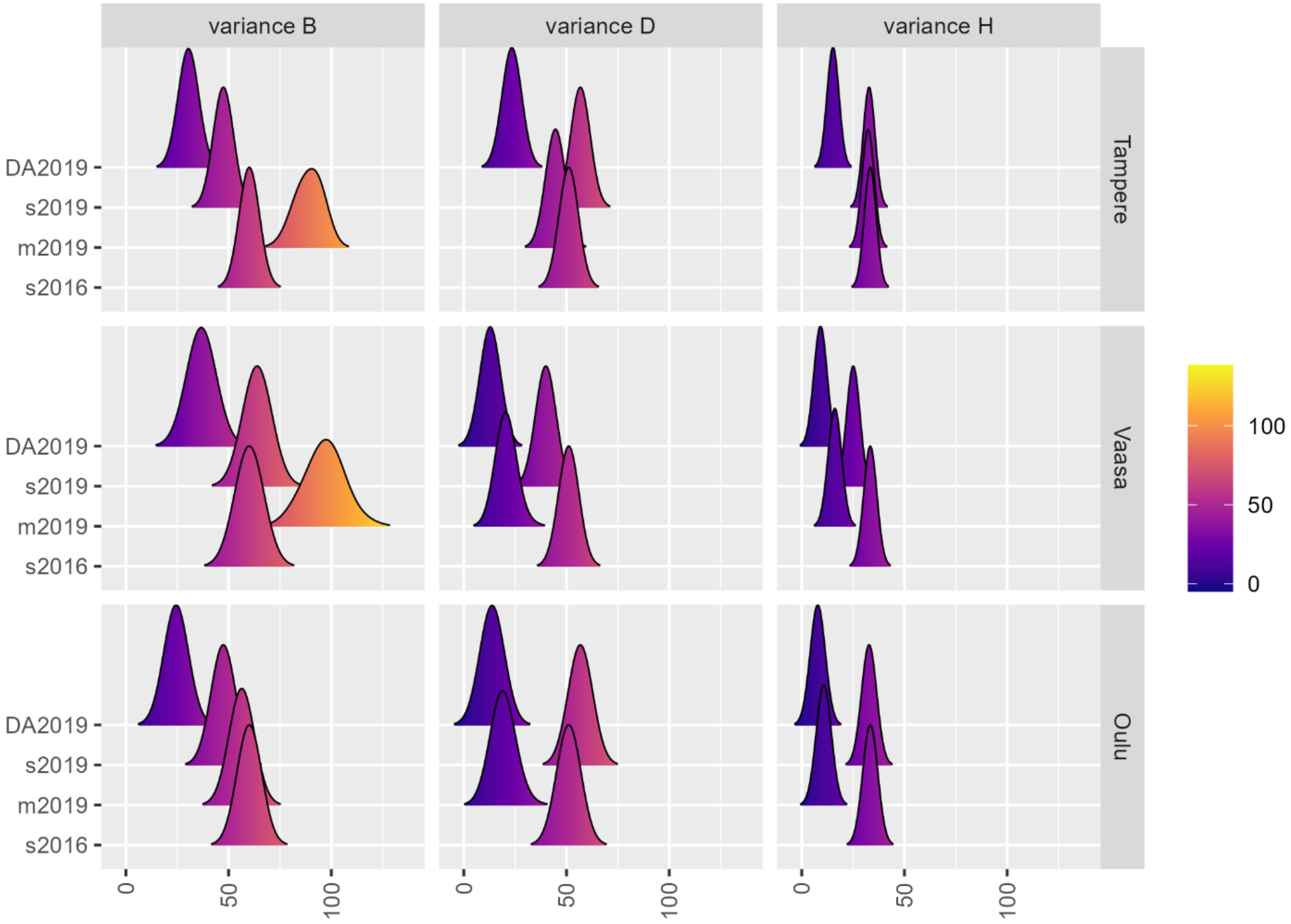
Distribution of the highest probability estimates for each pixel of stand average height (H), stand average diameter at breast height (D) and stand basal area (B) over the three tiles. The distributions were drawn from satellite based estimates (s2016, s2019), model based estimates (m2019) and data assimilation (DA2019).

Results



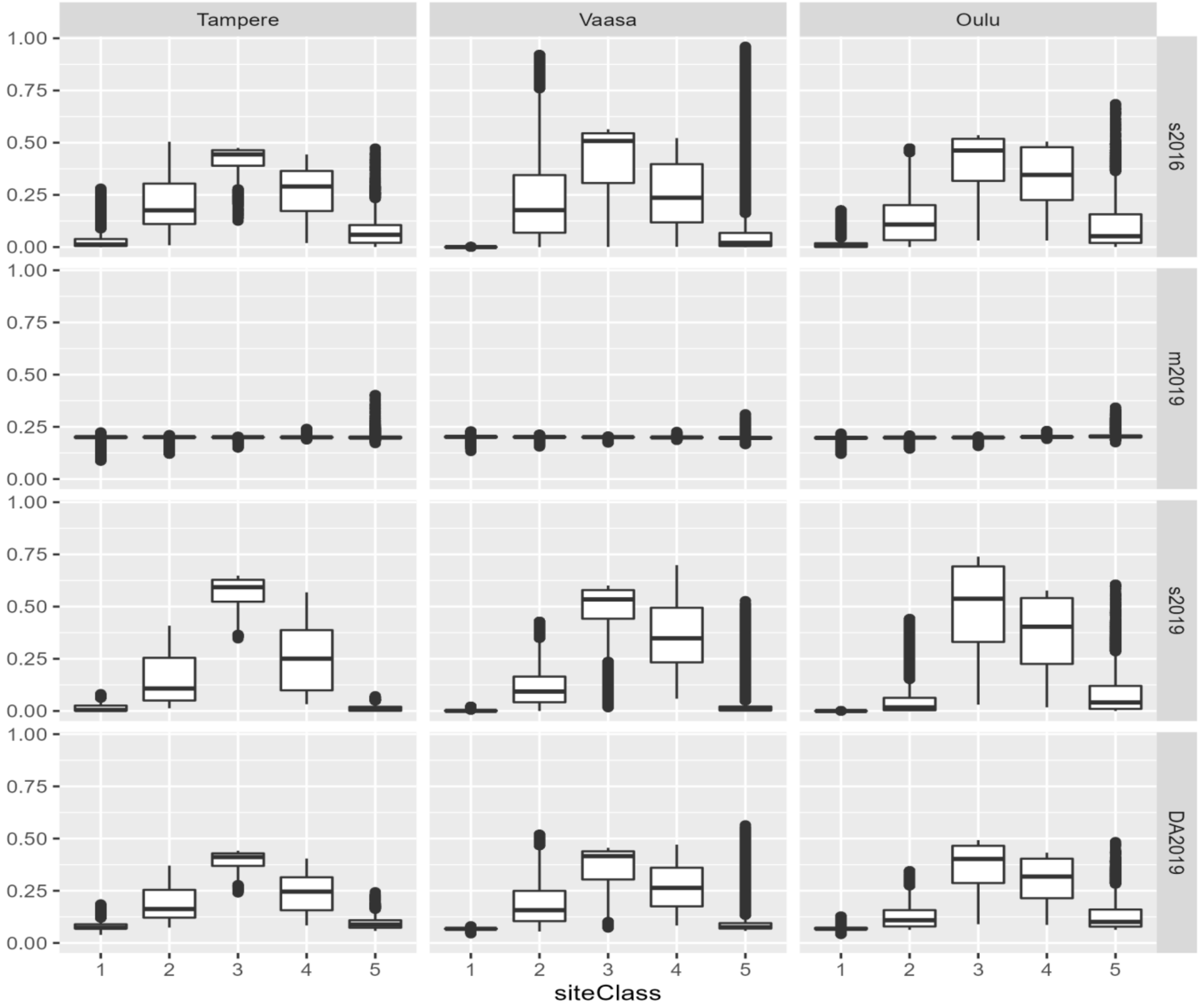
Distribution of deciduous species, pine and spruce percentage cover over the three tiles using the highest probability estimates for each pixel. The distributions were drawn from satellite based estimates (s2016, s2019), model based estimates (m2019) and data assimilation (DA2019).

Results



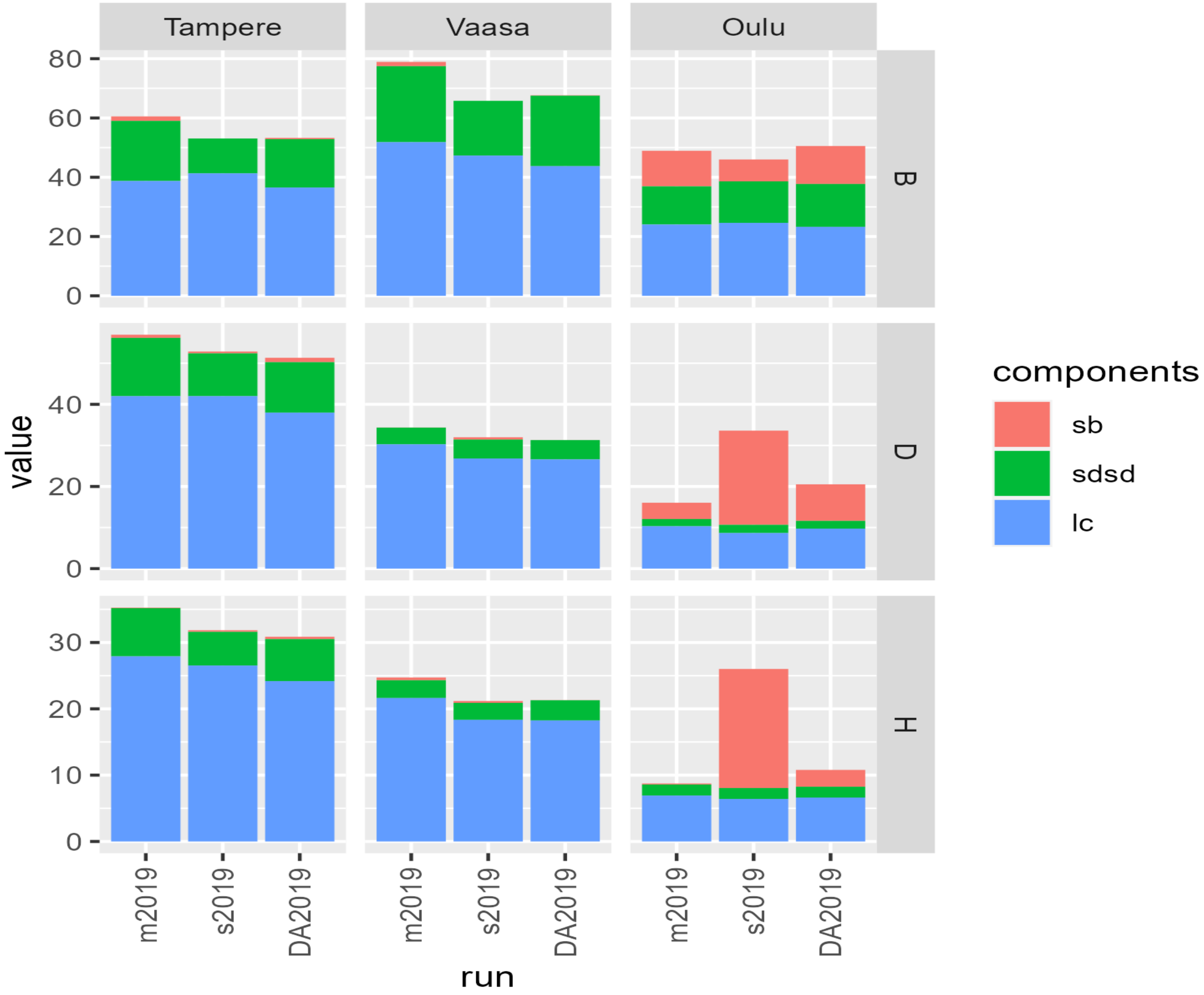
Variance distributions of the stand average height (H), diameter at breast height (D) and basal area (B) over the three tiles. The distributions were drawn from satellite based estimates (s2016, s2019), model based estimates (m2019) and data assimilation (DA2019).

Results



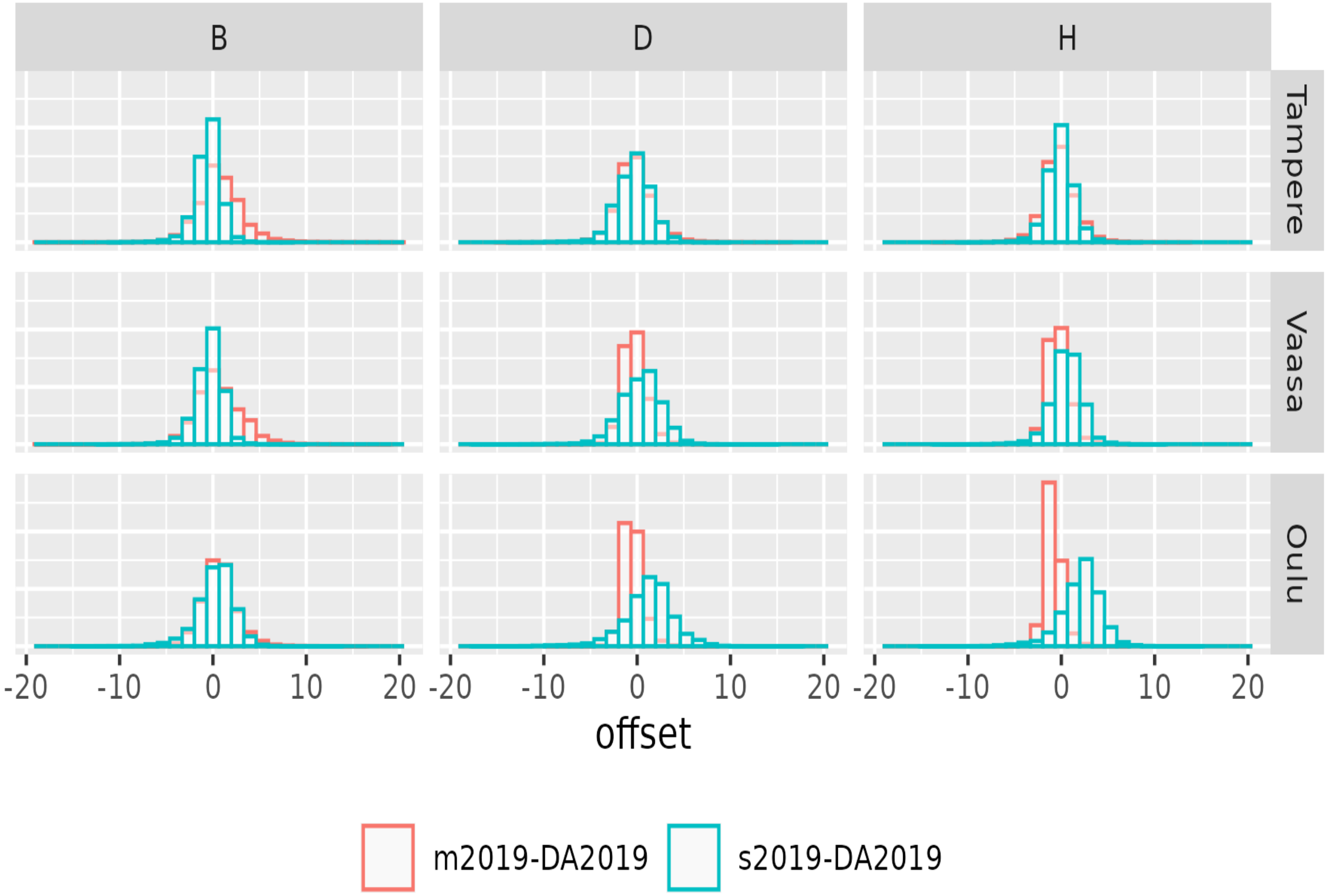
Distribution of site fertility class for the different tiles. Satellite based estimates for 2016 and 2019 (s2016 and s2019), model based estimates (m2019) and data assimilation results (DA2019) are reported.

Results validation



Mean squared errors (MSE) for the forest structural variable estimates based on model forecasts (m2019), Sentinel 2 data of 2019 (s2019) and data assimilation of m2019 and s2019 (DA2019). Mean squared error was decomposed in three components: bias (sb), data variability (sdsd) and lack of correlation (lc).

Results validation



Forest structural variables maps drawn using the *maximum a posteriori probability* of DA2019. The distributions report the deviation between m2019 (red) and s2019 (blue) from DA2019 estimates at pixel level.

Conclusions

- DA of forest structural variables **reduced the uncertainty** of the estimates and **improved the accuracy** of the forest structural variable estimates **reducing the impact of biased data**.

- DA is particularly suitable for Forest monitoring and forest modelling by **continuously updating the current state** of a forest every time new data become available

- Model emulators allowed to reduce the **computational load** of DA, making possible the processing of an enourmous ammount of data.

Knowledge gaps & priorities

DA can be extended to **any kind of model** and **any kind of data** (UAV, eddy covariance, lidar).

The **use of field measurements**, such as inventory campaign (NFI), is always desirable. The advantage is that we can identify the weakest components of the framework, i.e model predictions, satellite based estimates …

Develop in the framework routines that allow to identify and quantify **disturbances over large areas** (change detection algorithms). Integrate information about the **physiological status** of the forests, such as drought stress

Extend the applicability of our DA framework to new environments (i.e., Mediterranean, temperate, alpine, tropical forests).

A **more extensive application** of the framework using data of different uncertainties and longer period simulations is desirable to explore the full potential of the method.

Thanks for the attention!