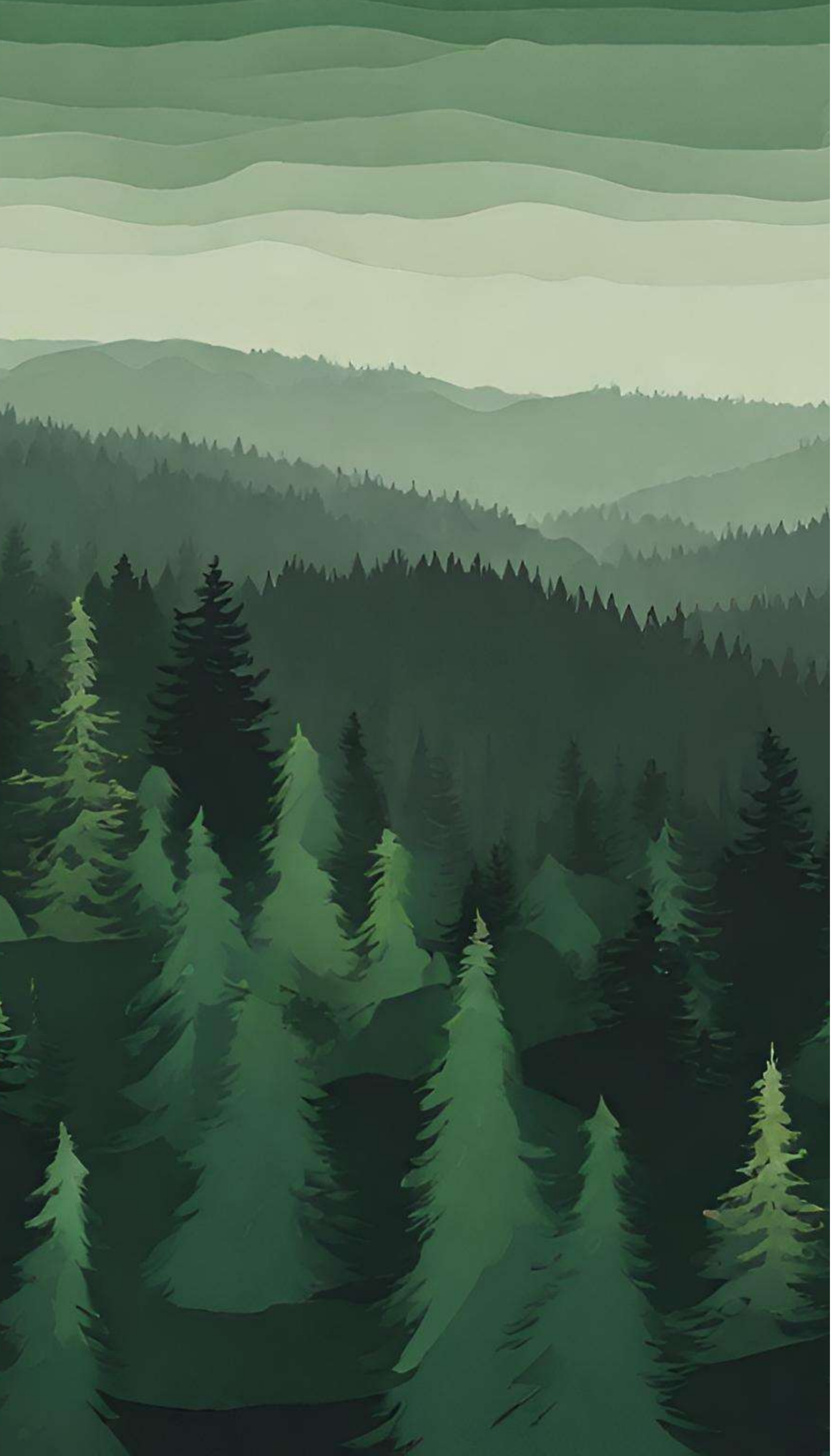
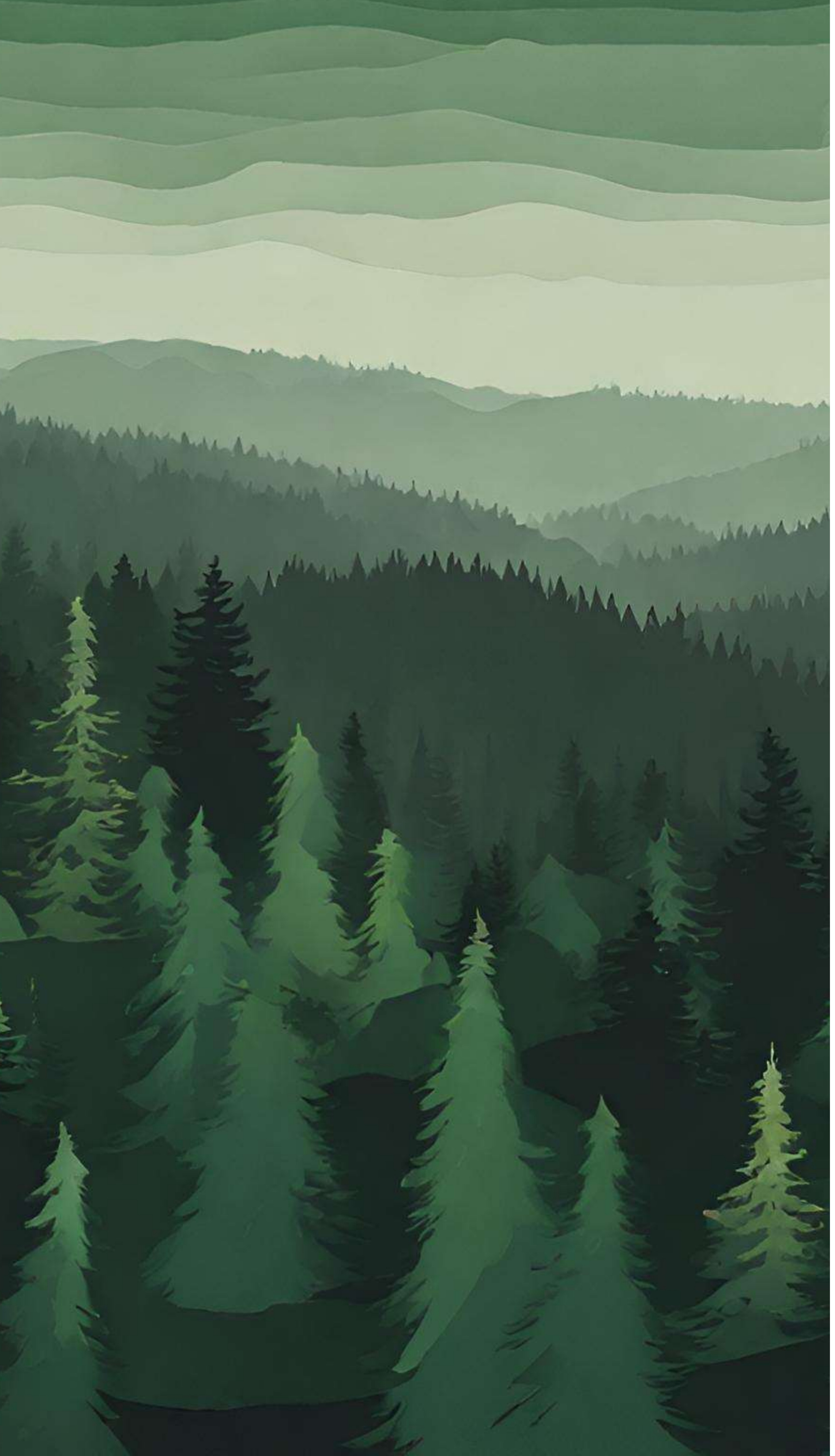


DETERMINING FOREST LANDSCAPE VULNERABILITY TO SPRUCE BUDWORM USING MACHINE LEARNING ALGORITHMS

Friday May 3rd
17th CFR CONFERENCE

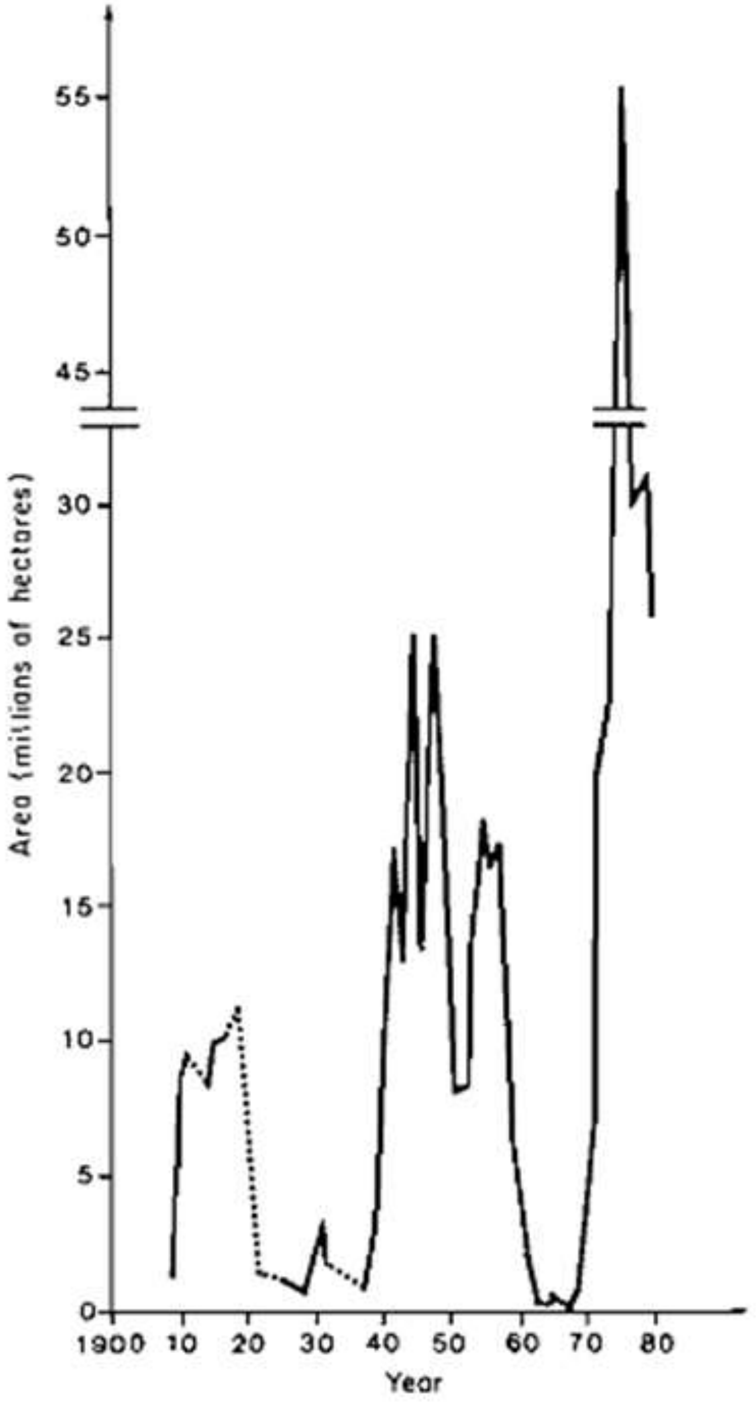
Rindra Ranaivomanana, PhD student, UQAM
Élise Filotas, Pr, TÉLUQ, Dpt Science et technologie
Mathieu Bouchard, Pr, ULaval, Dpt des sciences du bois et de la forêt



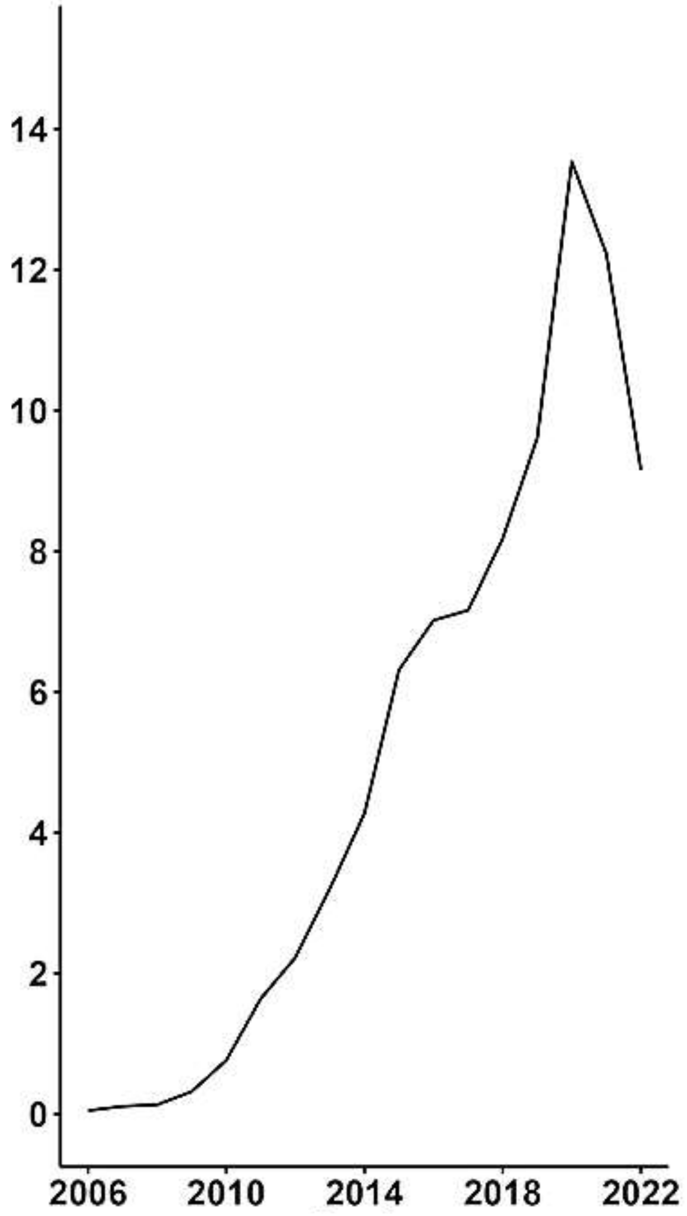
SBW AND ITS IMPACT

SBW

Spruce budworm (Choristoneura fumiferana (Clem.))



Defoliated area (millions of ha)
Past outbreaks 20th century
Blais (1983)



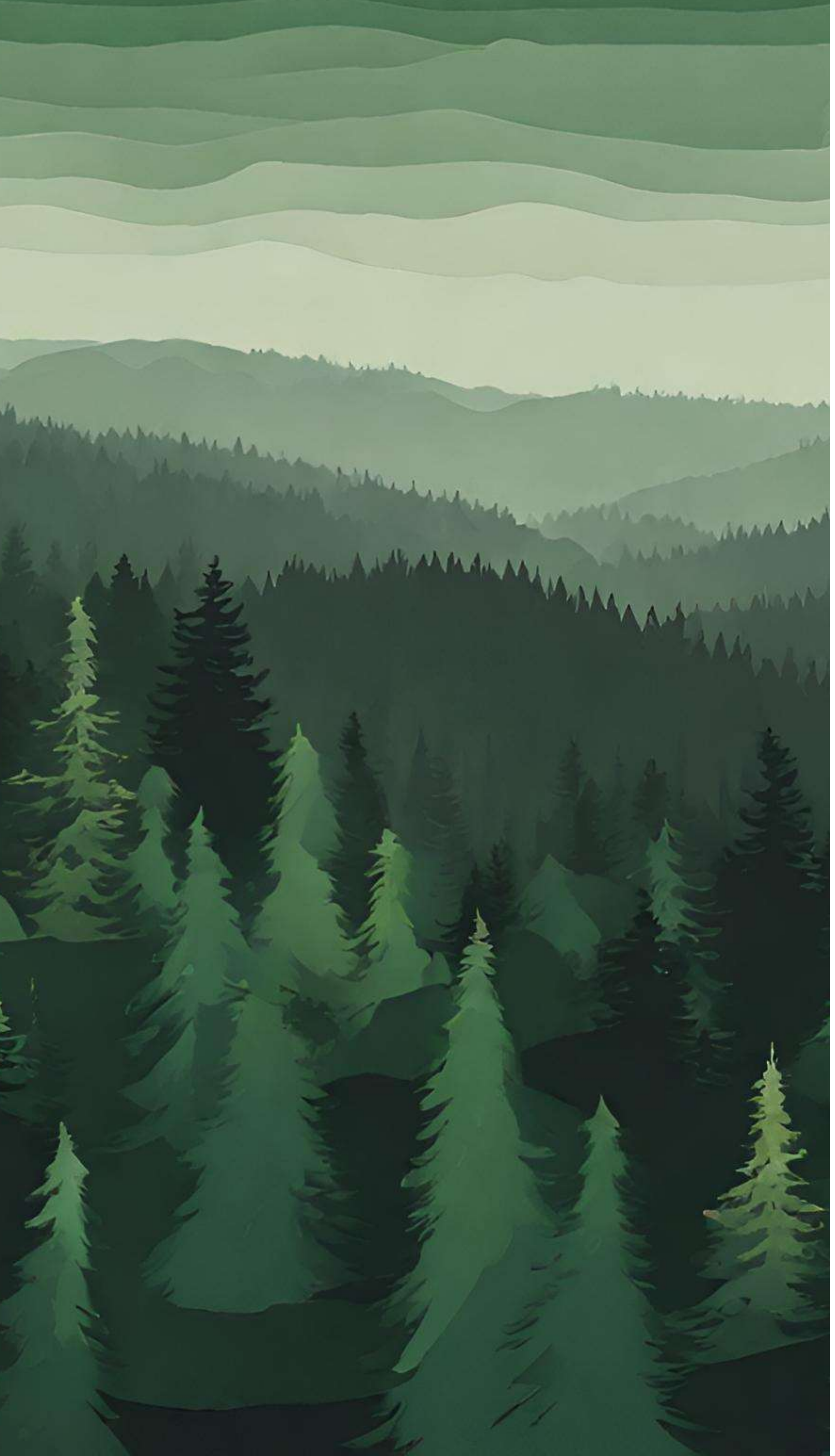
Defoliated area (millions of ha)
Actual outbreak
MFFP (2022)

Objectives

YEAR	BTK TREATMENT % (Areas with all defoliation levels)	BTK TREATMENT % (Areas with severe & moderate defoliation)
2022	10.87	31.00
2021	11.59	57.57
2020	1.33	2.89
2019	7.21	14.17

⟹ Project : Optimize suppression activities

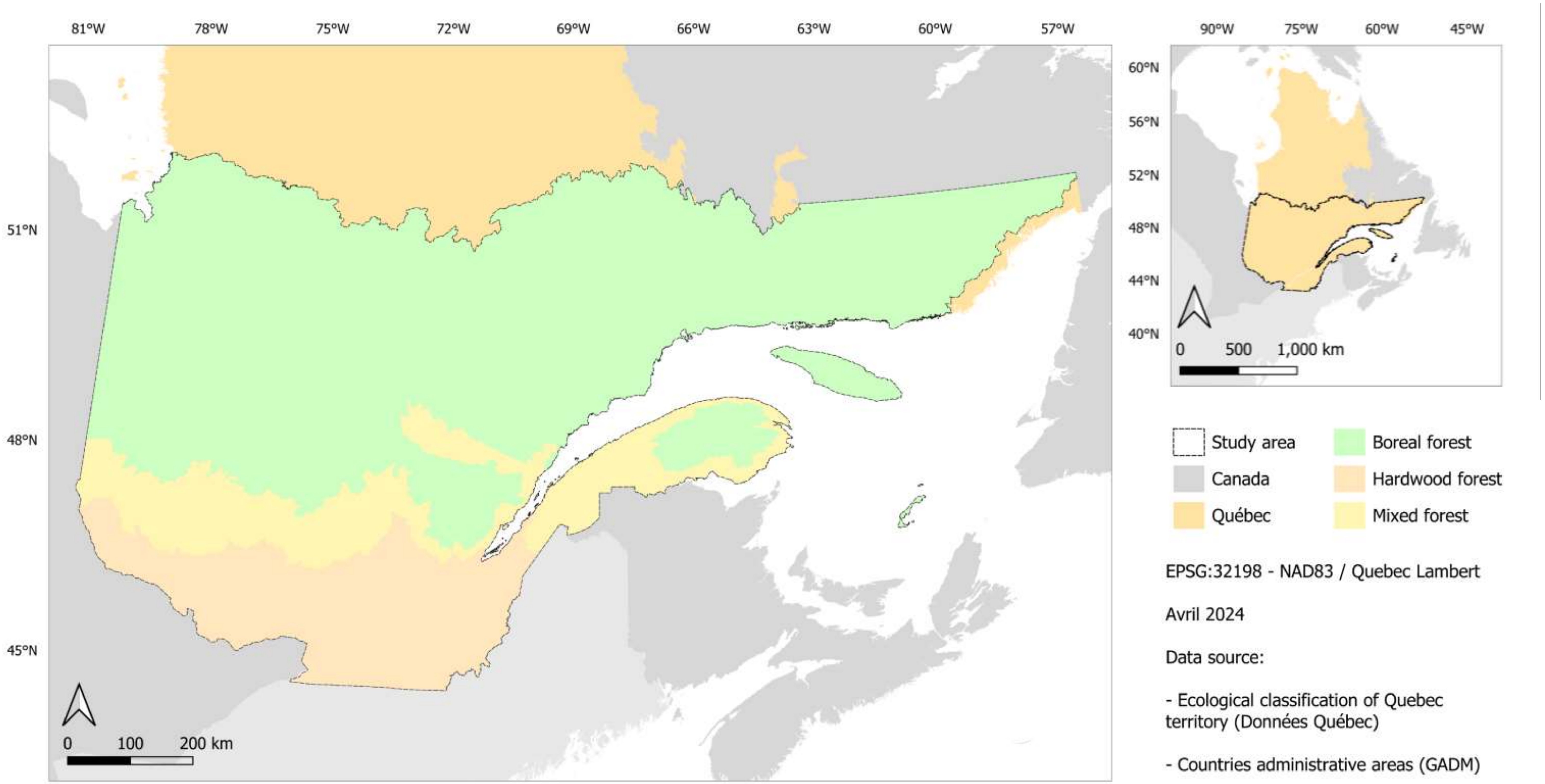
Identify vulnerable areas and factors influencing forest vulnerability at the stand and landscape scale with machine learning models



METHODOLOGY

Study area

South of Québec



Forest vulnerability

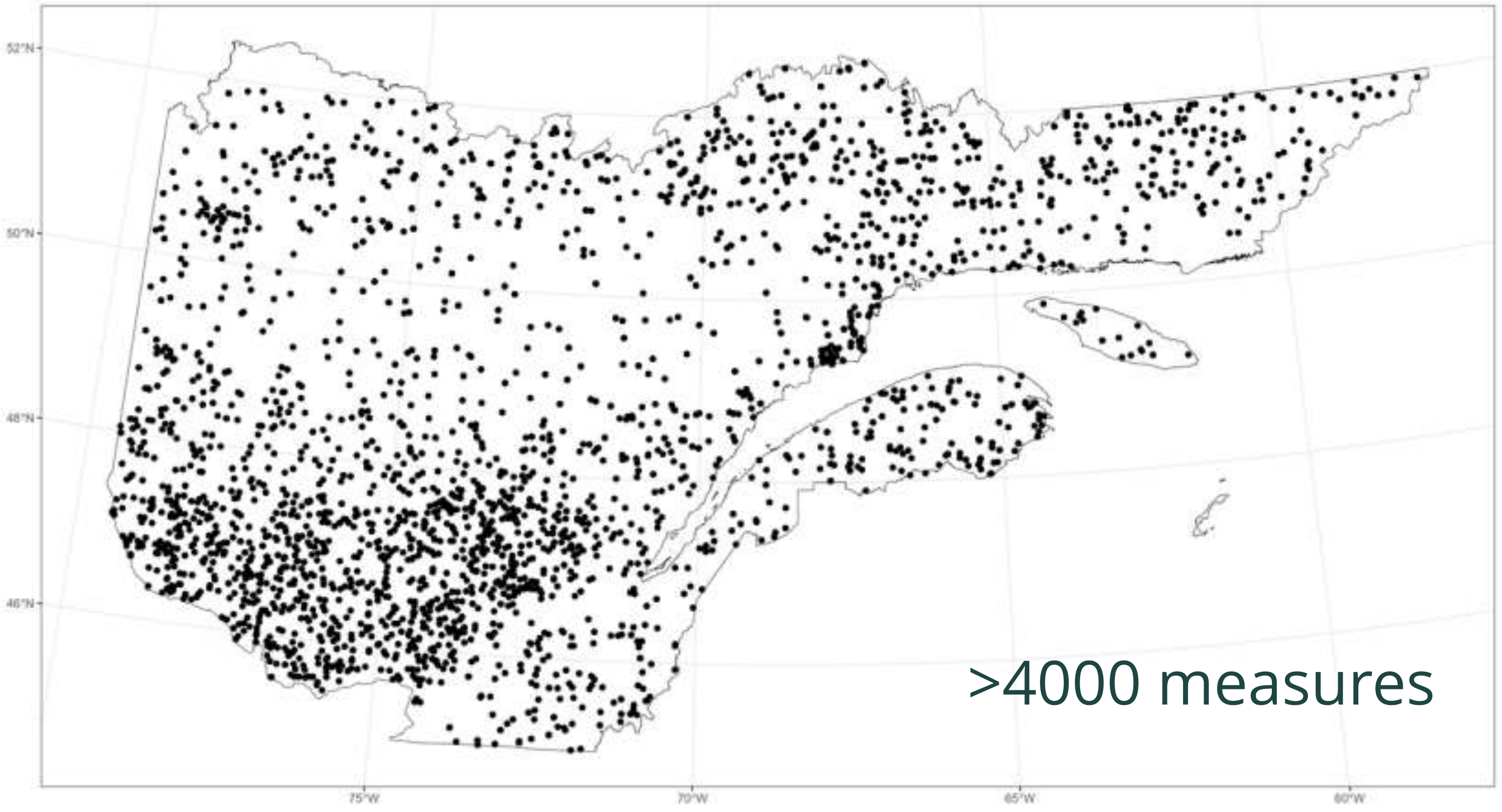
VULNERABILITY = BIOMASS LOSS OF 25% OR MORE

Response variable
in ML models

CLASS 1
Biomass loss

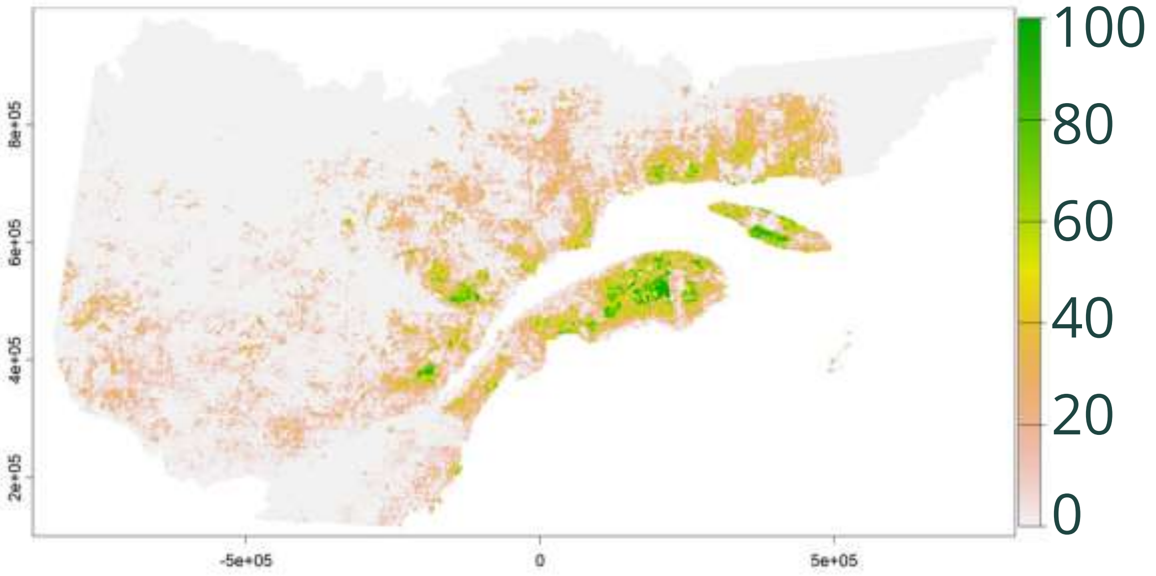
CLASS 0
No biomass loss

Permanent plots : last outbreak 1967-1991

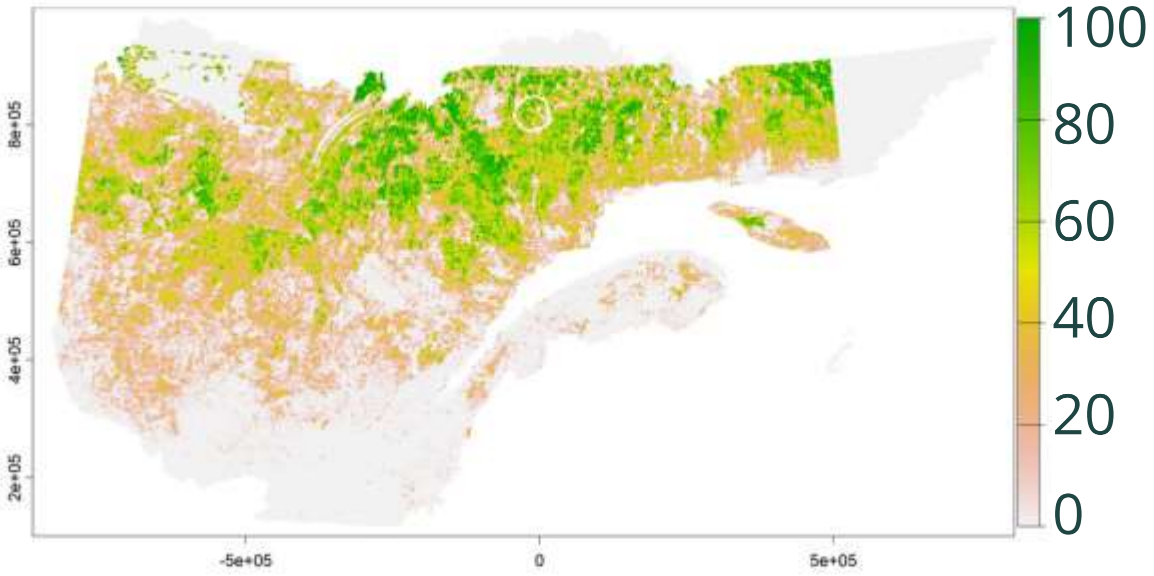


Ratio ≃1:8 → Imbalanced data

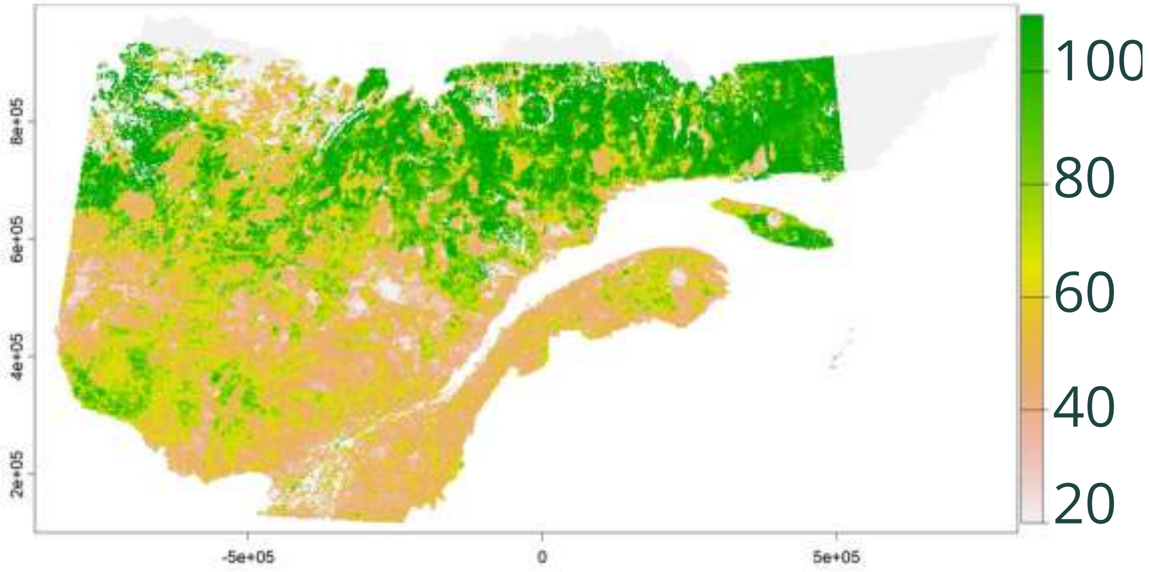
Explanatory variables : Composition variables



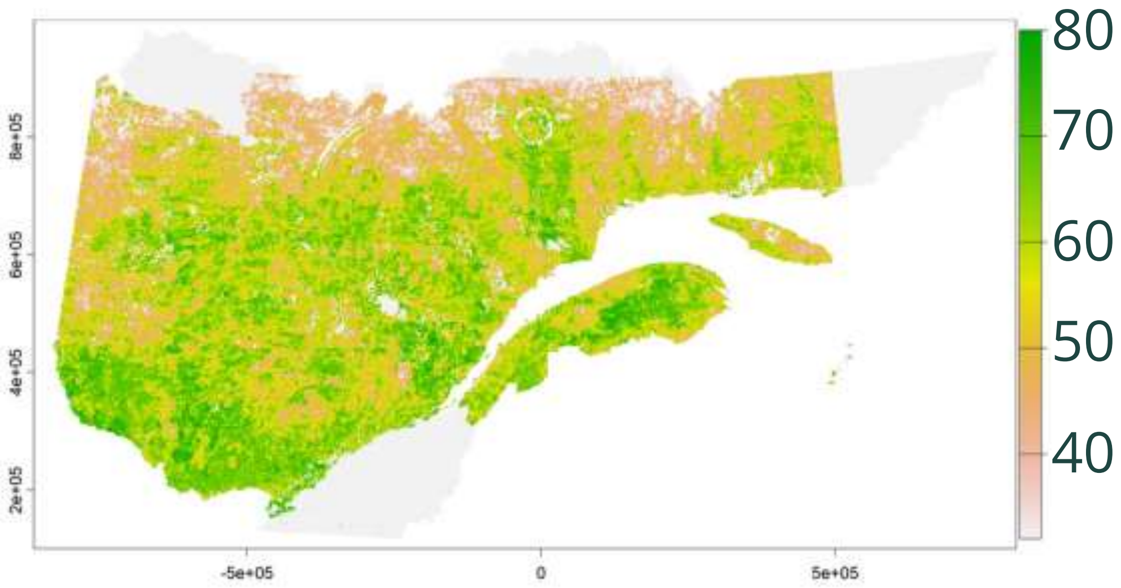
% Primary host
(balsam fir, white spruce)



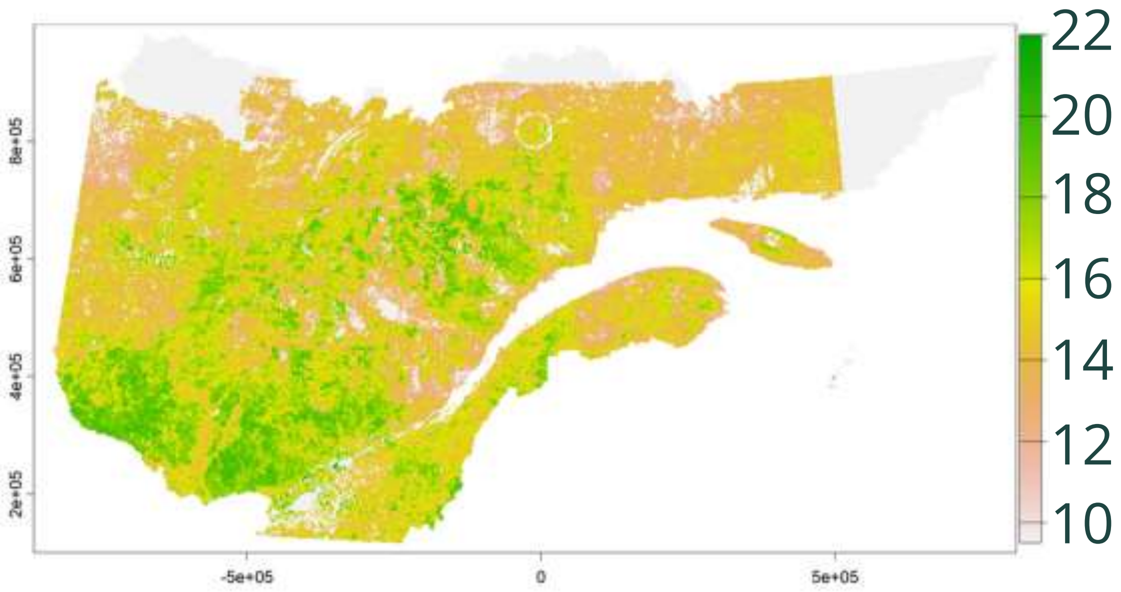
% Secondary host
(red spruce, black spruce)



Stand age



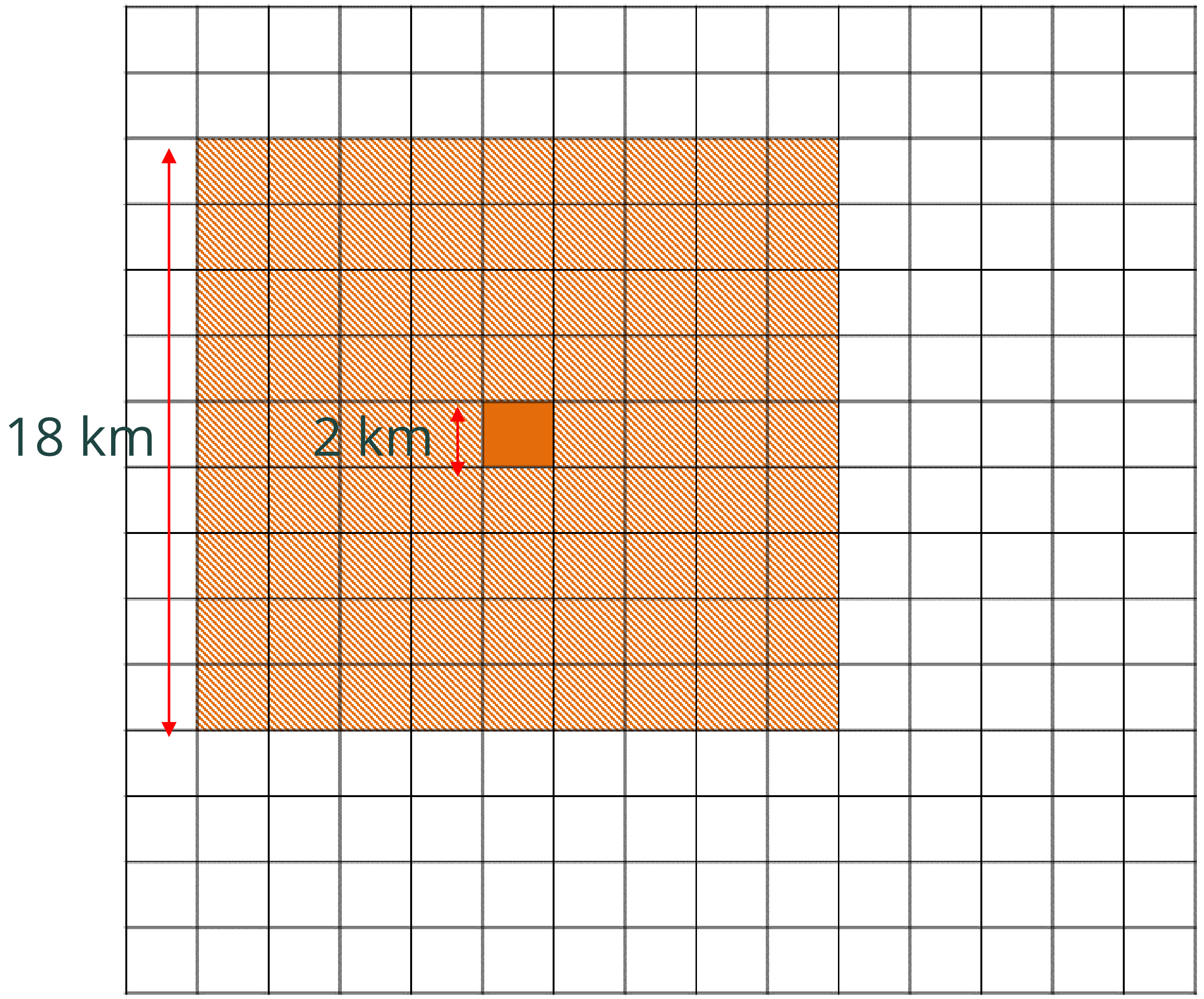
Stand density



Stand height

SIFORT
Cartes écoforestières

Explanatory variables : Landscape variables



SHDI

Shannon's diversity index

$$SHDI = -\sum_{i=1}^{m}(P_i * \ln P_i)$$

P_i : proportion of class i

Patches per class

$$NP = n_i$$

n_i : number of patches of class i

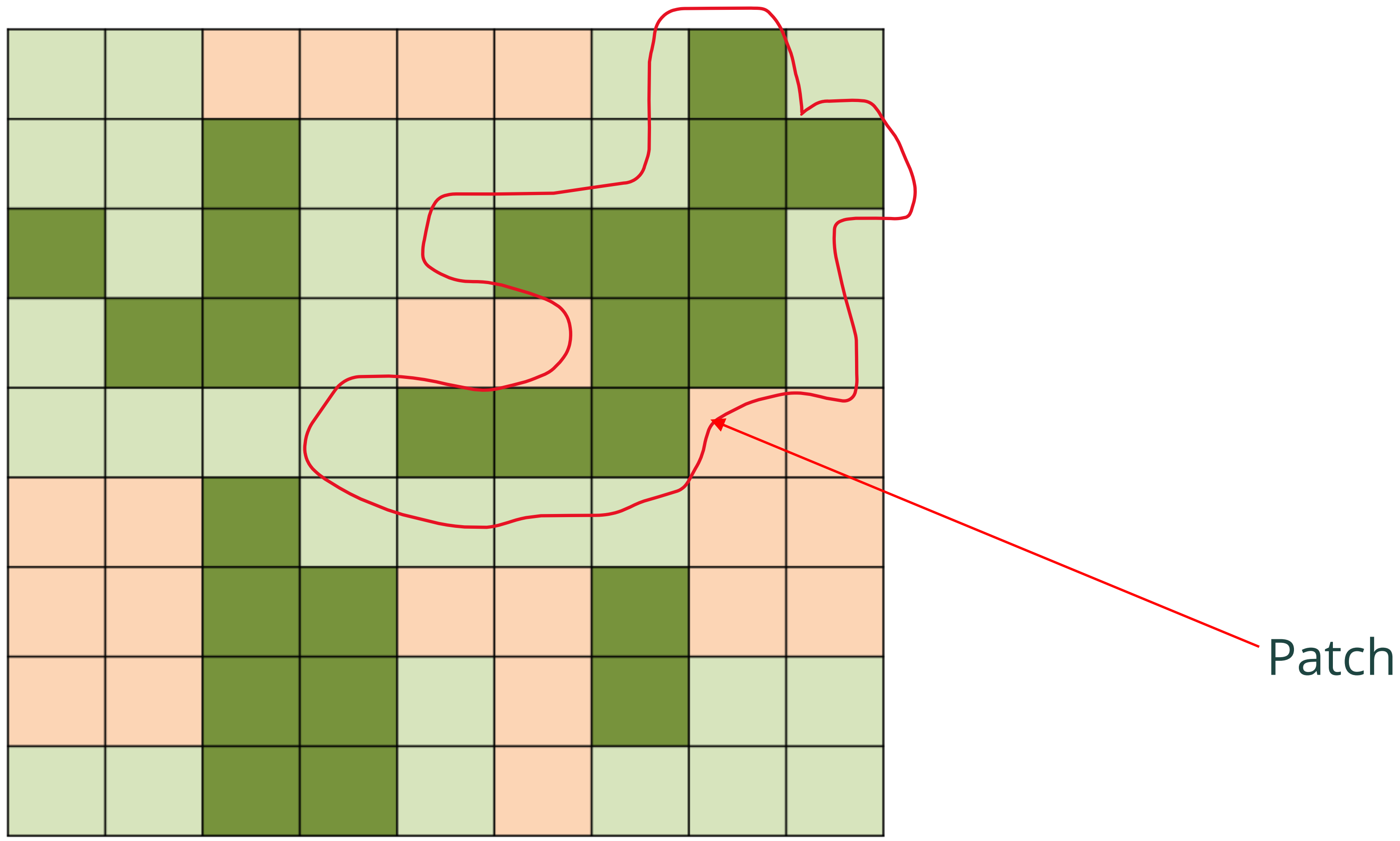
Area per class

$$CA = sum(AREA[patch_{ij}])$$

Patch_ij : area of patch j of class i

Hesselbarth et al. (2019)
Wang et al. (2014)
SIFORT / Cartes écoforestières

Explanatory variables : Landscape variables



SHDI

Shannon's diversity index

$$SHDI = -\sum_{i=1}^{m}(P_i * \ln P_i)$$

P_i : proportion of class i

Patches per class

$$NP = n_i$$

n_i : number of patches of class i

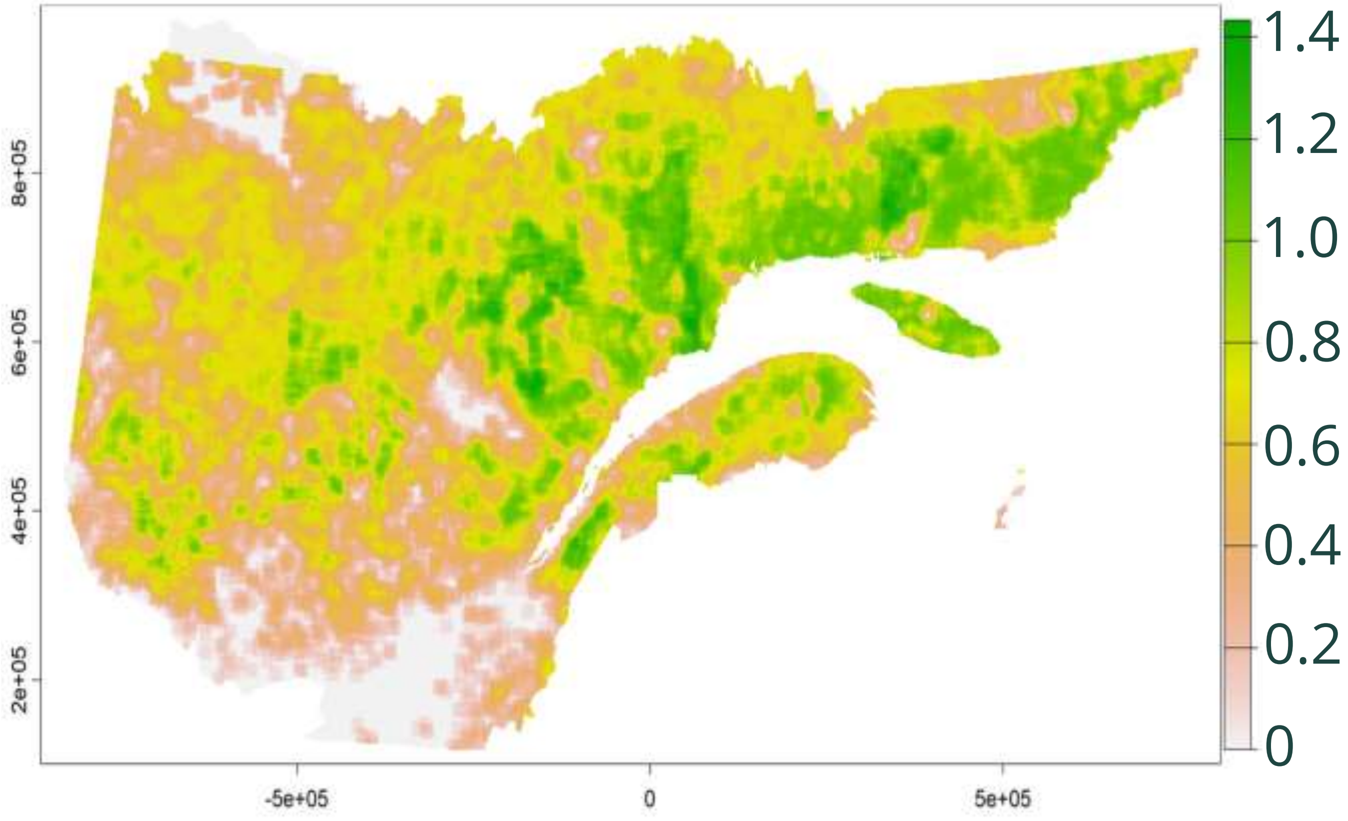
Area per class

$$CA = sum(AREA[patch_{ij}])$$

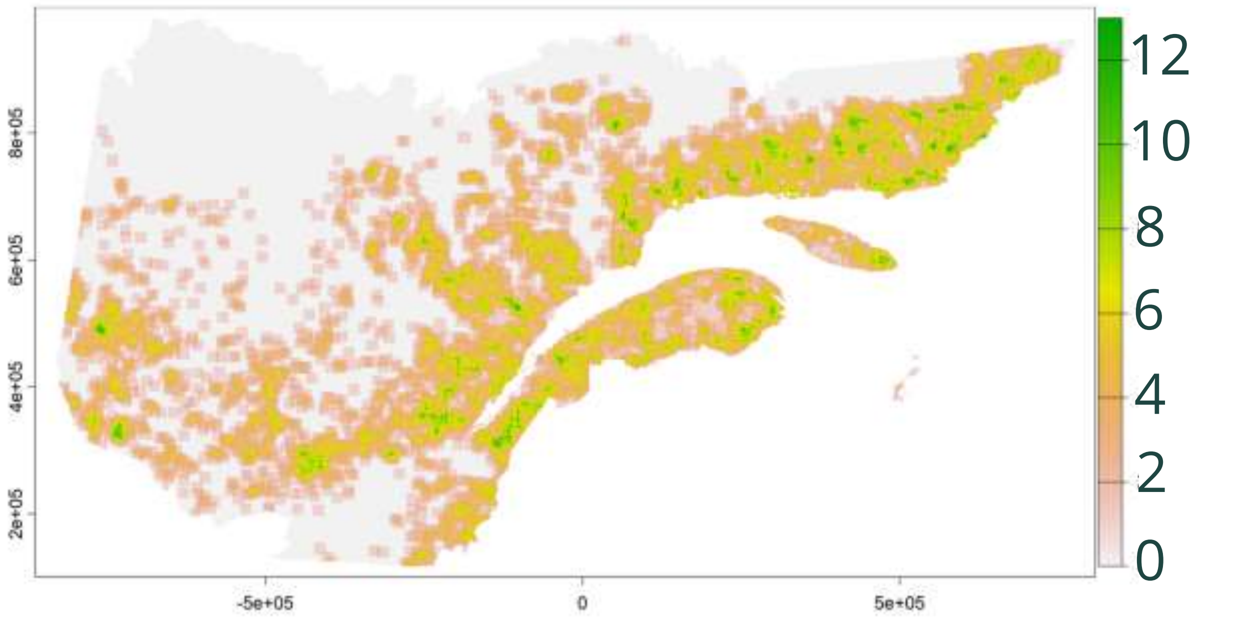
Patch_ij : area of patch j of class i

Hesselbarth et al. (2019)
Wang et al. (2014)
SIFORT / Cartes écoforestières

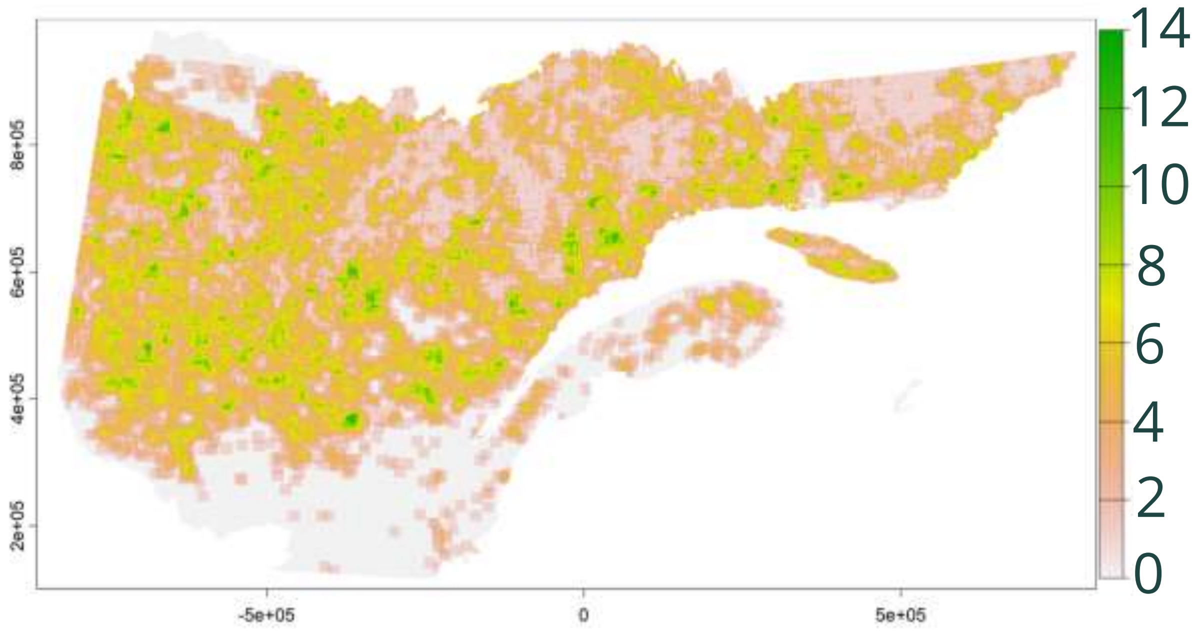
Explanatory variables : Landscape variables



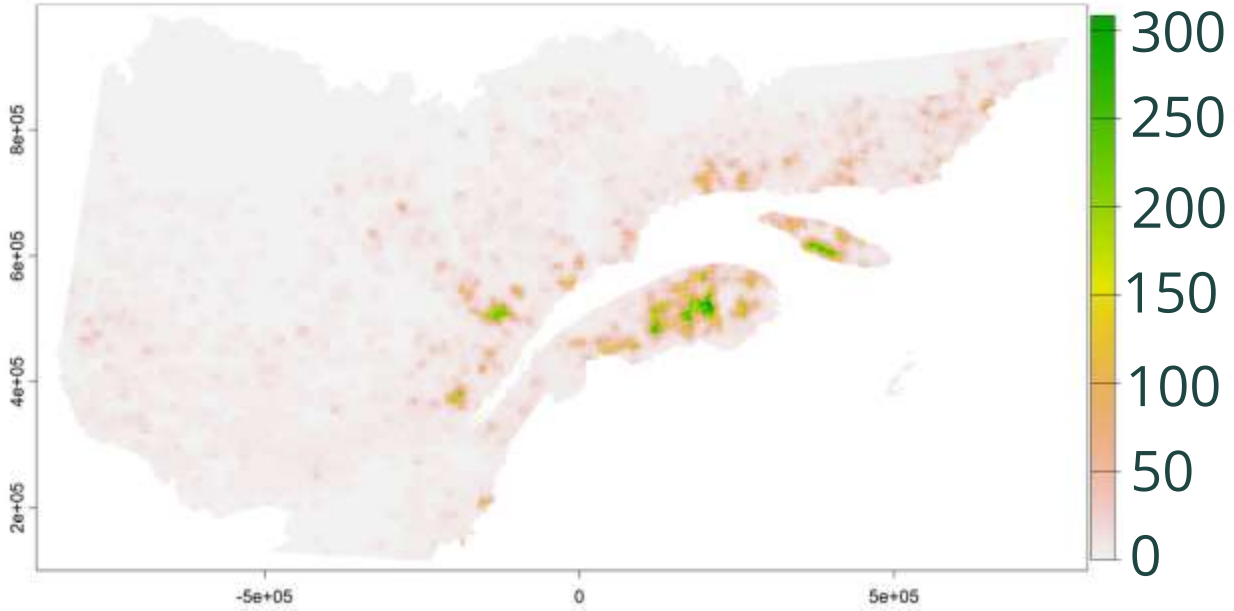
SHDI
(Shannon diversity index)



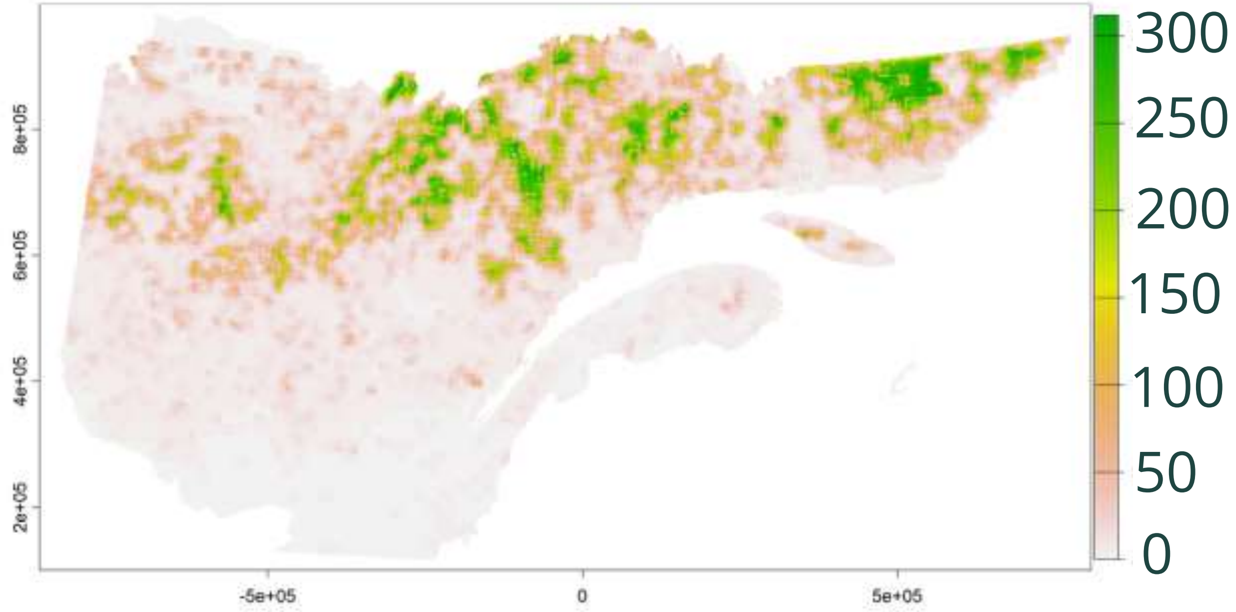
Primary Host N patches



Secondary Host N patches

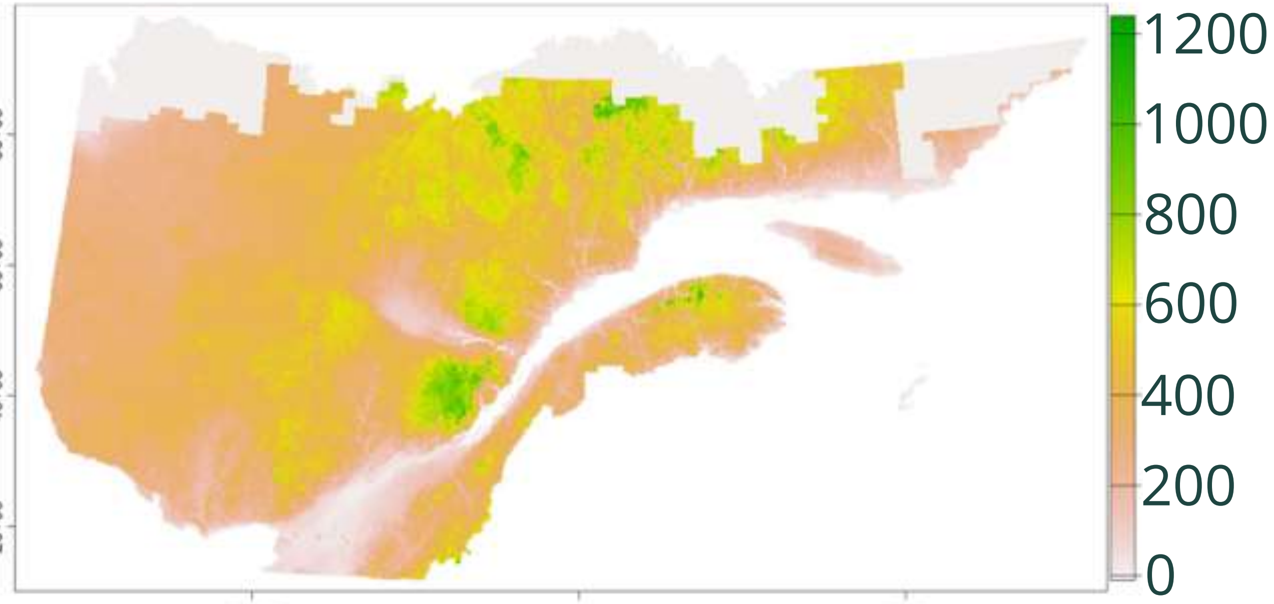


Primary Host Area

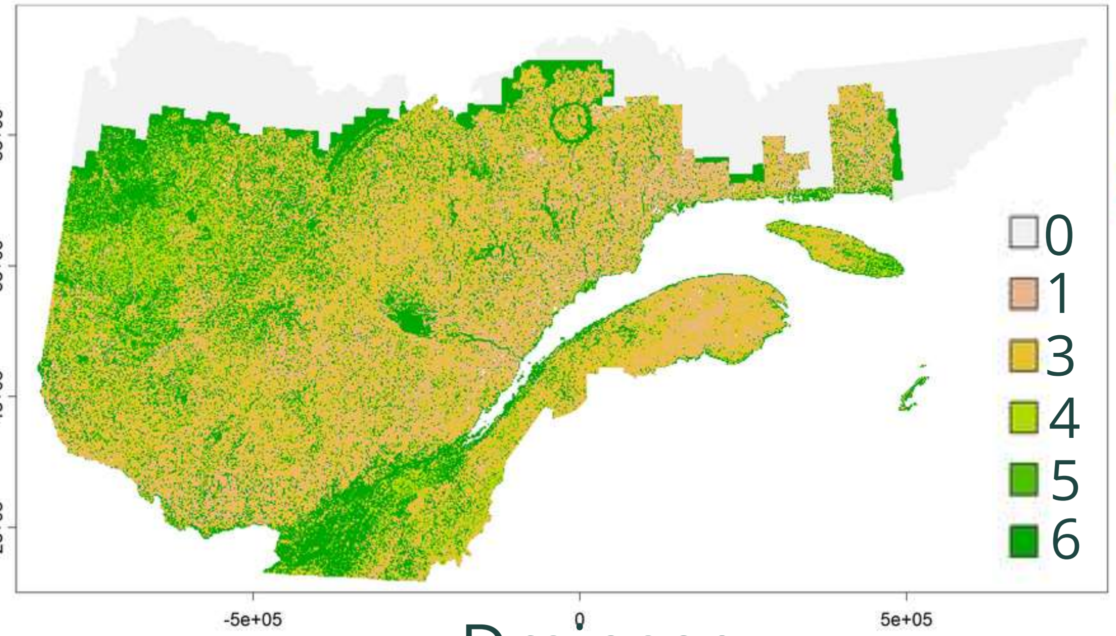


Secondary Host Area

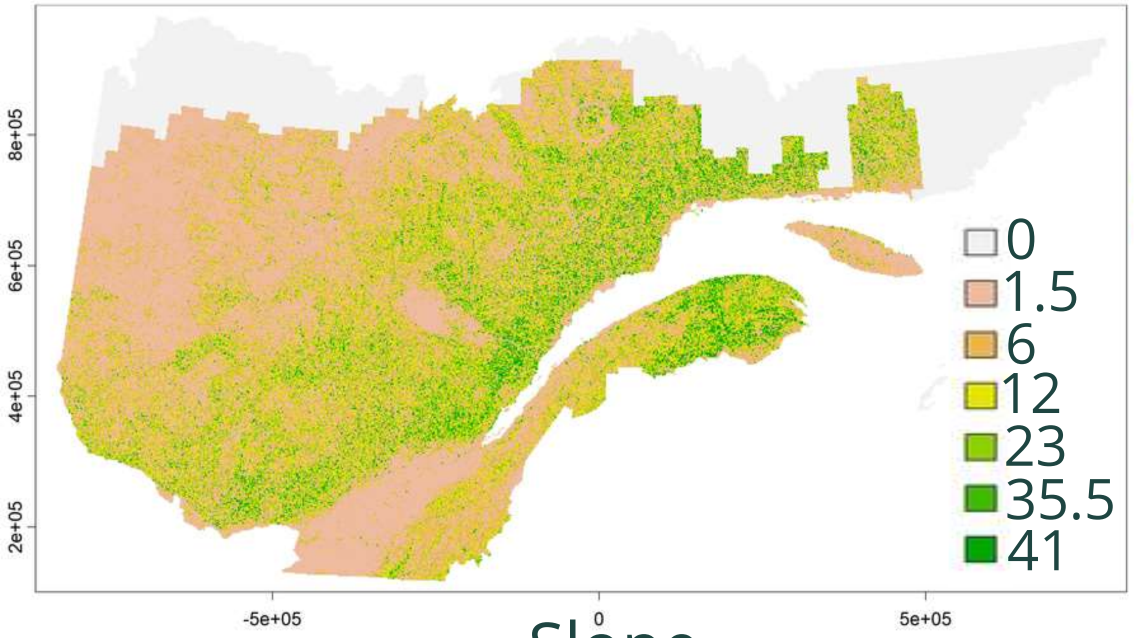
Explanatory variables : Abiotic variables



Elevation

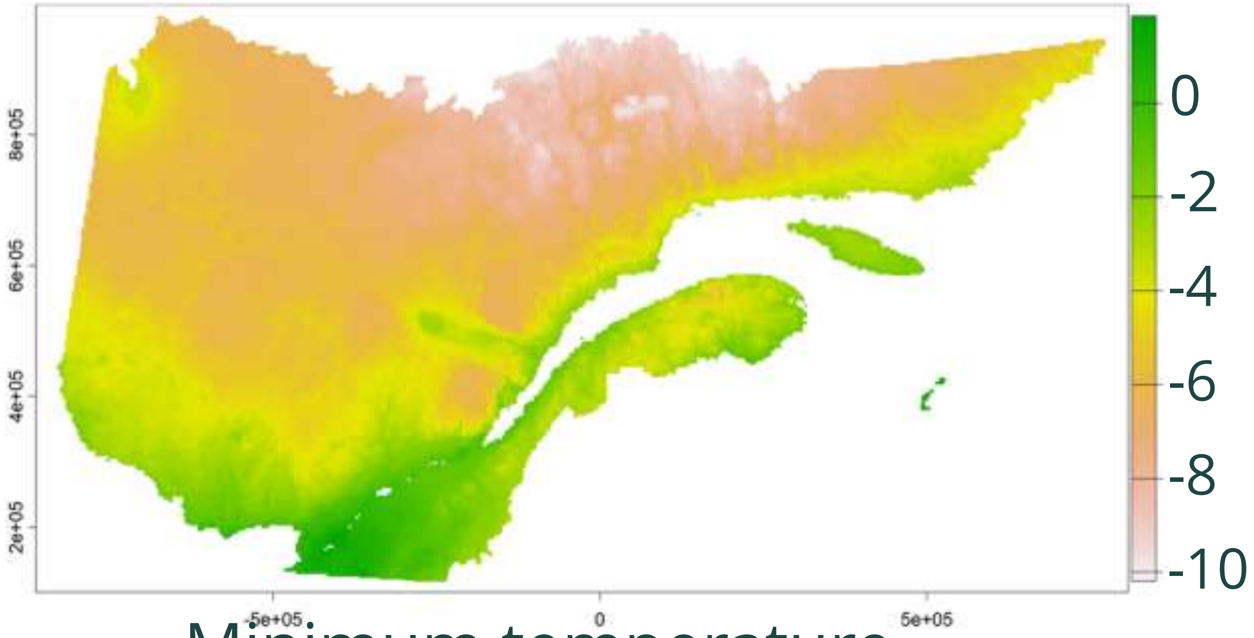


Drainage



Slope

Precipitations

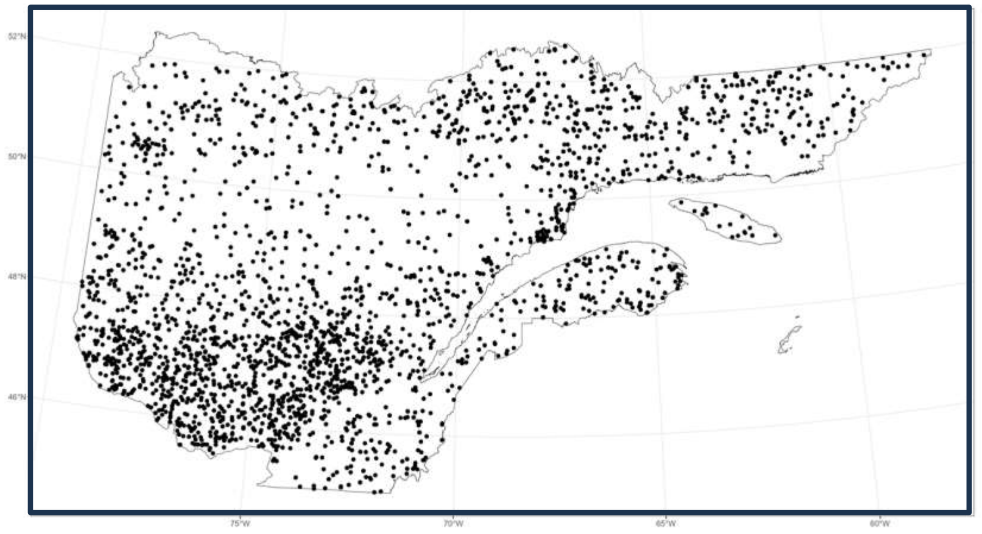


Minimum temperature

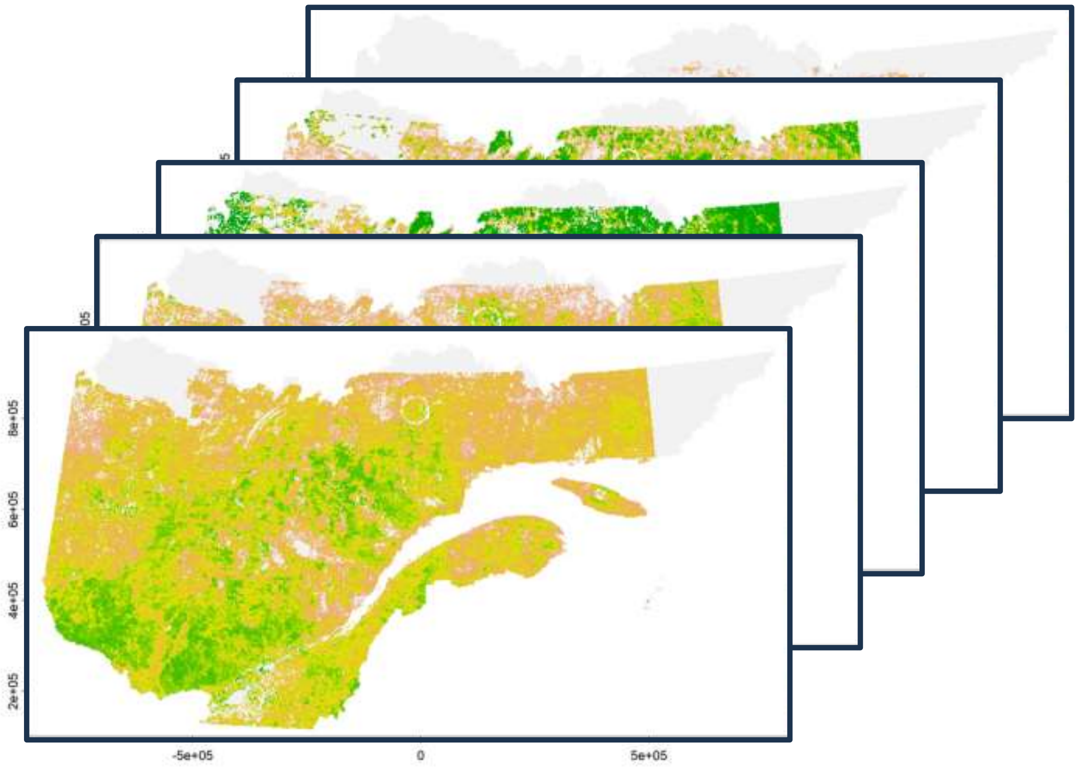
WorldClim
Cartes écoforestières
DEM

Modelling process and algorithms

Forest vulnerability



Explanatory variables



MODEL CALIBRATION AND EVALUATION
1967-1991

NAIVE BAYES
MLP
K-NEAREST
NEIGHBOR
RANDOM FOREST
SVM
XGBOOST

AUC PR CURVE
KAPPA

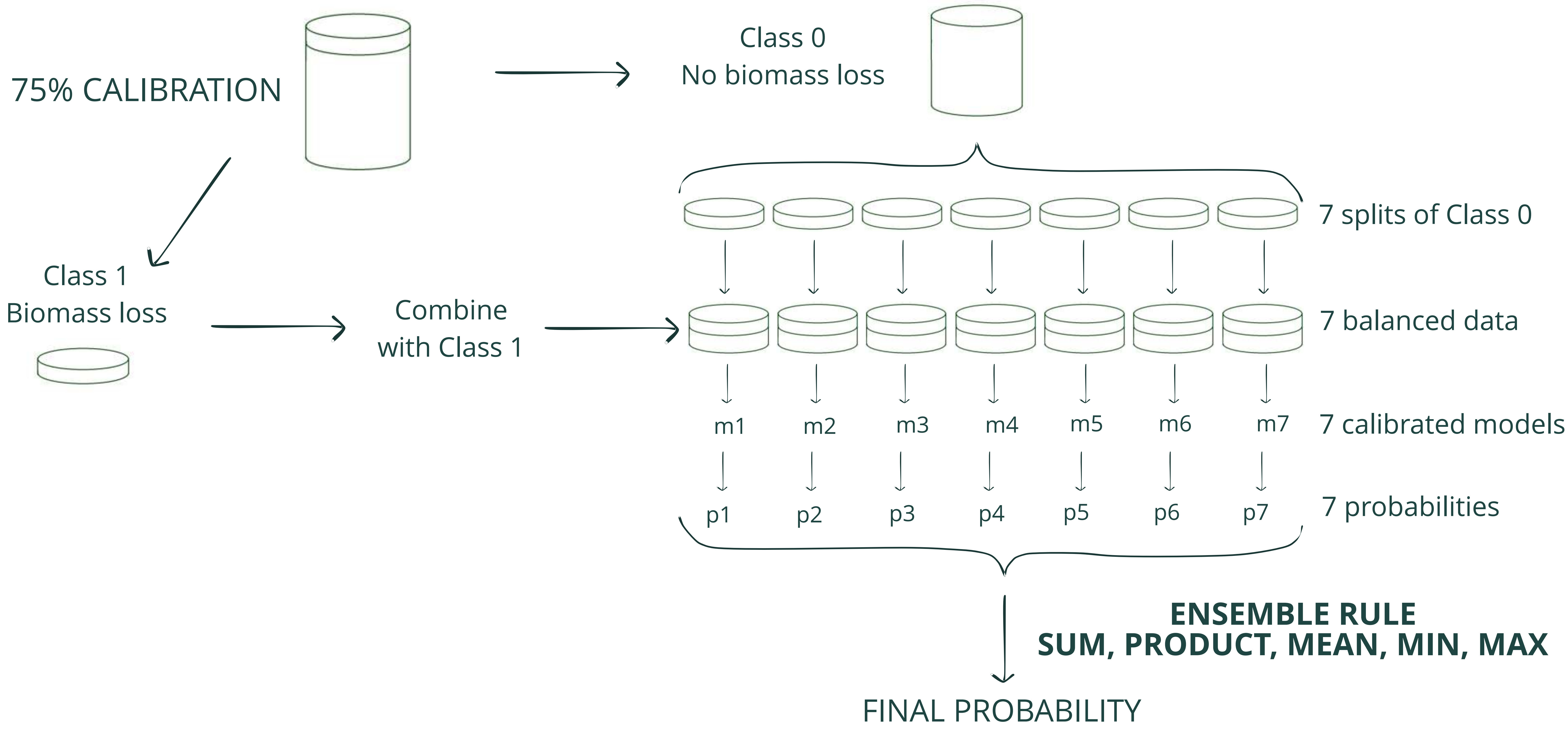
PREDICTION
2006-2022

PROBABILITY OF BIOMASS LOSS
VARIABLE CONTRIBUTION

Data partitioning

Sun et al. (2015)
Tahir et al. (2012)
Kittler et al. (1998)

Data balancing method



Sun et al. (2015)
Tahir et al. (2012)
Kittler et al. (1998)

Splitting balancing method

Undersampling : loss of data
Oversampling : generated data

Model performance metrics

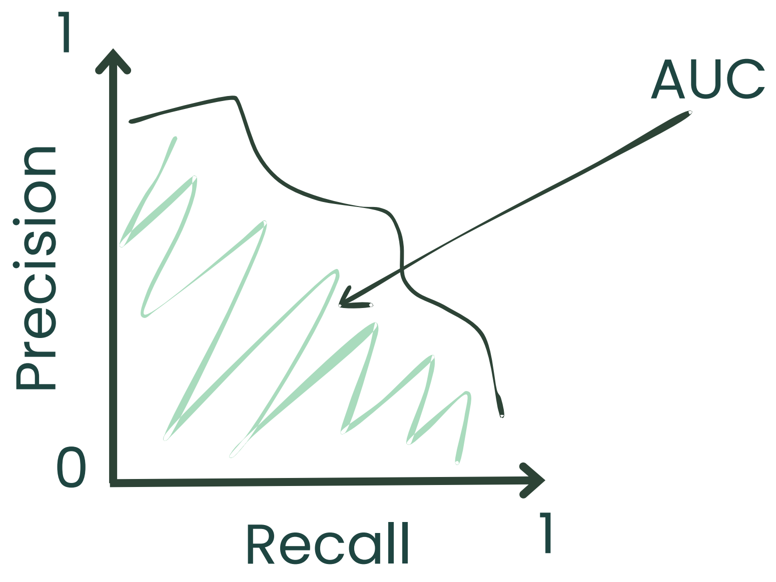
Kappa

Agreement between predicted and observed data

< 0	Poor
0.00 - 0.20	Slight
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Substantial
0.81 - 1.00	Almost perfect

AUC (Precision-Recall Curve)

Proportion of overall correctly classified Class1

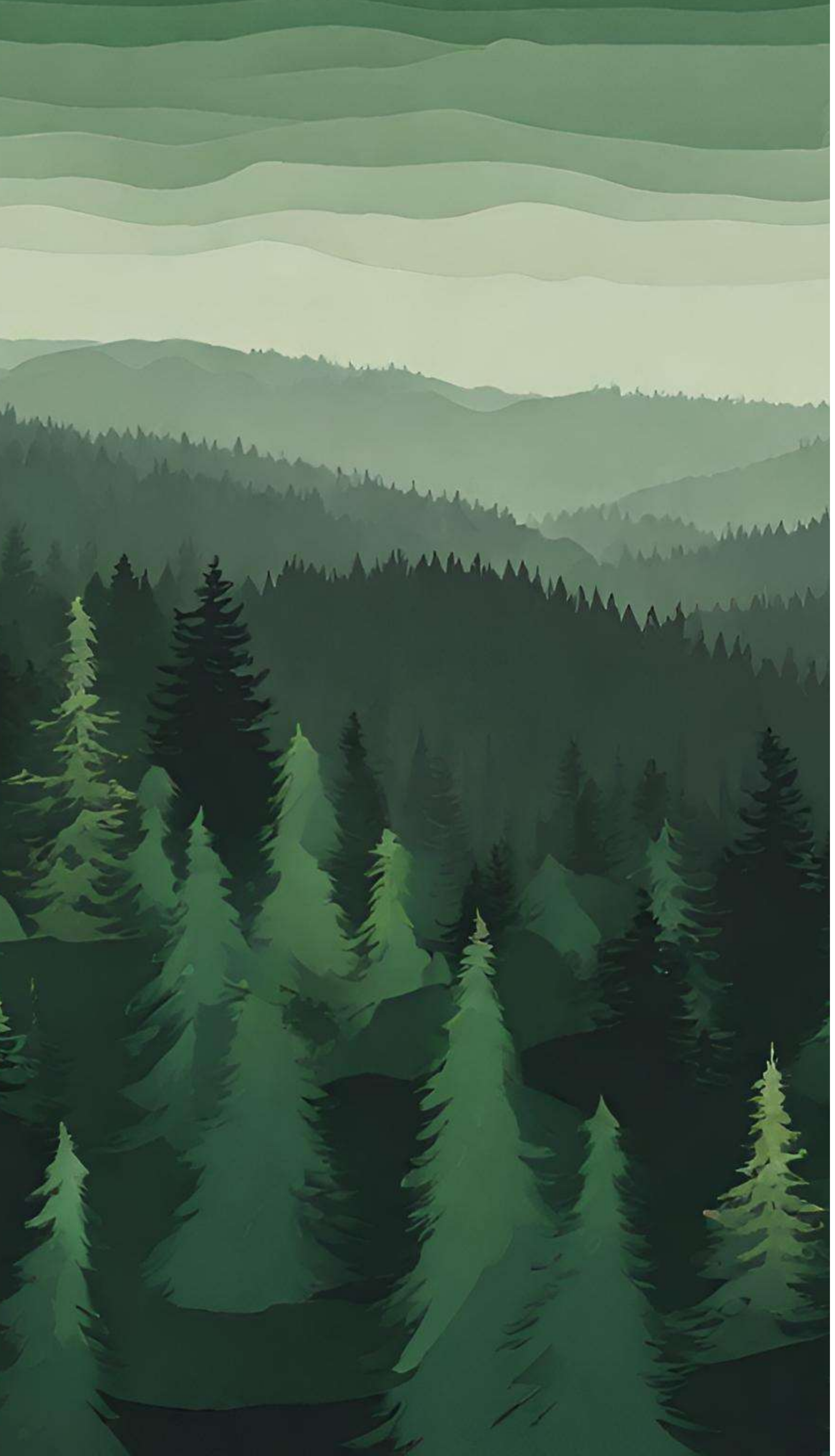


Proportion of correctly classified Class1 compared to all observed Class1

- Not a threshold dependant metric
- Model overall performance

→ **Metrics usually used for imbalanced data**

Landis et Koch (1977)



RESULTS

Model performance

Performance per algorithm

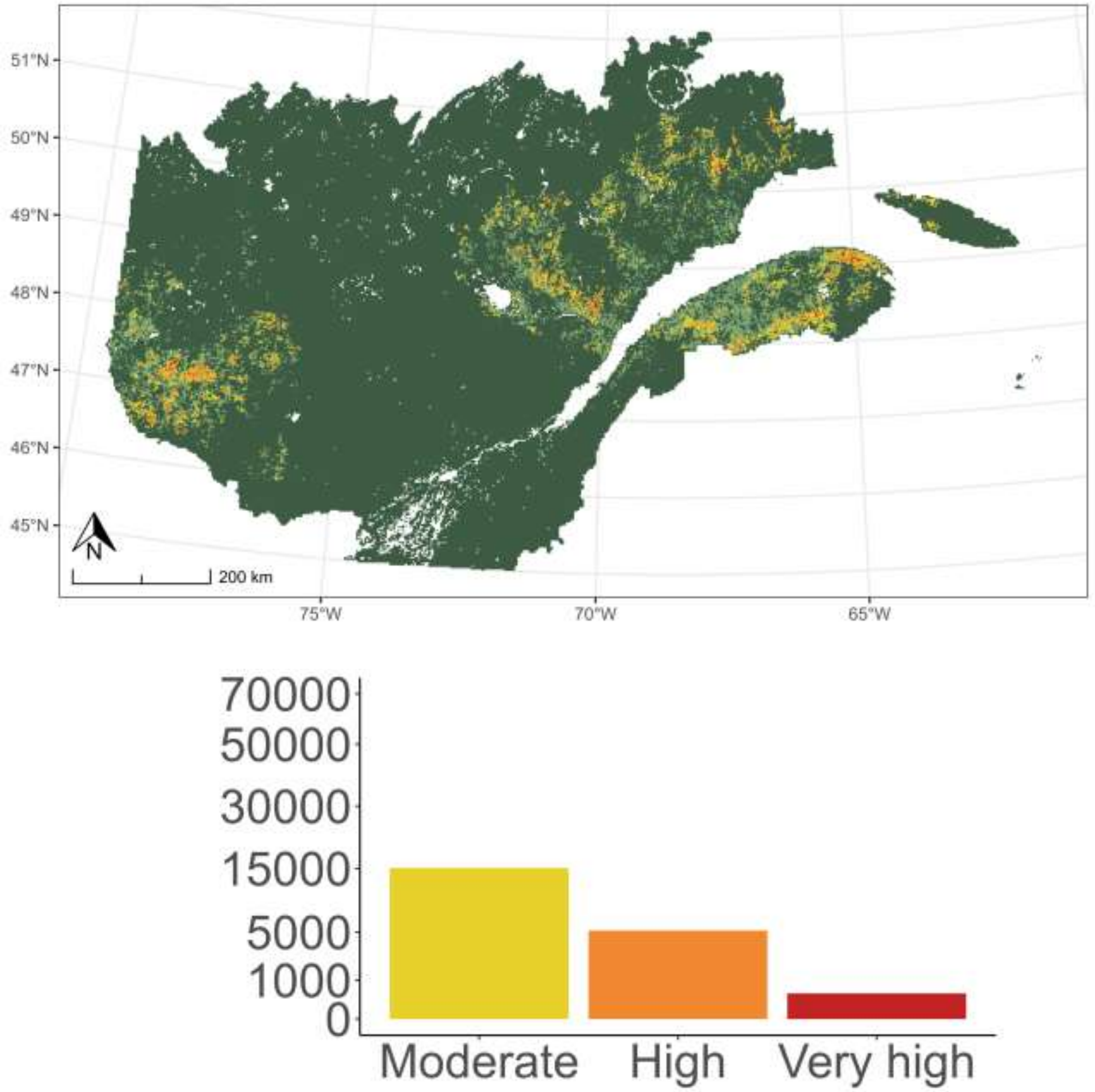
MODEL	RULE	KAPPA	AUC
MLP	Minimum	0.504	0.541
RF	Minimum	0.495	0.576
XGB	Minimum	0.486	0.511
SVM	Minimum	0.456	0.594
MLP	Mean	0.454	0.547
SVM	Mean	0.379	0.564
RF	Mean	0.345	0.571
XGB	Mean	0.341	0.526
GLM	Minimum	0.310	0.385
KNN	Minimum	0.310	0.432

Performance per ensemble rule

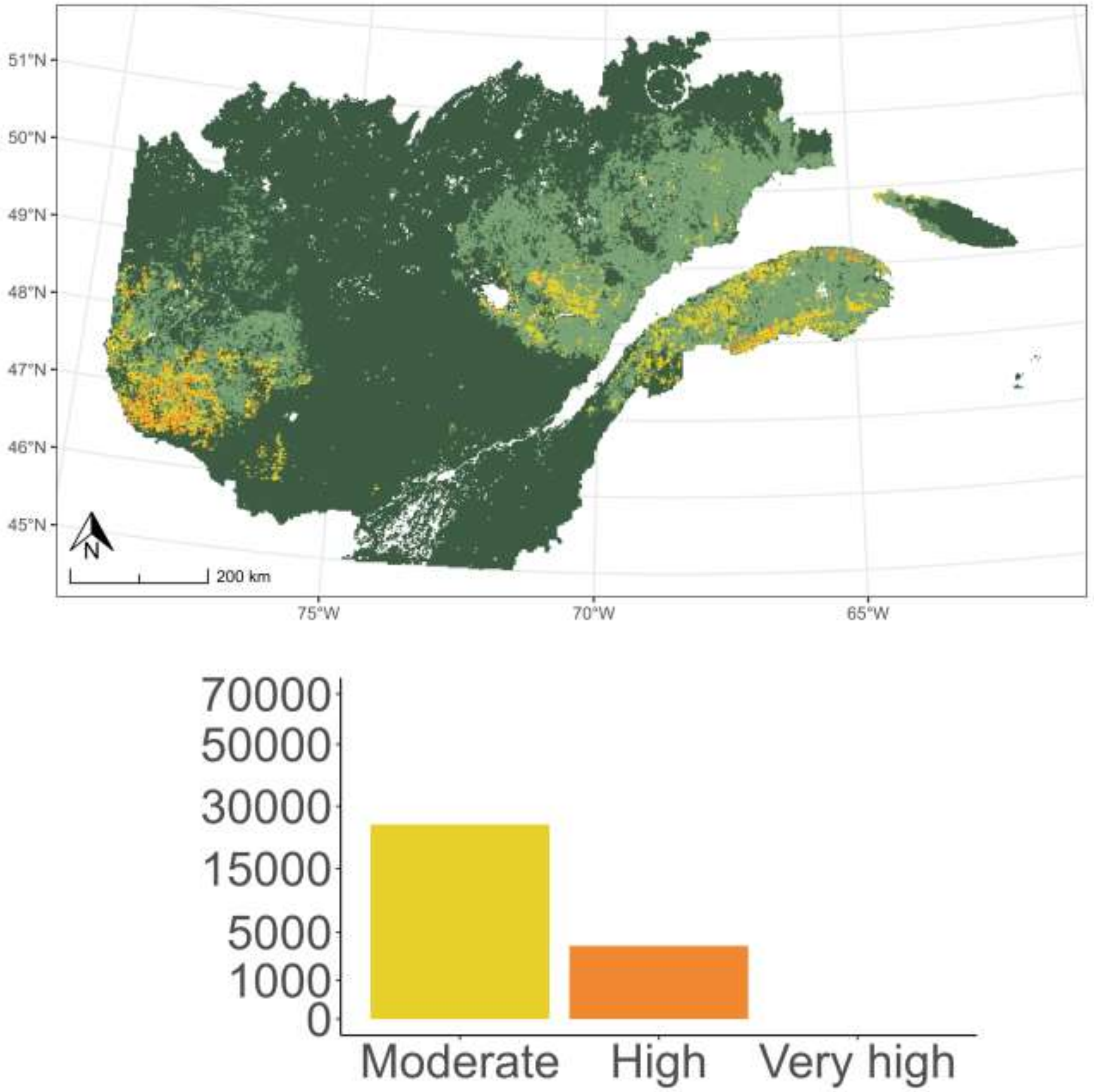
RULE	MEAN KAPPA	MEAN AUC
Minimum	0.385 ±0.126	0.491 ± 0.09
Mean	0.293 ± 0.096	0.485 ± 0.089
Product	0.241 ± 0.149	0.485 ± 0.089
Maximum	0.209 ± 0.056	0.412 ± 0.1
Sum	0.097 ± 0.058	0.485 ± 0.089

Biomass loss probability 2006-2022 : Minimum ensemble rule

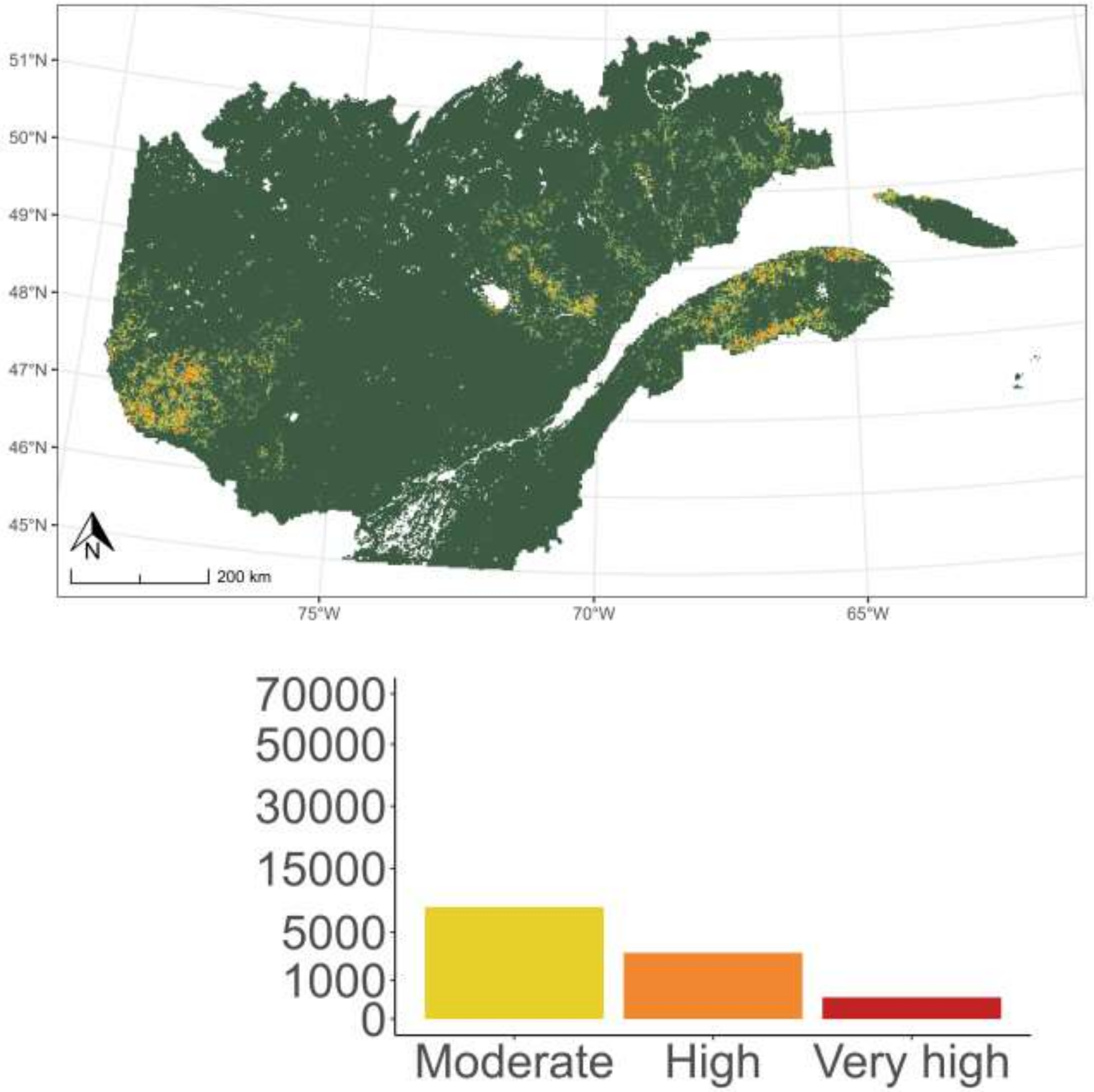
MLP



RF



XGBoost



Biomass loss probability

[0-0.2] Very low

[0.2-0.4[Low

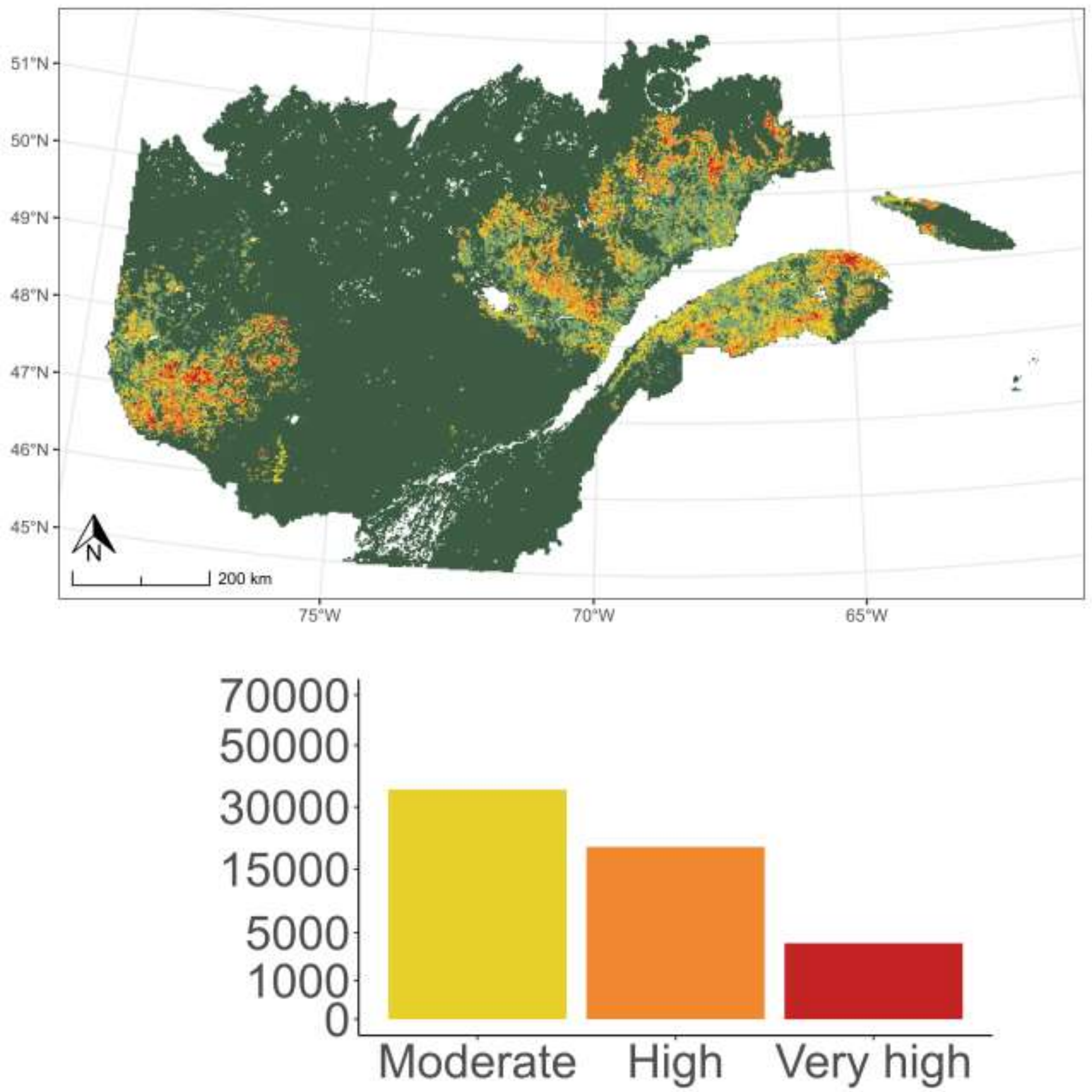
[0.4-0.6[Moderate

[0.6-0.8[High

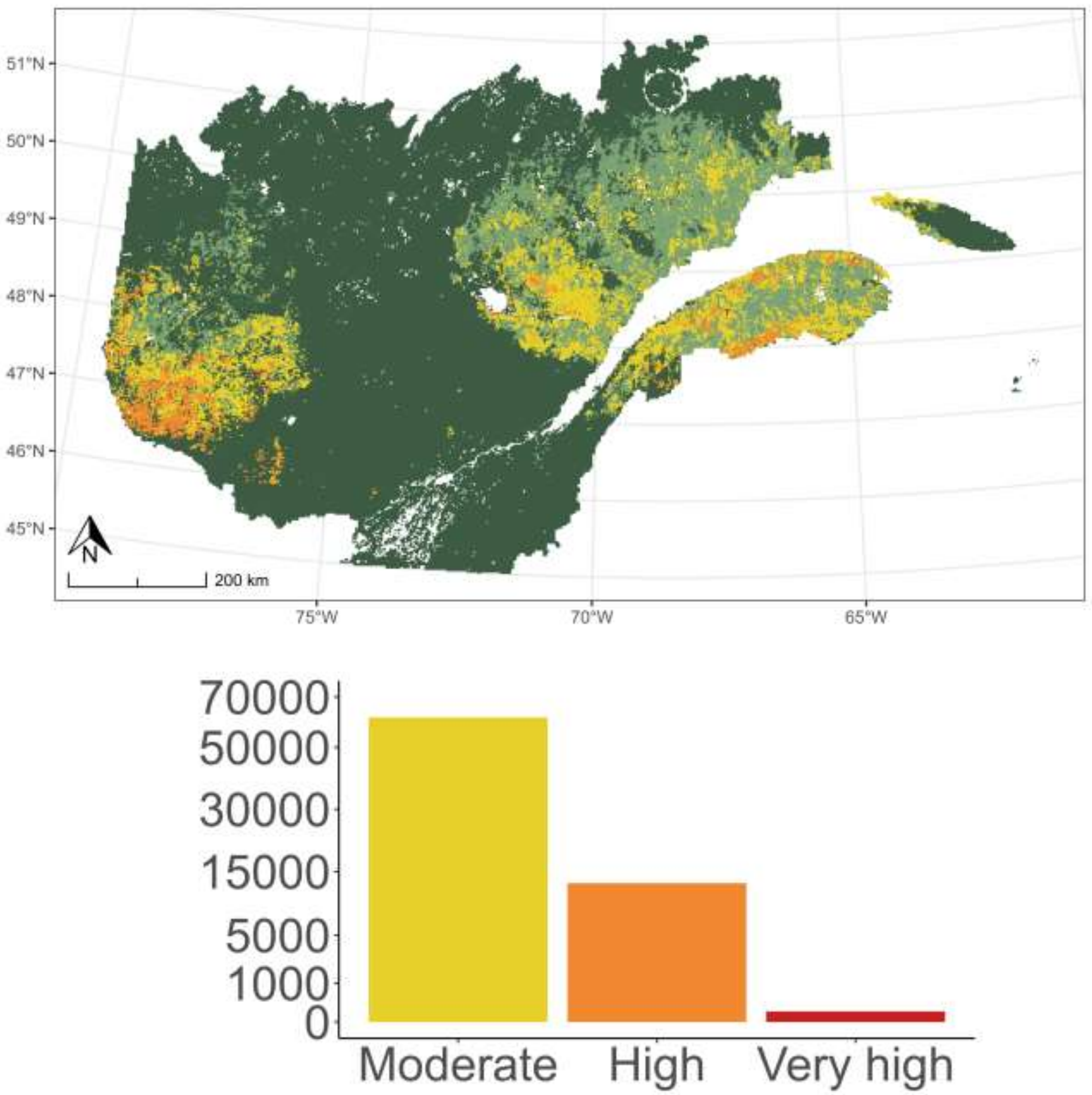
[0.8-1] Very high

Biomass loss probability 2006-2022 : Mean ensemble rule

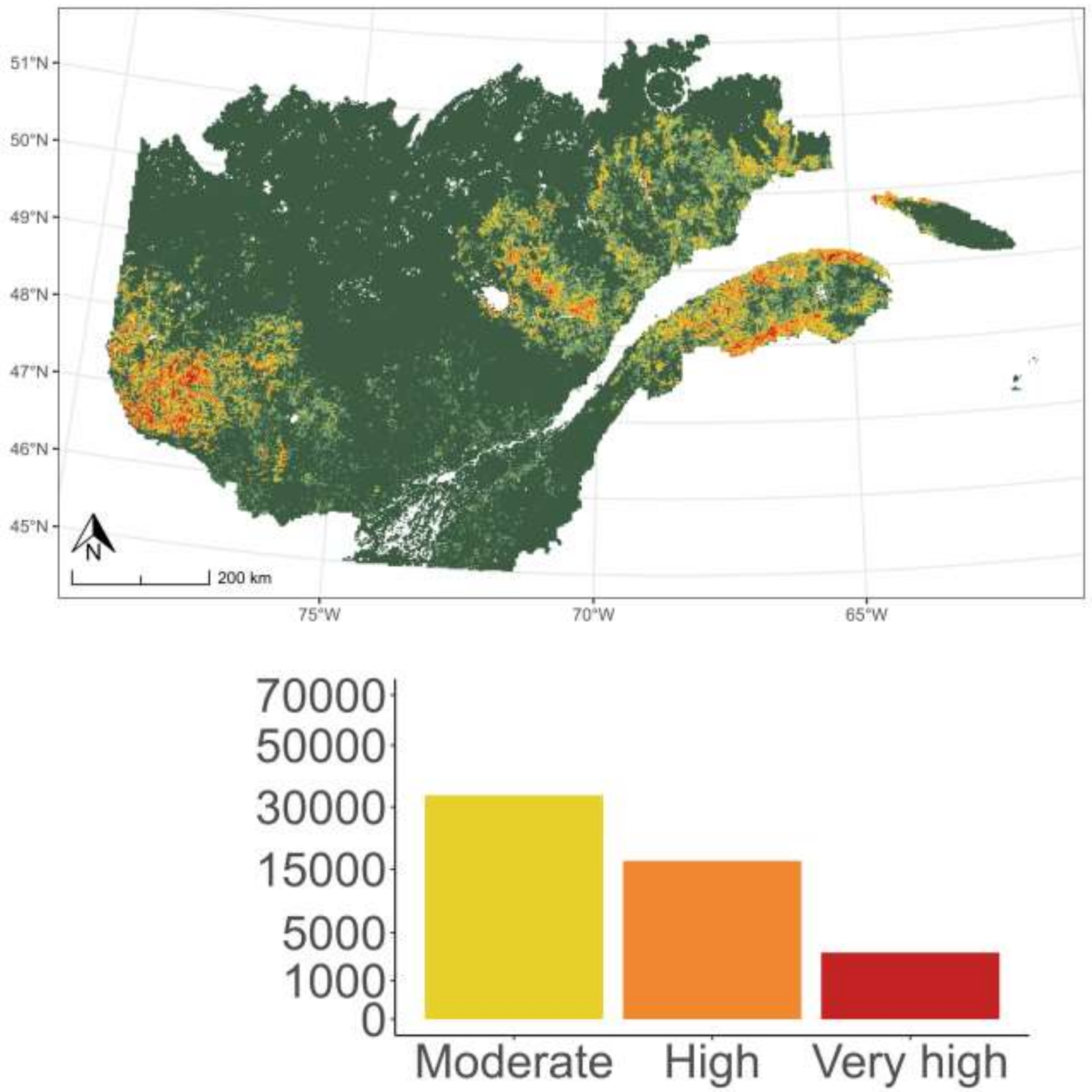
MLP



RF



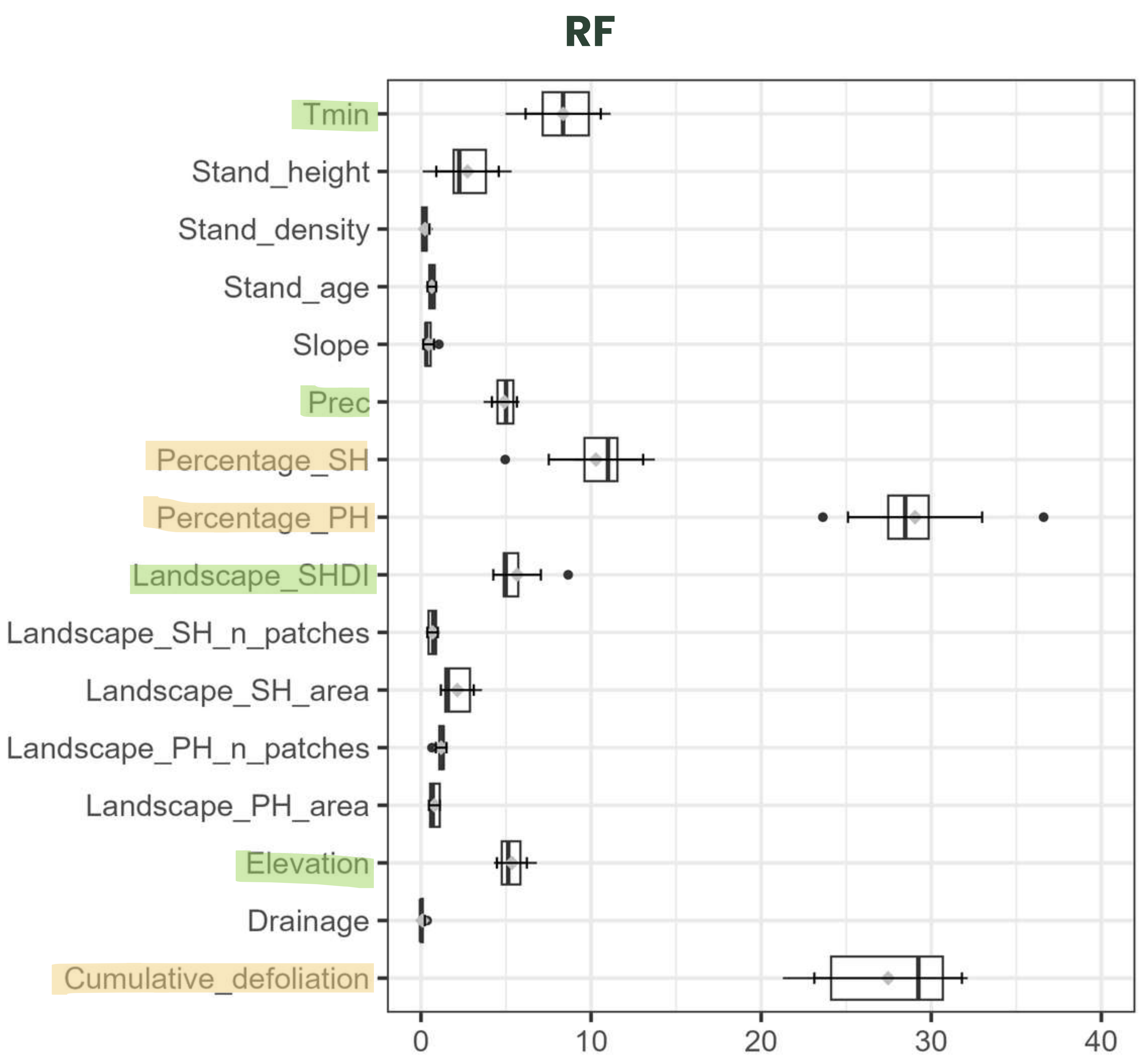
XGBoost

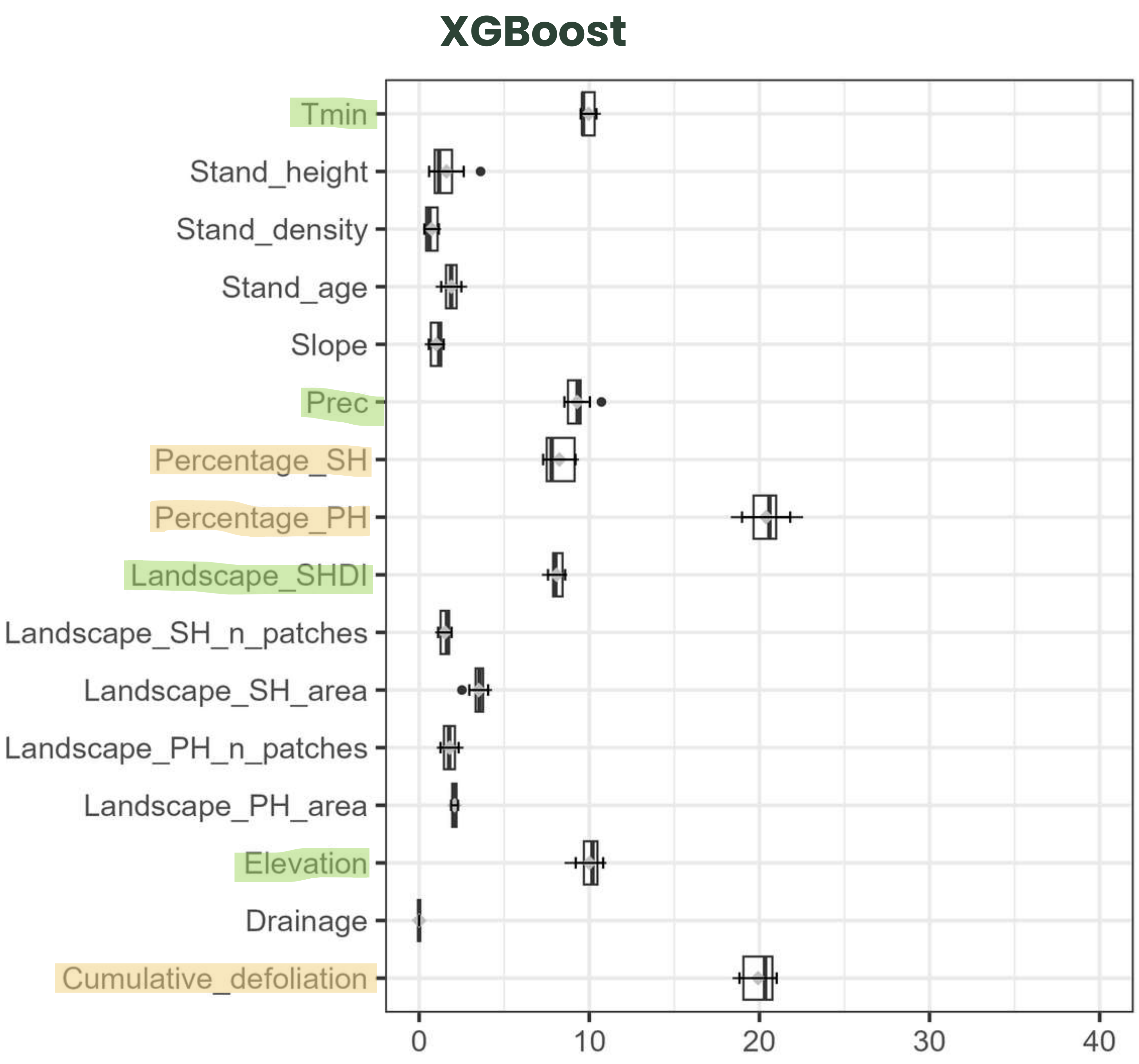


Biomass loss probability

- [0-0.2] Very low
- [0.2-0.4[Low
- [0.4-0.6[Moderate
- [0.6-0.8[High
- [0.8-1] Very high

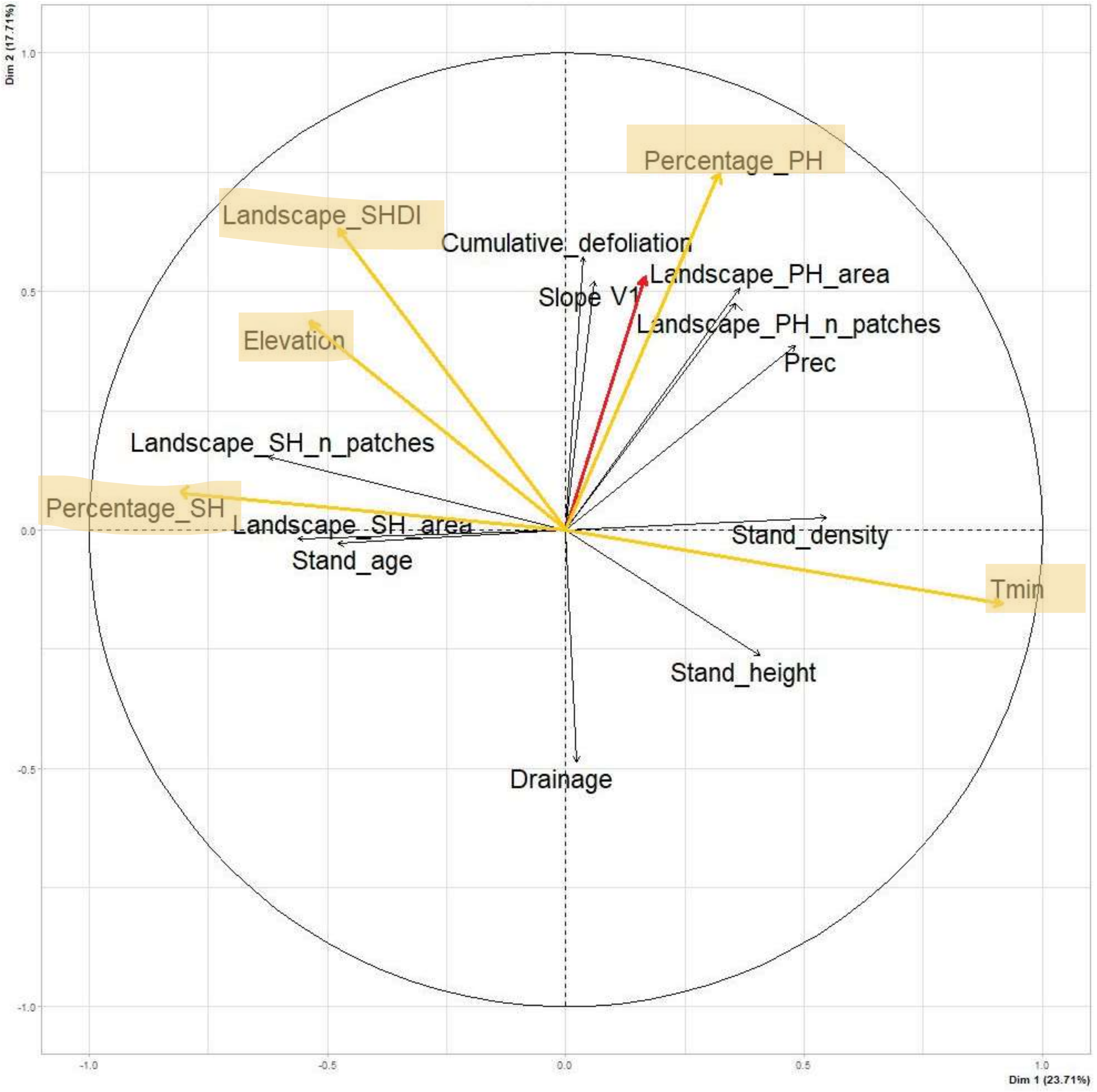
Variable contribution

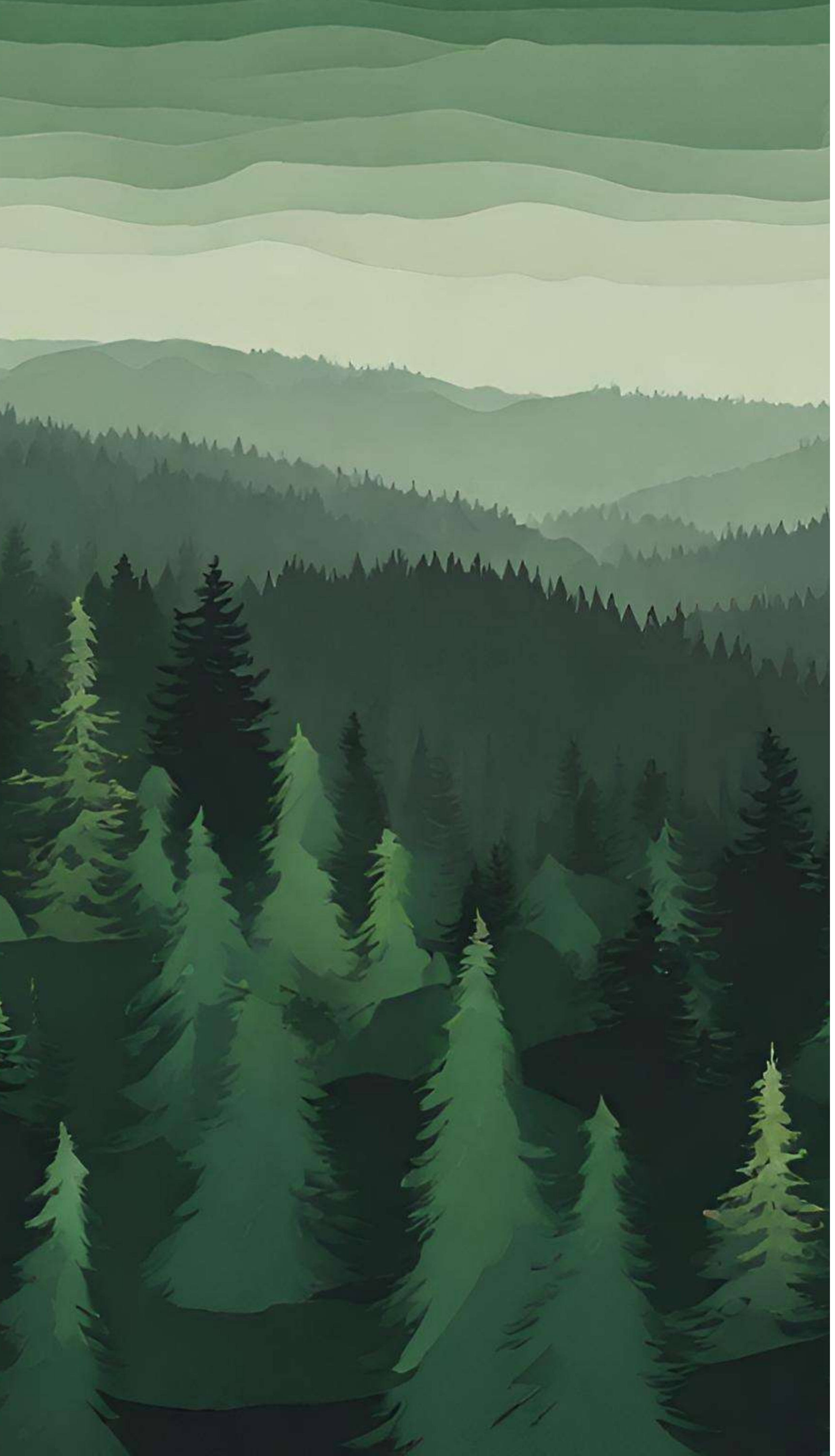




PCA on 2006–2022 data

MLP



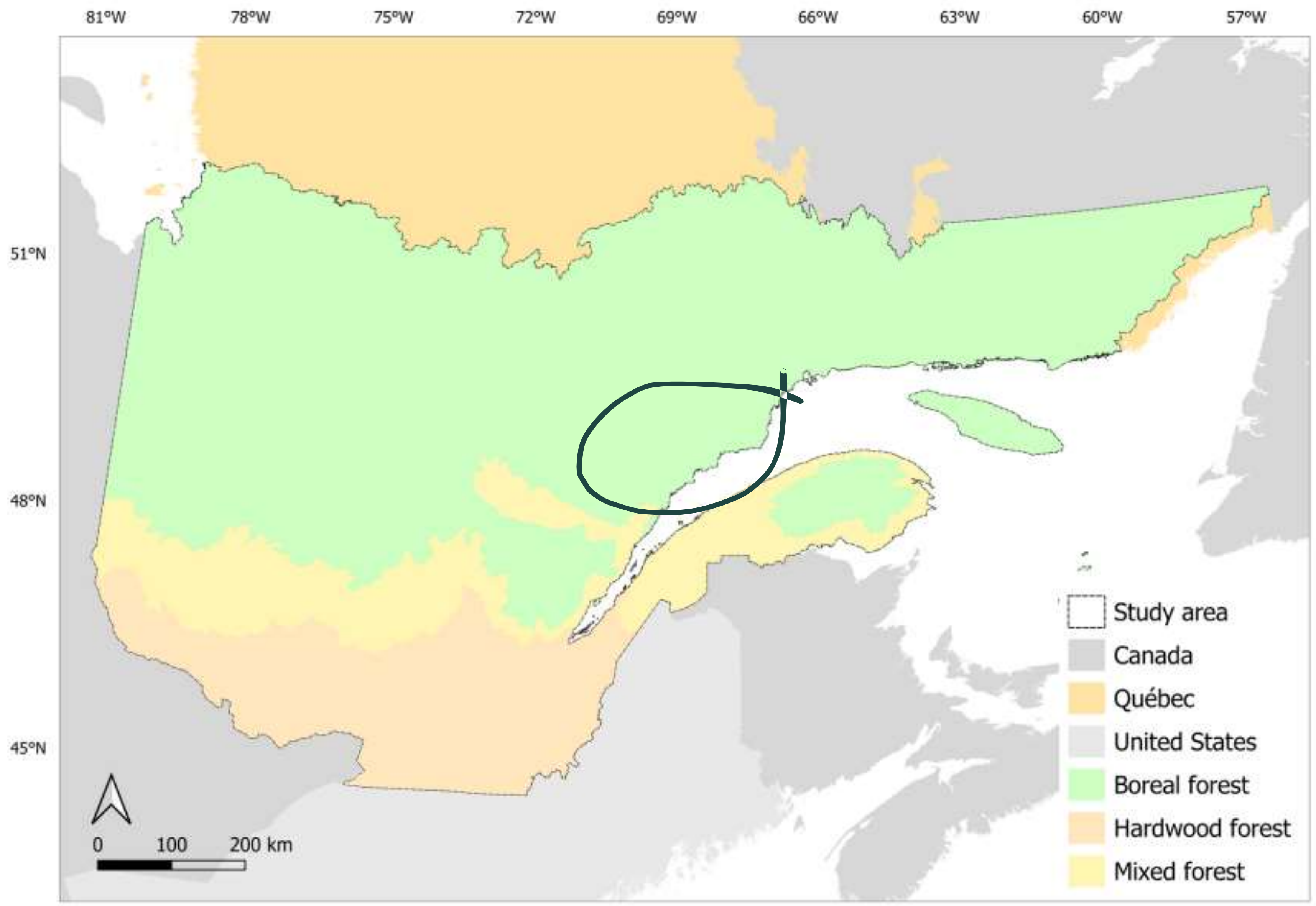


CONCLUSION

Research perspectives

- Re-evaluate the model in regions where the 2006 outbreak is done : Côte-Nord

⟹ Compare to real biomass loss



Conclusion

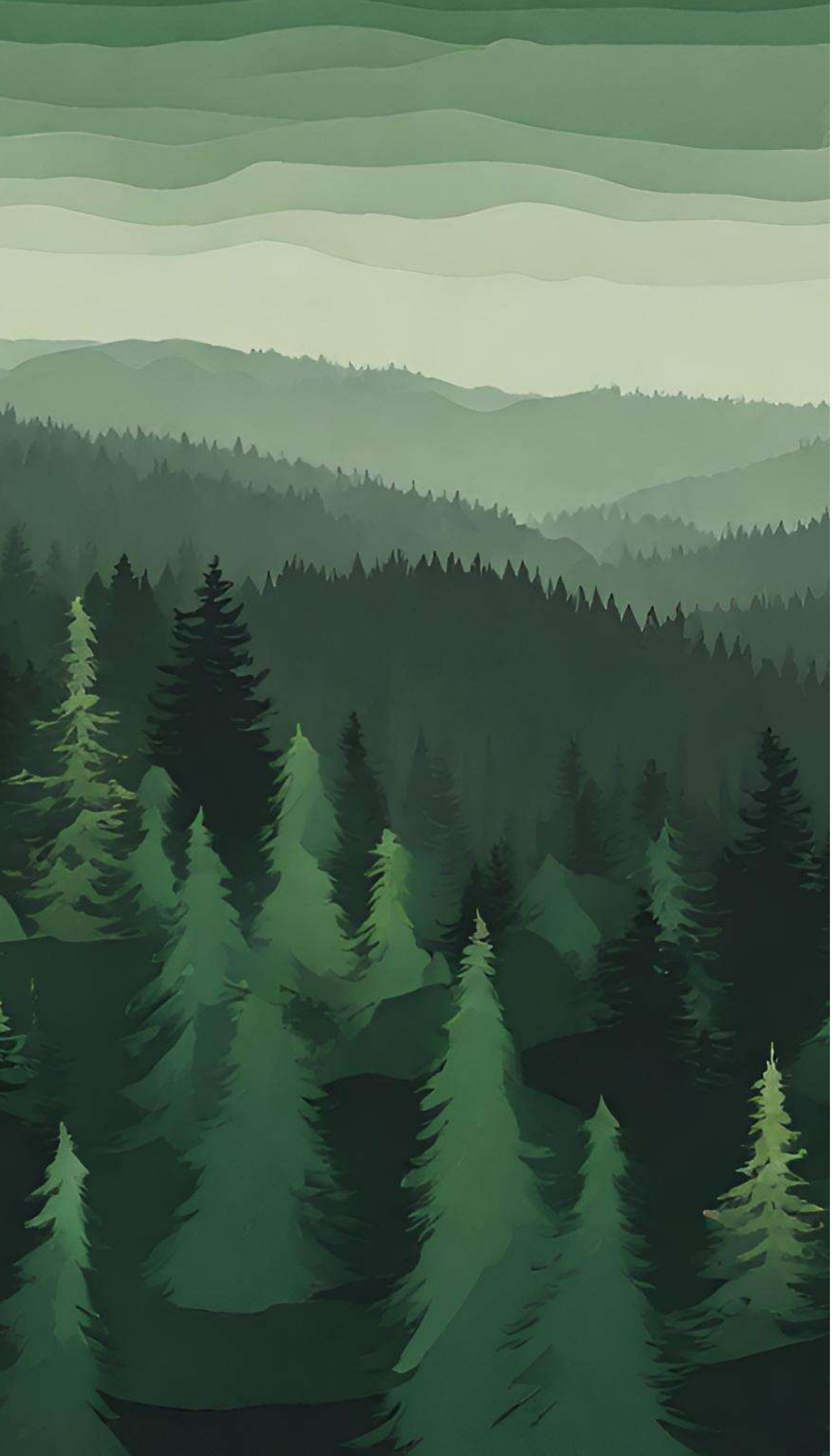
1. **ML algorithms : satisfactory prediction of biomass loss probability at the landscape scale**
   1. MLP
   2. XGBoost
   3. Random Forest
2. **Low interpretability : MLP algorithm**
3. **Regeneration/Plantation : not considered in biomass loss evaluation from permanent plots**

Conclusion

1. **Variables contribution :**
    1. Cumulative defoliation, Proportion of primary hosts
    2. Climate variables (Minimum temperature, Precipitations), Elevation, Proportion of secondary hosts, Landscape diversity
2. **Vulnerable areas :**
    1. Very high probability of biomass loss : 249 km2 $_{\pm 224}$ → 2282 km2 $_{\pm 2003}$
    2. High probability of biomass loss : 3501 km2 $_{\pm 1475}$ → 15633 km2 $_{\pm 4807}$

References

- Blais, J. R. (1983). Trends in the frequency, extent, and severity of spruce budworm outbreaks in eastern Canada. *Canadian Journal of Forest Research*, *13*(4), 539–547. https://doi.org/10.1139/x83-079
- Hesselbarth, M.H.K., Sciaini, M., With, K.A., Wiegand, K., Nowosad, J. 2019. *landscapemetrics*: an open-source R tool to calculate landscape metrics. Ecography, 42: 1648-1657 (ver. 0)
- Kittler, J., Hatef, M., Duin, R. P. W., & Matas, J. (1998). On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(3), 226–239. https://doi.org/10.1109/34.667881
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159. https://doi.org/10.2307/2529310
- Ministère des Forêts, de la Faune et des Parcs. (2022). *Aires infestées par la tordeuse des bourgeons de l'épinette au Québec en 2022* (p. 25).
- Tahir, M. A., Kittler, J., & Bouridane, A. (2012). Multilabel classification using heterogeneous ensemble of multi-label classifiers. *Pattern Recognition Letters*, *33*(5), 513–523. https://doi.org/10.1016/j.patrec.2011.10.019
- Wang, X., Blanchet, F. G., & Koper, N. (2014). Measuring habitat fragmentation: An evaluation of landscape pattern metrics. *Methods in Ecology and Evolution*, *5*(7), 634–646. https://doi.org/10.1111/2041-210X.12198



THANK YOU!

Friday May 3rd
17th CFR CONFERENCE

Rindra Ranaivomanana
eb191970@ens.uqam.ca
rindra.fanomezana@gmail.com