

Towards **temporal trends** in **plants** based on **massive opportunistic data**

Combining **occupancy models** and **deep learning algorithms**

Rencontre CiSStats-RESSTE - February 2026

Raphaël BENERRADI PhD student at Inria-LIRMM (Montpellier, France)

Supervision: Alexis Joly, Christophe Botella, Maximilien Servajean



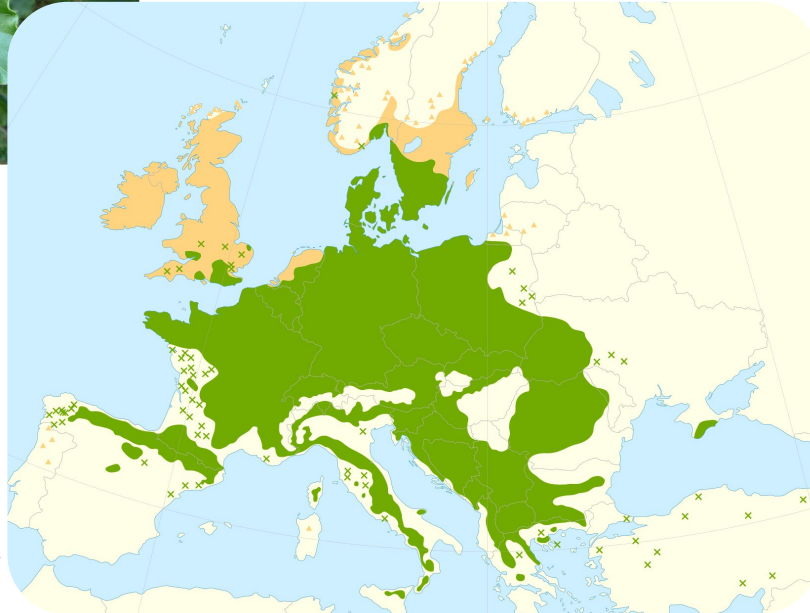
Pl@ntNet



LIRMM

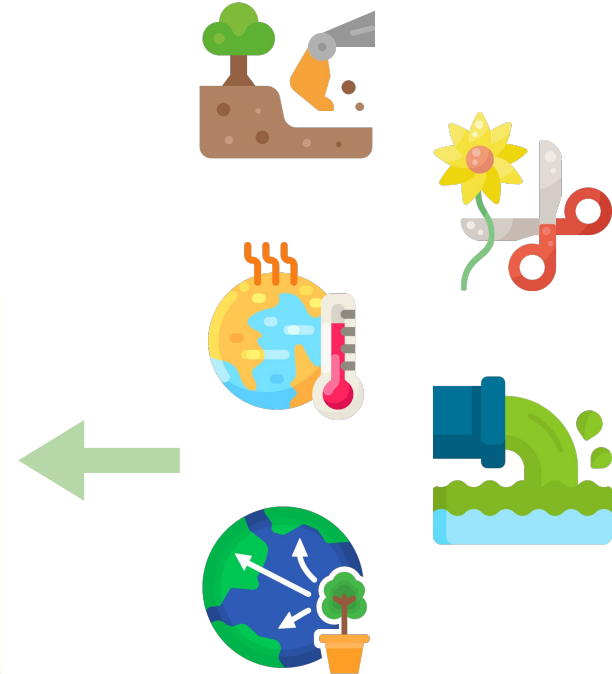


Species Distributions



**Distribution map of *Fagus sylvatica* -
European Beech**

Caudullo, G., Welk, E., San-Miguel-Ayanz, J.,
2017. Chorological maps for the main
European woody species. Data in Brief 12,
662-666. DOI: doi.org/10.1016/j.dib.2017.05.007



IPBES (2019): Global assessment report on
biodiversity and ecosystem services - ipbes.net

How do we get data?

500ENI

Réseau 500 ENI

500 Parcelles
Pour étudier les Effets Non Intentionnels
des Pratiques agricoles sur la biodiversité



Données

Relevés / parcelle / an*



Parcelles et paysages
Géographie
Climatologie
Gestion des terres



Pratiques agricoles*
Interventions culturales
Intrants (pesticides)



Oiseaux*
Parages



Coléoptères*
Bords de champs



Flore*
Bords de champs

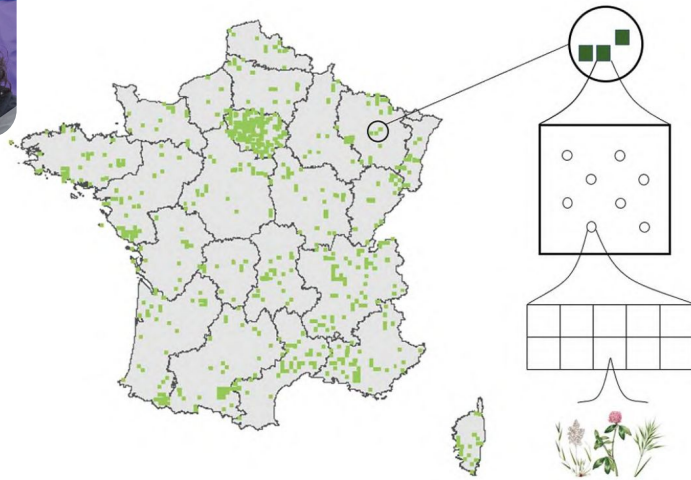


Vers de terre*
Dans la parcelle



© Vanessa Lainé
Vigie-flore bilan 2024

VigieFlore



Map of the 715 Vigie-flore squares sampled
between 2009 and 2023

Vigie-flore bilan 2023

Réseau 500 ENI - Biovigilance

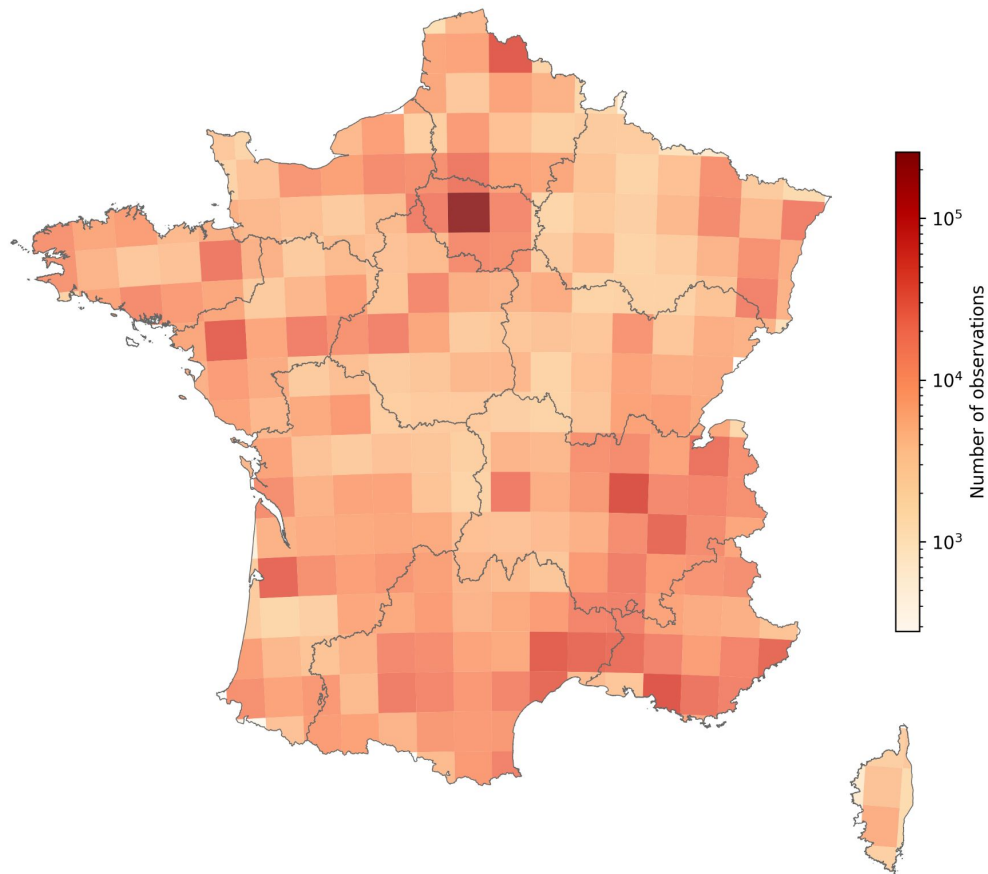
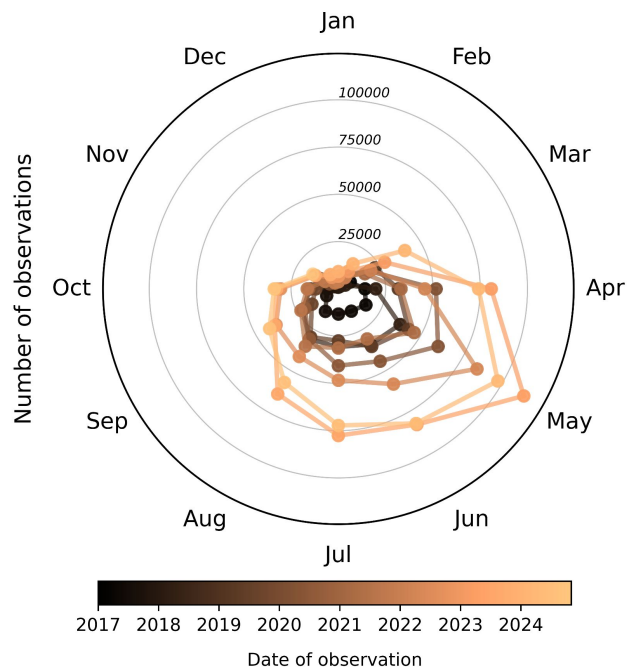
Citizen Science Programs - Opportunistic data



“Hey, today I found *this plant species* **here!**”

In **opportunistic data**, people observe

- **wherever** they want
- **whenever** they want
- **whatever** they want



Trend estimation methods on PO data

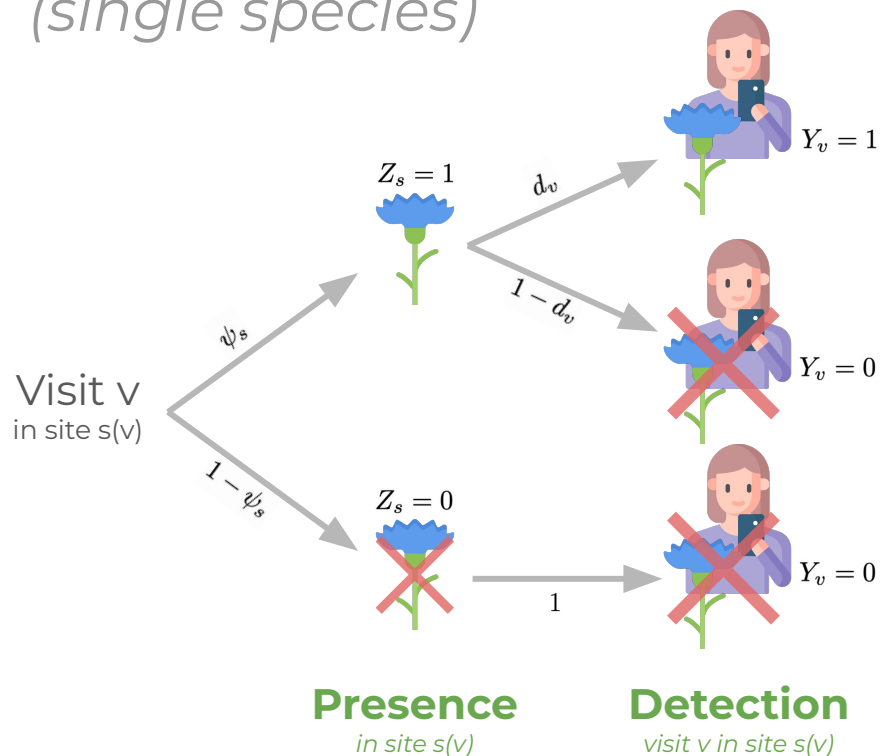
| Method | Poisson regression with TGB | Frescalo | Occupancy |
|----------------------------|-----------------------------|----------|-----------|
| Sampling bias | ✓ | ✓ | ✓ |
| Detection & reporting bias | ✗ | ✗ | ✓ |

Example of **Poisson** with TGB: $N_{\text{sp}_k, \text{cell}_i, t} \sim \mathcal{P}\left(N_{\text{TGB}} \times e^{\alpha_{k,(i)} + \beta_k t}\right)$

Frescalo: corrects sampling bias using information from “neighborhoods”

Occupancy models

(single species)



(Presence) $Z_s \stackrel{\text{ind.}}{\sim} \mathcal{B}(\psi_s)$

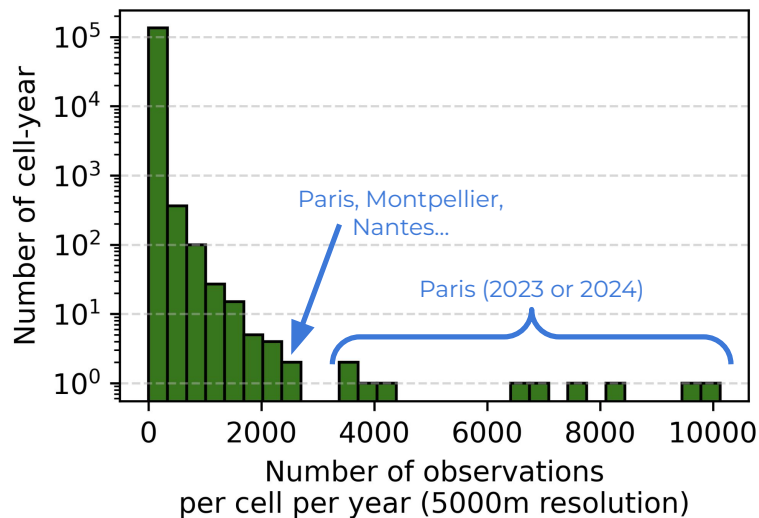
(Detection) $Y_v | Z_{s(v)} \stackrel{\text{ind.}}{\sim} \mathcal{B}(d_v \cdot Z_{s(v)})$

- Explicit likelihood to maximize
(only depending on ψ^s , d^v and y^v)
- Models for ψ^s and d^v could be anything,
e.g. for spatio-temporal sites:

$$\psi_s = \psi_{\text{cell}_i, t} = \text{sigmoid}(\alpha + \beta t)$$

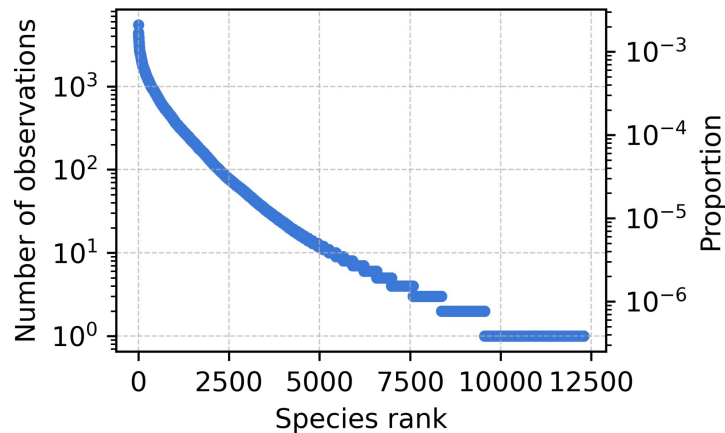
$$\text{or } \psi_s = \text{sigmoid}(\text{NN}(X_s))$$

Issues with Pl@ntNet data



In metropolitan France:

- 2.6 Million observations
- 1.4 Million annotated



➔ Scalability and reliability of the estimates

Fresco:

Inefficient implementation...

Occupancy:

- unmarked: memory issues
- Bayesian: computational time

New implementations

Fresco: Fast implementation in R...

But still quadratic → heavy for small resolutions

Occupancy:

- Sparse implementation in `scipy`
 - Gradient based optimization
 - MLE and penalized versions
- Stochastic Gradient Descent in `torch`
 - Suitable for large datasets
 - Possible extensions to neural networks

$$\begin{matrix} & \begin{matrix} \text{visits} \end{matrix} \\ \begin{matrix} \text{sites} \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & \dots & 1 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & \text{NA} & \text{NA} & \dots & \text{NA} \end{pmatrix} \end{matrix}$$

Tests on realistic simulated data

Dataset:

$$x_s \stackrel{\text{ind.}}{\sim} \mathcal{N}(0, 1) \ ; \ \text{logit}(\psi_s) = \alpha + \beta x_s \ ; \ \text{logit}(d_v) = \gamma + \delta x_s$$

$$\text{Nbr.Visits}_s \propto e^{1.5 x_s}$$

- Similar to **50 km cell size** and **~100 detections** out of 1 million visits

Fitting:

- MLE grid search
- MLE gradient based

➔ Are occupancy MLE estimates reliable?

Trend estimation reliability

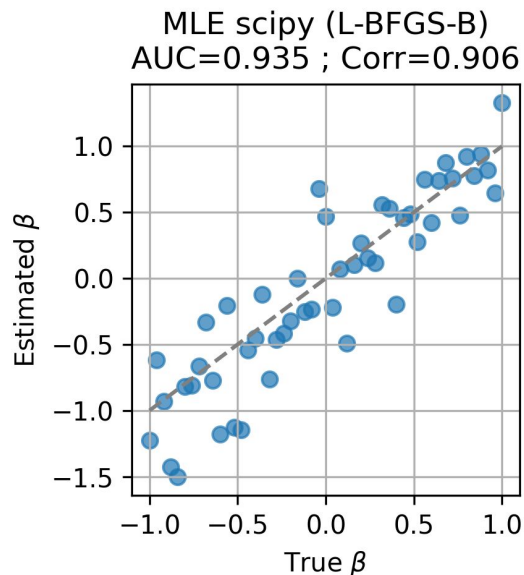
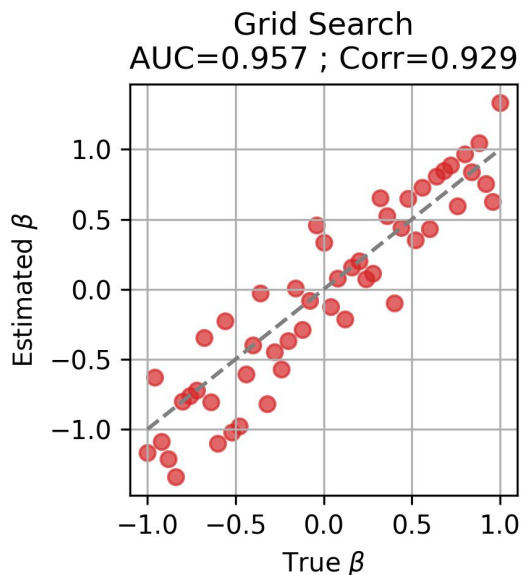
$x_s \stackrel{\text{ind.}}{\sim} \mathcal{N}(0, 1)$; $\text{logit}(\psi_s) = \alpha + \beta x_s$; $\text{logit}(d_v) = \gamma + \delta x_s$

$\text{Nbr.Visits}_s \propto e^{1.5 x_s}$

$\alpha = -1.5 - 0.5\beta$, $\gamma = -6.50$, $\delta = -1.00$

~ 100 detections out of 1 million visits

Realistic for
50 km cell size

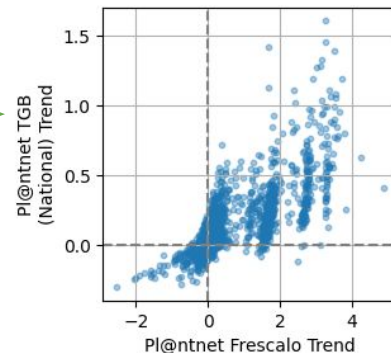


Real data - trend estimation (Pearson correlations)

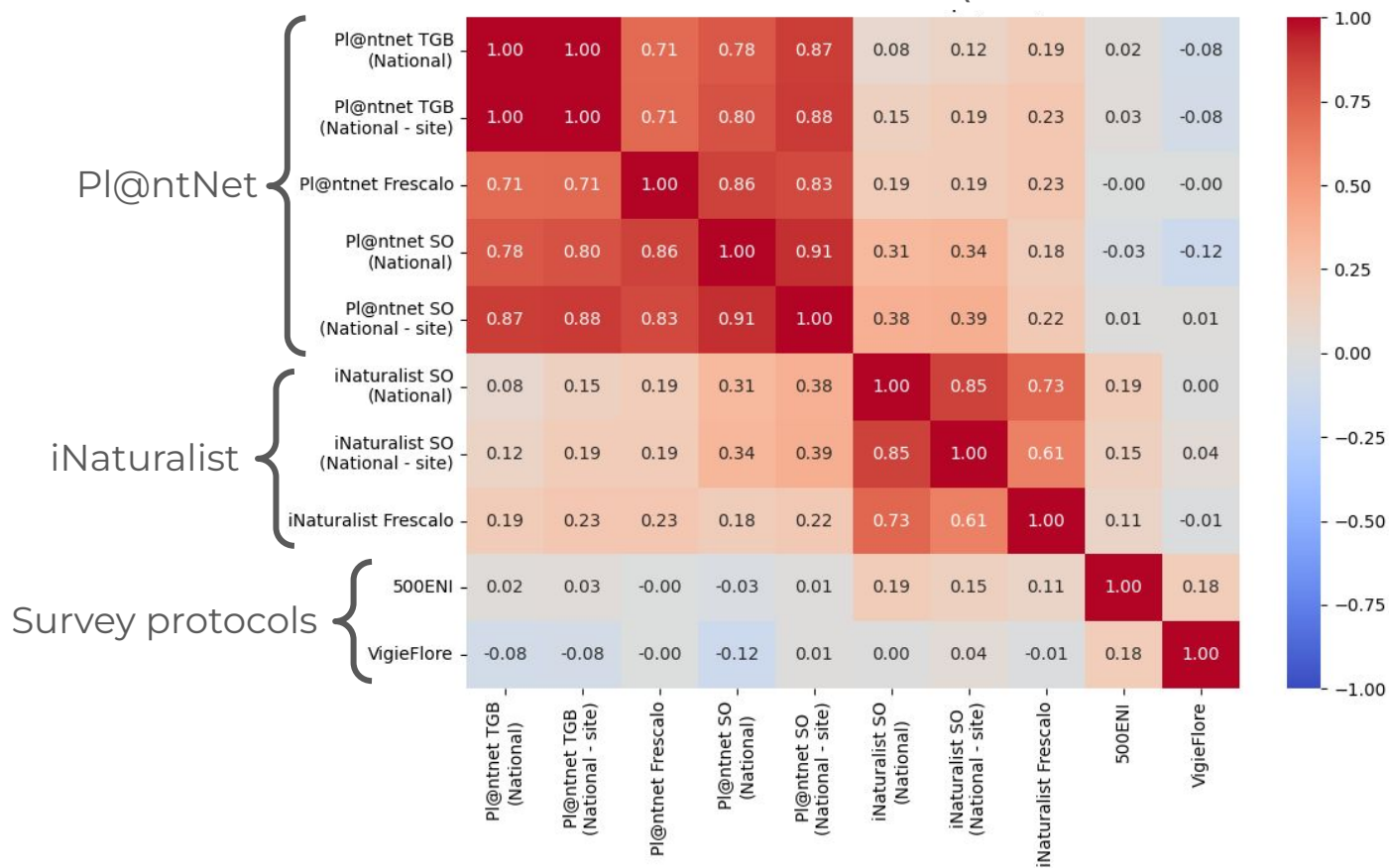
$$N_{\text{sp}_k, \text{cell}_i, t} \sim \mathcal{P}(N_{\text{TGB}} \times e^{\alpha_{k,(i)} + \beta_k t})$$

$$\psi_{\text{sp}_k, \text{cell}_i, t} = \text{sigmoid}(\alpha_{k,(i)} + \beta_k t)$$

| | | | | | | |
|--|--------------------------------|------|------|------|------|------|
| $\left\{ \begin{array}{l} \text{Pl@ntnet TGB (National)} \\ \text{Pl@ntnet TGB (National - site)} \end{array} \right.$ | Pl@ntnet TGB (National) | 1.00 | 1.00 | 0.71 | 0.78 | 0.87 |
| | Pl@ntnet TGB (National - site) | 1.00 | 1.00 | 0.71 | 0.80 | 0.88 |
| Pl@ntnet Frescalo | | 0.71 | 0.71 | 1.00 | 0.86 | 0.83 |
| $\left\{ \begin{array}{l} \text{Pl@ntnet SO (National)} \\ \text{Pl@ntnet SO (National - site)} \end{array} \right.$ | Pl@ntnet SO (National) | 0.78 | 0.80 | 0.86 | 1.00 | 0.91 |
| | Pl@ntnet SO (National - site) | 0.87 | 0.88 | 0.83 | 0.91 | 1.00 |
| | Pl@ntnet TGB (National) | | | | | |
| | Pl@ntnet TGB (National - site) | | | | | |
| | Pl@ntnet Frescalo | | | | | |
| | Pl@ntnet SO (National) | | | | | |
| | Pl@ntnet SO (National - site) | | | | | |



Real data - trend estimation (Pearson correlations)



thanks for your attention

Supervisors:



**Alexis
JOLY,**
Inria, Montpellier



**Maximilien
SERVAJEAN,**
LIRMM, Montpellier



**Christophe
BOTELLA,**
Inria Montpellier



Pl@ntNet



LIRMM



Bibliography

- [1] Botella, C., Joly, A., Monestiez, P., Bonnet, P., & Munoz, F. (2020). Bias in presence-only niche models related to sampling effort and species niches: Lessons for background point selection. *PLOS ONE*, 15(5), e0232078. <https://doi.org/10.1371/journal.pone.0232078>
- [2] Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, 19(1), 181-197. <https://doi.org/10.1890/07-2153.1>
- [3] Lasgorceux, F., Papaix, J., Bunz, Y., Combrisson, D., & Opitz, T. (2024). Space-time species distribution modeling for opportunistic presence-only data: A case study of passerines in a protected area. <https://doi.org/10.17180/ram8-mv76>
- [4] Hill, M. O. (2012). Local frequency as a key to interpreting species occurrence data when recording effort is not known. *Methods in Ecology and Evolution*, 3(1), 195-205. <https://doi.org/10.1111/j.2041-210X.2011.00146.x>
- [5] Eichenberg, D., Bowler, D. E., Bonn, A., Bruelheide, H., Grescho, V., Harter, D., Jandt, U., May, R., Winter, M., & Jansen, F. (2021). Widespread decline in Central European plant diversity across six decades. *Global Change Biology*, 27(5), 1097-1110. <https://doi.org/10.1111/gcb.15447>
- [6] Goury, R., Bowler, D. E., Harrower, C., Münkemüller, T., Vallet, J., Yearsley, J., Thuiller, W., & Pescott, O. L. (2025). A practical guide to species trend detection with unstructured data using local frequency scaling (Frescalo). <https://ecoevortexiv.org/repository/view/9467/>
- [7] MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Andrew Royle, J., & Langtimm, C. A. (2002). Estimating Site Occupancy Rates When Detection Probabilities Are Less Than One. *Ecology*, 83(8), 2248-2255. [https://doi.org/10.1890/0012-9658\(2002\)083%255B2248:ESORWD%255D2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083%255B2248:ESORWD%255D2.0.CO;2)
- [8] van Strien, A. J., van Swaay, C. A. M., & Termaat, T. (2013). Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology*, 50(6), 1450-1458. <https://doi.org/10.1111/1365-2664.12158>