

# Opportunistic data : definition(s) and thoughts about their use

Grégoire Lois, Vigie-Nature, CESCO, MNHN, OFB, ARB-IDF



**Geo**  
Nature

  
eBird  
The Cornell Lab of Ornithology

  
seek  
by iNaturalist

eBird

  
iNaturalist



  
Faune  
France  
Notez toutes vos observations  
naturalistes sur  
[www.faune-france.org](http://www.faune-france.org) ou  
son appli **NaturaList** !

  
Clicnat



**GBIF**  
Global Biodiversity  
Information Facility



Pl@ntNet

# How do opportunistic/unstructured data end up in a database ?

The observer identifies one or several taxa

Who ? (Not everyone although many species are known from a very large part of the entire population : naturalist' skills)

AND he/she is aware of one or several database initiative

How did this happen ? How is such knowledge distributed among citizens ?

AND he/she wishes to share/testify/record his/her observations

WHY ? Under which conditions ? Legitimacy ? Reputation ?

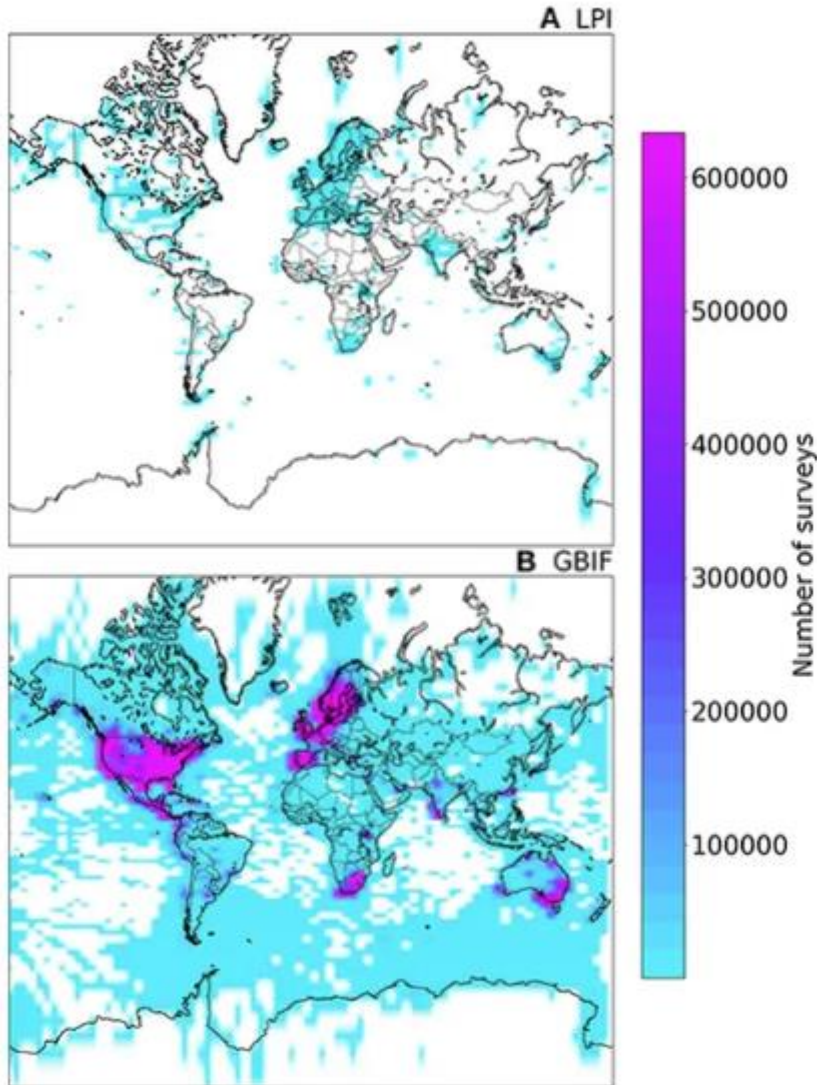
## Definitions :

Le terme « données opportunistes » renvoie au fait que les données récoltées ne le sont pas dans le cadre d'un projet ou d'un protocole spécifique. Il s'agit typiquement de données issues d'observations réalisées à l'occasion d'une randonnée, au cours d'une balade ou tout autre circonstance ne renvoyant pas à un projet particulier. <https://naturagis.fr/outils-donnees-naturalistes/definition-donnee-naturaliste/>

(...) opportunistic citizen data, that is, observations of species collected without standardized field protocol and without explicit sampling design (...) van Strien, 2013. <https://doi.org/10.1111/1365-2664.12158>

unstructured citizen science data : The majority of citizen science projects that gather ecological data can be characterized as unstructured (Pocock et al. 2017) and do not gather information on the observation process, meaning that there is no fully statistically defensible way of accounting for the biases inherent in the data collection. S. Kelling, 2019.  
<https://doi.org/10.1093/biosci/biz010>

# Structured : non opportunistic data sets



Structured

Ratio approx. 1/2000

Unstructured

# Semi-Structured (semi-opportunistic data sets ?) : datasets known to be complete checklists

## eBird

- 2 billions data with 150 millions checklists



Faune  
France

### GÉNÉRALES

Observateurs inscrits :	238 913
<div><span>Au total</span><span>Cette année</span><span>Ce mois</span></div>	
Total des contributions :	153 942 311
Nombre total de listes :	3 736 743
Nombre total de listes complètes :	2 066 951

} ?



# eBird

## Various ways of participating

### Top 100 des observateurs dans France par espèces

Testimony first	Listes complètes	Espèces (% du total)
1. Frederic Veyrunes	77	509 90.73%
2. Dimitri Davignon	1 341	490 87.34%
3. Paul Doniol-valcroze	307	488 86.99%
4. Julien Birard	1 810	481 85.74%
5. Laurent Veyrunes	45	476 84.85%
6. Bernard Bougeard	5	474 84.49%
7. Paul Dufour	1 083	470 83.78%
7. Benjamin Luneau	0	470 83.78%
9. Thomas Galewski	1 494	464 82.71%
10. Aymeric Le Calvez	1	461 82.17%
11. Boris Delahaie	952	460 82.0%
12. Valentin motteau	449	452 80.57%

### Top 100 des observateurs dans France par listes complètes

Contribution first	Listes complètes	Espèces (% du total)	Dernier ajout
1. Laurent Chevallier	14 648	371 66.13%	Martin-pêcheur d'Amérique
2. Jean-Marie Frenoux	10 233	272 48.48%	Effraie des clochers (22 févr.
3. Adrien Charbonneau	9 524	345 61.5%	Capucin bec-de-plomb (29 é
4. Daniel DE SOUSA	6 510	382 68.09%	Plongeon du Pacifique (18 ja
5. Adrien Mauss	6 476	422 75.22%	Plongeon du Pacifique (17 ja
6. Olivier Benoist	4 637	386 68.81%	Bihoreau gris (25 juil. 2024)
7. Thomas Bitsch	3 934	375 66.84%	Puffin cendré (25 août 2025)
8. Olivier SWIFT	3 569	280 49.91%	Gravelot à collier interrompu
9. Camille HELLIO	3 321	252 44.92%	Tichodrome échelette (18 ja
10. Nicolas Marmet	3 198	411 73.26%	Niverolle alpine (2 janv. 202
11. Thierry Besançon	2 985	329 58.65%	Aigle botté (31 août 2024)
12. Jean-Roch Poutrieux	2 620	278 49.55%	Bernache de Hutchins (20 de

# eBird



Faune  
France

Checklists are in birders habits.

Gamification of observations : Either you mention your lifelist, year list, garden list, window list, office list, trip list, WALK LIST, ...

All is needed is to ask for duration or to impose duration.

# True opportunistic data sets : no information regarding data collection



+++....





# Unstructured ?

**Records that do not follow standardized field protocol nor explicit sampling design.**

Vast majority of iNaturalist, Faune-France, Geonat-IDF, PlantNet, Xeno-Canto, ...

=> Some species are mentioned for one reason or the other during sampling occasion.

=> Complete species list cannot be assumed unless assumption is made

# True opportunistic data sets : case of iNaturalist



293.355.709  
OBSERVATIONS



550.711  
ESPÈCES



473.120  
IDENTIFICATEURS



4.006.791  
OBSERVATEURS

Rang	Utilisateur	Observations	Espèces
1	reiner	388.709	9.679
2	tonyrebelo	383.722	12.369
3	johndreynolds	337.842	14.511
4	ck2az	320.184	4.739
5	angel_fernandez_cancio	291.394	4.571
6	ku6777	231.022	6.178
7	kyle_campbell1	226.012	4.720
8	srall	214.968	4.353
9	mako252	177.375	6.856
10	astrobirder	177.149	12.964
11	erikamitchell	176.745	8.669
12	sambiology	171.213	12.045

Rang	Utilisateur	Observations	Espèces
1	silversea_starsong	131.578	26.120
2	cypselurus	70.925	17.230
3	felix_riegel	79.703	16.761
4	knightericm	39.911	16.257
5	josh_vandermeulen	64.757	16.146
6	dbeadle	42.474	15.658
7	trevvan	79.124	15.316
8	benoit_segerer	48.442	14.906
9	damontighe	100.077	14.544
10	johndreynolds	337.842	14.511
11	sheriff_woody_pct	99.129	14.493
12	jeanpaulboerekamps	74.237	14.430

No information on data collection  
Checklists are not available or even planned  
Common species are very often omitted (but might help to identify complete checklists if numerous species mentioned in one single occasion and list includes common species)



# Faune France

faune-iledefrance.org

CONSULTER

PARTICIPER

Bienvenue sur Faune Ile-de-France, votre base de données naturaliste

Au total

Cette année

Ce mois

Nombre d'observations : 7 137 232

Au total

Cette année

Ce mois

Nombre d'observations : 69 195

Au total

Cette année

Ce mois

Au total

Cette année

Ce mois

	observations	dernière contribution
1) Vincent Le Calvez	311 780	il y a 1 heure
2) Laurent Chevallier	233 317	avant hier
3) Nom de l'observateur non diffusé	164 003	hier
4) Christophe Bruneaux	135 535	hier
5) Christian Letourneau	131 964	il y a 53 minutes
6) Claude Hardel	104 004	il y a 108 jours
7) Isabelle Giraud	101 187	hier
8) Nicolas Marmet	97 155	il y a 6 jours
9) Stanislas Zeller	96 884	hier
10) Miguel Arrechea	96 662	hier

	observations	dernière contribution
1) Vincent Le Calvez	2 995	il y a 1 heure
2) Christophe Bruneaux	1 597	hier
3) Christophe Alexandre	1 590	il y a 5 minutes
4) Laurent Pascual-Le Tallec	1 334	hier
5) Miguel Arrechea	1 326	hier
6) Isabelle Giraud	1 031	hier
7) Léo Domingues-Haccart	1 024	hier
8) José Miguel Martins Da Silva	967	hier
9) Quentin Aracheloff	913	avant hier
10) Birgit Tollner	847	hier



# Faune France

faune-iledefrance.org

CONSULTER

PARTICIPER

Bienvenue sur Faune Ile-de-France, votre base de données naturaliste

Au total

Cette année

Ce mois

Nombre d'observations :

7 137 232

Au total

Cette année

Ce mois

Nombre d'observations :

69 195

Au total

Cette année

Ce mois

19%

Au total

Cette année

Ce mois

20%

	observations	dernière contribution
1) Vincent Le Calvez	311 780	il y a 1 heure
2) Laurent Chevallier	233 317	avant hier
3) Nom de l'observateur non diffusé	164 003	hier
4) Christophe Bruneaux	135 535	hier
5) Christian Letourneau	131 964	il y a 53 minutes
6) Claude Hardel	104 004	il y a 108 jours
7) Isabelle Giraud	101 187	hier
8) Nicolas Marmet	97 155	il y a 6 jours
9) Stanislas Zeller	96 884	hier
10) Miguel Arrechea	96 662	hier

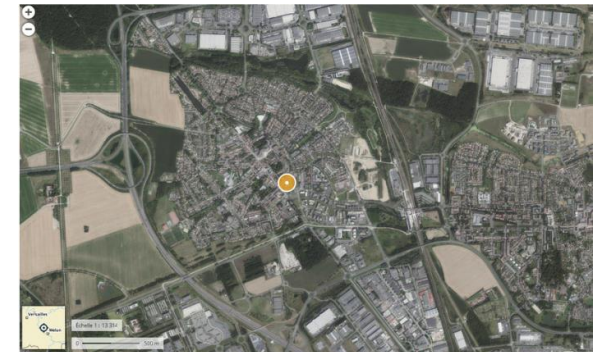
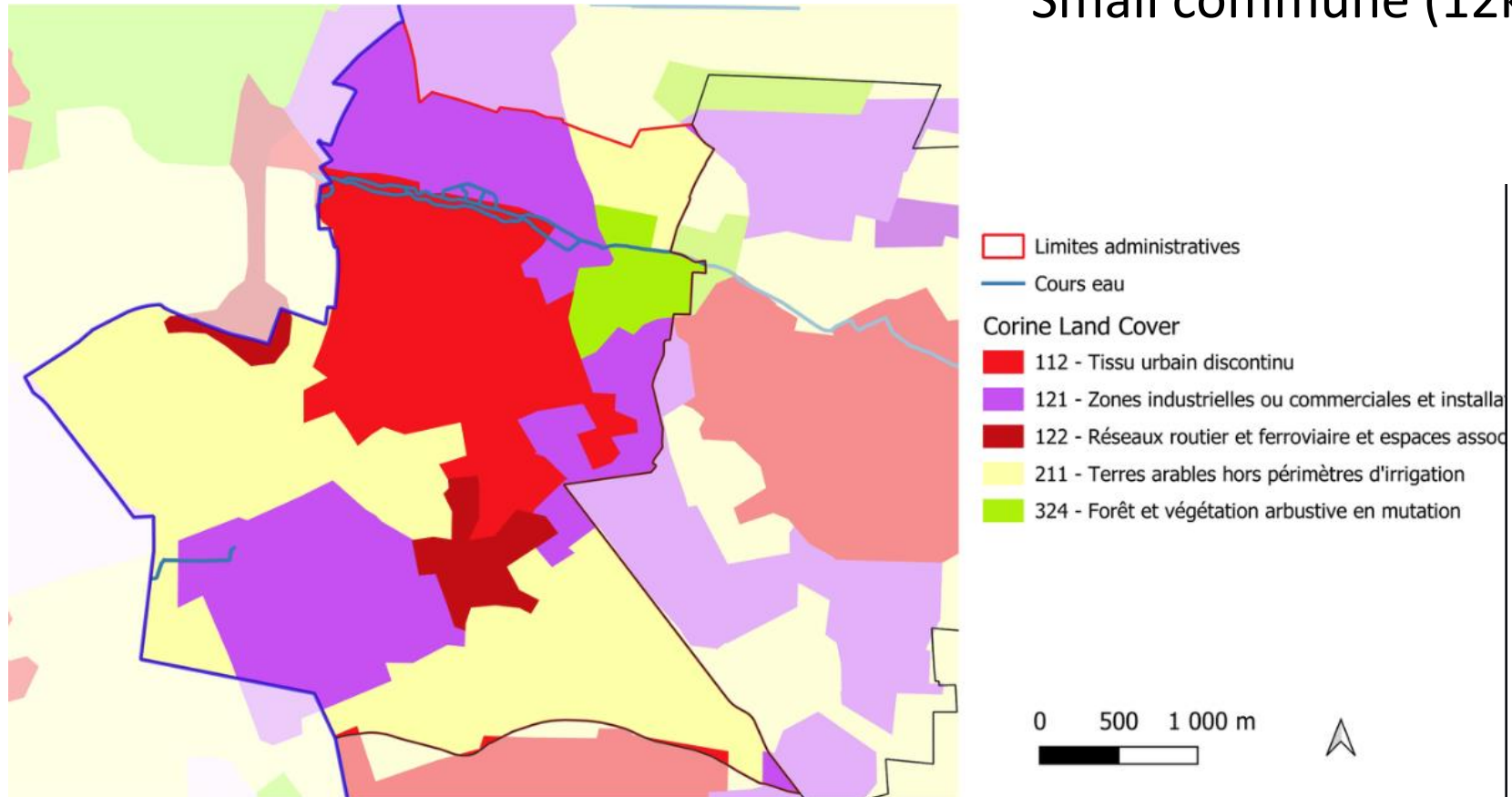
players

Sum

	observations	dernière contribution
1) Vincent Le Calvez	2 995	il y a 1 heure
2) Christophe Bruneaux	1 597	hier
3) Christophe Alexandre	1 590	il y a 5 minutes
4) Laurent Pascual-Le Tallec	1 334	hier
5) Miguel Arrechea	1 326	hier
6) Isabelle Giraud	1 031	hier
7) Léo Domingues-Haccart	1 024	hier
8) José Miguel Martins Da Silva	967	hier
9) Quentin Aracheloff	913	avant hier
10) Birgit Tollner	847	hier

# Exemple de complete checklist (eq. complete inventory) assumed by data owner

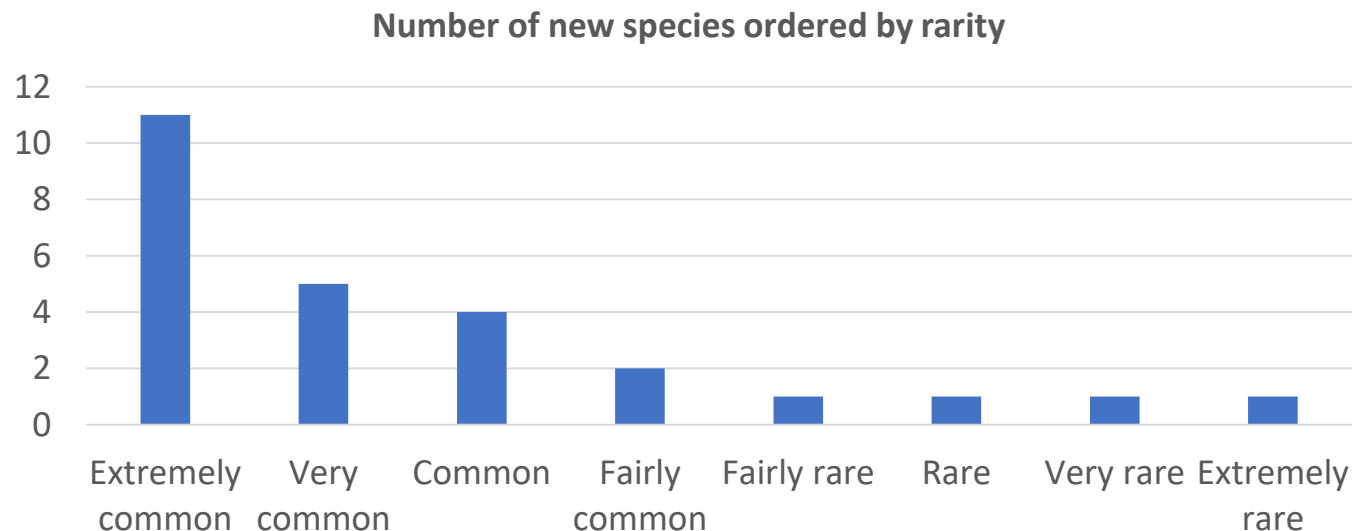
Small commune (12km<sup>2</sup>) in the suburbs of Paris



Exemple of complete checklist (eq. complete inventory) assumed by data owner

301 species listed by data owner over some 20 years (among 1600 possible in Ile-de-France)

141 species found during specific inventories in preseumably 'rich' areas during summer 2025 of which 31 where « new species »



Rarity status according to data owner (5 species are undefined).



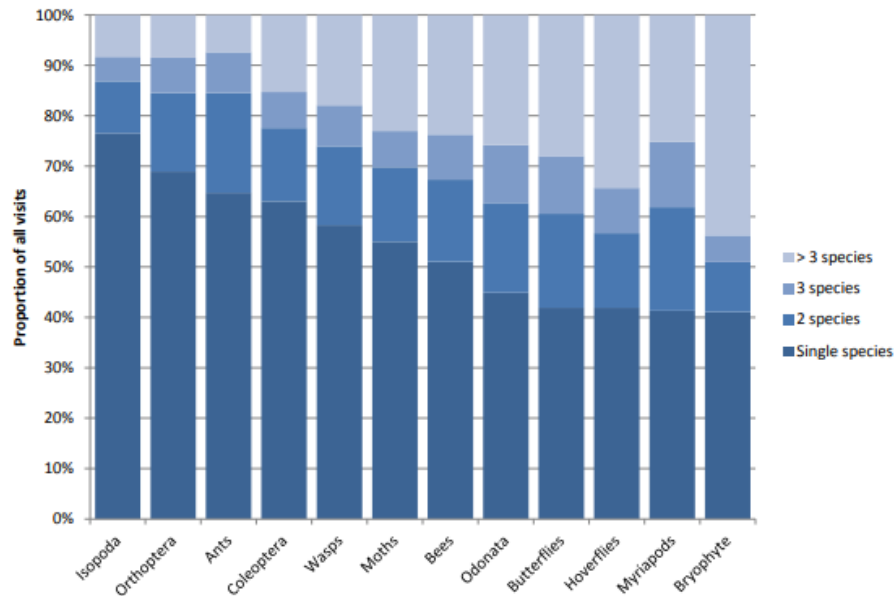
# Extracting trends from biological recording data

Nick Isaac  
Biological Records Centre  
Centre for Ecology & Hydrology

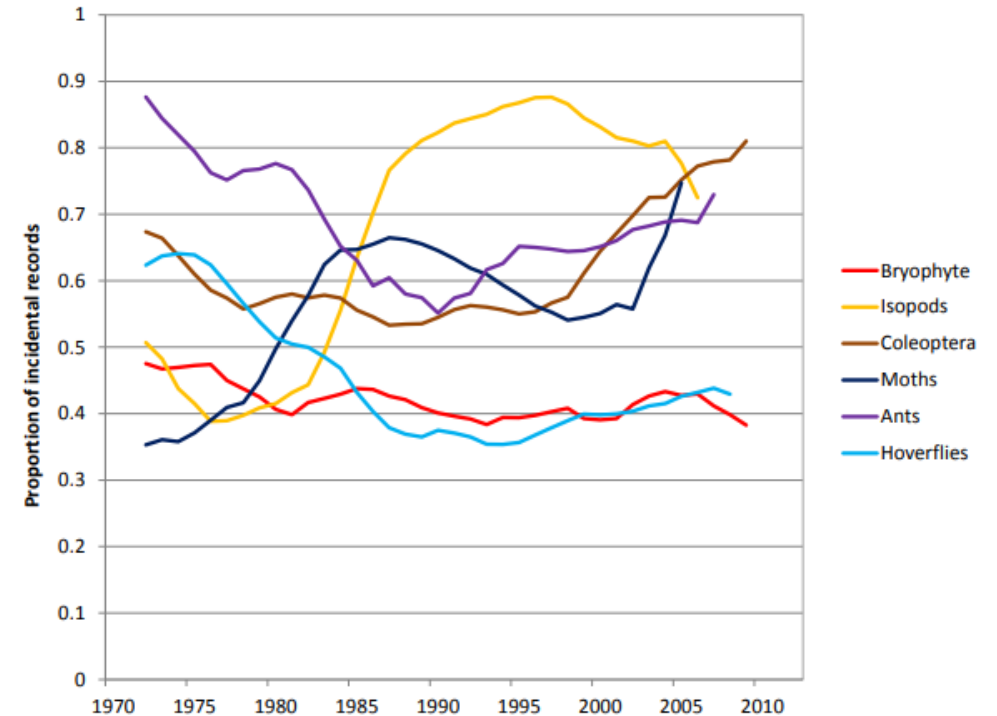
2012, UK only

## Most lists are incomplete

For most groups, ~50% of visits produce 'incidental records'



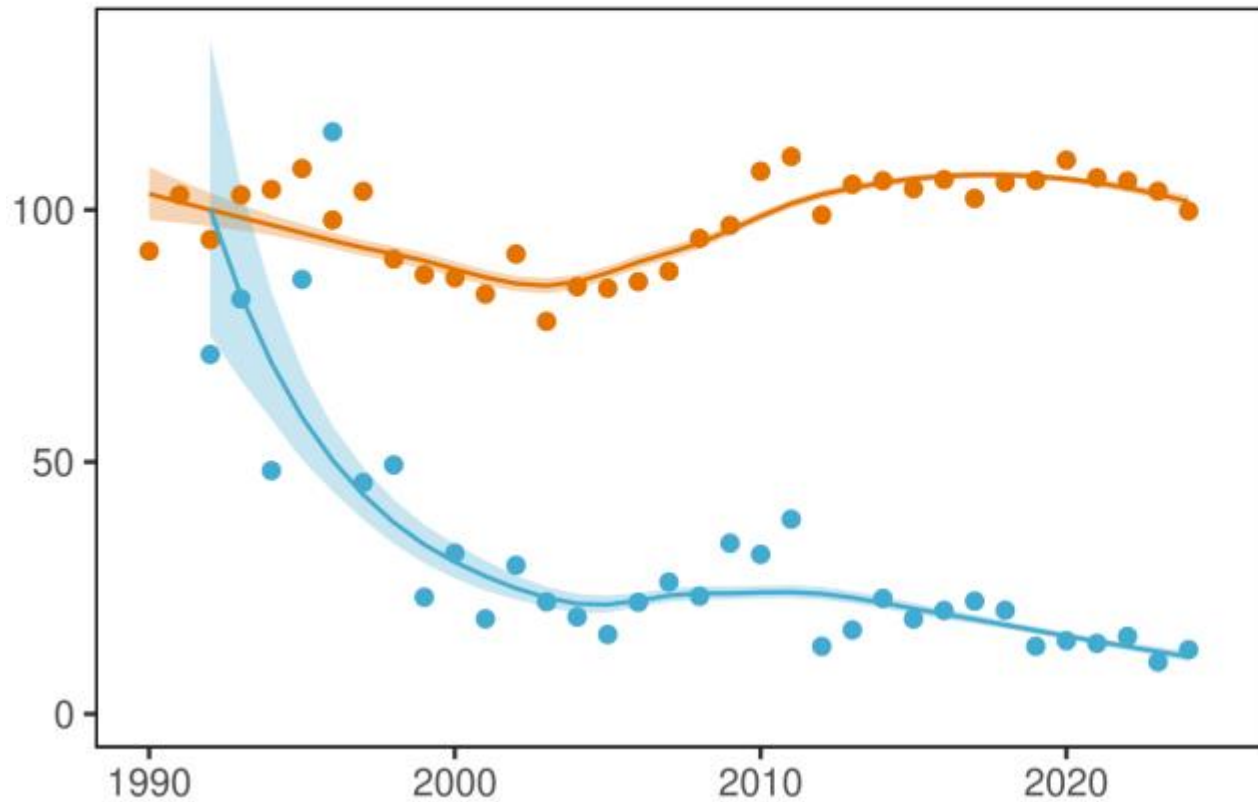
## Lists lengths are not constant over time



# Occupancy <> Population

*Ochlodes sylvanus*

Groot dikkopje



occupancy  
populatie

Van Swaay, 2025.

[https://www.ufz.de/export/data/10/299633\\_Chris%20van%20Swaay.pdf](https://www.ufz.de/export/data/10/299633_Chris%20van%20Swaay.pdf)

# Frescalo Application

- Are some common species systemically omitted ?
- Are some common species randomly omitted ?
- Can this be estimated ? Taken into account ?
- Does it impact « unbiased estimate of temporal trends » as it relies on « enough sampling to at least estimate species' local relative frequencies fairly accurately. » i.e. « temporal variation in effort is accounted for using an index of local recording completeness: the proportion of a suite of locally common species, sometimes referred to as “benchmark species” »? (Goury, 2025)

# Occupancy detection

- Is it feasible to model recorder behaviour ?
- Is it feasible to rebuild from opportunistic datasets (in which checklists are not flagged) some virtual checklists as occupancy models require detection/nondetection data collected during replicated visits ?