

ISSN 0154 - 2109



Agir pour
la biodiversité



LE GRAND-DUC N°91 (ANNEE 2023)

**L'avifaune de plateaux agricoles
granitiques
au sud de Clermont-Ferrand**

-

**Estimation des populations
d'oiseaux nicheurs communs
au printemps 2019**

**par la méthode du Distance Sampling et
comparaison de plusieurs méthodes et
durées d'inventaire**

Jean-Philippe Meuret
[monedula \(chez\) orange.fr](mailto:monedula@orange.fr)

François Guélin
[guelin.francois \(chez\) gmail.com](mailto:guelin.francois@gmail.com)

Résumé : Au printemps 2019, des inventaires par points d'écoute exploités selon la méthode du Distance Sampling ont permis d'estimer les populations d'une vingtaine d'espèces de passereaux nicheurs communs sur les 24 km² des plateaux granitiques de Cournols - Olloix - Montaigut-le-Blanc, secteur agricole de pâtures encore relativement préservé au sud du Puy-de-Dôme (63). Pour ces inventaires sur environ 100 points, de durées d'écoute de 5mn et 10mn, 2 équipes indépendantes ont travaillé l'une sur plans papier, l'autre avec l'application "Naturalist" pour smartphone, dans l'idée d'évaluer l'impact de ces variantes de protocole sur les estimations de population obtenues. Pour automatiser autant que possible le paramétrage, l'exécution et le dépouillement des très nombreuses variantes d'analyse des données de terrain suivant la méthode du Distance Sampling, un logiciel réutilisable ad-hoc a été développé. L'analyse finale des résultats obtenus semble montrer que les points d'écoute sur 10mn permettent d'obtenir des densités plus fortes et de meilleure qualité statistique que sur 5mn (même si cette comparaison est méthodologiquement sujette à caution), à l'inverse des relevés "au smartphone" sur 5mn, qui comparés aux relevés sur plans papier pour la même durée, sont entachés d'un déficit de données à courtes distances a priori problématique pour l'analyse des données en Distance Sampling.

Mots-clés : passereaux communs, échantillonnage, points d'écoute, plans papier, appli. Naturalist, distance sampling, optimisation, python, pyaudisam, plateau granitique, lande, pâturage extensif, France, Auvergne, Puy-de-Dôme - 63, Cournols - Olloix.

Introduction & objectifs

Les éléments qui nous ont conduits à lancer cette enquête et à décider des variantes de protocoles à utiliser sont multiples :

- contexte du déploiement par la LPO France et le MNHN du protocole EPOC depuis 2017,
- perspective du lancement en 2021 de l'enquête Atlas Oiseaux de France 2019-2024 aux objectifs quantitatifs ambitieux en partie basés sur le Distance Sampling,
- expérience collective des points d'écoute STOC-EPS de 5mn depuis 3 décennies,
- expérience de la pratique du Distance Sampling via des inventaires par points et par transects sur plans papier [GUÉLIN 2015] [GUÉLIN 2016] [GUÉLIN et al. 2018],
- expérience assez récente d'une partie des ornithologues de terrain auvergnats des inventaires par points d'écoute via l'application pour smartphone Naturalist, par "listes complètes avec géolocalisation précise de tous les contacts", avec une impression assez partagée : l'utilisation du smartphone pour noter et géolocaliser précisément les contacts n'est pas aussi naturelle et rapide qu'avec un crayon sur une copie papier d'image satellite de l'endroit ; elle semble mobiliser d'autres ressources mentales ce qui nuit probablement à la concentration auditive ; et induit sans doute une perte d'efficacité dans les relevés.

Nous avons donc voulu en avoir le cœur net, et évaluer aussi objectivement que possible l'impact sur l'efficacité des inventaires par points d'écoute :

- d'une part de l'utilisation de l'application "Naturalist" sur smartphone,
- d'autre part de la durée des inventaires,

en utilisant simplement comme critère de comparaison les estimations de densités obtenues par Distance Sampling.

Terrain d'étude et protocoles

A. Secteur d'étude et milieu naturel

La zone géographique choisie pour l'étude correspond à peu de choses près aux plateaux granitiques de Cournols - Olloix - Montaigut-le-Blanc, situés à environ 20km à vol d'oiseau au sud-ouest de Clermont-Ferrand, 63.

Elle couvre presque complètement les communes de Cournols au nord et Olloix au centre-ouest, la moitié ouest de la commune de Ludesse à l'est, et les 3/4 nord de la commune de Montaigut-le-Blanc situés au nord de la Couze Chambon. Elle est entaillée d'est en ouest par les gorges boisées de la Monne entre les communes de Cournols et Olloix.



Figure 2-1 : *Au nord-est, vers les gorges de la Monne en contrebas, le Puy de Peyronnière au centre, et la Montagne de la Serre à gauche en arrière plan (photo J.-P. Meuret)*

A une altitude moyenne de 750m, dominé par le Puy d'Olloix à l'ouest (1002 m) et le Puy de Gourdon (879 m) au sud-ouest, ces plateaux de l'étage collinéen constituent un isolat géologique au paysage assez original parsemé de chaos granitiques, de bosquets de pins et de landes à genets et genévriers, cette dernière essence par endroits en massifs denses et étendus.



Figure 2-2 : *Au centre, direction sud, chaos granitique, lande et pâtures, le Puy de Gourdon au fond à droite (photo J.-P. Meuret)*

En dehors des gorges de la Monne qui le traversent sur un axe nord-est / sud-ouest entre Cournols et Olloix, et des vallées boisées en aval des quelques petits ruisseaux qui y prennent source, il présente un habitat assez ouvert, où le pâturage extensif longtemps d'usage est maintenant en fort déclin, avec par endroits les restes d'un parcellaire de murs en pierres sèches, ailleurs des parcours sans clôtures.



Figure 2-3 : *Au centre, direction sud-est, chaos granitique, friche et parcours à moutons, vallée de la Couze Chambon au fond (photo J.-P. Meuret)*

La zone de l'étude tangente également à l'est, entre Ludesse et Chaynat, les parties hautes du secteur de grandes cultures de Plauzat.



Figure 2-4 : *Au sud-est, grandes cultures entre Ludesse et Chaynat (photo J.-P. Meuret)*

Le choix de ce secteur d'étude n'est pas un hasard : l'endroit, très apprécié de longue date des ornithologues auvergnats et en particulier clermontois, "subit" en effet depuis plusieurs décennies une pression d'observation ornithologique plus régulière et intense qu'ailleurs ; il a même fait l'objet de plusieurs séries d'IPA [DULPHY 1994] [DULPHY 2012], la dernière série à ce jour ayant eu lieu ... en même temps que la présente étude, afin de tenter des estimations (par "rétro-extrapolation") des populations d'oiseaux communs du passé en combinant les IPA et le Distance Sampling [DULPHY et al. 2023].

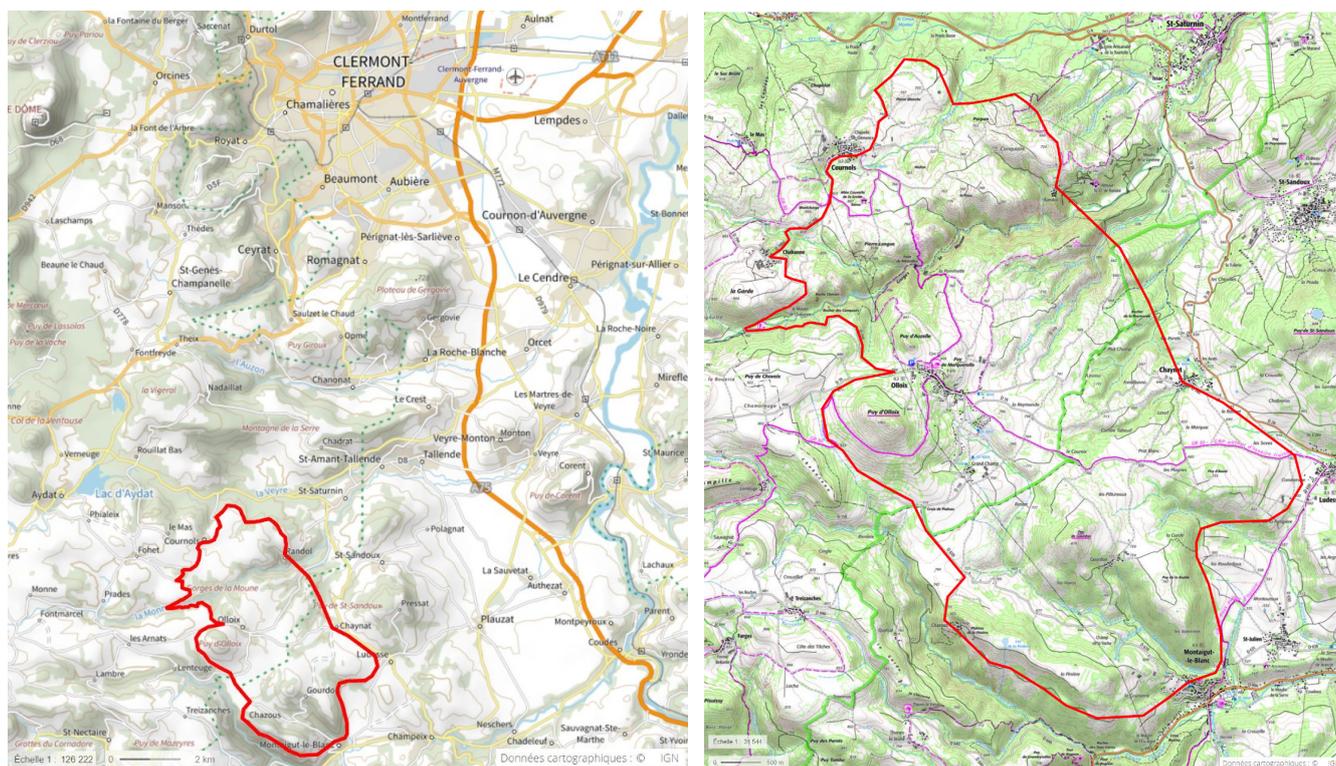


Figure 2-5 : Localisation de la zone d'étude : situation d'ensemble à gauche, détail à droite (© geoportail.fr)

B. Méthode et protocoles

1. Des points d'écoute placés aléatoirement (ou presque)

Les points d'écoute ont été d'abord répartis dans le secteur d'étude selon une **grille régulière** et complète à **mailles de 500 mètres**, le point de référence (le plus au nord) ayant été placé arbitrairement aux coordonnées "UTM 31" 503,5 km est, 5055,8 km nord :

- c'est en effet une manière simple de placer aléatoirement les points vis à vis des éléments géologiques, paysagers ou d'origine anthropique,
- qui permet de couvrir en grande partie la surface d'étude tout en limitant autant que possible les doubles-comptages d'oiseaux chanteurs pour beaucoup des espèces ciblées, à l'exception des Grive draine, Coucou gris, Huppe fasciée, Pigeon ramier, ... etc, à chants plus portants ; en ce sens, et avec l'expérience des STOC-EPS et des inventaires auvergnats par Distance Sampling "par points" [GUÉLIN 2015] [GUÉLIN 2016] [GUÉLIN et al. 2018], ce "**rayon de perception auditive et visuelle**" de **250 mètres** autour de chaque point semble un compromis acceptable dans le milieu naturel ciblé, plutôt ouvert.

Ensuite, pour la simplicité des analyses et la solidité des résultats - qui seront obtenus en première approche selon le bon vieux principe "d'extrapolation après échantillonnage" - nous avons choisi de cibler un seul milieu naturel type, celui couvrant la grande majorité de la zone d'étude : un **milieu ouvert** de prairies bordées de petites haies et / ou de murets, pâtures extensives à petits bosquets de pins et chaos granitiques, friches basses, landes à genets et / ou genévriers ; nous avons donc éliminé de la grille complète initiale les points dont le "rayon de perception" ne correspondait pas en grande majorité au milieu cible (forêt), même au prix d'un déplacement du point d'une distance inférieure ou égale à 100m ; certains **points difficiles d'accès** (forte pente, ...) ont également été éliminés.

Enfin, certains points ont également été **déplacés de 100 mètres au maximum** :

- d'une part, comme suggéré ci-dessus, pour améliorer la proportion couverte du milieu cible quand c'était possible,
- d'autre part pour faciliter les 2 accès nécessaires (voir ci-dessous) durant l'enquête : malgré des biais certains (haies au bord des chemins, lisières de landes, ...), nous avons tenu à respecter les éventuelles cultures (pas de piétinement), à éviter les difficultés de franchissement de certaines clôtures, et enfin le dérangement des troupeaux (de même que le dérangement de l'ornithologue par les troupeaux !) ; ce déplacement bien a sûr été laissé à l'appréciation individuelle des ornithologues "en charge", à condition que les 2 passages soient faits au même endroit à quelques mètres près.

Au total, **108 points ont été retenus**, offrant ainsi une couverture quasi-exhaustive d'environ 88% pour le milieu naturel ciblé (de surface estimée à **24 km²**, comparée à celle cumulée des 108 cercles de 250 mètres de rayon).

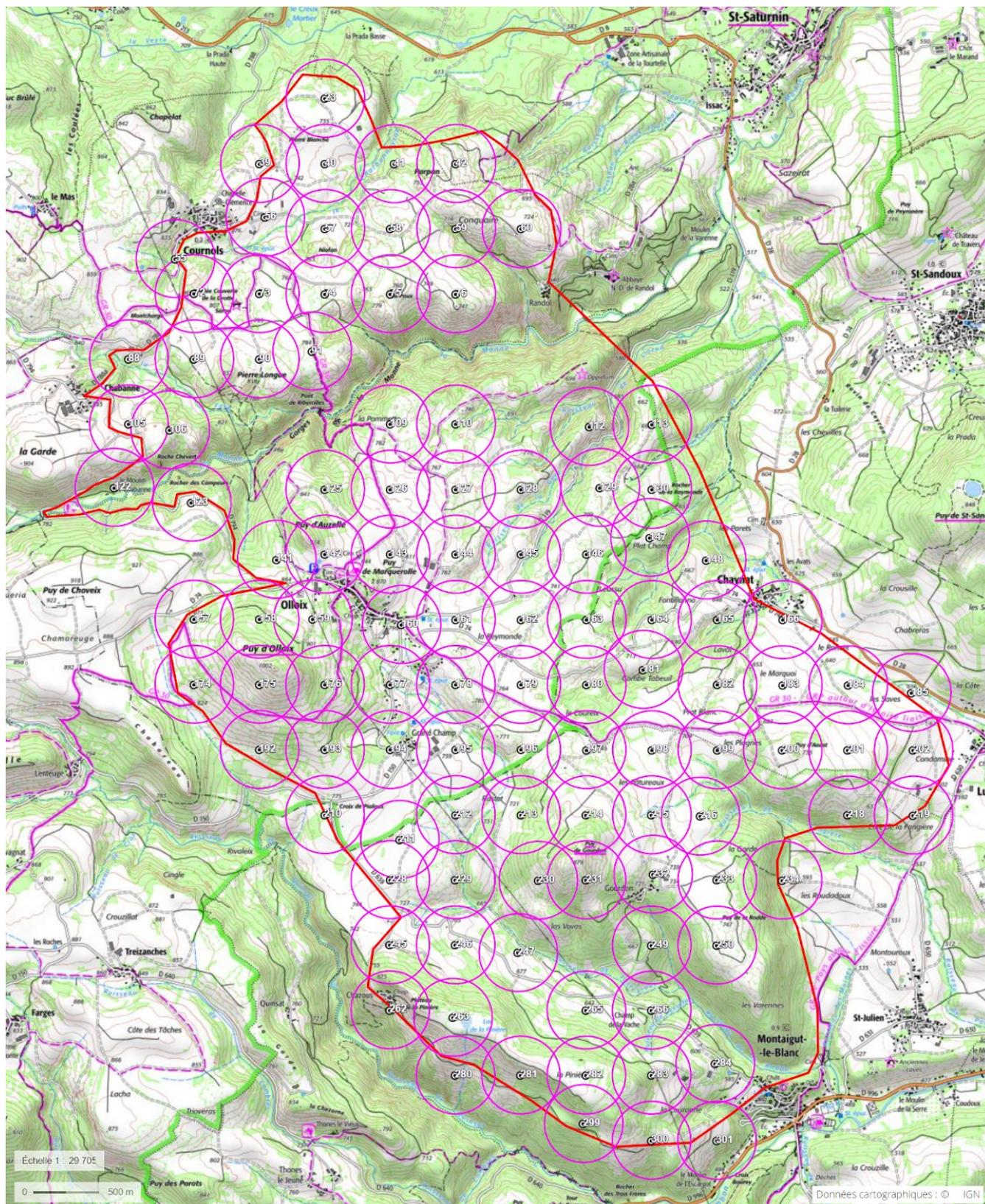


Figure 2-6 : Distribution géographique initiale des points d'écoute, avec cercles de rayon 300m (© geoportail.fr)

2. Deux passages dans la saison

Afin de tenter de tenir compte des décalages de périodes de reproduction entre les différentes espèces ciblées, et de l'enchaînement, pour certaines espèces, de 2 nidifications, associées à 2 cycles de détectabilité par les chants, nous avons choisi d'effectuer 2 passages sur chaque point, à protocole et ornithologue identique (voir ci-dessous) : un premier passage entre le 6 avril et le 15 mai, le second entre le 16 mai et le 15 juin. Ce n'est pas idéal bien sûr, mais il serait difficile d'aller plus loin pour mieux coller aux périodes de reproductions de chaque espèce ciblée : les inventaires doivent en effet être effectués par météo favorable, et par une majorité de bénévoles qui ne peuvent s'y livrer qu'en fin de semaine !

3. Tôt le matin par météo favorable

Les inventaires des points ont été faits en grande majorité (dans la mesure du possible) :

- par le même ornithologue (à protocole identique) pour les 2 passages,
- par météo favorable (pas de pluie, ni de vent gênant),
- tôt le matin, aux horaires habituels d'un STOC-EPS : entre 1h et 4h après le lever du jour (pour éviter le chorus des chants de l'aube, et utiliser la période optimale pour les chants).

4. Deux méthodes d'inventaire

Afin d'évaluer l'impact de l'utilisation de l'application "Naturalist" pour smartphone sur l'efficacité des inventaires par points d'écoute exploités en Distance Sampling, chaque point a en outre été inventorié selon 2 méthodes, par 2 ornithologues différents :

- méthode "classique" (dite "**Papyrus**"), déjà éprouvée, de relevé des contacts sur plan papier (copie de photo satellite en noir et blanc, sur feuille A4 ou A3, couvrant un rayon d'au moins 250m), à l'aide d'un stylo ou crayon, permettant de placer aussi précisément que possible chacun des contacts,
- méthode "smartphone" (dite "**Naturalist**"), de relevé des contacts via l'application Naturalist (Cf. Annexe 1), qui permet de positionner chaque contact sur un fond de photographie satellite / d'avion IGN OrthoPhoto 5m (peu précis, mais seul disponible hors réseau 4G à l'époque, suffisant pour l'enquête avec un peu d'habitude) ou Google (haute résolution, mais nécessite du réseau 4G).

5. Deux durées d'inventaire

Afin d'évaluer l'impact de la durée des points d'écoute avec relevé des distances sur les résultats de leur exploitation selon la technique du Distance Sampling, nous avons également fait le petit effort supplémentaire (5mn par point) de doubler la durée de **5 minutes** retenue par le protocole EPOC du MNHN, tout en "marquant" les contacts spécifiquement obtenus pendant les 5 premières minutes ; dans l'objectif d'obtenir 2 jeux de données distincts exploitables en Distance Sampling : le premier à partir des contacts des 5 premières minutes, le second à partir de tous contacts obtenus en **10 minutes**.

Concrètement,

- pour la méthode "Papyrus", il a "suffi" de changer de couleur de stylo ou crayon pour noter les contacts obtenus après les 5 premières minutes,
- pour la méthode "Naturalist", aucun effort supplémentaire n'a été nécessaire, puisque les données enregistrées sont automatiquement horodatées, et donc facilement "attribuables" a posteriori aux 5 premières minutes ou aux 5 suivantes).

6. Et quelques autres règles ...

Habituellement, l'estimation des populations d'oiseaux nicheurs cible plutôt les couples : pour les espèces à forte activité de chant, il est beaucoup plus aisé de repérer les mâles, et c'est à eux qu'on s'intéresse en priorité ; on approximera alors le nombre de couple par le nombre de mâles estimé. Pour les autres espèces, plus discrètes, ou moins aisées à "sexer" plus généralement, on ne peut guère que compter des individus ; l'estimation de leur nombre total donnera donc celle de la population entière, soit encore, en première approximation, la moitié du nombre de couples nicheurs. Enfin, les juvéniles de l'année ne doivent pas être inclus dans les calculs : il est important de les séparer des adultes lors du relevé.

Au total, par mesure de simplification lors des inventaires, nous avons ainsi choisi de compter, et donc de **noter séparément**, pour toutes les espèces :

- les **mâles adultes**,
- les **"autres" adultes** (femelles ou individus non sexés),
- les juvéniles.

Les oiseaux en vol n'ont pas été pris en compte dans les calculs, hormis pour quelques espèces au comportement nuptial particulier : Alouettes des Champs et Lulu, Verdier d'Europe (chant en vol).

Quelle que soit la méthode d'inventaire, les contacts sont obtenus **essentiellement "à l'oreille"** (chants majoritairement, cris aussi) ; d'autres le sont également "à la vue", mais sans utilisation des jumelles, sauf pour confirmer l'identification ou le sexe d'un oiseau repéré à l'oreille ou à l'œil nu (pas de "scan" aux jumelles !).

Afin de s'assurer autant que possible que la règle fondamentale du Distance Sampling "tous les oiseaux sont détectés à distance zéro" est respectée, l'observateur doit prendre soin, lors de son **arrivée** - forcément discrète - sur chaque **point d'écoute**, de bien **noter TOUS les oiseaux qui fuient**, même si c'est juste AVANT le début de la période de 5 ou 10mn d'écoute ! Pratiquement, en mode Naturalist, il ne doit pas démarrer le formulaire dans l'appli. avant d'être arrivé réellement au point d'écoute, ou tout au plus à quelques mètres, de manière à ne pas biaiser la trace GPS qui permettra plus tard de calculer automatiquement la géolocalisation de ce point. En mode Papyrus, c'est moins important, mais la durée d'écoute doit cependant bien rester proche de 5 ou 10mn !

7. En résumé

- 108 points d'écoutes répartis aléatoirement et régulièrement tous les 500m dans le secteur d'étude, en évitant les endroits ne correspondant pas au milieu ouvert cible,
- deux méthodes appliquées par 2 équipes indépendantes sur chaque point : "Papyrus" et "Naturalist",
- deux durées d'inventaires "imbriquées" sur chaque point : 5mn et 10mn,
- deux passages printaniers pour chaque point et chaque équipe : avant mi-mai (mais pas trop tôt), après mi-mai (mais pas trop tard),
- séparation des mâles adultes, des autres adultes, et des juvéniles,
- par météo favorable, entre 1h et 4h après le lever du jour.

Bilan des inventaires de terrain : quelques constats et analyses simples

Les petits aléas de la météo et des disponibilités des bénévoles impliqués n'ont pas grevé la **couverture** de la **zone** ciblée, finalement très **bonne**, voire **excellente** pour l'équipe "Papyrus" ; les points inventoriés ayant produits des données exploitables se répartissent ainsi (par rapport à l'objectif initial de 108 points) :

Tableau 3-1 : Nombre d'inventaires effectués par chaque équipe à chaque passage (comparés au plan initial)

Nombre de points	1er passage (avril)	2nd passage (mai)
Papyrus (9 observateurs)	104 (~96 %)	104 (~96 %)
Naturalist (8 observateurs)	96 (~89 %)	94 (~87 %)

Les **2 équipes d'observateurs**, de taille et d'expérience équivalentes sur ce genre d'inventaire, semblent à première vue **légèrement déséquilibrées** en ce qui concerne le **nombre d'espèces** notées, tous observateurs et tous points cumulés ; le doublement de la durée d'inventaire et l'avancée dans la saison ne font en plus qu'accentuer ce phénomène.

Tableau 3-2 : Nombre d'espèces contactées selon les durées et méthodes d'inventaire

Nombre d'espèces	1er passage (avril)			2nd passage (mai)		
	5 mn	10 mn	de 5 à 10 mn	5 mn	10 mn	de 5 à 10 mn
Papyrus	57	61	+7,0 %	55	59	+7,3 %
Naturalist	63	69	+9,5 %	63	72	+14,3 %
de Papyrus à Naturalist	+10,5 %	+13,1 %		+14,5 %	+22,0 %	

Mais l'analyse rapide des données (Cf. Annexe 2 : tableau des nombres de données par espèce et par méthode et durée d'inventaire) montre que cela n'a a priori **aucune influence sur les analyses** en Distance Sampling, car :

- aucune des espèces les plus communes et nombreuses, seules potentiellement "analysables" par la méthode du Distance Sampling, n'est concernée,
- les espèces non (ou faiblement) notées par l'une ou l'autre des équipes sont soit d'apparition très peu probable (effet "coup de chance" bien réparti entre les 2 équipes), soit intentionnellement ignorées par l'équipe Papyrus (c'est confirmé explicitement par un observateur au moins), car non ciblées par l'enquête (hérons, rapaces, hirondelles, martinets, moineaux, espèces non nicheuses, ... etc), mais tout de même notées par l'équipe Naturalist, probablement influencée par l'application Naturalist, qui insiste à chaque fois sur la notion de liste "complète" (équipe qui aurait donc peut-être pu être plus efficace en appliquant la même règle que l'équipe Papyrus ?).

... à l'exception notable de *Phylloscopus Bonelli* (Pap / Nat : 4 / 29 sur 5mn, 11 / 39 mâles sur 10mn => facteur 3 !), pour laquelle on verra que les données Naturalist sont exploitables en Distance Sampling, mais pas les données Papyrus (trop peu nombreuses).

A l'échelle individuelle (des observateurs), on retrouve bien sûr ce léger déséquilibre (les données sources sont les mêmes), mais on vérifie surtout l'évidence que plus on fait d'inventaires, plus on observe d'espèces ; c'est l'effet "prime au temps passé" (à probabilité de détection équivalente, on détecte plus d'espèces si l'on observe plus longtemps).

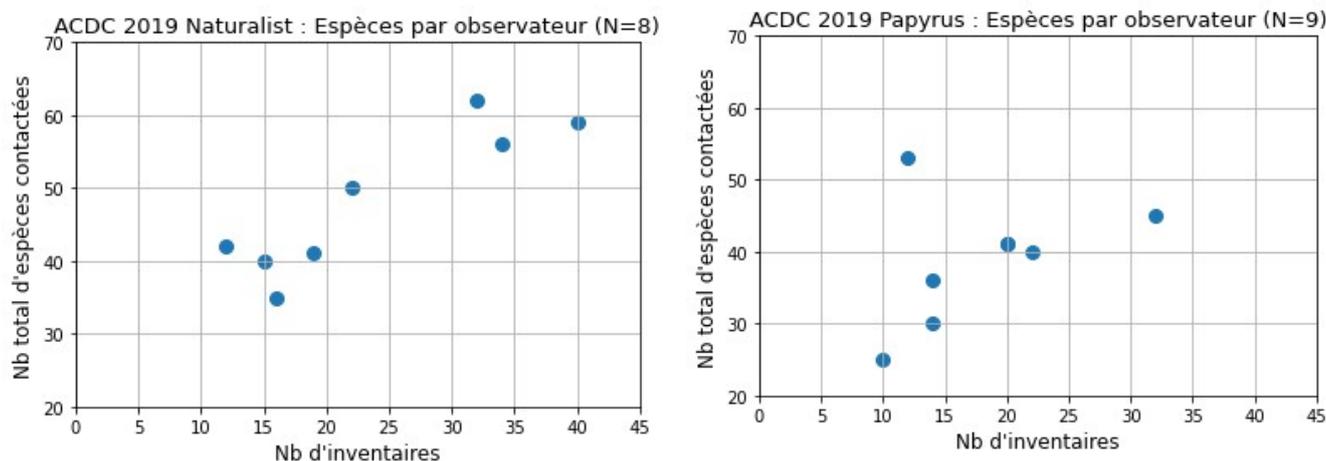


Figure 3-1 : Nombre total d'espèces contactées par observateur (chaque point des 2 graphiques correspond à un seul observateur) pour chaque méthode d'inventaire

De manière anecdotique également, indépendamment de la méthode de relevé, on constate que les inventaires de 10mn permettent de détecter 7 à 15% d'espèces en plus (en comparaison des inventaires de seulement 5mn) : là encore, c'est l'effet "prime au temps passé". Mais notons que le gain n'est pas énorme, ce qui montre bien que le "rendement" temporel des inventaires de 10mn n'est pas aussi bon que celui des inventaires de 5mn.

Ce déséquilibre sur le nombre d'espèces notées, sans aucune importance donc (hors *Phylloscopus Bonelli*), en croise un autre beaucoup plus notable et intéressant, et probablement lié cette fois à la **méthode de notation** des contacts sur le terrain (papier / appli. smartphone) : que ce soit pour l'ensemble des adultes contactés, ou seulement pour les mâles, on constate :

- un **léger déficit** de données "Naturalist" par rapport aux données "Papyrus" dans les **5 premières minutes**, et l'**opposé** dans les **5 minutes qui suivent** (dernière ligne des tableaux ci-dessous) ;
- un **surcroît de contacts** important (facteur 2) avec la méthode **Naturalist** si on compare les 5 premières minutes à la totalité des 10 minutes (colonnes "de 5 à 10mn" des tableaux ci-dessous).

Tableau 3-3 : Nombre d'individus adultes contactées selon les durées et méthodes d'inventaire

Nombre d'adultes	1er passage (avril)			2nd passage (mai)		
	5 mn	10 mn	de 5 à 10 mn	5 mn	10 mn	de 5 à 10 mn
Papyrus	1326	1791	+35,1 %	1493	1973	+32,2 %
Naturalist	1231	2095	+70,2 %	1301	2120	+63,0 %
de Papyrus à Naturalist	-7,2 %	+17,0 %		-12,9 %	+7,5 %	

Tableau 3-4 : Nombre de mâles adultes contactés selon les durées et méthodes d'inventaire

Nombre de mâles	1er passage (avril)			2nd passage (mai)		
	5 mn	10 mn	de 5 à 10 mn	5 mn	10 mn	de 5 à 10 mn
Papyrus	1195	1615	+35,2 %	1354	1783	+31,7 %
Naturalist	1024	1679	+64,0 %	1139	1790	+57,2 %
de Papyrus à Naturalist	-14,3 %	+4,0 %		-15,9 %	+0,4 %	

Cela semble bel et bien confirmer l'impression des observateurs d'un "débordement" supérieur (difficulté de noter tous les contacts) dans les 5 premières minutes avec la saisie par appli. smartphone ... et mettre en outre en lumière un probable biais de "rattrapage de saisie" après la 5ème minute ... des données "en excès" durant les 5 premières ! Par ailleurs, finalement, sur 10mn, en considérant uniquement les mâles, les 2 méthodes s'équilibrent, avec des effectifs très proches.

Nous avons évoqué le léger déséquilibre des équipes Naturalist et Papyrus en ce qui concerne le nombre d'espèces contactées ; lorsqu'on examine un autre indicateur, celui du nombre moyen de contacts par inventaire (= par passage sur chaque point) et par observateur, on constate aussi un léger déséquilibre global : si l'on ignore un observateur "extra-terrestre" côté Papyrus (moyenne proche de 35 contacts par inventaire de 5mn !), l'équipe Naturalist contacte en moyenne légèrement plus d'oiseaux sur 10mn que l'équipe Papyrus ; sur 5mn en revanche, c'est assez équilibré.

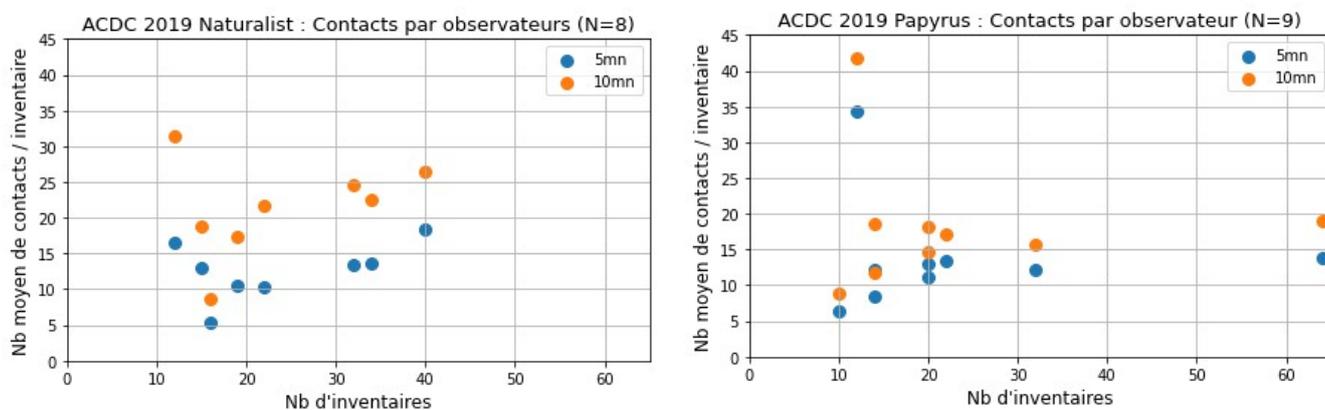


Figure 3-2 : Nombre moyen d'individus adultes contactés par inventaire de 5 et 10mn (chaque point des graphiques correspond à un seul observateur)

Et côté Naturalist comme côté Papyrus, ce sont les observateurs ayant obtenu les plus grands nombres moyens de contacts par inventaire qui ont effectué le plus d'inventaires ; l'explication ne serait donc pas là.

Peut-être s'agit probablement là encore d'espèces intentionnellement ignorées par l'équipe Papyrus, car non ciblées par l'enquête, alors que l'équipe Naturalist aurait été incitée par l'application à les noter quand même ?

Sur le sujet de la chronologie des relevés de contacts dans le temps à l'intérieur de chaque inventaire, on constate qu'en moyenne (N = 190 points), la **distance** des oiseaux notés **augmente avec le temps** (durant des points d'écoute), particulièrement fortement pendant les 2 premières minutes, et plus légèrement après : les observateurs se concentrant très probablement d'abord sur les oiseaux relativement proches.

N.B. Cette analyse temporelle n'a été faite que sur les relevés effectués via la **méthode "Naturalist"**, la seule pour laquelle on dispose de la minute précise de saisie de chaque donnée.

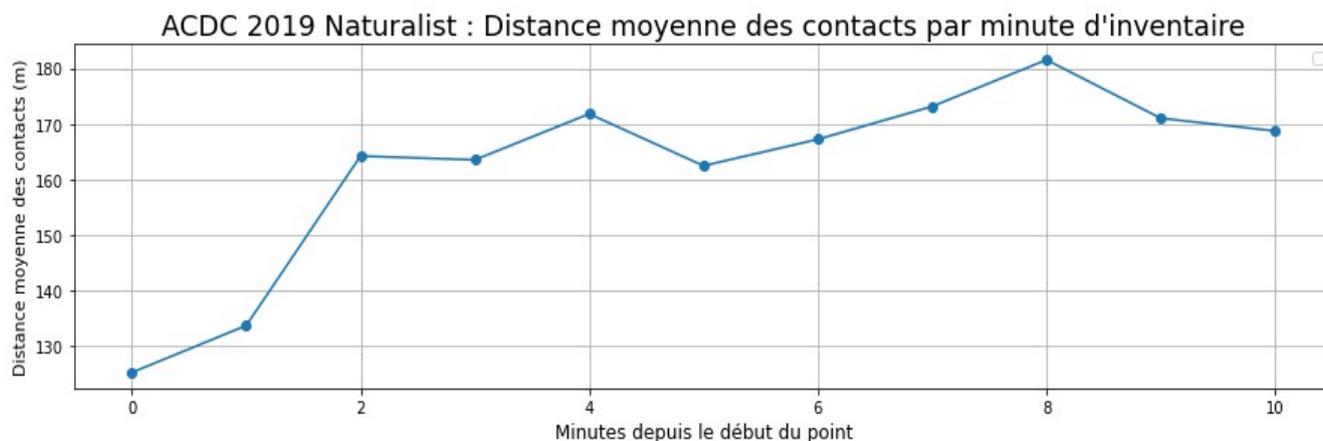


Figure 3-3 : *Distance moyenne des contacts par minute durant les inventaires (en abscisse, 0 = la 1ère minute, 1 = la 2ème ... etc)*

Et ceci est d'autant plus vrai que la quantité de données est beaucoup plus forte dans les 5 premières minutes !

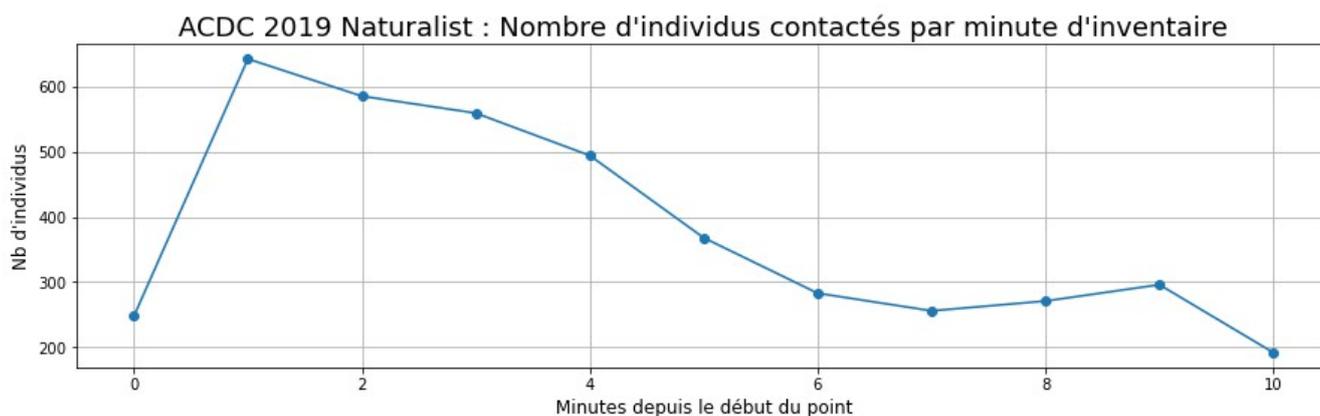


Figure 3-4 : *Nombre d'individus contactés par minute durant les inventaires (en abscisse, 0 = la 1ère minute, 1 = la 2ème ... etc)*

Enfin, sur la distribution des distances des contacts, on constate des différences importantes entre les méthodes Naturalist et Papyrus :

- **Naturalist :**
 - il y a un déficit relatif de données à courtes distances (moins de 10m, et à un degré moindre sous les 40m), avant comme après les 5 premières minutes d'inventaire,
 - comme déjà vu ci-dessus (Cf. Tableau 3-4), il y a une plus forte augmentation des nombres de contacts entre les 5 première minutes, et les 5 suivantes si l'on compare les

méthodes Naturalist et Papyrus ; mais on voit aussi que cette tendance est relativement indépendante de la distance (augmentation Naturalist toujours supérieure),

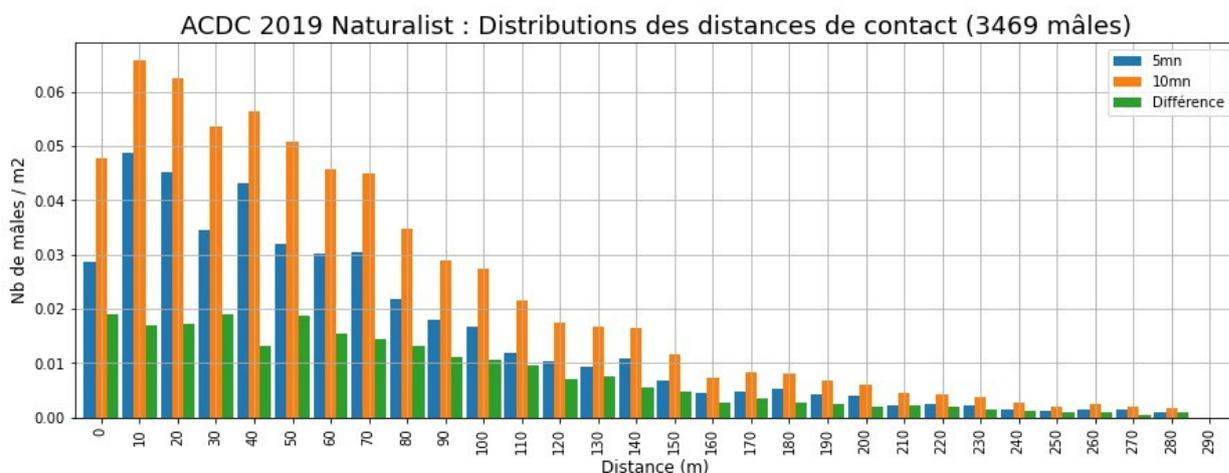


Figure 3-5 : *Naturalist : Distribution des distances de contacts (nombre de mâles par unité de surface concernée ; en abscisse, 0 correspond à la tranche de distance [0, 10m[, 10 à [10, 20m[, ... etc)*

- **Papyrus :** les distributions des distances de contacts pour les durées d'inventaires de 5mn et de 10mn sont très régulières et beaucoup plus cohérentes l'une avec l'autre (comparé à celles de la méthode Naturalist).

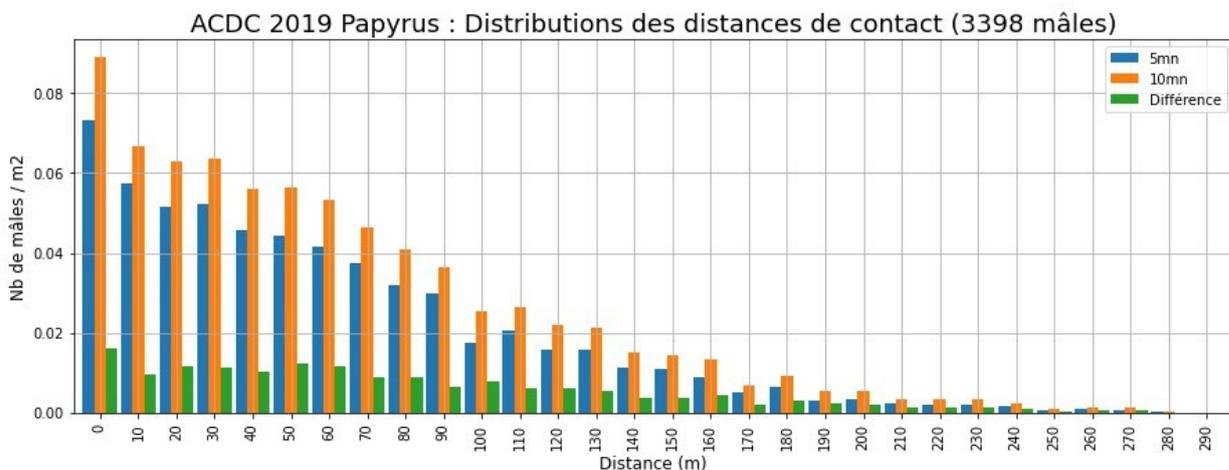


Figure 3-6 : *Papyrus : Distribution des distances de contacts (nbre de mâles par unité de surface concernée ; en abscisse, 0 correspond à la tranche de distance [0, 10m[, 10 à [10, 20m[, ... etc)*

Nous n'irons pas plus loin sur ces premiers constats à ce stade, mais on voit bien qu'il y a là matière à d'intéressantes analyses.

Choix des échantillons et méthode d'analyse

A. Outils logiciels et automatisation

Les analyses dont les résultats font l'objet du présent article ont toutes été faites à l'aide du logiciel open source "**pyaudisam**" en version 0.9.3 (Cf. Annexe 1) développé spécialement pour automatiser l'exécution de très nombreuses variantes d'analyses, optimiser les troncatures en distance, et aider au tri et à la sélection dans la masse des résultats produits ; pour chaque analyse individuelle, ce logiciel utilise le "moteur de calcul" MCDS V6 du logiciel Distance V7 sous Windows (Cf. Annexe 1), toujours en mode Conventional Distance Sampling (pas de co-variable) : de ce côté, aucune différence donc avec des analyses "manuelles" que nous menons habituellement avec le logiciel Distance.

Ce travail de développement pour l'automatisation a été entamé sur le constat que la quantité d'analyses à mener pour les combinaisons d'échantillons envisagées prenait des proportions rendant difficile, fastidieuse et sujette à trop d'erreurs de "manipulation" ... leur exécution "en manuel" avec le logiciel Distance :

- 30 espèces,
- 2 durées d'inventaires (5mn et 10mn),
- 3 "méthodes" d'inventaire (Naturalist, Papyrus, et la combinaison des 2, par fusion des 2 jeux de données),
- au strict minimum, une dizaine de combinaisons de paramètres de modélisation par analyse (modèles Half-Normal, Hazard-Rate et Uniform, séries d'ajustement Cosinus et Polynôme, plusieurs essais de troncatures à gauche et à droite, et/ou intervalles de découpage de l'espace des distance)

... soit un **minimum de 1800 analyses à effectuer et dépouiller** manuellement !

Le fonctionnement de "**pyaudisam**" peut être résumé simplement de la manière qui suit :

- Dans le **premier mode** nommé "**pré-analyse**",
 - en entrée, on lui fournit
 - l'ensemble des **données individualisées** pour la méthode choisie (espèce, 1er ou 2nd passage, présence / absence dans les 5 premières minutes, et dans les 5 suivantes, sexe et âge de l'individu, distance observateur - oiseau
 - la liste des **échantillons à pré-analyser** : les combinaisons espèce x durée d'inventaire x sexe et âge x passage
 - le logiciel effectue ensuite une série d'**analyses simples** pour chacun de ces échantillons, sans **aucune troncature**, en essayant les différents modèles et ajustements disponibles par ordre de préférence jusqu'à en trouver un qui donne un résultat,
 - en sortie, il produit un **tableau à 1 ligne par échantillon** (Cf. Annexe 4) présentant notamment des **histogramme effectifs / distances** (3 largeurs de tranches) et les **courbes de modélisations calculées par MCDS**, qui permettent d'évaluer visuellement la probabilité d'obtenir des densités exploitables et les troncatures éventuelles à prévoir,
- dans le **second mode** nommé "**analyse**",
 - en entrée, on lui fournit
 - le même ensemble choisi des **données individualisées** que pour les pré-analyses,
 - la liste des **échantillons à analyser** (a priori réduite des échantillons jugés inexploitables lors des pré-analyses),

- les **combinaisons de paramètres d'analyse** à couvrir pour chaque échantillon (modèle x série d'ajustement x troncature en distance à gauche x à droite x nombre de tranches de distance pour la modélisation ; N.B. un mode "automatique" existe pour ces 3 derniers paramètres, via une technique d'optimisation numérique qui détermine automatiquement les troncatures à utiliser),
- le logiciel enchaîne ensuite automatiquement les grandes étapes suivantes :
 - pour les combinaisons de paramètres "**à troncatures automatiques**", détermination automatique des distances de troncatures à gauche et à droite par une technique simple **d'optimisation** sur un **critère de "qualité combinée"** (fonction des indicateurs produits par MCDS et des paramètres d'analyses : Cf. Annexe 5 pour les détails) ;
 - exécution des **analyses** individuelles correspondant à chacune des combinaisons de paramètres (à troncatures maintenant toutes fixées) pour chaque échantillon (comme lors d'analyses manuelles via le logiciel Distance 7, c'est son "moteur" de calcul MCDS V6 qui est utilisé, en mode "Conventional Distance Sampling") ;
 - **classement et filtrage automatique** des résultats selon leur "qualité combinée", pour chaque échantillon, suivant une suite d'étapes proche de ce que ferait "manuellement" un analyste humain cherchant le "meilleur résultat" d'analyse parmi ceux correspondant à de multiples combinaisons de paramètres (Cf. Annexe 3 pour le détail des étapes) ;
- en sortie, il produit un rapport présentant les résultats de chaque analyse conservée lors du filtrage ci-dessus, entre autre sous la forme d'un **tableau de synthèse** trié par échantillon et dans l'ordre décroissant du critère de "qualité combinée" ;
- la **dernière étape, manuelle** celle-ci, consiste à examiner, pour chaque échantillon, les résultats des quelques "meilleures" analyses retenues, et à choisir laquelle conserver finalement : le filtrage automatique décrit ci-dessus n'étant pas parfait, il convient de **vérifier systématiquement** que la "**meilleure**" analyse proposée par l'automate est effectivement "meilleure" que les suivantes du classement ... et d'en choisir une autre si ce n'est pas le cas.

On notera l'importance, dans le processus de tri et filtrage automatique décrit ci-dessus, de l'**indicateur / critère de "qualité combinée"** calculé à partir des indicateurs produits par MCDS et des paramètres d'analyses (Cf. Annexe 5 pour tous les détails) : c'est lui qui concentre l'essentiel de "l'intelligence" de l'automate de tri et filtrage cherchant à imiter celle d'un analyste humain expérimenté.

B. Méthode d'analyse

1. Préparation des données de terrain

Préalablement aux analyses DS, les données de terrain ont dû être mises en forme :

- pour la méthode Papyrus, les distances observateur - oiseau ont été calculées à la main sur les plans papier (double-décimètre et règle de trois pour l'échelle), puis ensuite individualisées (passage de la donnée potentiellement multi-individus au contact individuel) via un code Python développé pour cela dans le passé,
- pour la méthode Naturalist, les distances observateur - oiseau ont été calculées automatiquement via un code Python également (à partir de la géolocalisation de chaque

donnée, et de celle de chaque point d'écoute), et ensuite individualisées comme les données Papyrus.

2. Pré-analyses "de défrichage"

Dans le but principal d'examiner les histogrammes de distance pour chacun des échantillons x durée d'inventaire x méthode d'inventaire disponibles à l'issue de l'enquête de terrain, on a d'abord effectué des pré-analyses automatiques (voir ci-dessus dans le chapitre "Outils logiciels") sur les échantillons présentant suffisamment de données pour avoir la moindre chance d'être exploitable en Distance Sampling.

Le seuil de sélection a ainsi été fixé à "**au moins 15 individus** (mâles et autres) **sur 5mn** pour l'une au moins des 2 méthodes d'inventaire" selon le raisonnement et les compromis suivants :

- nous utilisons habituellement un seuil autour de 20 individus (selon notre expérience lors des précédentes enquêtes Distance Sampling faites en Auvergne [GUÉLIN 2015] [GUÉLIN 2016] [GUÉLIN et al. 2018]),
- nous voulions faire des analyses DS sur les jeux de données de durée 10mn, mais aussi de durée 5mn,
- les jeux de durée 10mn sont approximativement de taille double de ceux de durée 5mn,
- pour trouver la limite en cette matière complexe, il faut la franchir pour ensuite être en capacité de s'en rendre compte (échec des analyses DS, ou résultats de trop mauvaise qualité statistique),
- nous voulions un règle simple.

Par ailleurs, nous avons naturellement choisi comme stratégie de choix des modèles x séries d'ajustement notre ordre de préférence habituel : d'abord Half-Normal, ensuite Uniform, et enfin Hazard-Rate (le modèle Negative Exponential, disponible dans MCDS, a été écarté pour son inadéquation manifeste au sujet d'étude, puisque de fonction "Probabilité de détection" à dérivée non nulle en 0) ; pour chaque modèle, les séries Cosine ont d'abord été testées, et ensuite Polynomial.

C'est à dire finalement l'ordre suivant : Half-Normal & Cosine, Half-Normal & Polynomial, Uniform & Cosine, Uniform & Polynomial, Hazard-Rate & Cosine, Hazard-Rate & Polynomial.

3 - Choix des échantillons pour les analyses

Il serait tentant de s'en remettre strictement aux résultats des aux pré-analyses, en ne retenant, en toute logique, que les échantillons x durée d'inventaire x méthode d'inventaire pour lesquels on a obtenu des résultats (non échec des pré-analyses) et en outre au moins corrects (critère du Khi2, de Kolmogorov-Smirnov, variabilité DCv des densités estimées < 30%), statistiquement et graphiquement (allure des courbes modélisées comparées aux histogrammes en distances).

Mais ce serait oublier que les pré-analyses n'introduisent pas de troncature à gauche et/ou à droite : les pré-analyses peuvent échouer complètement pour un échantillon x méthode d'inventaire (aucun modèle ou ajustement ne convient), ou bien donner des résultats incorrects statistiquement ou graphiquement, alors que dans certains cas, l'application d'une troncature à gauche et/ou à droite peut résoudre simplement un biais dans les données.

En pratique, nous avons donc appliqué une nouvelle fois une règle à seuil : "**au moins 30 individus sur 10mn** pour l'une au moins des 2 méthodes d'inventaire" (le raisonnement et les compromis sont les mêmes que pour les pré-analyses) :

- nous utilisons habituellement un seuil autour de 20 individus (selon notre expérience lors des précédentes enquêtes Distance Sampling faites en Auvergne [GUÉLIN 2015] [GUÉLIN 2016] [GUÉLIN et al. 2018]),
- nous voulons faire des analyses DS sur les jeux de données de durée 10mn, mais aussi de durée 5mn,
- nous acceptons un peu de "déchet" pour certains échantillons (échec des analyses DS, ou résultats de trop mauvaise qualité statistique) : pour trouver la limite, il faut parfois la franchir !
- les jeux de durée 10mn sont approximativement de taille double de ceux de durée 5mn,
- nous voulons un règle simple.

Ensuite, nous avons appliqué une deuxième sélection, pour tenir compte du milieu naturel ciblé cette fois : nous n'avons globalement gardé que les **espèces estimées communes** dans le milieu naturel cible ou **plus ou moins typiques** de ce milieu, et à territoire de taille raisonnable (difficulté de l'estimation de la distance d'un contact ... croissante avec la distance).

Dans les espèces à effectif suffisant (première règle), nous avons ainsi écarté volontairement *Passer domesticus*, *Streptopelia decaocto* (trop liés à l'habitat humain), *Corvus corone*, *Sturnus vulgaris*, *Garrulus glandarius* et *Troglodytes troglodytes* (trop "forestiers" ou à trop grand territoire).

Nous avons en revanche conservé ***Cuculus canorus*** et ***Upupa epops***, malgré leur territoire habituellement assez étendus : il nous a paru intéressant d'essayer tout de même la méthode sur ces espèces en déclin.

Quatre espèces que nous aurions en toute logique aimé pouvoir analyser, vu le milieu naturel ciblé et les rapports historiques de suivi du secteur [DULPHY 1994] [DULPHY 2012], n'ont pas été suffisamment contactées selon le critère d'abondance énoncé ci-dessus : *Hippolais polyglotta*, *Sylvia borin*, *Picus viridis*, *Carduelis carduelis*.

Oriolus oriolus, pourtant "hors des clous" pour le faible nombre de contacts sur le seul passage considéré (le second) a également été conservé, par erreur en quelque sorte, suite à l'oubli que seul le second passage devait être considéré pour ce migrateur relativement tardif ; mais on verra que la qualité statistique des estimations obtenues n'est pas si mauvaise dans tous les cas, et qu'il n'est pas forcément inutile de conserver des cas limites dans le jeu de données.

En ce qui concerne la période temporelle retenue pour l'exploitation des données, nous l'avons logiquement restreinte au **second passage pour les migrants "tardifs"** de la liste (*Oriolus oriolus*, *Streptopelia turtur* et *Lanius collurio*), pour tenter d'éviter de prendre en compte les oiseaux en transit migratoire ; pour toutes les autres espèces, les données des 2 passages ont été utilisées ... ce qui laisse matière à discussions pour certaines espèces chantant en migration, comme *Sylvia atricapilla*, *Phylloscopus collybita*, ... etc, mais reste assez cohérent avec les approches adoptées depuis longtemps pour le STOC-EPS, et plus récemment pour les EPOC et EPOC-OdF (Cf. Annexe 1).

Quant à la cible des estimations de densités - couples nicheurs ou individus présents en période de reproduction - nous avons utilisé :

- uniquement les données de **mâles adultes** pour les espèces à chant régulier, facile à détecter et spécifique des mâles (objectif : la densité de mâles, indicateur approximatif direct de la densité de couple nicheurs),
- pour *Jynx torquilla*, dont les 2 sexes chantent de manières assez difficilement séparables (pour des équipes composées d'une majorité de non-experts auditifs de l'espèce), nous avons également exploité les données notées "mâle chanteur", mais cette fois en ciblant une densité en nombre d'individus, bien conscients de cette difficulté,
- toutes les données d'adultes (mâles, femelles et indéterminés) pour les autres espèces.

Finalement, voici le tableau des espèces et types d'effectifs retenus pour les analyses :

- la colonne Passage indique si l'on a retenu la totalité des individus contactés lors de l'étude (a+b = les 2 passages réunis) ou seulement ceux d'un des passages (ici, b = uniquement le 2ème passage, pour les espèces à arrivée tardive) ; N.B. Lorsque les 2 passages ont été exploités, les analyses ont bien été menées en tenant compte du fait que chaque point avait été inventorié 2 fois (à environ 1 mois d'intervalle), et donc en leur attribuant un effort de 2 (contre 1 lorsqu'un seul passage a été exploité).
- la colonne Adulte indique si l'on a retenu tous les adultes contactés (m+a = mâles et autres) ou seulement les mâles (m),
- en regard, les effectifs des échantillons correspondants selon la méthode et la durée d'inventaire.

Tableau 4-1 : Echantillons retenus pour les analyses

Espèce	Passage	Adulte	Nombre d'individus contactés (à l'ouïe ou la vue)			
			Papyrus		Naturalist	
			5mn	10mn ▼	5mn	10mn
<i>Sylvia atricapilla</i>	a+b	m	340	427	270	403
<i>Turdus merula</i>	a+b	m	323	422	234	405
<i>Sylvia communis</i>	a+b	m	178	244	126	201
<i>Alauda arvensis</i>	a+b	m	170	217	157	236
<i>Luscinia megarhynchos</i>	a+b	m	159	213	108	158
<i>Cuculus canorus</i>	a+b	m	121	161	93	142
<i>Fringilla coelebs</i>	a+b	m	136	160	86	147
<i>Columba palumbus</i>	a+b	m	101	156	109	168
<i>Phylloscopus collybita</i>	a+b	m	100	131	84	135
<i>Parus major</i>	a+b	m	83	130	67	126
<i>Lullula arborea</i>	a+b	m	72	105	69	119
<i>Lanius collurio</i>	b	m	67	95	43	83
<i>Emberiza cirrus</i>	a+b	m	66	87	72	115
<i>Turdus philomelos</i>	a+b	m	45	54	29	52
<i>Erithacus rubecula</i>	a+b	m	43	52	50	85

<i>Carduelis chloris</i>	a+b	m	40	51	18	34
<i>Streptopelia turtur</i>	b	m	47	51	19	31
<i>Emberiza citrinella</i>	a+b	m	35	50	44	66
<i>Dendrocopos major</i>	a+b	m+a	30	48	24	49
<i>Prunella modularis</i>	a+b	m	31	47	21	47
<i>Saxicola rubicola</i>	a+b	m	36	46	42	68
<i>Carduelis cannabina</i>	a+b	m	27	41	25	39
<i>Cyanistes caeruleus</i>	a+b	m	18	33	33	57
<i>Turdus viscivorus</i>	a+b	m	24	32	30	41
<i>Anthus trivialis</i>	a+b	m	26	32	19	30
<i>Upupa epops</i>	a+b	m	24	32	16	23
<i>Emberiza calandra</i>	a+b	m	26	29	32	45
<i>Jynx torquilla</i>	a+b	m	17	27	21	31
<i>Oriolus oriolus</i>	b	m	10	16	4	11
<i>Phylloscopus bonelli</i>	a+b	m	4	11	29	37

4. Sélection des variantes d'analyse à effectuer et des paramètres d'optimisation

Nous avons choisi de mener toutes les analyses correspondant aux variantes suivantes :

- 3 "méthodes" d'inventaire : Naturalist (Nat), Papyrus (Pap), et la combinaison des 2, par fusion des 2 jeux de données (Nat+Pap),
- 2 durées d'inventaires : 5mn et 10mn,
- 30 espèces : pour chacune,
 - 1 seul période temporelle : 1er et 2ème passage (a+b) ou seulement 2ème passage (b),
 - 1 seul type d'effectif cible : mâles seuls (m), ou tous les adultes (m+a),
- 3 modèles Half-Normal, Hazard-Rate et Uniform (le modèle Negative Exponential a été écarté pour les mêmes raisons que lors des pré-analyses),
- 2 séries d'ajustement (des modèles) : Cosine et Polynomial,
- troncatures à gauche et/ou à droite (indépendamment),
 - d'abord, toutes les combinaisons de (soit $2 \times 2 = 4$ variantes):
 - pas de troncature,
 - calculé automatiquement (par optimisation) de celle qui donne le meilleur indicateur "qualité combinée" (tous autres paramètres fixés par ailleurs),
 - ensuite toutes les combinaisons de (soit 1×3 variantes),
 - 20m à gauche,
 - 100m, 200m, 300m à droite,
 - enfin, pour certaines espèces ...
 - à chants de grande portée, troncature à droite en conséquence (*Cuculus canorus* : 1000m, *Turdus viscivorus*, *Columba palumbus*, *Upupa epops* : 500m)
 - à notoire défaut de données à courte distance, troncature à gauche en conséquence (*Alauda arvensis*, *Prunella modularis* : 50m),
 - très difficiles à localiser au loin (*Alauda arvensis* : 300m).
- intervalles de découpage de l'espace des distance (2 variantes):
 - soit à la décision de MCDS,
 - soit calculé automatiquement (par optimisation), retenant celle qui donne le meilleur indicateur "qualité combinée" (tous autres paramètres fixés par ailleurs).

N.B. A partir du moment où on active le calcul automatique des troncatures (par optimisation), il est probablement illusoire de "forcer" des variantes d'analyses fixant ces troncatures à des valeurs données, puisque l'optimiseur explore intégralement les espaces de distances possibles (dans les limites de taux d'outliers fixés) et qu'ensuite, les analyses dont les résultats ne sont pas d'une qualité "combinée" suffisante sont éliminées (pour mémoire, l'indicateur de qualité "combinée" intègre le taux d'outliers éliminés par les troncatures).

Pour le calcul automatique des distances de troncature à gauche et à droite, et du nombre de tranches de distance pour la modélisation MCDS, les paramètres suivants ont été utilisés :

- critère cible : maximiser l'indicateur de qualité "combinée" (Cf. Annexe 5),
- limites sur la troncature à gauche : au minimum, 0 m ; au maximum, 1er quantile à 5% = distance en dessous de laquelle il y a 5% de contacts (autrement dit, pas plus de 5% "d'outliers" éliminés par la troncature à gauche ; par "outliers" on entend ici des données qu'on élimine avant analyse en considérant qu'elles sont aberrantes d'une manière ou d'une autre, et donc détériorent la qualité du jeu de données),
- limites sur la troncature à droite : au maximum, la distance du contact le plus éloigné ; au minimum, la plus grande des 2 distances suivantes : celle obtenue par application de la règle de Tuckey à droite (quartile à 75% + 1.5*(quartile à 75% - quartile à 25%)), celle du dernier quantile à 5% = distance au dessus de laquelle il y a 5% de contacts (en tout cas, pas plus de 5% d'outliers éliminés par la troncature à droite),
- limites sur le nombre de tranches de distances pour le fitting du modèle par MCDS : entre $\frac{2}{3}$ et $\frac{3}{2}$ de la racine carrée du nombre de mesures de distances (= d'individus contactés),
- chaque optimisation est faite 2 fois, et seule le meilleur résultat suivant le critère cible est conservé (car il y a un part d'aléatoire dans la stratégie d'exploration de l'espace des possibles pour les troncatures et nombre de tranches de distance pour le fitting),
- l'optimiseur ne fait pas plus de 200 essais d'analyse (soit il "converge" avant, soit il s'arrête et rend son meilleur résultat obtenu),
- l'algorithme d'optimisation utilisé est SRACOS (Cf. Annexe 1).

5. Exécution des analyses

Au total, pour chacune des 3 "méthodes" d'inventaire,

- ce sont donc environ 5500 analyses qui ont été exécutées au moyen du logiciel pyaudisam (une fois éliminées les variantes en doubles),
- parmi lesquelles environ 2500 nécessitant au préalable des optimisations pour calculer automatiquement les distances de troncatures et "tranches de distance pour modélisation" (optimisations impliquant environ 850 000 analyses dédiées, selon les paramètres d'optimisations choisis).

Sur le plan pratique, sur un micro-ordinateur portable récent assez haut de gamme Core i7 à 6 cœurs "hyper-threading" sous Windows 10, les calculs ont duré, pour chacune des 3 "méthodes" d'inventaire :

- entre 16h30 et 18h pour les analyses d'optimisation des troncatures (les 850 000),
- entre 4 et 5mn pour les analyses finales, une fois les troncatures et "tranches de distance pour modélisation" calculées par optimisation (les 5500).

6. Filtrage des meilleures analyses et génération des rapports

Cela a également été fait automatiquement avec pyaudisam, pour les 3 "méthodes" d'inventaire (étapes filtrage et rapports), avec les **paramètres de filtrage** suivants :

- pour les analyses à troncatures strictement identiques, on n'a conservé que les **N = 3** ayant les meilleurs AIC,
- on n'a conservé que les analyses "tronquées en distance" à moins de **X = 7,5%** (d'outliers en somme),
- seules on été conservées les analyses ayant une valeur de critères de "qualité combinée" supérieure ou égale à **Y = 0,2** (rappel : c'est un indicateur normalisé dans l'intervalle [0, 1],
- enfin, seules les **P = 12** (au plus) meilleures analyses restantes ont été conservées, pour faciliter l'étape manuelle suivante.

7. Vérification et correction manuelle de la pré-sélection automatique

Cette dernière étape est nécessaire pour vérifier la qualité de la (pré-)sélection effectuée automatiquement par le logiciel, et la corriger au besoin (voir ci-dessus les critères utilisés pour ce jugement).

Nous n'avons été amené à faire des corrections que dans 2 cas :

- *Oriolus oriolus*, Papyrus, 10mn : il y a clairement trop peu de données et l'histogramme des distance est déséquilibré ; le logiciel a du mal à décider : les indicateurs statistiques, au demeurant très proches, ne désignent pas la meilleure analyse ; nous en choisissons la 4ème, à la courbe modélisée plus conforme à la forme de l'histogramme,
- *Carduelis chloris*, Naturalist, 5mn : là encore, assez peu de donnée (quoiqu'un peu plus) : la 4ème modélisation (Half-Normal) est clairement meilleure au vue de l'histogramme que la première modélisation (Uniform) choisie par le logiciel, même si les indicateurs statistiques, au demeurant très proches, disent le "contraire".

Ces 2 cas de difficulté de l'automatisme à répondre au mieux sont très probablement liés au faible nombre de données des jeux correspondants (même si quelques autres échantillons aussi dans ce cas ont été mieux classés automatiquement). Mais ils pointent probablement aussi la nécessité de chercher à améliorer le classement pour ces cas limites.

Pour tous les autres échantillons et méthodes d'analyse, la sélection automatique a été conservée telle quelle, après vérification des critères visuels habituels, ou constat que de toute manière, les densités et fourchettes estimées étaient équivalentes. Soit environ **99% de bon classement automatique** (178 / 180).

Résultats

A - Valeurs et qualité des densités estimées

Le tableau 5-1 suivant donne les résultats obtenus pour chaque échantillon, chaque méthode et chaque durée d'inventaire :

- **l'estimation de densité** en gras dans les colonnes "Estim", précédée de la borne inférieure de l'intervalle de confiance à 95% (colonne "Min"), et suivie de sa borne supérieure (colonne "Max") (unité = nombre de mâles (m) ou d'individus (m+a) **par km²**) ; elle est de plus surlignée en nuances de couleurs indiquant la **dispersion** des estimations de densités au sein des 4 meilleures analyses, sur une échelle correspondante allant du rouge (énorme dispersion) au vert (pas de dispersion), en passant par l'orange et le jaune,
- **l'indicateur de qualité** est dans les colonnes "Quali" : c'est tout simplement l'indicateur de "**qualité combinée**" précédemment décrit (**sans unité, dans l'intervalle [0, 1]**, 0 voulant dire "affreusement mauvais", et 1 "excellentissime", surligné en nuances de couleurs sur une échelle correspondante allant du rouge au vert en passant par l'orange et le jaune).

Sur la "**dispersion**" au sein des densités estimées pour les 4 meilleures analyses (la meilleure et les 3 suivantes), précisons les choses :

- elle est calculée selon la méthode suivante : $2 * \text{écart-type}(\text{densités}) / \text{moyenne}(\text{densités})$ (le facteur 2 correspond à un intervalle de confiance à 95% dans l'hypothèse - non vérifiée - d'une distribution normale des densités ; un facteur de 1 correspondrait à un intervalle de confiance à 68%, et de 3 à 99%),
- puis multipliée par 100, pour se ramener à un **% de la moyenne des densités**,
- en raisonnant de manière similaire à ce que l'on pratique habituellement avec le "coefficient de variation de la densité" (DCv) calculé par MCDS pour chaque analyse, on considère qu'**au delà de 30%**, il y a une dispersion qui pose problème,
- ce que traduit l'échelle de couleur appliquée au fond des colonnes "Estim", qui passe à l'orange aux alentours de 30%,
- mais (pour 1 échantillon, une durée et une méthode d'inventaire donnés) **un indicateur de dispersion rouge n'est absolument pas bloquant** : il rappelle seulement que les estimations de densités des 4 meilleures analyses étaient plutôt dispersées, et donc que l'analyse retenue à l'étape 7 (Cf. Choix des échantillons et méthode d'analyse) l'a été à l'issue d'une vérification plus attentive et d'une éventuelle correction "manuelle", puisqu'il y avait là "hésitation" de l'automatisme entre des modélisations donnant des résultats différents, mais de qualités combinées très proches ; ce n'est qu'un **problème potentiel**, dans la mesure où l'analyste humain ne prendrait peut-être pas 2 fois de suite la même décision de correction (ou pas) de la sélection automatique si ses critères de choix ne sont parfaitement objectivés, ou bien encore où 2 analystes humains différents n'auraient peut-être pas pris la même décision ...

Sur l'indicateur de **qualité "combinée"**, l'examen détaillé des rapports d'analyse (tableaux d'indicateurs statistiques et de résultats, graphiques de modélisation de la probabilité de détection, ... etc) montre qu'il n'y a concrètement jamais de véritable problème de qualité au dessus des environs de **0,45 ou 0,50**.

Comme annoncé, en conformité avec leur indicateur de qualité de valeur nulle, les densités estimées pour Oriolus oriolus Naturalist 5mn et Phylloscopus Bonelli Papyrus 5mn - barrées dans le tableau - sont à ignorer.

Tableau 5-1 : Estimations de densités obtenues, avec bornes de l'intervalle de confiance à 95% + indicateurs de qualité et de dispersion

Echantillon			Durée 5mn												Durée 10mn											
Espèce	Passage	Adulte	Nat				Pap				Nat+Pap				Nat				Pap				Nat+Pap			
			Min	Estim	Max	Quali	Min	Estim	Max	Quali	Min	Estim	Max	Quali	Min	Estim	Max	Quali	Min	Estim	Max	Quali	Min	Estim	Max	Quali
Sylvia atricapilla	a+b	m	24,03	30,51	38,74	0,76	35,09	46,06	60,47	0,94	27,68	33,09	39,55	0,91	30,71	37,89	46,74	0,66	41,45	51,82	64,79	0,95	33,79	39,46	46,10	0,92
Turdus merula	a+b	m	12,06	15,34	19,51	0,72	30,27	36,65	44,37	0,96	24,67	29,91	36,26	0,95	25,30	30,75	37,37	0,86	39,25	45,60	52,97	0,98	36,52	42,17	48,70	0,96
Alauda arvensis	a+b	m	7,82	11,98	18,34	0,85	11,40	14,58	18,65	0,93	9,80	12,64	16,31	0,97	11,45	15,96	22,25	0,90	13,17	16,37	20,34	0,93	12,10	15,21	19,12	0,97
Sylvia communis	a+b	m	31,40	44,31	62,53	0,90	38,64	52,04	70,09	0,93	34,17	43,07	54,29	0,96	40,99	54,34	72,03	0,93	44,08	57,88	76,01	0,94	43,93	54,06	66,52	0,96
Columba palumbus	a+b	m	1,91	2,78	4,03	0,90	4,35	6,26	9,00	0,90	3,01	3,87	4,98	0,95	2,55	3,32	4,31	0,95	6,35	8,10	10,33	0,92	4,48	5,46	6,66	0,96
Luscinia megarhynchos	a+b	m	3,98	5,76	8,36	0,90	9,28	11,52	14,30	0,94	7,49	9,24	11,40	0,96	4,98	6,79	9,24	0,92	12,76	16,11	20,33	0,97	9,42	11,41	13,81	0,96
Cuculus canorus	a+b	m	0,48	0,65	0,88	0,95	0,89	1,22	1,67	0,82	0,97	1,23	1,56	0,86	0,71	0,90	1,14	0,98	1,36	1,77	2,31	0,82	1,15	1,48	1,90	0,82
Fringilla coelebs	a+b	m	4,25	5,96	8,38	0,88	12,66	15,51	19,01	0,91	7,18	8,97	11,21	0,92	7,02	9,16	11,95	0,91	14,33	17,65	21,74	0,92	9,04	10,97	13,32	0,92
Phylloscopus collybita	a+b	m	5,26	7,22	9,90	0,89	9,78	13,22	17,87	0,82	9,25	11,64	14,66	0,98	8,39	10,91	14,18	0,91	12,74	16,60	21,62	0,94	13,47	16,53	20,28	0,98
Parus major	a+b	m	5,14	7,41	10,69	0,79	9,77	14,04	20,19	0,80	8,90	12,64	17,93	0,79	10,17	13,10	16,89	0,96	14,12	19,31	26,41	0,85	12,79	16,63	21,62	0,84
Lullula arborea	a+b	m	4,07	6,20	9,44	0,83	5,13	7,27	10,30	0,88	5,56	7,40	9,84	0,89	5,87	7,99	10,88	0,95	7,22	9,80	13,30	0,79	7,09	9,45	12,59	0,91
Emberiza cirulus	a+b	m	7,35	11,28	17,31	0,84	6,21	8,93	12,83	0,90	5,88	8,17	11,35	0,87	12,10	15,76	20,52	0,90	7,46	10,22	13,99	0,93	10,31	12,68	15,58	0,88
Erithacus rubecula	a+b	m	4,99	7,91	12,54	0,82	10,42	19,33	35,87	0,70	7,27	11,42	17,93	0,85	7,79	11,16	15,98	0,90	9,78	16,68	28,45	0,78	9,81	14,40	21,13	0,87
Saxicola rubicola	a+b	m	8,76	14,22	23,10	0,81	4,37	7,32	12,28	0,81	6,96	10,06	14,53	0,90	11,45	16,41	23,53	0,91	5,28	7,71	11,27	0,88	8,95	12,23	16,73	0,95
Emberiza citrinella	a+b	m	3,65	5,76	9,09	0,84	3,24	5,43	9,10	0,79	3,79	5,54	8,11	0,86	4,90	7,16	10,47	0,87	4,45	7,01	11,02	0,87	5,30	7,33	10,15	0,92
Cyanistes caeruleus	a+b	m	5,30	9,17	15,87	0,78	4,41	8,58	16,68	0,63	5,85	10,20	17,77	0,76	9,98	14,99	22,52	0,87	8,29	13,97	23,52	0,72	10,34	15,02	21,83	0,81
Turdus philomelos	a+b	m	0,92	1,38	2,07	0,87	1,28	1,86	2,70	0,88	1,12	1,48	1,96	0,92	1,44	2,15	3,22	0,85	1,64	2,26	3,10	0,89	1,73	2,19	2,76	0,94
Emberiza calandra	a+b	m	2,17	3,48	5,58	0,83	1,82	3,63	7,23	0,63	2,02	3,18	4,99	0,85	2,51	3,96	6,23	0,85	1,75	3,50	7,00	0,62	2,20	3,30	4,95	0,87
Turdus viscivorus	a+b	m	0,44	0,68	1,08	0,83	0,52	0,90	1,57	0,66	0,53	0,79	1,18	0,88	0,62	0,96	1,48	0,88	0,88	1,54	2,68	0,71	0,89	1,28	1,84	0,88
Prunella modularis	a+b	m	1,19	2,23	4,17	0,67	2,19	3,51	5,62	0,80	2,16	3,23	4,83	0,88	3,05	4,46	6,52	0,80	2,95	4,40	6,57	0,79	3,29	4,59	6,39	0,86
Phylloscopus bonelli	a+b	m	0,87	1,51	2,62	0,77	0,65	0,92	1,94	0,90	0,53	0,86	1,41	0,81	0,83	1,39	2,35	0,77	0,34	0,67	1,33	0,50	0,78	1,25	2,00	0,81
Carduelis cannabina	a+b	m	3,69	6,78	12,44	0,59	5,74	8,73	13,27	0,86	6,40	9,19	13,19	0,85	6,40	10,38	16,83	0,72	9,77	15,52	24,66	0,86	6,94	9,92	14,17	0,87
Jynx torquilla	a+b	m	0,56	1,03	1,89	0,68	0,77	1,30	2,21	0,74	1,30	2,12	3,44	0,82	1,04	1,71	2,79	0,60	0,99	1,63	2,66	0,80	1,31	1,94	2,87	0,77
Carduelis chloris	a+b	m	1,13	2,46	5,35	0,41	2,96	4,72	7,51	0,86	2,36	3,58	5,42	0,90	1,62	2,52	3,92	0,81	3,70	5,59	8,44	0,90	3,05	4,30	6,06	0,94
Anthus trivialis	a+b	m	0,52	0,92	1,62	0,70	1,91	3,43	6,16	0,72	1,16	1,77	2,71	0,86	0,78	1,34	2,29	0,77	2,03	3,23	5,13	0,83	1,58	2,34	3,44	0,88
Upupa epops	a+b	m	0,04	0,08	0,15	0,61	0,25	0,41	0,66	0,76	0,17	0,29	0,50	0,80	0,07	0,13	0,25	0,68	0,33	0,52	0,82	0,78	0,28	0,44	0,69	0,82
Lanius collurio	b	m	10,12	17,42	29,98	0,72	13,53	19,56	28,27	0,87	15,54	23,20	34,62	0,87	16,72	23,47	32,95	0,92	20,19	27,42	37,24	0,87	19,87	25,52	32,77	0,96
Oriolus oriolus	b	m	0,06	0,06	1,91	0,00	0,27	0,57	1,20	0,52	0,27	0,48	0,84	0,71	0,29	0,63	1,36	0,47	0,34	0,59	1,04	0,48	0,29	0,48	0,81	0,76
Streptopelia turtur	b	m	0,45	0,89	1,79	0,54	2,97	4,38	6,47	0,88	1,66	2,32	3,23	0,78	0,87	1,48	2,52	0,74	3,32	4,76	6,84	0,89	2,53	3,46	4,72	0,92
Dendrocopos major	a+b	m+a	0,69	1,50	3,25	0,38	0,70	1,21	2,11	0,67	1,05	1,64	2,57	0,85	1,60	2,34	3,43	0,86	1,27	2,09	3,44	0,80	1,72	2,42	3,40	0,90
Moyenne qualité tous échantillons						0,74				0,78			0,87				0,84				0,83					0,89
Médiane qualité tous échantillons						0,80				0,82			0,87				0,87				0,86					0,91

Tableau 5-2 : Echelle de couleurs pour la dispersion (fond des colonnes "Estim" du tableau 5-1 ci-dessus)



B - Comparaison des densités estimées selon les méthodes et durées d'inventaires

Les tableaux 5-3, 5-4 et 5-5 suivants donnent, pour chaque échantillon, les écarts d'estimations de densité :

- pour chacun des **3 jeux de données** Naturalist, Papyrus, et Naturalist+Papyrus, **entre les 2 durées d'inventaire**,
- pour la durée d'inventaire de **5mn**, **entre les 3 jeux de données** Naturalist, Papyrus, et Naturalist+Papyrus,
- pour la durée d'inventaire de **10mn**, **entre les 3 jeux de données** Naturalist, Papyrus, et Naturalist+Papyrus.

Dans ces tableaux, par définition, $X / Y \% = 100 * (X - Y) / Y = \% \text{ d'augmentation / diminution}$ pour passer de Y à X, où X et Y sont les estimations de densité à comparer.

Tableau 5-3 : Ecarts d'estimations de densité entre durées d'inventaire 5mn et 10mn

Espèce	Passage	Adulte	Nat 10mn / 5mn %	Pap 10mn / 5mn %	Nat+Pap 10mn / 5mn %
Sylvia atricapilla	a+b	m	24	12	19
Turdus merula	a+b	m	100	24	41
Alauda arvensis	a+b	m	33	12	20
Sylvia communis	a+b	m	23	11	26
Columba palumbus	a+b	m	19	29	41
Luscinia megarhynchos	a+b	m	18	40	23
Cuculus canorus	a+b	m	39	45	20
Fringilla coelebs	a+b	m	54	14	22
Phylloscopus collybita	a+b	m	51	26	42
Parus major	a+b	m	77	38	32
Lullula arborea	a+b	m	29	35	28
Emberiza cirrus	a+b	m	40	14	55
Erithacus rubecula	a+b	m	41	-14	26
Saxicola rubicola	a+b	m	15	5	22
Emberiza citrinella	a+b	m	24	29	32
Cyanistes caeruleus	a+b	m	63	63	47
Turdus philomelos	a+b	m	56	22	48
Emberiza calandra	a+b	m	14	-4	4
Turdus viscivorus	a+b	m	40	70	61
Prunella modularis	a+b	m	100	25	42
Phylloscopus bonelli	a+b	m	-8	111	45
Carduelis cannabina	a+b	m	53	78	8
Jynx torquilla	a+b	m	65	25	-8
Carduelis chloris	a+b	m	35	18	20
Anthus trivialis	a+b	m	46	-6	32

Upupa epops	a+b	m	69	27	49
Lanius collurio	b	m	35	40	10
Oriolus oriolus	b	m	91	-10	1
Streptopelia turtur	b	m	65	9	49
Dendrocopos major	a+b	m+a	57	72	47
Moyenne tous échantillons			46	29	30
Médiane tous échantillons			41	25	30

Tableau 5-4 : Ecarts d'estimations de densité entre méthodes d'inventaire (durée 5mn)

Espèce	Passage	Adulte	Pap / Nat %	Nat+Pap / Nat %	Nat+Pap / Pap %
Sylvia atricapilla	a+b	m	51	8	-28
Turdus merula	a+b	m	139	95	-18
Alauda arvensis	a+b	m	22	6	-13
Sylvia communis	a+b	m	17	-3	-17
Columba palumbus	a+b	m	125	39	-38
Luscinia megarhynchos	a+b	m	100	60	-20
Cuculus canorus	a+b	m	89	90	1
Fringilla coelebs	a+b	m	160	50	-42
Phylloscopus collybita	a+b	m	83	61	-12
Parus major	a+b	m	89	71	-10
Lullula arborea	a+b	m	17	19	2
Emberiza cirrus	a+b	m	-21	-28	-8
Erithacus rubecula	a+b	m	144	44	-41
Saxicola rubicola	a+b	m	-49	-29	37
Emberiza citrinella	a+b	m	-6	-4	2
Cyanistes caeruleus	a+b	m	-6	11	19
Turdus philomelos	a+b	m	34	7	-20
Emberiza calandra	a+b	m	4	-9	-12
Turdus viscivorus	a+b	m	32	16	-12
Prunella modularis	a+b	m	57	45	-8
Phylloscopus bonelli	a+b	m	-79	-43	170
Carduelis cannabina	a+b	m	29	36	5
Jynx torquilla	a+b	m	26	105	62
Carduelis chloris	a+b	m	153	92	-24
Anthus trivialis	a+b	m	274	93	-48
Upupa epops	a+b	m	422	278	-28
Lanius collurio	b	m	12	33	19
Oriolus oriolus	b	m	74	45	-17
Streptopelia turtur	b	m	390	159	-47
Dendrocopos major	a+b	m+a	-19	10	35

Moyenne tous échantillons	79	45	-4
Médiane tous échantillons	43	37	-12

Tableau 5-5 : Ecart de estimations de densité entre méthodes d'inventaire (durée 10mn)

Espèce	Passage	Adulte	Pap / Nat %	Nat+Pap / Nat %	Nat+Pap / Pap %
Sylvia atricapilla	a+b	m	37	4	-24
Turdus merula	a+b	m	48	37	-8
Alauda arvensis	a+b	m	3	-5	-7
Sylvia communis	a+b	m	7	-1	-7
Columba palumbus	a+b	m	144	65	-33
Luscinia megarhynchos	a+b	m	137	68	-29
Cuculus canorus	a+b	m	98	65	-17
Fringilla coelebs	a+b	m	93	20	-38
Phylloscopus collybita	a+b	m	52	51	0
Parus major	a+b	m	47	27	-14
Lullula arborea	a+b	m	23	18	-4
Emberiza cirrus	a+b	m	-35	-20	24
Erithacus rubecula	a+b	m	50	29	-14
Saxicola rubicola	a+b	m	-53	-25	59
Emberiza citrinella	a+b	m	-2	2	5
Cyanistes caeruleus	a+b	m	-7	0	8
Turdus philomelos	a+b	m	5	1	-3
Emberiza calandra	a+b	m	-12	-17	-6
Turdus viscivorus	a+b	m	60	33	-17
Prunella modularis	a+b	m	-1	3	4
Phylloscopus bonelli	a+b	m	-52	-10	85
Carduelis cannabina	a+b	m	50	-4	-36
Jynx torquilla	a+b	m	-5	13	19
Carduelis chloris	a+b	m	122	71	-23
Anthus trivialis	a+b	m	141	74	-28
Upupa epops	a+b	m	293	233	-15
Lanius collurio	b	m	17	9	-7
Oriolus oriolus	b	m	-19	-23	-6
Streptopelia turtur	b	m	222	134	-27
Dendrocopos major	a+b	m+a	-11	3	16
Moyenne tous échantillons			48	29	-5
Médiane tous échantillons			30	11	-7

Le tableau graphique suivant résume ces 3 mêmes comparaisons, dans le même ordre (1 par ligne) :

- chaque point correspond à 1 échantillon (1 ligne dans les tableaux précédents),
- la couleur de chaque point figure la moyenne géométrique des qualités des 2 estimations (RacineCarrée(qualité X x qualité Y), selon la même échelle exactement que pour le tableau de résultats bruts 5-1,
- les axes X et Y représentent la densité estimée en individus / km² (mâles adultes ou tous adultes selon l'échantillon "m" ou "m+a"), pour les 2 variantes comparées (Cf. titre de chaque graphique),
- l'échelle logarithmique choisie pour ces axes X et Y permet d'obtenir une répartition spatiale des points plus lisible que l'échelle linéaire : la droite de régression (en pointillés rouge) est donc courbe, mais cela n'enlève rien à sa linéarité !
- la droite $x = y$ est figurée sur chaque sous-graphique, pour simplifier la comparaison visuelle : les points sur cette droite indiquent des densités comparées identiques.

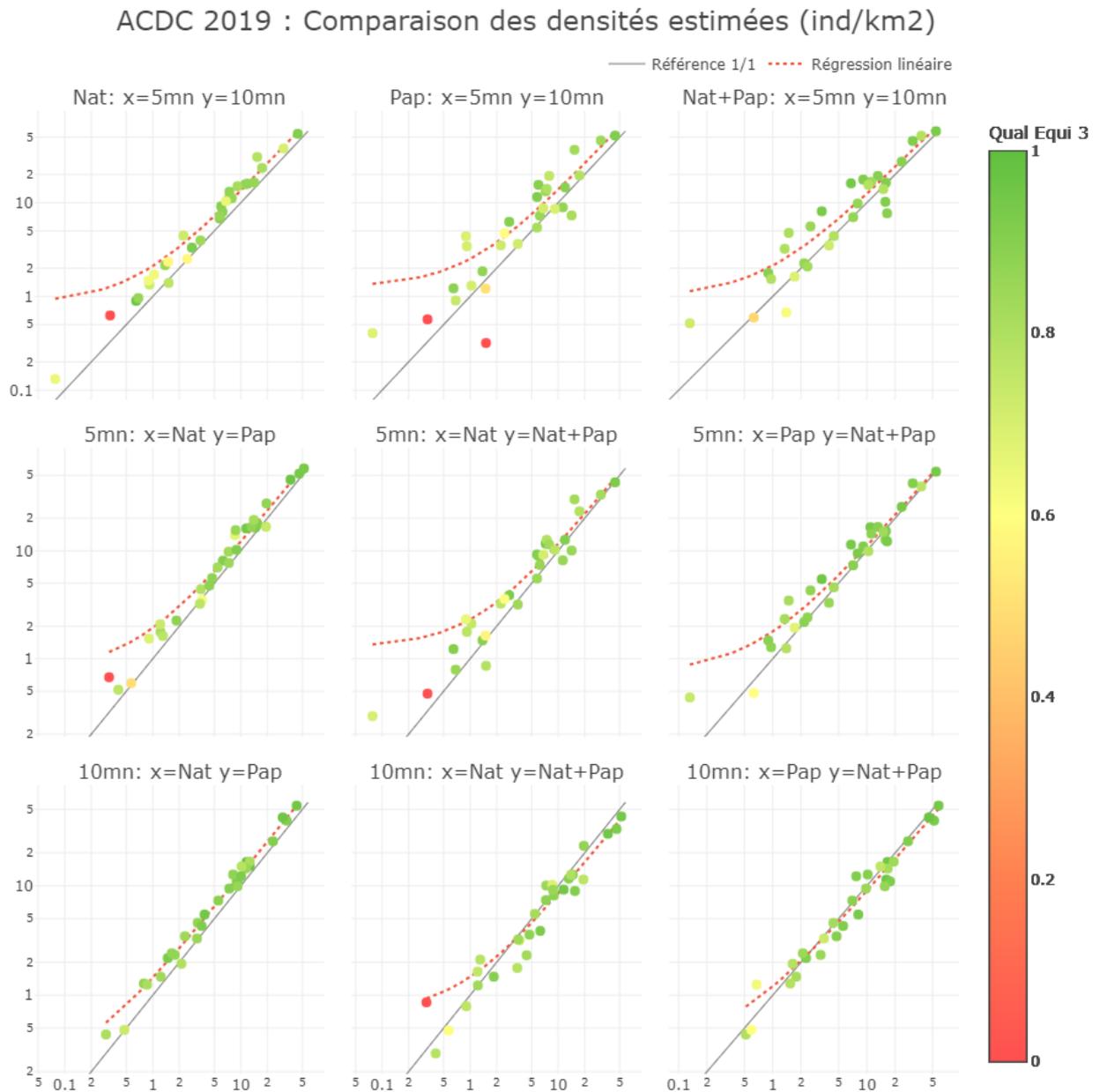


Figure 5-1 : Ecart d'estimations de densité entre toutes les variantes de durées d'inventaire (5mn et 10mn), et de jeux de données (Naturalist, Papyrus et Naturalist + Papyrus)

Discussion

A - Observations sur la distribution des distances de contact

Avant d'aborder les estimations de densités proprement dites, on peut constater pour certaines espèces à peu près les mêmes biais dans les données de terrain que lors des autres enquêtes [GUÉLIN 2015] [GUÉLIN 2016] [GUÉLIN et al. 2018] :

- trop de contacts à courtes distances** : *Turdus merula*, *Parus major* (l'hypothèse d'espèces fuyant systématiquement à l'arrivée de l'observateur sur le point d'écoute, ou au contraire restant sur place, mais dans tous les cas, de manière peu discrète, paraît assez peu probable, puisqu'elle remettrait en cause l'hypothèse fondamentale du Distance Sampling qui veut que l'on contacte systématiquement tous les individus à distance zéro : ces cas de prétendue "surcharge" à faible distance seraient donc les seuls où cette hypothèse serait respectée, et dans tous les autres cas, elle ne le serait pas ? mais peut-être a-t-on affaire ici plus simplement à un **biais d'arrondi à 0 des très courtes distances** ("angle rounding") ? ou alors au fait que les fuyards sont essentiellement détectés à vue, alors que la majorité des contacts se fait à l'ouïe, et donc qu'on tenterait vainement d'appliquer un modèle de détection "ouïe" sur des données "vue"),

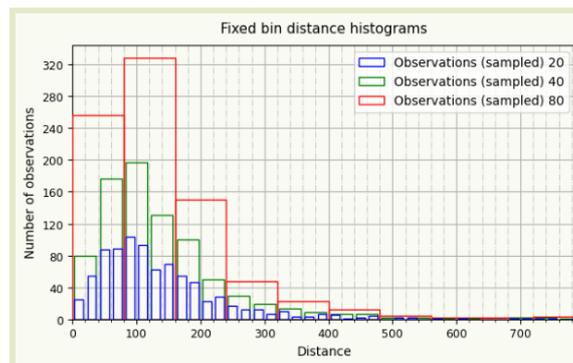


Figure 6-1 : *Turdus merula* Naturalist + *Papyrus* 10mn : Exemple de biais spécifique dans la distribution des distances de contact, avec une nette "surcharge" à courtes distances

- pas assez de contacts à courtes distances** : *Alauda arvensis*, *Sylvia communis*, *Turdus philomelos* et *viscivorus*, et à un degré moindre *Columba palumbus*, *Cuculus canorus* (hypothèse opposée : espèces fuyant systématiquement à l'arrivée de l'observateur sur le point d'écoute, ou au contraire restant sur place à proximité de l'observateur, mais dans tous les cas, **de manière très discrète**, au moins dans les premières minutes ; ce qui impliquerait que l'hypothèse fondamentale du Distance Sampling "tous les oiseaux sont détectés à distance zéro" n'est pas respectées pour ces espèces ?).

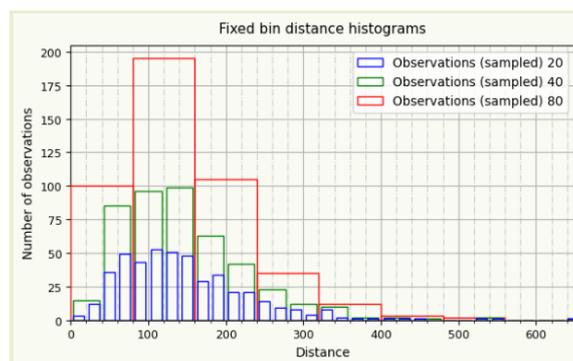


Figure 6-2 : *Alauda arvensis* Naturalist + *Papyrus* 10mn : Exemple de biais spécifique dans la distribution des distances de contact, avec un net déficit à courtes distances

Pour d'autres espèces, même abondantes ou assez communes, selon qu'on examine les données Naturalist ou Papyrus, et/ou les durées 5mn ou 10mn, on observe des biais opposés ou inexistantes pour l'une des méthode d'inventaire / durées ... signe que l'espèce n'est manifestement pas en cause (*Sylvia atricapilla*, *Lullula arborea*, ...).

B - Qualité des estimations de densités obtenues

Globalement, la qualité des estimations produites est **bonne à très bonne**, si l'on se fie à l'indicateur de qualité combinée : Cf. Tableau 5-1, colonnes "Quali", dernières lignes "moyenne" et "médiane", et coloration d'ensemble "très verte" (148 / 180 analyses, soit 82 %, au dessus de 0,75),

- pour les 3 jeux de données Naturalist, Papyrus et Naturalist + Papyrus,
- pour les 2 durées d'inventaire 5mn et 10mn.

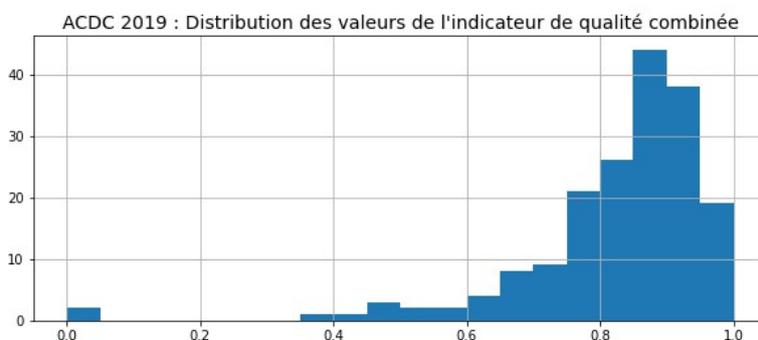


Figure 6-3 : Distribution des valeurs de l'indicateur de qualité combinée

Pour l'illustration, voici 2 exemples (parmi bien d'autres) de ces modélisation de (bonne à) très bonne qualité combinée.

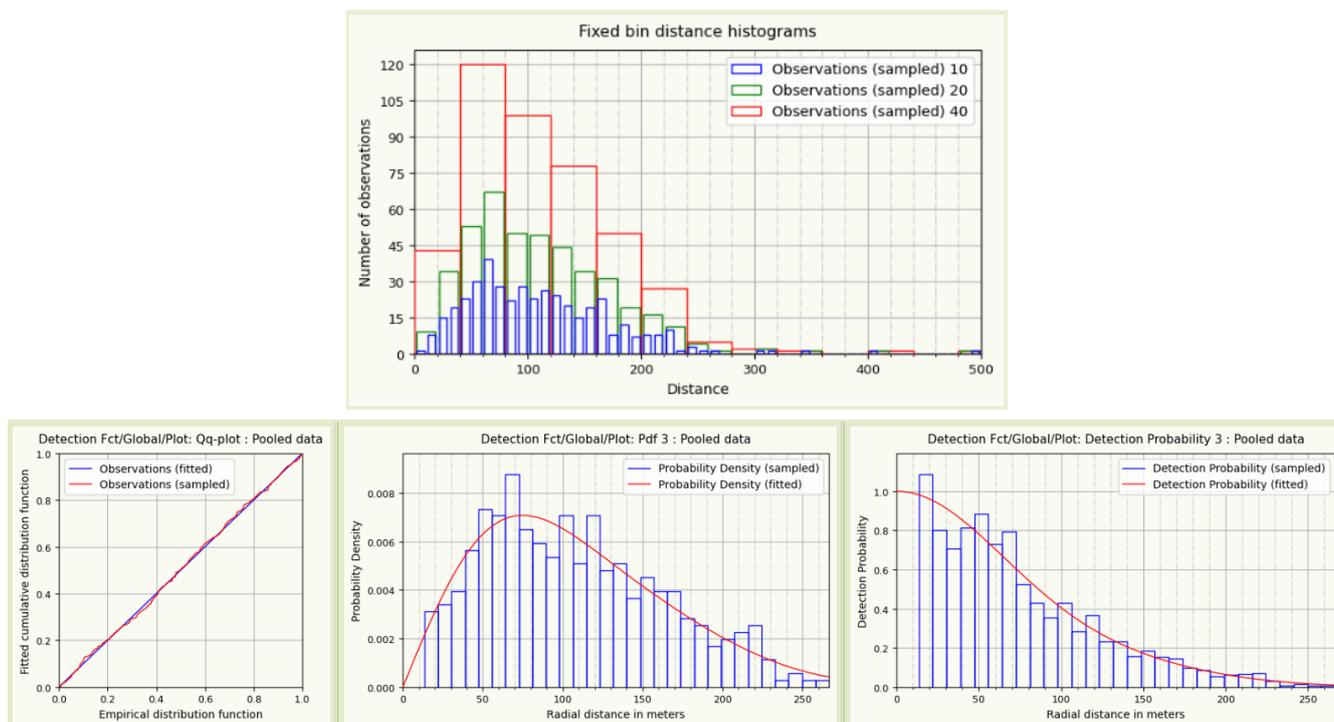


Figure 6-4 : *Sylvia atricapilla* Papyrus 10mn : Exemple d'une modélisation de très bonne qualité combinée (parmi beaucoup d'autres).

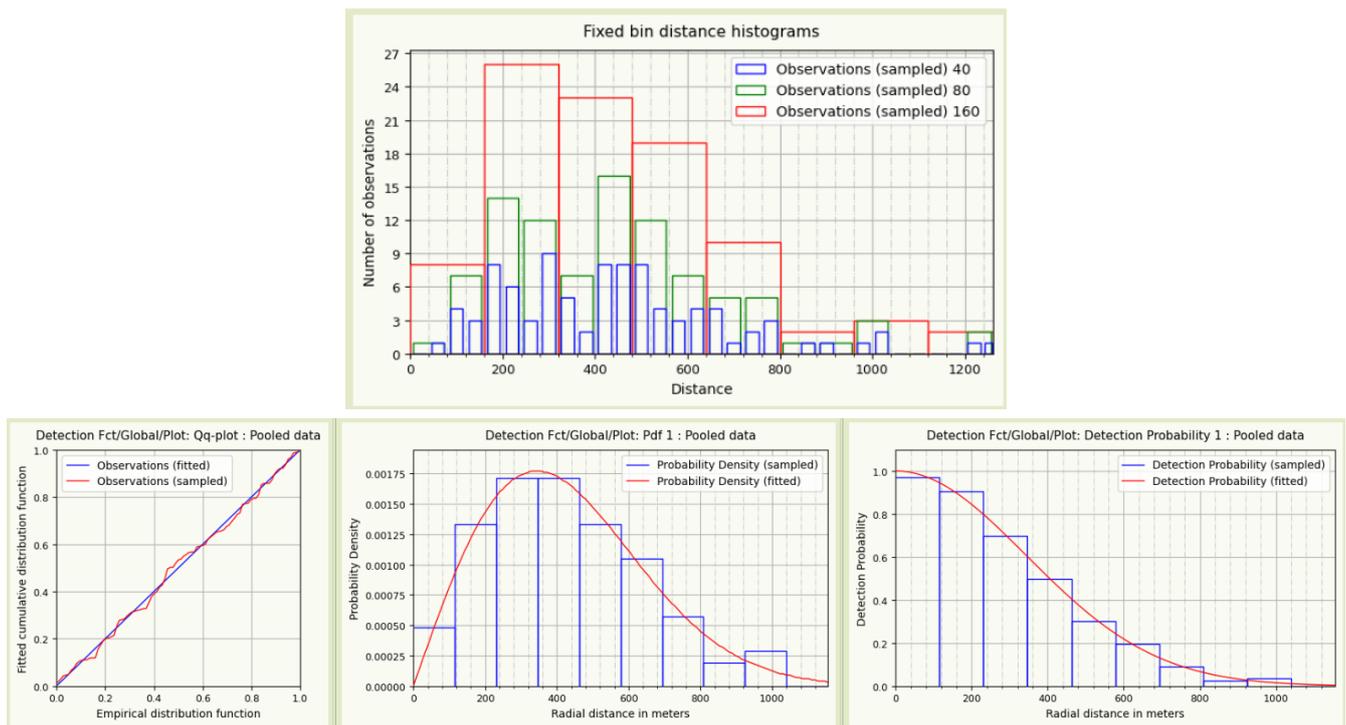


Figure 6-5 : *Cuculus canorus* Naturalist 5mn : Autre exemple d'une modélisation de très bonne qualité combinée (parmi beaucoup d'autres).

Dans le détail, sans surprise, la qualité se dégrade :

- lorsque les effectifs contactés sur le terrain sont insuffisants,
- ou lorsque la distribution des distances de contact est défectueuse : pas assez ou trop de contacts dans certaines tranches de distances,
- mais le plus souvent les 2 à la fois.

Les 2 cas les plus critiques (indicateur qualité combinée à 0) sont sans surprise *Oriolus oriolus* Naturalist 5mn et *Phylloscopus Bonelli* Papyrus 5mn, pour lesquels le nombre de contacts (4 !) ne permet bien sûr pas d'analyse DS fiable : on ne retiendra aucune estimation de densité pour ces 2 cas.

Les indicateurs de qualité combinée, respectivement de 0,38 et 0,41, de *Dendrocopos major* Naturalist 5mn et *Carduelis chloris* Naturalist 5mn ne sont pas très bons : le déficit de données et la distribution défectueuse des distances de contact dans les 2 cas entraînent une certaine sur-modélisation que l'indicateur de qualité combinée ne "pénalise" pas assez ; et le coefficient de variation de la densité DCv dépasse les 30% !

Mais ne serait-ce pas plus simplement que ces 2 espèces présentent une **répartition réellement hétérogène** sur la zone étudiée ? Ce serait assez logique pour le Pic épeiche, assez forestier et donc naturellement hors-liste des espèces des milieux naturels ciblés, et de fait peu contacté suite à l'exclusion des zones boisées du secteur d'étude ; et assez cohérent avec les observations de terrain pour le Verdier, effectivement peu abondant et seulement présent en certains endroits, peut-être en lien avec une exigence particulière de l'espèce sur les sites de nidification.

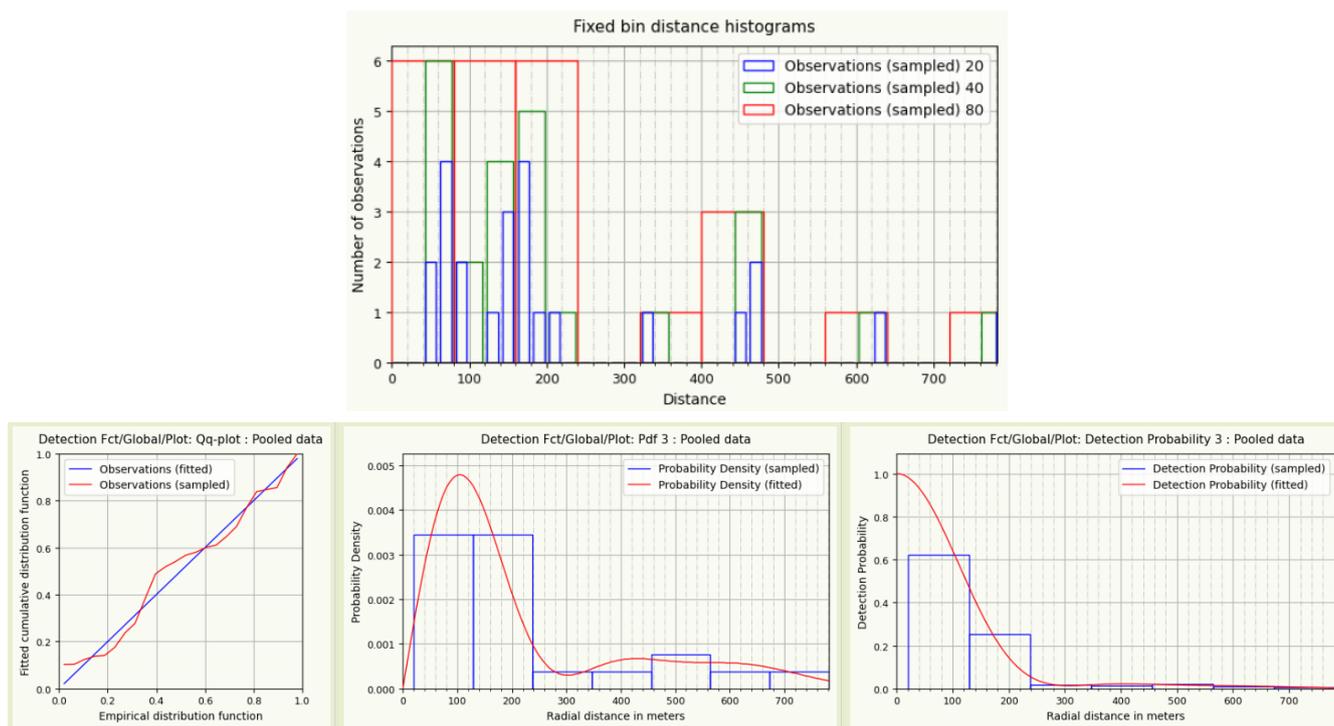


Figure 6-6 : *Dendrocopos major* Naturalist 5mn : Distribution des distances et probabilité estimée de détection : exemple de déficit de données et de distribution des distances déficiente entraînant une sur-modélisation pas assez pénalisée par l'indicateur de qualité combinée

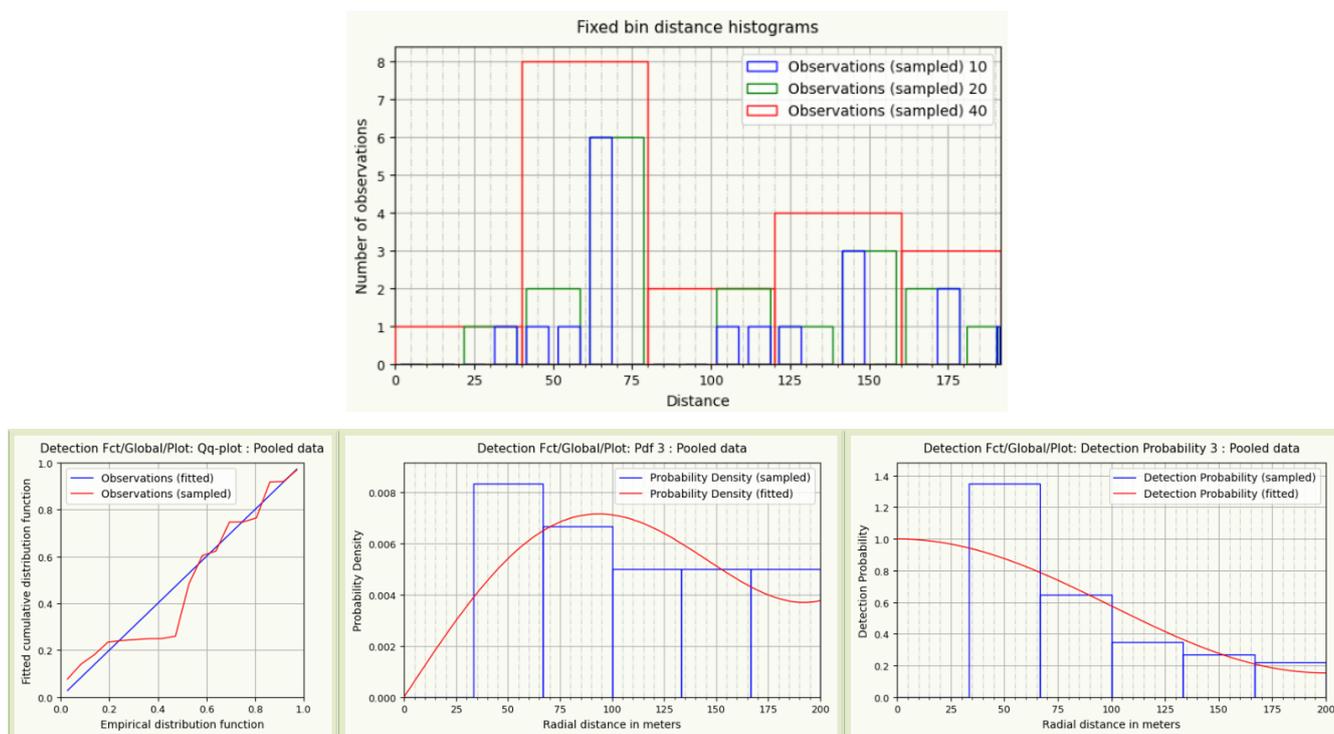


Figure 6-7 : *Carduelis chloris* - Naturalist - 5mn : Distribution des distances et probabilité estimée de détection : exemple de déficit de données et de distribution des distances déficiente entraînant une sur-modélisation pas assez pénalisée par l'indicateur de qualité combinée

Le constat est similaire pour *Oriolus oriolus* Naturalist 10mn, *Oriolus oriolus* Papyrus 10mn et *Phylloscopus Bonelli* Papyrus 10mn, qui ont des indicateurs de qualité combinée légèrement meilleurs, entre 0,45 et 0,50, grâce à un nombre de contacts moins déficitaire (mais encore insuffisant) ; le coefficient de variation de la densité DCv dépasse par ailleurs toujours le seuil des 30%.

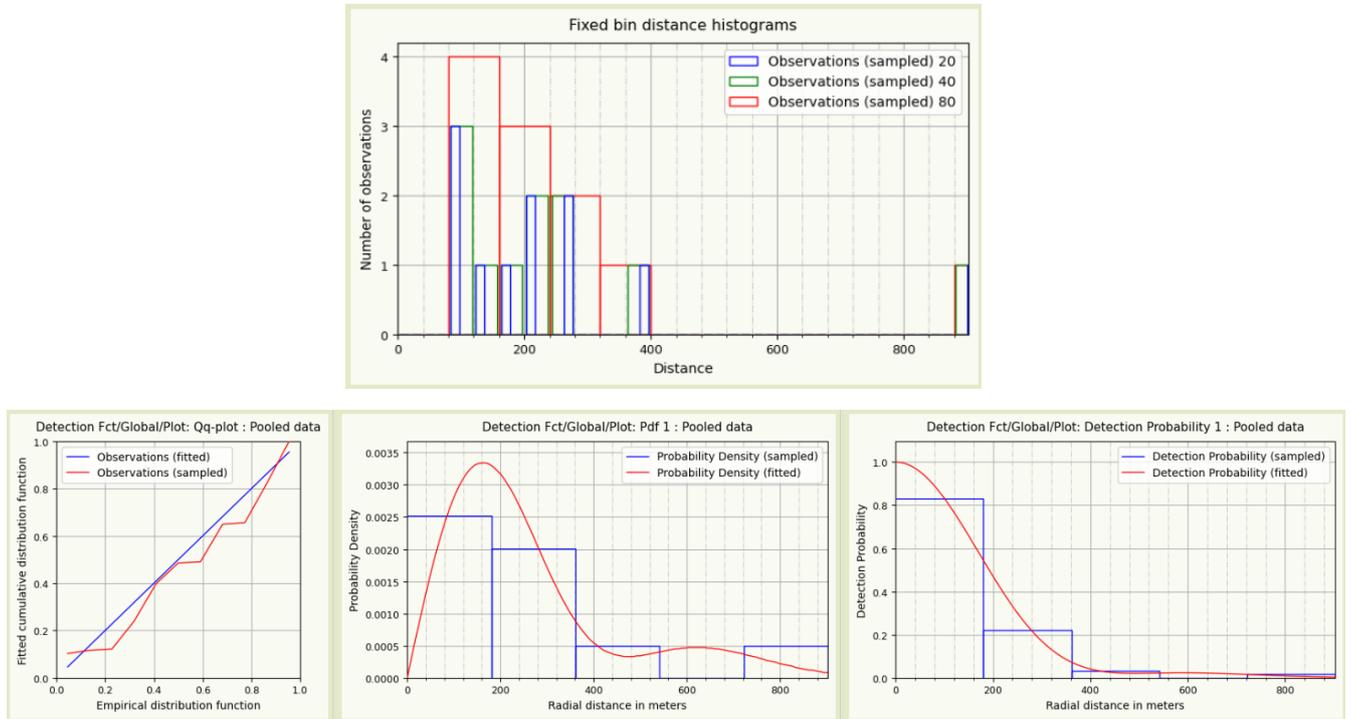


Figure 6-8 : *Oriolus oriolus* Naturalist 10mn : Distribution des distances et probabilité estimée de détection : pas assez de données, distribution des distances peu régulière => piètre modélisation.

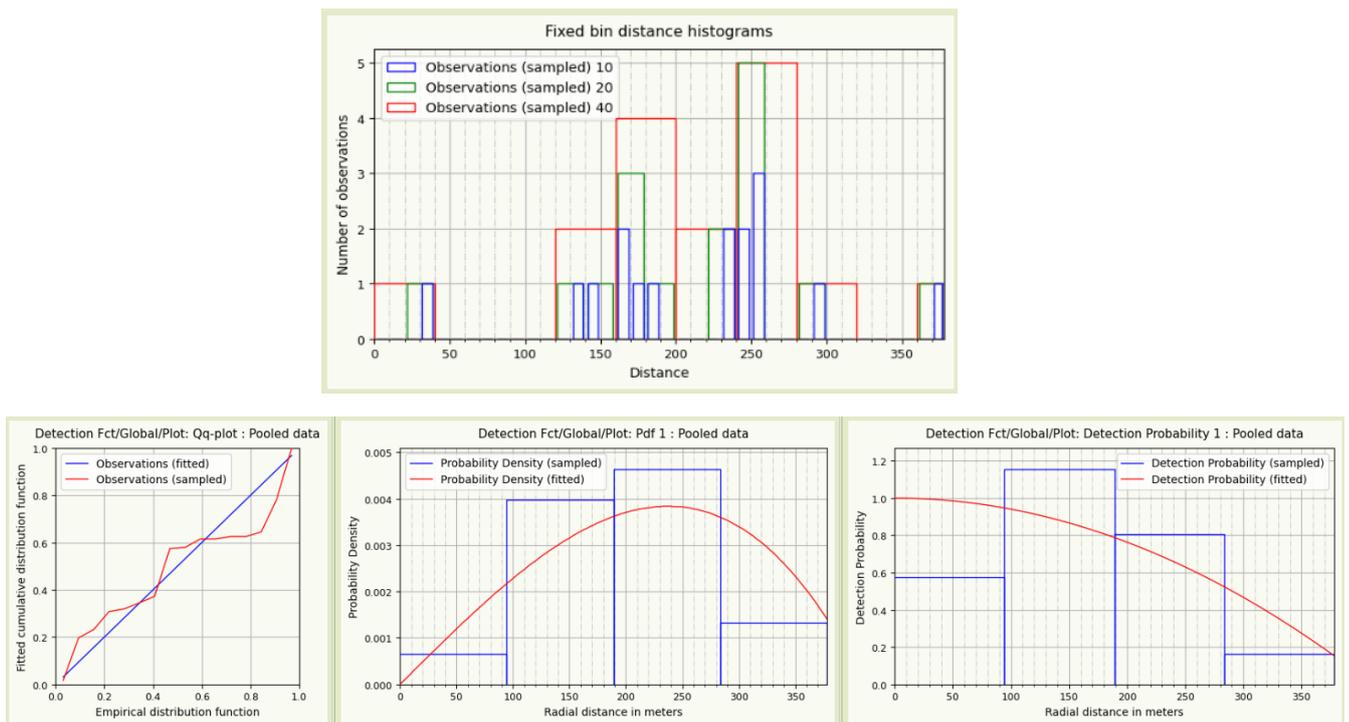


Figure 6-9 : *Oriolus oriolus* Papyrus 10mn : Distribution des distances et probabilité estimée de détection : pas assez de données, distribution des distances très irrégulière => piètre modélisation.

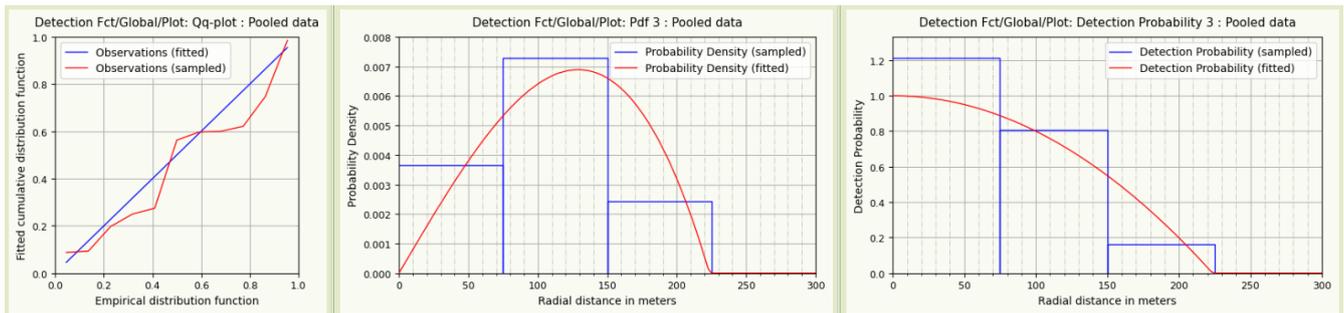
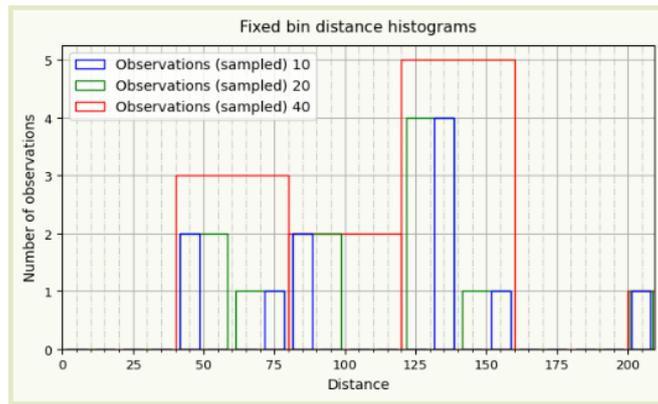


Figure 6-10 : *Phylloscopus Bonelli Papyrus 10mn* : Distribution des distances et probabilité estimée de détection : pas assez de données, distribution des distances très irrégulière => piètre modélisation.

Au total, ce sont donc 7 échantillons sur 180 (chiffre assez anecdotique finalement) pour lesquels les densités estimées sont à ignorer ou à prendre avec des pincettes.

Sur les biais observés pour certaines espèces dans la distribution des distances de contact, on constate :

- que l'automate sélectionne en général très correctement les meilleures modélisations (du point de vue d'un analyste humain) dans le cas du manque de données à courtes distances (*Alauda arvensis*, *Sylvia communis*, ...) : presque systématiquement, un troncature à gauche est automatiquement appliquée pour cela, et quand elle ne l'est pas, les meilleures densités avec troncature sont très proches,

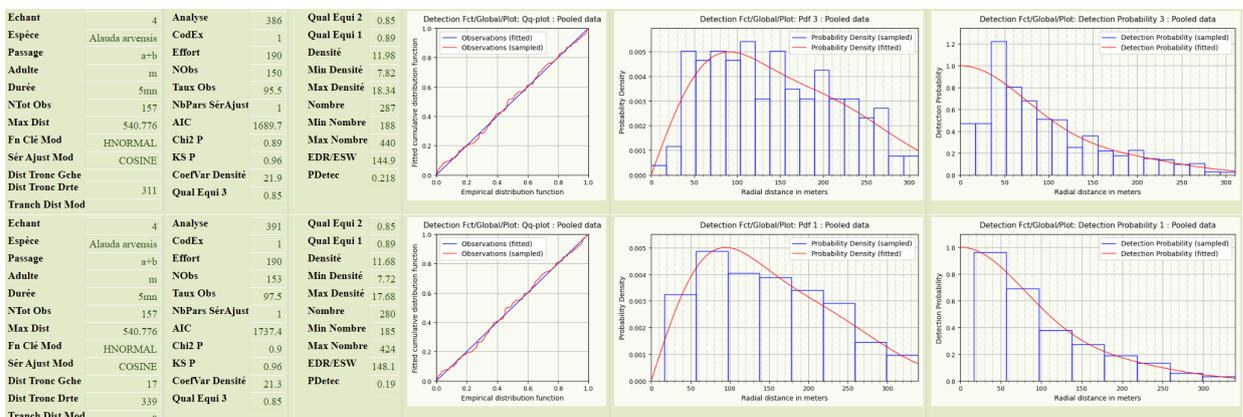


Figure 6-11 : Exemple avec *Alauda arvensis Naturalist 5mn* : les 2 meilleures analyses - la 1ère sans troncature à gauche, la 2nde avec - produisent des densités très proches (Diagramme Quantile-Quantile et Probabilité estimée de détection).

- mais qu'il hésite dans certains cas avec les échantillons présentant trop de données à courtes distances (Turdus merula, ...) entre des modélisations avec et sans troncature à gauche ... tout comme l'analyste humain : faute d'éléments supplémentaires pour décider différemment, nous avons conservé le classement automatique pour ces quelques cas.

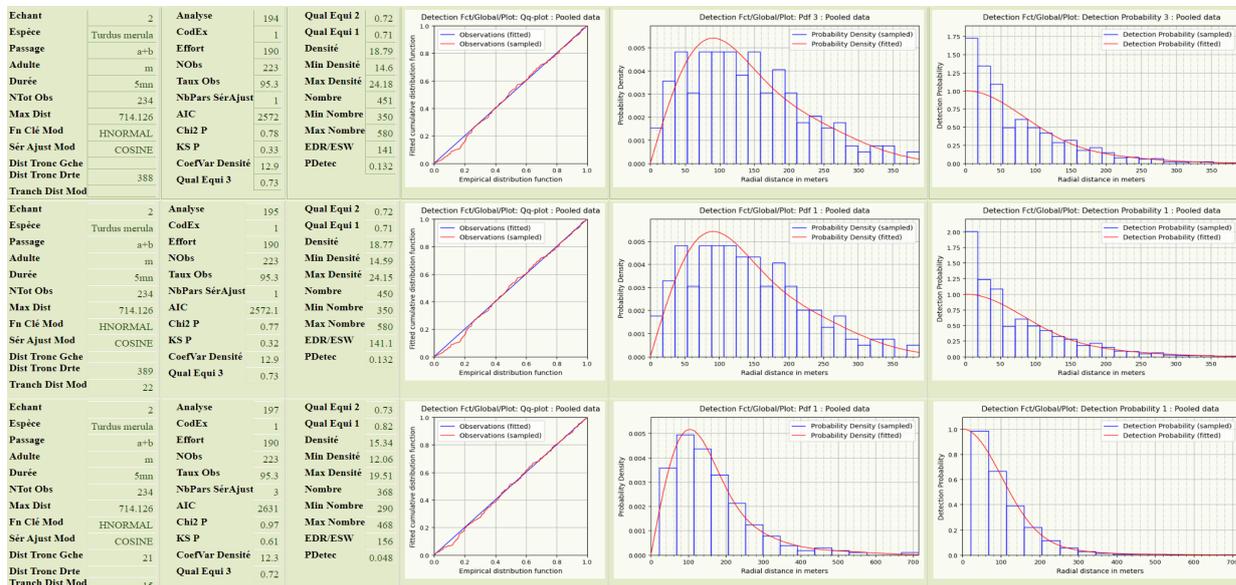


Figure 6-12 : Exemple avec *Turdus merula* Naturalist 5mn : les 3 meilleures analyses - les deux premières sans troncature à gauche, la 2nde avec - produisent des densités différentes d'environ 15% (Diagramme Quantile-Quantile et Probabilité estimée de détection), mais on ne sait pas décider laquelle des 2 1ères ou de la 3ème est la meilleure.

Enfin, rappelons qu'une **vingtaine** de cas sur 180 présente une **dispersion** notable voire importante dans les meilleures densités obtenues (la meilleure, qu'on a finalement retenue, et ses 3 suivantes) ; comme déjà dit, ce n'est pas bloquant, mais le symptôme d'une **hésitation** là encore dans la modélisation DS, hésitation qui incite à rester prudent sur les estimations de densités correspondantes. Ce sont toutefois ces "meilleures" densités que nous conservons en conclusion : le fond coloré des colonnes "Estim" du tableau 5-1 permet de souligner la prudence à observer (cas en orange voire en rouge).

C. Comparaison de la qualité des estimations de densités obtenues selon les combinaisons de méthodes et durées d'inventaires

Tout d'abord, globalement, la **qualité** combinée des estimations produites **s'améliore** (colonnes "Quali" du tableau 5-1) :

- lorsque l'on **pass**e d'inventaires **de 5mn à 10mn**,
- lorsque l'on **pass**e du jeu de données **Naturalist au jeu Papyrus** (sauf pour les inventaires de 10mn : globalement, qualité combinée similaire), puis aux 2 jeux cumulés.

Intuitivement, il n'est en effet pas surprenant que lorsque l'on augmente la taille des jeux de données (5mn => 10mn, ou cumul des jeux Naturalist et Papyrus), on améliore la qualité combinée des résultats des analyses que l'on peut faire dessus, puisque l'on a probablement amélioré ses qualités statistiques intrinsèques (loi des grands nombres).

La comparaison entre les méthodes Naturalist et Papyrus est plus intéressante et confirme ce qu'on soupçonnait déjà en examinant les quantités brutes de données : à équipe d'observateurs équivalentes,

- on fait des inventaires de **meilleure qualité** combinée avec la méthode **Papyrus** sur les **5 premières minutes**,
- sur 10mn, les 2 méthodes semblent équivalentes, encore une fois si l'on s'intéresse uniquement à la qualité combinée des analyses DS.

D. Comparaison des valeurs de densité obtenues selon les combinaisons de méthodes et durées d'inventaires

Si l'on s'intéresse maintenant aux **valeurs de densités** estimées, comparons d'abord les inventaires de 5 et 10 mn, pour les 3 jeux de données Naturalist, Papyrus, et les 2 cumulés (Tableau 5-3).

N.B. Quelle que soit méthode de notation Naturalist ou Papyrus, les jeux de données 5mn et 10mn ne sont pas statistiquement indépendants puisque le premier est inclus dans le second : les comparaisons statistiques effectuées ci-dessous sont donc en toute rigueur sujettes à caution ; mais cela ne doit pas masquer les faits constatés.

En moyenne / médiane, **le passage de 5mn à 10mn augmente les densités de 40-45% pour Naturalist, et de 25-30% pour Papyrus :**

- sur l'augmentation de densité en elle-même, indépendamment de la méthode d'inventaire, on ne peut que constater qu'il y a effectivement augmentation générale du nombre de contacts (augmentation d'ailleurs systématique à toutes les distances) (Figures 6-13 et 6-14),
- quant à la différence d'augmentation entre les 2 méthodes, la même explication reste valable, vu la différence d'augmentation des nombres de contacts (Tableau 3-4) ; mais ce phénomène est très probablement renforcé par un surcroît de contacts enregistrés aux très courtes distances après la 5ème minute avec la méthode Naturalist (Figures 6-13 et 6-14 : dans les 10 premiers mètres, +65% avec N=9/15, à comparer aux +22% avec N=23/28 de la méthode Papyrus), ce qui ne peut que renforcer le respect de la règle fondamentale du Distance Sampling qui veut que tous les oiseaux à distance zéro sont contactés, et donc améliorer les estimations de densités ; avec la méthode Papyrus en revanche, les contacts

supplémentaires après la 5ème minutes sont beaucoup plus également répartis en distance, sans pic à très courte distance,

ACDC 2019 Naturalist : Surcroît de contacts après 5mn selon la distance (3469 mâles)

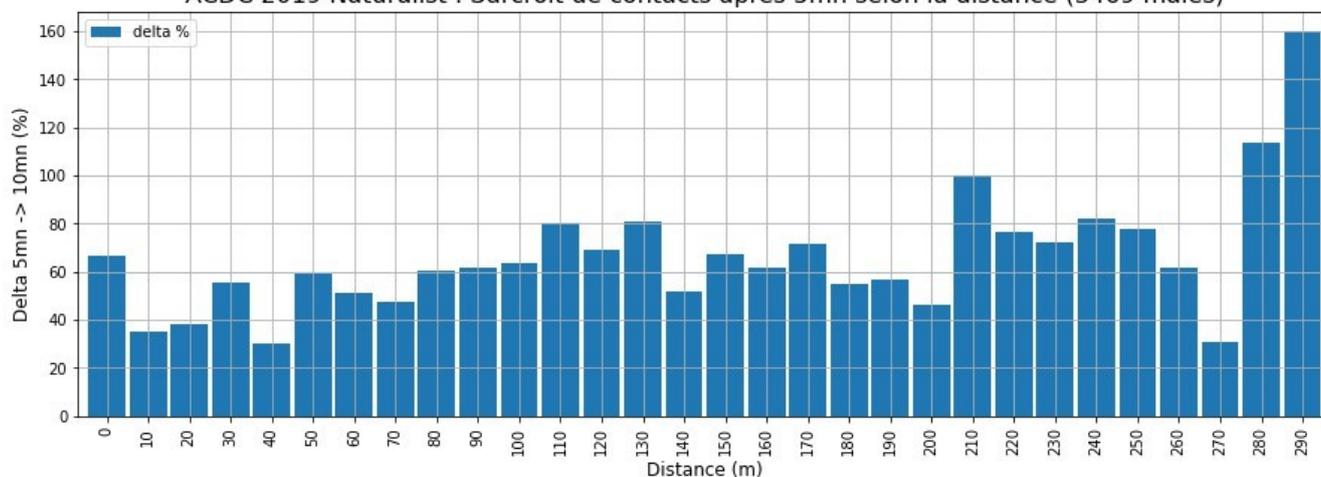


Figure 6-13 : Naturalist : Accroissement du nombre de contacts (mâles uniquement, N = 3469) en fonction de la distance entre les durées 5mn et 10mn : après la 5ème minute, les observateurs semblent compléter fortement leurs inventaires avec des oiseaux à toutes distances ; avec un effort légèrement supérieur au delà de 100m, très supérieur au delà de 280m ; et surtout presque 2 fois supérieur à moins de 10m par rapport au distances de 10 à 30m (en abscisse, 0 correspond à la tranche de distance [0, 10m[, 10 à [10, 20m[, ... etc)

ACDC 2019 Papyrus : Surcroît de contacts après 5mn selon la distance (3398 mâles)

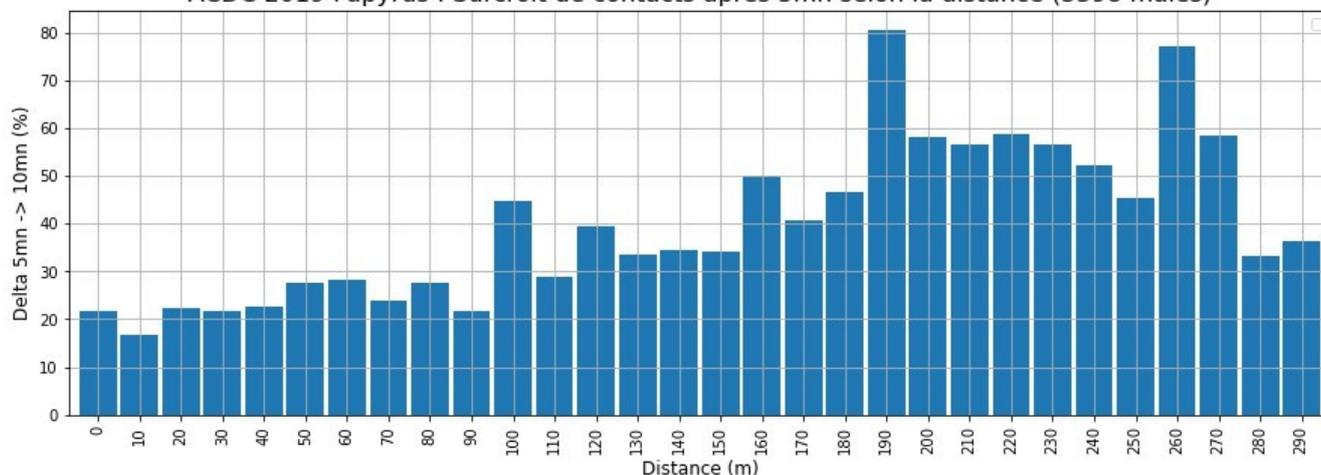


Figure 6-14 : Papyrus : Accroissement du nombre de contacts (mâles uniquement, N = 3398) en fonction de la distance entre les durées 5mn et 10mn : après la 5ème minute, les observateurs semblent compléter leurs inventaires avec les oiseaux plus éloignés (en abscisse, 0 correspond à la tranche de distance [0, 10m[, 10 à [10, 20m[, ... etc)

Et toujours en moyenne / médiane, le passage de 5mn à 10mn augmente les densités de 30% pour les 2 jeux de données cumulés, de manière très similaire au jeu de données Papyrus seul (la meilleure qualité des données Papyrus semble gommer les biais supplémentaires soupçonnés en Naturalist, et l'emporte lors du cumul).

Si, pour expliquer cette augmentation des densités estimées avec la durée d'inventaire, on peut invoquer l'**augmentation de la probabilité** (avec le temps) de contacter des **oiseaux chantant moins souvent**, on ne peut pas exclure non plus totalement une certaine proportion de **doubles comptages** (augmentation de la probabilité de déplacement de certains oiseaux), même si, connaissant ce risque (surtout sur 10mn), nous avons été très vigilants sur ce point.

Comparons ensuite les 2 méthodes Papyrus et Naturalist.

En moyenne / médiane, la méthode **Papyrus** donne des **densités plus élevées** (et de qualité combinée légèrement meilleure ou équivalente) que la méthode Naturalist : de **45-75% en 5mn** (Tableau 5-4), de **30-50% en 10mn** (Tableau 5-5) :

- cette fois, la comparaison des nombres bruts de mâles dans les 2 jeux de données (Tableau 3-4) ne permet pas d'expliquer la majorité de cette différence (10-15% de mâles en moins en Naturalist 5mn, moins de 5% en moins en 10mn),
- mais l'examen des distributions de distances de contacts (mâles uniquement, normalisées par la surface concernée : Figure 3-5) montre un **net déficit de données aux courtes distances** avec la méthode Naturalist, pour les inventaires de 5mn (- 61 % par rapport à Papyrus), mais aussi pour ceux de 10mn (- 46 % par rapport à Papyrus, pas de "rattrapage 10mn" cette fois) ; autrement dit, **les oiseaux à faible distance sont nettement moins détectés** "en mode" Naturalist "qu'en mode" Papyrus ; et donc que **la règle fondamentale du Distance Sampling "Probabilité de détection à distance zéro = 1" n'est au minimum pas correctement respectée avec la méthode Naturalist** ; une bonne partie de la différence moyenne dans les estimations de densités peut probablement s'expliquer ainsi ! (par ailleurs, on notera que l'indicateur de qualité combinée ne semble pas mettre ce biais en lumière : c'est une chose qu'il faudrait améliorer dans pyaudisam).

ACDC 2019 : Surcroît de contacts Papyrus comparé à Naturalist selon la distance (3398 / 3469 mâles)

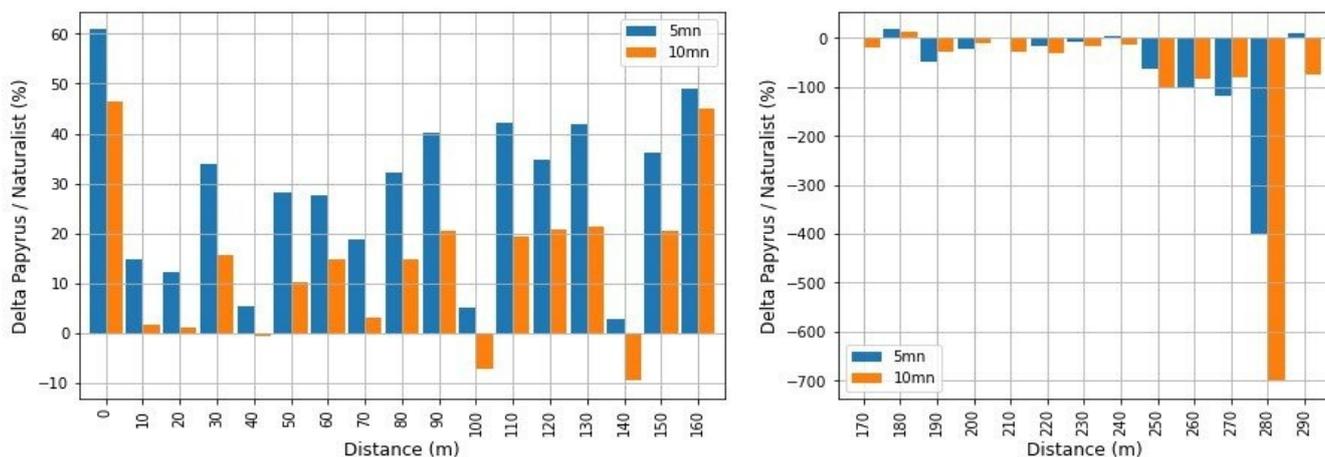


Figure 6-15 : *Ecarts de distribution des distances de contacts (mâles uniquement) entre les méthodes Papyrus (référence) et Naturalist, indépendamment pour les durées 5mn et 10mn (en abscisse, 0 correspond à la tranche de distance [0, 10m[, 10 à [10, 20m[, ... etc ; en ordonnée, par tranche de distance, $100 * (Nb \text{ mâles Papyrus} - Nb \text{ mâles Naturalist}) / Nb \text{ mâles Papyrus} =$ pourcentage de contacts en plus en mode Papyrus par rapport au mode Naturalist, 0% = égalité Papyrus / Naturalist)*

Lorsque l'on examine quelques espèces très abondantes pour lesquelles il n'y a pas de déficit systématique de données à courtes distances (toutes méthodes et durées d'inventaire : Cf. Observations sur la distribution des distances de contact), on obtient :

- Turdus merula : cohérent avec le constat global,
- Sylvia atricapilla : cohérent avec le constat global en 5mn, mais constat opposé en 10mn,
- Fringilla coelebs : constat opposé sur 5mn et 10mn.

Mais on parle ici d'effectifs très faibles même pour ces espèces très abondantes, vu les distances considérées (au maximum 15 pour Turdus merula, 10 pour Sylvia atricapilla, et moins de 5 pour Fringilla coelebs) : difficile donc d'en tirer une quelconque conclusion à l'échelle spécifique.

Comment expliquer ce déficit global de données à courtes distances pour la méthode Naturalist ? On ne peut qu'imaginer quelques hypothèses :

- en supposant que les 2 équipes ont par ailleurs appliqué le protocole exactement de la même manière (ce qui est probable, mais difficile à vérifier) :
 - le "**débordement**" de l'observateur moyen en mode Naturalist en début d'inventaire (premières minutes) lui fait perdre certains contacts à courte distance, par simple oubli face à l'avalanche ?
 - **penché sur son smartphone**, l'observateur regarde moins ce qui se passe autour de lui, en particulier en début d'inventaire, quand il est débordé, **ratant de ce fait certains contacts visuels** ?
- si en revanche on peut soupçonner des différences entre équipes dans l'exécution du protocole :
 - **déficit dans l'enregistrement des contacts proches** (oiseaux fuyant) à l'arrivée de l'observateur sur chaque point inventorié (une analyse par observateur des distributions des distances de contact permettrait peut-être vérifier cette hypothèse).

Comparons enfin les densités estimées en **cumulant** les jeux de données **Naturalist et Papyrus** avec celles obtenues avec le jeu **Papyrus**.

N.B. Les analyses effectuées sur ce jeu de données fusionné (Naturalist + Papyrus) n'ont pas tenu compte du fait que les 2 jeux de données n'avaient pas été collectés **strictement** de la même manière (le protocole est cependant strictement le même : seuls diffèrent les outils utilisés pour "enregistrer" les données : papier et crayon, ou application Naturalist) : faute de connaître et de savoir appliquer cette technique, nous n'avons pas introduit la co-variable "méthode d'inventaire" dans les paramètres de MCDS. Les comparaisons effectuées ci-dessous sont donc en toute rigueur statistiquement sujettes à caution ; cependant les constats effectués sont le reflet d'une réalité de terrain.

En moyenne / médiane, le jeu "**cumulé**" donne des **densités légèrement inférieures** à celles du jeu Papyrus (4-12 % pour inventaires de 5mn, 5-7 % pour 10mn : resp. tableaux 5-4 et 5-5) : c'est intuitivement assez logique :

- les densités obtenues sont logiquement intermédiaires (sorte de "moyenne"),
- la meilleure qualité d'ensemble du jeu Papyrus compense les défauts du jeu Naturalist,
- si les densités estimées sont finalement si proches en moyenne de celles du jeu Papyrus, c'est peut-être simplement le cumul de cette meilleure qualité du jeu Papyrus et du doublement du nombre de données ?

En résumé,

- les densités estimées sont de **bonne à très bonne qualité** à l'aune de l'indicateur de **qualité combinée** et à l'allure graphique des modélisations retenues (fonction probabilité de détection notamment), hormis pour un petit nombre (7 / 180) d'échantillons, pour lesquels le nombre de contacts est insuffisant (limite aux alentours de 20) et / ou leur distribution en distance est irrégulière ou déficitaire,
- les densités estimées sur **10mn** sont de **meilleure qualité combinée** et plus fortes que sur 5mn (plus de données, rattrapage du retard des 5 premières minutes en mode Naturalist),
- les densités estimées par la méthode **Papyrus** sont, comparées à celles des jeux Naturalist :
 - de **meilleure** qualité combinée,
 - **plus fortes** (entre 30 et 75 % selon les cas),
- les densités estimées par **cumul des jeux Naturalist et Papyrus** sont, si on les compare à celles du jeu Papyrus seul,
 - de **meilleure** qualité combinée,
 - **légèrement inférieures** (de 4 à 12% selon les durées d'inventaire),
- la méthode **Naturalist** semble pénalisée par un **déficit important de données à courtes distances**, biais qui semble provoquer une forte sous-estimation des densités, si on les compare à celles de la méthode Papyrus ; **c'est un problème, puisque l'hypothèse fondamentale du Distance Sampling "probabilité de détection = 1 à distance 0" semble ne pas être respectée** ; l'indicateur de qualité combinée ne met cependant pas en lumière ce biais (piste d'amélioration pour pyaudisam).

Mais à choisir, quel jeu de densités retenir comme le "meilleur" ?

- à coup sûr, un jeu estimé à partir d'inventaires de 10mn (pas d'argument objectif contre),
- mais ensuite : Papyrus seul, ou bien Naturalist + Papyrus ?
 - les différences sont certes faibles en moyenne (5-7%), mais pas toujours dans le détail des espèces (-38 à +59%, si l'on excepte *Phylloscopus Bonelli* à + 85%)

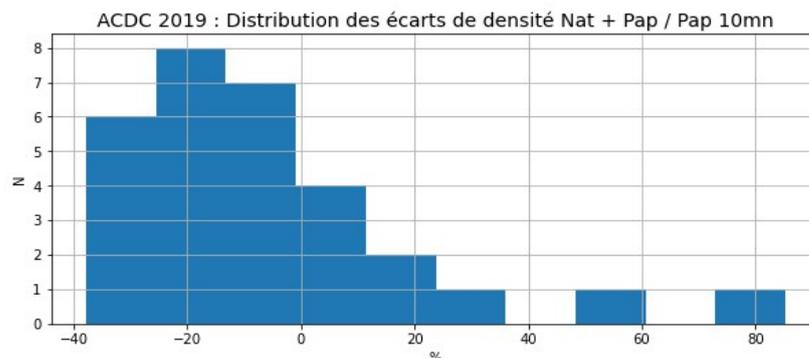


Figure 6-16 : Distribution des écarts de densités estimées entre les jeux de données *Naturalist* + *Papyrus* 10mn et *Papyrus* seul 10mn (moyenne -5%, 25% ≤ -22%, médiane = -7%, 25% ≥ 3%)

- n'aurions nous pas intérêt à finalement ignorer le jeu *Naturalist*, s'il y a un risque que sa mauvaise qualité (biais important à courtes distances) ne compense pas son apport indiscutable en terme de quantité de données ?

E. Et sur un plan plus strictement ornithologique ...

Pour une **analyse ornithologique** des densités obtenues lors de cette étude, et le croisement de ces résultats avec d'autres relevés quantitatifs effectués par le passé sur la même zone, on se reportera à [DULPHY et al. 2023].

Conclusion

Les analyses selon la méthode du **Distance Sampling** ont toutes été faites aussi **automatiquement** que possible à l'aide du logiciel "**pyaudisam**", spécialement développé "au dessus de" MCDS 6 - le moteur de calcul du logiciel Distance - face au constat qu'il y avait une **grande quantité de variantes de jeux de données et de paramètres d'analyse à tester** :

- 30 espèces retenues ici,
- 2 durées d'inventaires (5mn et 10mn),
- 3 "méthodes" d'inventaire (Naturalist, Papyrus, et la combinaison des 2, par fusion des 2 jeux de données),
- au strict minimum, une dizaine de combinaisons de paramètres de modélisation par analyse (modèles Half-Normal, Hazard-Rate et Uniform, séries d'ajustement Cosinus et Polynôme, plusieurs essais de troncatures à gauche et à droite, et/ou intervalles de découpage de l'espace des distances),

... soit un strict minimum de 1800 analyses à effectuer et dépouiller manuellement (et probablement plus en fait) !

Face à la difficulté particulière du paramétrage des troncatures en distance à gauche et à droite, une technique de **détermination automatique des "meilleures" troncatures** a même été intégrée au logiciel, pour compléter les choix de l'analyse humain.

Et le paramétrage et l'exécution d'une grande quantité d'analyses DS sont une chose, le **dépouillement** des non moins nombreux résultats en est une autre, au moins aussi **consommatrice de temps et sujette à erreurs** face à la masse : pour essayer de simplifier et d'accélérer ce travail, il a également été intégré au logiciel un système :

- d'évaluation objective de la qualité des modélisations DS (**indicateur** de nature essentiellement "statistique de la modélisation" dit de "**qualité combinée**"),
- de **filtrage et sélection automatique** d'un petit nombre de **résultats**, les "**meilleurs**" selon cet indicateur qualité, pour chaque échantillon traité (espèce, durée, méthode, type d'effectif), et selon un enchaînement d'étapes très similaire à ce qu'aurait fait un analyste humain, et une dernière étape "manuelle" de correction si nécessaire (dernier mot à l'analyste humain),
- de présentation synthétique de ces meilleurs résultats.

Cet effort de développement informatique équivalent à 1 année de travail à temps plein, étalé sur 2 années et demie, a effectivement permis d'**alléger** énormément le **travail** nécessaire pour **paramétrer, exécuter et dépouiller** toutes ces analyses (de l'ordre de 2 ou 3 semaines, pour finalement 16 500 analyses), mais aussi de **réduire** considérablement les possibilités d'**erreurs humaines** lors du **paramétrage** initial des analyses et lors des opérations nécessaires à la collecte des résultats et à l'extraction finale de la "**substantifique moelle**" d'une telle masse de résultats (gain très difficile à estimer, mais la charge mentale associée a été fortement allégée !). Et bien entendu, ce **logiciel générique** pourra servir à "exploiter en Distance Sampling" toutes les études à base de points d'écoute avec notation de distance qui seront réalisées à l'avenir.

Les résultats obtenus concernent donc **180 échantillons** (30 espèces, 2 durées et 3 "méthodes" d'inventaire), pour lesquels on peut dire globalement que :

- les **densités estimées sont de bonne à très bonne qualité** à l'aune de l'indicateur de qualité combinée et à l'allure graphique des modélisations retenues (hormis pour environ 1% des

échantillons, caractérisés par des jeux de données trop déficitaires et mal répartis en distances pour permettre des analyses DS au moins correctes),

- les densités estimées sur **10mn** sont de **meilleure qualité** combinée et **plus fortes que** sur **5mn** (mais les 2 jeux de données 5mn et 10mn n'étant pas indépendants, la comparaison statistique des densités correspondantes reste sujette à caution),
- les densités estimées par la méthode **Papyrus** sont de **meilleure qualité** combinée, et **plus fortes de 30 à 75%** comparées à celles obtenues avec la méthode Naturalist,
- les densités estimées par **cumul des jeux Naturalist et Papyrus** sont d'une encore **meilleure qualité** combinée, mais légèrement inférieures en valeurs (quelques pourcents) au densités Papyrus,
- **l'impression** pré-existante à cette étude de **plus grande difficulté** à effectuer des inventaires de terrain via un smartphone (appli. **Naturalist**) que via des plans imprimés (Papyrus) semble **confirmée** par les faits sur une durée de **5mn** (durée choisie par le MNHN pour le protocole EPOC proche de celui que nous avons utilisé) : non seulement le jeu de données Naturalist 5mn est déficitaire en données comparé au jeu Papyrus 5mn, mais il comporte un **déficit important de données à courtes distances** (source probable de la forte sous-estimation des densités obtenues) que nous ne pouvons expliquer qu'en invoquant le débordement ressenti par les observateurs dans les premières minutes (les observateurs Papyrus l'on beaucoup moins ressenti) ; c'est par ailleurs un **problème** notable, puisque **l'hypothèse n°1 du Distance Sampling** (probabilité de détection = 1 à distance 0) semble ne **pas** être **respectée** ; ajoutons que l'indicateur de qualité combinée "automatique" gagnerait à être amélioré de manière à mettre en lumière ce biais important.

Ces constatations nous amènent donc à **retenir** comme **résultats ornithologiques** de l'enquête ACDC les **densités** estimées en **cumulant** les jeux de **données Naturalist et Papyrus** des inventaires de **10mn**.

Notons toutefois que :

- le choix entre des densités légèrement plus fortes du jeu Papyrus 10mn et la qualité combinée globale légèrement meilleure du jeu "cumulé" n'est pas si évident : faute de meilleure idée, nous avons tranché en faveur du meilleur critère de qualité combinée, et donc de la meilleure qualité statistique des modélisations. Mais nous ne savons en fait pas dire si les densités obtenues avec Papyrus seul sont moins proches de la réalité ;
- les analyses effectuées sur le jeu de données fusionné (Naturalist + Papyrus) n'ont pas tenu compte du fait que les 2 jeux de données n'avaient pas été collectés strictement de la même manière ; faute de connaître et de savoir appliquer cette technique, nous n'avons pas introduit la co-variable "méthode d'inventaire" dans les paramètres de MCDS ; et nous ne connaissons pas l'impact que cela aurait eu sur les résultats. Cependant, le faible écart avec les densités "Papyrus 10mn", notre second choix, et le constat que les densités "Naturalist + Papyrus 10mn" (même entachées de cette erreur méthodologique) apparaissent de fait comme une moyenne pondérée par l'indicateur de qualité entre les densités "Naturalist 10mn" et les densités "Naturalist + Papyrus 10mn", nous rassurent quant à l'ampleur de l'erreur commise sur les densités : probablement faible (quelques pourcents tout au plus).

Si l'on s'intéresse uniquement à la **durée des inventaires**, en laissant de côté le problème de non indépendance statistique des jeux de données analysés (ce qui est **très discutable**), on peut également en déduire que si l'enquête avait été effectuée uniquement avec des EPOC (protocole du MNHN), c'est à dire des **inventaires Naturalist de 5mn**, les **densités** obtenues en Distance Sampling auraient été de moins bonne qualité statistique, mais surtout de **30 à 75% plus faibles** selon les espèces !

Ce constat, qui mériterait toutefois d'être vérifié par d'autres enquêtes similaires, semble poser question quant à l'exploitation des données des EPOC selon la méthode du Distance Sampling !

Mais ce n'est sans doute pas un hasard si, pour l'Atlas des Oiseaux de France 2019 - 2024 (Cf. Annexe 1), la LPO et le MNHN ont ajouté pour le volet quantitatif de l'enquête ... un protocole "**EPOC OdF**" à base d'EPOC ... à **3 répliques** : outre le fait que ces EPOC OdF sont, comme pour la présente enquête, à localisation imposée et "aléatoire", la présence de ces 3 répliques vise sans doute à gommer :

- le **biais de sous-détection des oiseaux qui chantent moins fréquemment**, biais que nous soupçonnions et qui nous a conduit à allonger les inventaires à 10mn,
- le **biais de débordement de l'observateur** : il demeure lors du 1er EPOC, mais se trouve probablement fortement atténué lors des répliques, l'observateur mémorisant de mieux en mieux la répartition des chanteurs autour du point d'écoute à mesure de le temps passe, la plupart d'entre eux ne se déplaçant probablement que très peu.

Il est également probable que ces répliques permettront d'appliquer d'autres méthodes d'exploitation des données : mais cela est une autre histoire.

A posteriori, on peut donc regretter de ne pas avoir pensé à faire 2 répliques plutôt que de prolonger chaque inventaire de 5mn jusqu'à 10mn ; car outre l'atténuation des 2 biais évoqués ci-dessus, cette technique des répliques

- élimine le problème de non indépendance des échantillons de 5mn et de 10mn (puisque les données du premier sont un sous-ensemble des données du second), problème qui limite fortement, dans la théorie statistique, les possibilités de comparaisons des résultats produits avec ces 2 échantillons,
- diminue fortement le risque de double comptage lors d'inventaires de 10mn (risque soupçonné, mais difficile à évaluer objectivement).

Au vu de la qualité statistique des analyses sur les jeux de données des inventaires de 10mn, on peut également avoir un autre regret : celui d'avoir utilisé des critères trop drastiques pour la sélection des échantillons ; on aurait sans doute pu obtenir des estimations de densités de bonne qualité pour d'autres espèces intéressantes dans le milieu naturel ciblé (*Hippolais polyglotta*, *Sylvia borin*, *Picus viridis*, *Carduelis carduelis*, *Coturnix coturnix*, *Pica pica*). *Oriolus oriolus*, espèce retenue par "erreur" au mépris des critères édictés, montre que cet espoir n'est sans doute pas vain !

Mais les données brutes de l'étude sont à disposition de tous : avis aux amateurs (Cf. Annexe 1)

!

Remerciements

Ils vont d'abord naturellement aux 20 observateurs de terrain sans l'investissement de qui cette étude n'aurait simplement pas été possible :

- équipe Papyrus : Alex Clamens, Anne Citron, Camille Fasolin, François Guélin, Gilles Saulas, Jean-François Carrias, Jean-Pierre Dulphy, Sandra Robert, Sylvain Sainnier, Thibault Brugerolle,
- équipe Naturalist : Clément Rollant, Cyril Brunel, Cyrille Jallageas, Hugo Samain, Jean-Philippe Meuret, Matthieu Clément, Patrick Mougel, Pierre Turret, Romain Riols, Sandra Robert.

Qu'ils soient chaleureusement remerciés pour ces quelques **350 heures** passées et **3500 km de trajets** (points d'écoute, déplacements et saisie des données Papyrus), et aussi pour leurs compétences ornithologiques pas si courantes (les oiseaux à l'oreille !) qui ont permis l'exécution de protocoles de terrain exigeants.

Merci ensuite à François Guélin, initiateur des études de Distance Sampling en Auvergne, pour son travail de codirection de cette étude, pour ses encouragements et son aide tout au long de son déroulement, pour ses conseils lors de l'exploitation des données et la rédaction de cet article, et enfin pour son travail de relecture.

Merci enfin à Jean-Pierre Dulphy et Jean Nabias pour leur relecture attentive et leurs commentaires et critiques bienvenus.

Bibliographie

[DULPHY 1994] DULPHY J.-P., 1994. - Etude de l'avifaune par I.P.A. sur les plateaux granitiques au sud de Clermont-Ferrand. *Le Grand-Duc*, 44 : 12-17

[DULPHY 2012] DULPHY J.-P., 2012. - Etude de l'avifaune sur les plateaux granitiques au sud de Clermont-Ferrand : 20 ans après ! *Le Grand-Duc*, 80 : 43-49

[DULPHY et al. 2023] DULPHY J.-P., GUÉLIN F., 2023. - L'avifaune de plateaux agricoles granitiques au sud de Clermont-Ferrand : trente ans après, estimation des populations de quelques espèces des années 1990 par Distance Sampling et Indices Ponctuels d'Abondance. *Le Grand-Duc*, 91 : 84-96

[GUÉLIN 2015] GUÉLIN F., 2015. - Estimation des populations d'espèces d'oiseaux prairiales (Pipits, Alouettes) du massif du Sancy par différentes méthodes d'extrapolation. *Le Grand-Duc*, 83 : 2-23

[GUÉLIN 2016] GUÉLIN F., 2016. - Test de la méthode du Distance Sampling pour évaluer les populations de pipits et d'alouettes du plateau du Guéry dans le Sancy (63). *Le Grand-Duc*, 84 : 77-85

[GUÉLIN et al. 2018] GUÉLIN F., CARRIAS J.-F., CLAMENS A., MEURET J.-P. & SAULAS G., 2018 - Évaluation des populations de passereaux communs nicheurs d'un plateau agricole du Puy-de-Dôme (France) par la méthode du « distance sampling ». *Alauda* 86 (4) : 279-298

ANNEXES

ANNEXE 1 : Liens

- le **STOC-EPS** est le volet "points d'écoute" du programme de Suivi Temporel des Oiseaux Communs coordonné par le MNHN, la LPO et l'OFB : <https://www.vigienature.fr/fr/suivi-temporel-des-oiseaux-communs-stoc>
- les EPOC et EPOC-OdF sont 2 protocoles "points d'écoute" du projet d'Atlas des Oiseaux de France porté par la LPO en partenariat scientifique avec le MNHN et l'OFB : <https://www.oiseauxdefrance.org/get-involved>
- **Naturalist** (https://www.faune-france.org/index.php?m_id=20015) est une application pour smartphone
 - Android : <https://play.google.com/store/apps/details?id=ch.biolovision.naturalist>,
 - et Apple : <https://apps.apple.com/kh/app/naturalist/id1175280268> (plus limitée),éditée et maintenue par **Biolovision Sàrl** <https://www.biolovision.net/> sur commande depuis les années 2000 de multiples organismes et associations d'étude et de protection de la nature en Europe, dont la Ligue pour la Protection de Oiseaux en France
- **Distance** (<http://distancesampling.org/>) est un logiciel fonctionnant sous **Windows** et permettant de mener des analyses selon de multiples variantes de "la" méthode du **Distance Sampling** (https://en.wikipedia.org/wiki/Distance_sampling) ; il fournit une interface homme / machine et des outils de pré et post-traitement facilitant beaucoup l'utilisation de "moteurs de calcul en ligne de commande", qui effectuent - eux - les calculs proprement dit une fois les données d'entrée mises en forme ; l'un de ces moteur de calcul est "**MCDS**", qui implémente la variante historique "Conventional Distance Sampling" de la méthode (avec co-variable ou pas)
- **toutes les données des inventaires** (anonymisées, individualisées, avec les distances observateur - oiseau), la cartographie, les paramètres d'analyse, les rapports de synthèse, les notebooks Python utilisés et les fichiers sources des tableaux et figures de ce document ... **sont en accès libre** : <https://sylbor63.pagesperso-orange.fr/ds/acdc19/materiau-public.zip> (pour plus de détails, pour refaire les analyses et vérifier, ou faire d'autres analyses ... etc)
- **pyaudisam** (<https://pypi.org/project/pyaudisam/>, <https://github.com/denmedius/pyaudisam>), alias "**Python Automated Distance Sampling**", est un module et un programme écrit avec le langage **Python 3** (<https://www.python.org/>) pour l'automatisation des analyses en Distance Sampling via MCDS et de leur dépouillement
- **SRACOS** : l'algorithme d'**optimisation sans dérivation** utilisé par pyaudisam (via le module zoopt <https://github.com/polixir/ZOOpt>) pour le calcul automatique des distances de troncature à gauche et à droite, et le nombre de tranches de distances pour la modélisation dans MCDS <http://lamda.nju.edu.cn/yuy/GetFile.aspx?File=papers/aaai17-sracos-with-appendix.pdf>)

ANNEXE 2 : Nombres de données par espèce, durée, méthode, passage**Tableau A2-1** : Nombres de données par espèce (toutes), passage et durée d'inventaire
(source : classeur ACDC2019-bilan-final.ods, feuille "EspècesTousAd")

Méthode	Papyrus				Naturalist			
	a		b		a		b	
	05mn	10mn	05mn	10mn	05mn	10mn	05mn	10mn ▼
Turdus merula	159	213	191	242	122	233	141	243
Sylvia atricapilla	177	217	171	221	130	189	147	224
Alauda arvensis	81	102	95	124	83	113	81	135
Sylvia communis	92	121	88	126	69	102	62	112
Lanius collurio	0	1	73	103	0	0	48	98
Columba palumbus	52	77	56	87	62	104	59	88
Luscinia megarhynchos	72	97	87	116	50	72	58	87
Fringilla coelebs	53	67	86	99	39	72	53	84
Cuculus canorus	62	85	61	79	42	67	53	77
Emberiza cirrus	31	41	38	49	38	64	40	67
Phylloscopus collybita	66	84	39	52	46	78	41	60
Lullula arborea	37	54	41	57	37	71	40	59
Passer domesticus	11	11	5	5	28	47	48	57
Parus major	52	91	46	64	49	98	27	55
Corvus corone	7	8	17	22	20	38	33	48
Erithacus rubecula	27	34	23	29	30	59	26	41
Emberiza citrinella	19	26	20	28	20	35	29	41
Streptopelia turtur	1	3	47	51	4	4	19	35
Saxicola rubicola	17	24	25	29	23	47	25	34
Cyanistes caeruleus	17	27	8	17	25	48	15	31
Carduelis cannabina	21	29	13	24	19	34	15	29
Emberiza calandra	10	13	17	17	11	20	21	26
Turdus philomelos	21	24	24	30	19	35	11	23
Sturnus vulgaris	3	10	9	25	20	35	11	22
Dendrocopos major	22	34	8	14	14	27	10	22
Turdus viscivorus	15	21	14	16	26	42	12	20
Garrulus glandarius	27	33	9	11	22	35	11	20
Prunella modularis	22	37	10	11	15	29	7	19
Phylloscopus bonelli	0	3	4	8	14	18	15	19
Hippolais polyglotta	0	0	16	20	0	1	12	18
Troglodytes troglodytes	5	11	9	10	6	17	11	16
Upupa epops	14	17	10	16	5	9	11	15
Carduelis carduelis	7	7	6	6	7	9	10	15

Anthus trivialis	18	22	10	12	9	16	10	14
Pica pica	8	9	7	9	10	15	6	14
Carduelis chloris	21	26	21	27	12	23	7	13
Streptopelia decaocto	10	16	12	14	10	12	7	13
Picus viridis	4	7	5	9	7	13	6	13
Oriolus oriolus	13	17	10	16	14	19	4	11
Sylvia borin	3	5	7	9	2	2	7	11
Jynx torquilla	10	19	9	10	17	21	4	10
Coturnix coturnix	0	0	14	16	0	1	5	8
Serinus serinus	2	3	0	2	2	4	4	6
Hirundo rustica	1	1	0	0	4	10	4	5
Certhia brachydactyla	5	6	2	7	4	12	0	4
Buteo buteo	0	0	2	2	1	5	3	4
Regulus ignicapilla	0	0	0	1	0	0	2	4
Apus apus					0	0	3	4
Aegithalos caudatus	1	1	3	3	0	5	2	3
Phoenicurus ochruros	5	5	1	2	5	8	2	3
Milvus migrans	1	1	1	1	3	12	2	3
Alectoris rufa	0	0	1	1	0	0	2	3
Falco tinnunculus	1	2	0	0	0	3	1	3
Sitta europaea	3	3	0	0	1	2	2	3
Phasianus colchicus	1	1	0	0	0	1	3	3
Milvus milvus					5	7	1	3
Coccothraustes coccothraustes					1	1	0	3
Columba oenas					0	0	3	3
Motacilla flava	2	3	5	5	1	1	2	2
Regulus regulus	1	1	0	0	0	0	2	2
Motacilla alba	0	0	3	3	2	3	1	1
Corvus corax	0	0	1	1	3	4	0	1
Phylloscopus trochilus	4	7	0	0	1	5	1	1
Pyrrhula pyrrhula	3	3	0	0	3	4	0	1
Dendrocopos minor	1	1	0	0	1	2	0	1
Periparus ater					2	3	1	1
Circaetus gallicus					1	2	1	1
Emberiza cia					1	1	0	1
Poecile palustris					1	1	0	1
Circus pygargus					0	0	0	1
Pernis apivorus					0	0	1	1
Strix aluco					0	0	0	1
Anthus pratensis	1	1	0	0	8	10	0	0
Dryocopus martius	1	1	0	0	0	7	0	0

Locustella naevia	0	1	0	0	1	1	0	0
Lophophanes cristatus					1	4	0	0
Ardea cinerea					2	2	0	0
Phylloscopus sibilatrix					1	1	0	0
Columba livia f. domestica	4	4	9	9				
Corvus monedula	1	1	2	2				
Acrocephalus scirpaceus	0	0	1	1				
Muscicapa striata	0	0	0	1				
Phoenicurus phoenicurus	0	0	0	1				
Saxicola rubetra	0	0	1	1				
Dendrocopos medius	1	1	0	0				
Otus scops	0	1	0	0				

ANNEXE 3: Fonctionnement résumé de pyaudisam

Appliqué à notre étude, le fonctionnement du logiciel "**pyaudisam**" peut être résumé de la manière suivante.

Dans le **premier mode** dit de "**pré-analyse**",

- en entrée, on lui fournit
 - l'ensemble choisi des **données individualisées** (données Papyrus, données Naturalist, ou la fusion des 2 ensembles), chaque donnée précisant : l'espèce, le passage saisonnier concerné (1er ou 2nd), sa présence / absence dans les 5 premières minutes, ou dans les 5 suivantes, le "sexe et âge" de l'individu contacté (Mâle adulte, autre adulte = femelle ou indéterminé, ou juvénile), et la distance observateur - oiseau
 - la liste des **échantillons à pré-analyser** : les combinaisons espèce x durée d'inventaire (5mn ou 10mn) x sexe et âge (Mâle adulte, autre adulte, ou juvénile) x passage
- le logiciel effectue ensuite une série d'analyses simples pour chacun de ces échantillons, sans aucune troncature, en suivant la stratégie suivante : essayer successivement les modèles disponibles et leurs séries d'ajustement associées, dans un ordre spécifié, et en s'arrêtant à la première analyse qui aboutit sans erreur (la série d'analyses effectuées est donc souvent limitée à une seule) ; si aucune analyse n'aboutit, on considère que l'échantillon est simplement inexploitable en Distance Sampling
- en sortie, il produit un rapport de synthèse présentant l'essentiel des résultats chiffrés et graphiques de l'analyse faite (avec succès) pour chaque échantillon, sous la forme d'un tableau à 1 ligne par échantillon (Cf. Annexe 4) ; dans ce tableau, le principal, c'est l'histogramme des distances (en fait plusieurs superposés, avec différentes largeurs de tranches) et les courbes de modélisations calculées par MCDS, qui permettent assez rapidement d'évaluer pour chaque échantillon :
 - les chances d'obtenir des densités de bonne qualité statistique,
 - éventuellement les troncatures spécifiques qu'il serait raisonnable / intéressant d'essayer lorsque l'on va vraiment lancer les analyses avec de multiples combinaisons de paramètres (troncature à gauche en particulier, par ex. pour les espèces à faible nombre de données à courte distance : *Alauda arvensis*, *Prunella modularis*, ...),

Dans le **second mode** dit "**d'analyse**",

- en entrée, on lui fournit
 - l'ensemble choisi des **données individualisées** (le même que pour le mode "pré-analyse" : les données des échantillons choisis en seront extraites automatiquement)
 - la liste des **échantillons à analyser** : comme dans le mode "pré-analyse", mais à priori un sous-ensemble, les pré-analyses étant faites pour filtrer les échantillons non exploitables,
 - les **combinaisons de paramètres d'analyse** à couvrir pour chaque échantillon : cette fois, c'est différent du mode pré-analyse, il faut spécifier ce que l'on veut faire : modèle x série d'ajustement x troncature en distance à gauche x troncature en distance à droite x nombre de tranches de distance à considérer pour la modélisation (N.B. pour ces 3 derniers paramètres, un mode "automatique" permet de ne pas avoir à spécifier les valeurs, et de laisser le logiciel les chercher lui-même par optimisation numérique de la qualité statistique des résultats de l'analyse DS),
- le logiciel enchaîne ensuite automatiquement les grandes étapes suivantes :
 - pour les combinaisons de paramètres "**à troncatures automatiques**", détermination automatique des distances de troncatures à gauche et à droite par une technique

- d'optimisation** (lancement de nombreuses analyses en faisant varier les dites troncatures dans des limites paramétrées par le nombre de données ignorées, considérées comme des "outliers", ... et en ne retenant que le meilleur jeu de troncatures (à gauche et/ou à droite) selon un **indicateur / critère de "qualité combinée"** à partir des indicateurs produits par MCDS et des paramètres d'analyses (voir ci-dessous pour un 1er survol, et à l'annexe 5 pour tous les détails) ;
- exécution des analyses individuelles correspondant à chacune des combinaisons de paramètres (à troncatures maintenant toutes fixées) pour chaque échantillon : comme lors d'analyses manuelles via le logiciel Distance 7, c'est son "moteur" de calcul MCDS V6 qui est utilisé, en mode "Conventional Distance Sampling" (pas de co-variable) ; les résultats ainsi produits alimentent ainsi une table dont chaque ligne correspond à une analyse, c'est à dire à l'une des combinaisons de paramètres pour chaque échantillon ; ces résultats sont ceux produits directement par MCDS V6, et complétés par l'indicateur de "qualité combinée" spécifique à pyaudisam
 - tri et filtrage automatique des résultats selon leur "qualité combinée" : pour chaque échantillon, suivant une suite d'étapes proche de ce que ferait "manuellement" un analyste humain cherchant le "meilleur résultat" d'analyse parmi ceux correspondant à de multiples combinaisons de paramètres,
 - ne conserver que les analyses à **résultats non "identiques", à un epsilon près** (pour les critères statistiques bruts en sortie de MCDS, les densités et probabilité de détection calculées, ainsi que leurs intervalles de confiance)
 - ne conserver que les analyses **sans erreur** rapportée par MCDS (analyses à "warnings" conservées),
 - pour les analyses à **troncatures strictement identiques**, ne conserver que les **N** ayant les **meilleurs AIC** (N paramétrable),
 - pour les analyses à **troncatures proches**, ne conserver que celle (**1 seule**) ayant le **meilleur Khi2**, le **meilleur KS** à Khi2 identique, le **meilleur DCv** à Khi2 et KS identiques (tests du Khi2 et de KS = Kolmogorov-Smirnov, DCv = variabilité à 95% de la densité estimée),
 - ne conserver que les **analyses "tronquées en distance" à moins de X %** (i.e. pour lesquelles les troncatures en distance à gauche et/ou à droite n'ont pas éliminé plus de X % des oiseaux contactés au total) (X paramétrable),
 - ne conserver que les analyses restantes ayant les **meilleurs critères de "qualité combinée"**, pourvu qu'ils dépassent la valeur minimale Y (paramétrable),
 - ne conserver enfin que les (au plus) **P** (paramétrable) meilleures analyses, pour limiter la taille du rapport, mais en conservant tout de même suffisamment de matière pour manuellement vérifier la sélection automatique et la corriger si besoin (Cf. "dernière étape manuelle" ci-dessous) ;
 - en sortie, il produit un rapport (Cf. Annexe 6) présentant notamment une synthèse des résultats chiffrés et graphiques de chaque analyse conservée lors du filtrage ci-dessus, sous la forme d'un tableau de synthèse (trié par échantillon et dans l'ordre décroissant du critère de "qualité combinée") ;
 - la **dernière étape, manuelle** celle-ci, consiste à examiner, pour chaque échantillon, les résultats des quelques "meilleures" analyses retenues, et à choisir laquelle conserver finalement : le filtrage automatique décrit ci-dessus n'étant pas parfait, il convient de vérifier systématiquement que la "meilleure" analyse automatiquement proposée est effectivement "meilleure" que les suivantes du classement ; pour la très grande majorité des échantillons, c'est bien le cas, et de toute manière, les écarts de critères qualité et de résultats sont faibles dans ce classement ; mais parfois, l'analyste humain doit préférer une autre "meilleure"

analyse, pour corriger les imperfections et les manques du logiciel, en lui-même, mais aussi vis à vis des biais et manques du côté des données de terrain, grâce à sa connaissance du secteur géographique inventorié, des milieux naturels présents et des espèces ciblées, lui donnent des arguments pour parfois corriger l'automatisme ; les cas suivants semblent les plus fréquents :

- l'histogramme des distance n'est pas régulier, présente des "trous" et / ou des "bosses", et le logiciel a sélectionné des analyses où il y a manifestement sur-modélisation, c'est à dire une modélisation qui cherche à s'approcher au mieux de l'enveloppe de l'histogramme, alors que ces "trous" et / ou "bosses" devraient être ignorées,
- préférence naturelle pour une modélisation en demi-normale,
- ...

Soulignons l'importance, dans le processus de tri et filtrage automatique décrit ci-dessus, de **l'indicateur / critère "qualité combinée"** calculé à partir des indicateurs produits par MCDS et des paramètres d'analyses (Cf. Annexe 5 pour tous les détails) : construit dans l'objectif idéal d'automatiquement orienter autant que possible la sélection des meilleures analyses, il simule en quelque sorte et de manière très simple les préférences d'un analyste humain expérimenté (selon l'état actuel de nos connaissances pratiques du Distance Sampling) :

- le modèle Half-Normal est préférable, suivi de peu par le modèle Uniform ; le modèle Hazard-Rate est moins recherché (pour s'éloigner souvent des réalités de terrain ornithologique : "épaule" longue à courte distance),
- la sur-modélisation ("over-fitting") des histogrammes de distance est à éviter autant que possible, donc les analyses avec un minimum de paramètres d'ajustement du modèle sont préférées,
- les troncatures éliminant peu ou pas de données sont préférées,
- les analyses à faible variabilité de l'estimation de densité (intervalle de confiance à 95%) sont préférées.

ANNEXE 4: Extrait d'un rapport de synthèse des pré-analyses de pyaudisam

Les rapports de synthèse des pré-analyses produits par pyaudisam présentent, pour chaque échantillon "pré-analysé" (espèce, passage(s) : a, b ou a+b, durée d'inventaire : 5mn ou 10mn, type d'effectif : m, ou m+a) :

- un tableau à 3 colonnes permettant d'identifier l'échantillon, la pré-analyse, les paramètres de la première modélisation DS ayant fonctionné pour cet échantillon, quelques chiffres-clés décrivant l'échantillon (nombre d'individus considéré, distance max. de contact, effort DS), et enfin les résultats principaux de la pré-analyse (densité estimée et intervalle de confiance à 95%, nombre estimé et intervalle de confiance à 95%, EDR,) ainsi que quelques uns des indicateurs habituels de qualité statistique de la modélisation DS (AIC, Khi2, KS, CoefVar Densité)
- 3 graphiques permettant de juger rapidement "à l'oeil" de l'exploitabilité de l'échantillon (à gauche : histogrammes des distances de contact, avec 3 largeurs différentes de tranches de distances), et de la qualité de cette première modélisation DS ayant fonctionné (au milieu : densité de probabilité de détection modélisée, superposée à l'histogramme des données de terrain, le tout en fonction de la distance ; à droite : intégrale correspondante, modélisée, superposée à l'histogramme des données de terrain, le tout en fonction de la distance).

Les unités sont précisées en entête du rapport (non visibles dans l'extrait ci-dessous : ici, distances en m, et surfaces en km²).

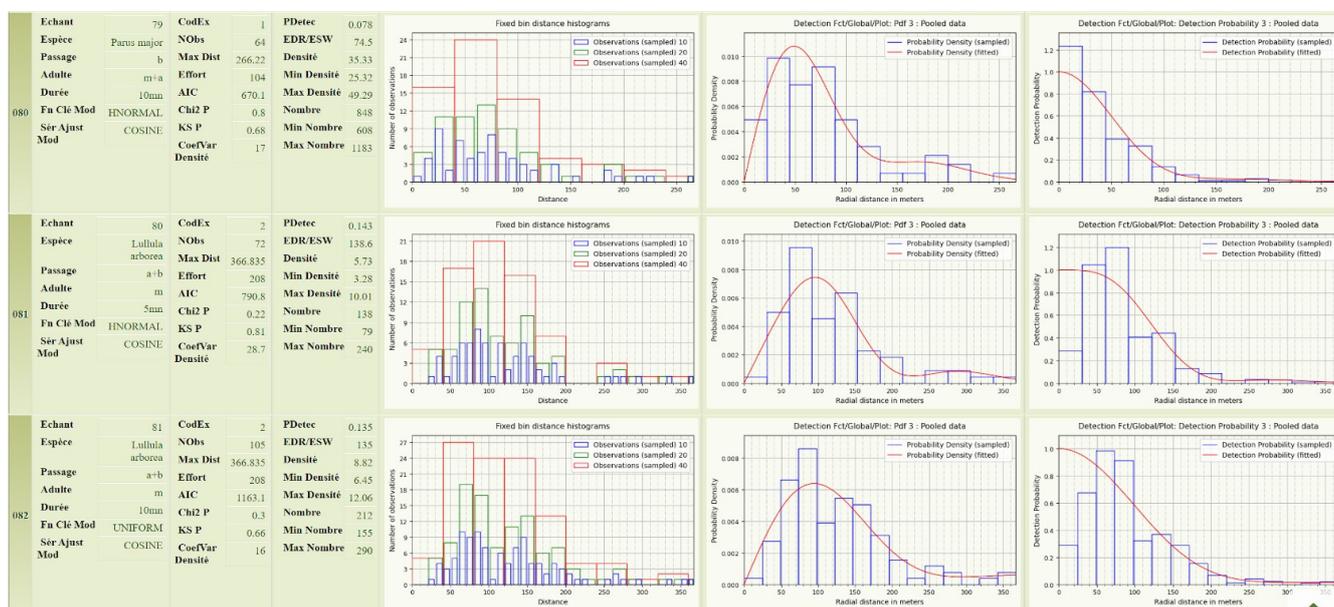


Figure A4-1 : Extrait pour l'exemple d'un rapport de pré-analyse pyaudisam : 3 lignes du rapport, pour 3 échantillons et 2 espèces

En cliquant dans la colonne de gauche (sur le numéro de ligne du tableau), on accède à une page de détails de la pré-analyse concernée, en tout points identique à ce que le logiciel Distance 7 produirait (il s'agit d'ailleurs de l'intégralité des informations produites par MCDS, comme dans Distance 7).

ANNEXE 5: critère de qualité "combinée"

Ce critère, calculé par pyaudisam entre 0 (mauvais) et 1 (excellent), combine après normalisation si nécessaire, certains paramètres d'analyse, ainsi que les critères et résultats bruts produits par MCDS :

Tableau A5-1 : *Détail sur la formule d'obtention du "critère de qualité combinée", des composantes impliquées et de leur méthode d'obtention*

Critère / résultat MCDS	Méthode de normalisation (= passage dans [0, 1])
Test du Khi2 (Khi2)	déjà dans [0, 1]
Test de Kolmogorov-Smirnov (KS)	déjà dans [0, 1]
Test de Cramer-von-Mises à pondération uniforme (CvM-Uw)	déjà dans [0, 1]
Tests de Cramer-von-Mises à pondération cosinusoïde (CvM-Cw)	déjà dans [0, 1]
Taux d'individus (oiseaux) conservé après troncatures	simple division par 100
Modèle utilisé Half-Normal, Uniform ou Hazard-Rate (le 1er considéré comme le plus conforme à la réalité ornithologique de terrain, le 2ème un peu moins, le 3ème encore moins)	Half-Normal => 1,0 Uniform => 0,9 Hazard-Rate => 0,6
Variabilité à 95% de la densité d'oiseaux estimée (DCv)	$\exp(-63 * DCv^{2.8})$
Nombre de paramètres de la série d'ajustement du modèle (polynomiale ou cosinusoïde)	$\exp(-0.17 * NbParams^2)$

La valeur finale est le produit de ces 8 valeurs normalisées, lui-même élevé à la puissance 1/8 (conservation de dimensionnalité).

ANNEXE 6: Extrait d'un rapport d'analyses de pyaudisam

Les rapports de synthèse des analyses produits par pyaudisam présentent, pour chaque échantillon analysé (espèce, passage(s) : a, b ou a+b, durée d'inventaire : 5mn ou 10mn, type d'effectif : m, ou m+a), les "meilleures" analyses à l'aune de l'indicateur de qualité "combinée" (Cf. annexe 5).

Ils ont une forme assez similaire aux rapports de pré-analyse (Cf. Annexe 4), si ce n'est que :

- on n'a plus 1 ligne par échantillon pré-analysé (toutes espèces contactées, tous passages, ...), mais 10 lignes par échantillon sélectionné (les 30 espèces, l'unique combinaison de passages b ou a+b, l'unique combinaison de type d'effectif m ou m+a, les 2 combinaisons de durées 5mn et 10mn), correspondant au 10 meilleures analyses,
- par ligne, le tableau à 3 colonnes donne plus de détails : en particulier les troncatures utilisées, et l'indicateur de qualité "combinée" ("Qual Equi 3" ici), critère de classement des analyses par échantillon (la meilleure en premier, en haut)
- des 3 graphiques permettant de juger rapidement "à l'oeil" de la qualité des résultats de l'analyse, le premier, à gauche a été remplacé logiquement par le diagramme quantile - quantile de comparaison modélisation - données de terrain ; pas de changement en revanche au milieu et à droite : respectivement : densité de probabilité de détection modélisée, superposée à l'histogramme des données de terrain, le tout en fonction de la distance, et à droite : intégrale correspondante, modélisée, superposée à l'histogramme des données de terrain, le tout en fonction de la distance).

De même, les unités sont précisées en entête du rapport (non visibles dans l'extrait ci-dessous : ici, distances en m, et surfaces en km²).

En cliquant dans la colonne de gauche (sur le numéro de ligne du tableau), on accède à une page de détails de la pré-analyse concernée, en tout points identique à ce que le logiciel Distance 7 produirait (il s'agit d'ailleurs de l'intégralité des informations produites par MCDS, comme dans Distance 7).

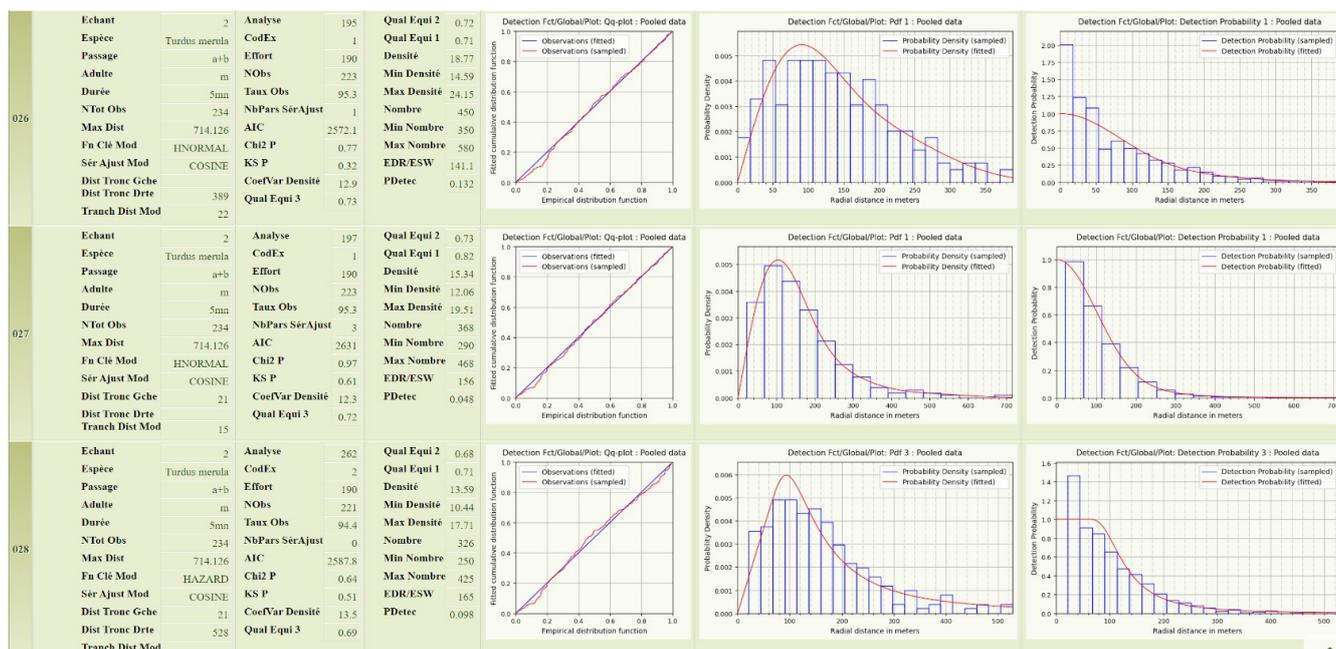


Figure A6-1 : Extrait pour l'exemple d'un rapport d'analyse pyaudisam : 3 lignes du rapport, pour les 3 meilleures analyses de l'échantillon Turdus merula Mâles Naturalist 5mn

Sous ce tableau de synthèse, 2 autres tableaux listant les analyses dans le même ordre présentent respectivement

- une sélection des colonnes de paramètres, indicateurs statistiques et résultats de chaque analyse (sélection enrichie comparée à celle du tableau de synthèse ci-dessus, soit environ 45 colonnes),
- la totalité de colonnes paramètres, indicateurs statistiques et résultats de chaque analyse, telles que produites par MCDS (environ 120 colonnes), en cas de besoin pointu.