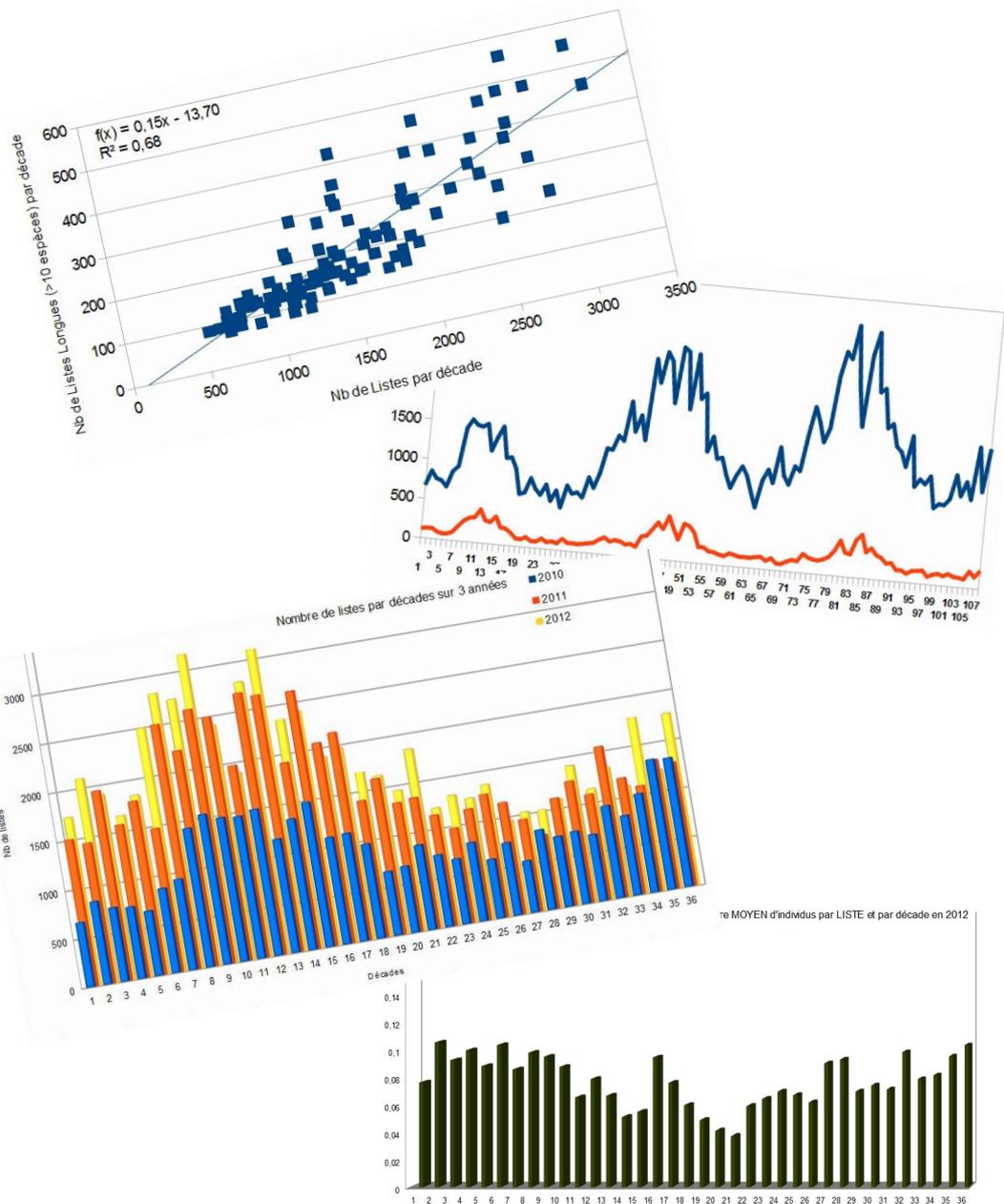




Méthodologie d'exploitation phénologique des données VisioNature.

F. Guélin

LE GRAND-DUC N°81 (ANNEE 2013)



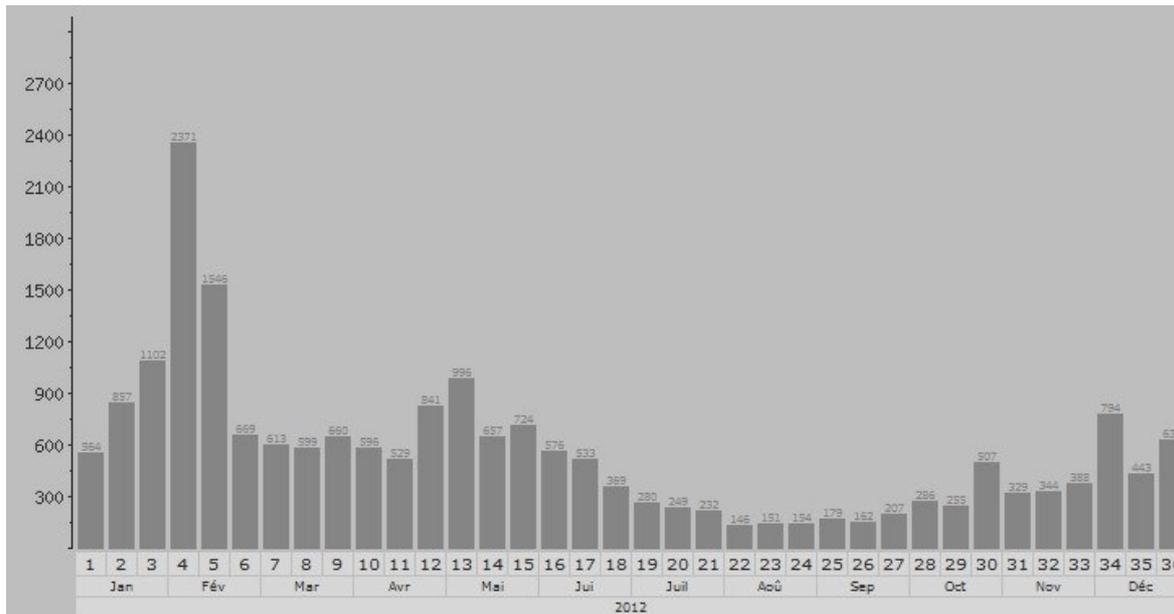
Préambule : l'arrivée dans les foyers ornithologiques du portail Internet VisioNature (www.faune-auvergne.org, abrégé ici « FA ») a révolutionné la récolte des données ornithologiques, et cette énorme masse de données offre des possibilités d'exploitation encore mal utilisées.

Les courbes automatiques de phénologie (« Quand la voir ? ») sont évidemment très pratiques, mais posent une question de fond : que représentent-elles et sont-elles exploitables de manière brute ?

Nous proposons un mode d'exploitation plus fiable des données phénologiques, en prenant en compte la pression d'observation.

Quels sont les biais d'interprétation des courbes phénologiques brutes ?

Prenons pour exemple, la courbe phénologique concernant le Merle noir avec cumul des individus par décade, sur l'année 2012, extraite de FA à partir d'une requête de tri multicritère (différente des courbes automatiques de la rubrique « Quand la voir ? » qui prennent en compte le nombre total de données sur toutes les années de la base, par décade):



L'augmentation régulière du nombre de données de janvier et février, également en mai, reflète-t-elle une meilleure détection (visuelle ou sonore) ? Une augmentation réelle du nombre d'oiseaux ? Une augmentation de la pression d'observation ? Ou bien une combinaison de ces trois paramètres ?

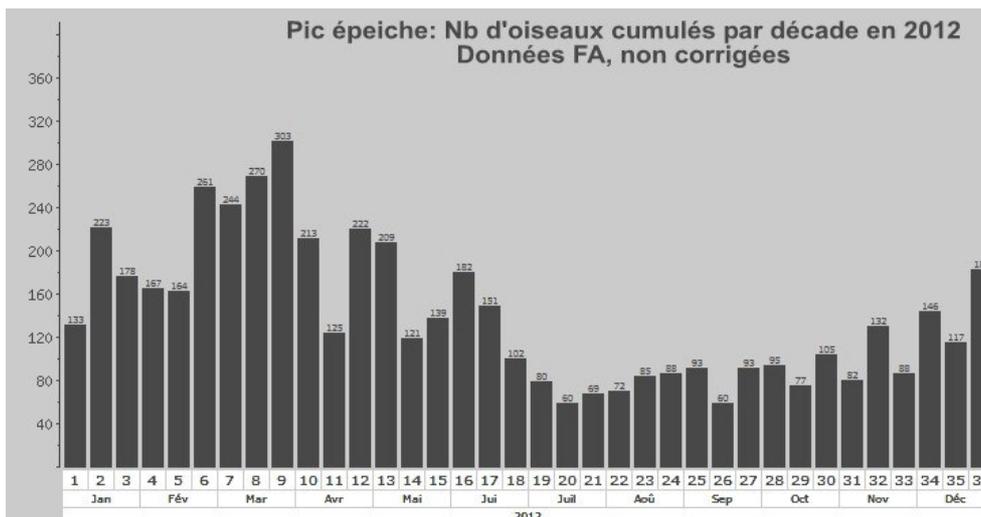
La différence de détection est un paramètre qu'on ne pourra pas étudier, car la saisie des données n'est pas conçue pour faire la distinction des différents modes de contact, notamment le chant par rapport aux autres contacts.

Mais la question de la part de la pression d'observation est plus accessible à l'analyse.

L'augmentation printanière (mars à mai) du nombre cumulé d'individus de Merle noir par décade est probablement très liée à une prospection accrue des naturalistes. En revanche, le nombre élevé en janvier et février est peut-être une conséquence du retour du chant (le métabolisme de l'ornithologue au sortir de l'hiver est saturé d'endorphines à l'écoute de ce chant et il saisit la donnée avec fébrilité: « premier chant Chez Moi ! »).

Les données de fin décembre sont du même tonneau (humm ...) : les dernières décades de l'année voient doubler le nombre de merles. Ce doublement des individus observés pendant les fêtes de fin d'année pourrait paraître également prévisible, conséquence d'un biais d'observation que tout le monde a en tête sans oser le formuler. Mais la notation des oiseaux dans les mangeoires de manière systématique est aussi une explication plus « sobre » :

Deuxième exemple : le Pic épeiche :



Cette espèce n'est pas choisie au hasard : elle est très sédentaire, et le nombre de données cumulées par décade devrait être assez constant (sauf au moment de l'envol des jeunes où le nombre d'individus devrait logiquement augmenter ?). L'augmentation du nombre de données des pics épeiches en mars est presque certainement un artefact (le Pic épeiche étant une des espèces les plus sédentaires de notre région), conséquence du biais de la pression d'observation, biais que nous devons essayer de supprimer ou de minorer.

Mode d'évaluation de la pression de prospection

La pression de prospection est évaluable : il s'agit de trouver une unité de mesure qui reflète la présence ornithologique sur le terrain. Nous n'utiliserons que la base de données concernant les oiseaux (pas celle des autres taxons vertébrés).

Cette présence ornithologique est mesurable par le nombre de « listes » d'oiseaux qui se retrouvent dans vos carnets de notes (virtuels ou papier) : une « liste » est une suite d'espèces observées à une date donnée, un endroit donné, par un observateur unique : par exemple, l'extrait de carnet de notes ci-contre comporte 6 listes différentes, de longueur variable. Il reflète assez fidèlement l'effort d'observation.

Nous avons donc en première étape calculé le nombre de listes par décades pour les années 2010, 2011 et 2012 (le site Faune-Auvergne.org a été ouvert début 2009).

Cette première étape est techniquement très longue :

→ extraction des données de chaque année depuis faune-auvergne, dans une base de données capable de traiter des centaines de milliers de données (Microsoft Access), car les tableurs classiques ne suffisent pas (nombre arrondi de données exploitables : 175 000 en 2010, 261 000 en 2011 et 277 000 en 2012).

→ traitement de ces centaines de milliers de données pour obtenir un tableau simple donnant, pour chacune des 36 décades annuelles, le nombre total de listes saisies par l'ensemble des ornithologues.

dimanche 6 janvier 2013

Le Jardin des Guélin / Roche-Blanche (La) (63)

- 6 Mésanges charbonnières (*Parus major*) [François Guélin]
- 1 Mésange noire (*Periparus ater*) [François Guélin]
- 10 Moineaux domestiques (*Passer domesticus*) [François Guélin]
- 1 Pie bavarde (*Pica pica*) [François Guélin]
- 1 Pinson des arbres (*Fringilla coelebs*) [François Guélin]
- 1 Rougegorge familier (*Erythacus rubecula*) [François Guélin]
- 1 Troglodyte mignon (*Troglodytes troglodytes*) [François Guélin]
- 8 Verdiers d'Europe (*Carduelis chloris*) [François Guélin]

Ravin de Macon / Roche-Blanche (La) (63)

- 1 Grive draine (*Turdus viscivorus*) [François Guélin]
Ma remarque : cht
- 1 Pic épeiche (*Dendrocopos major*) [François Guélin]
Mon code atlas : 5
- 2 Pigeons ramiers (*Columba palumbus*) [François Guélin]
Mon code atlas : 4

samedi 5 janvier 2013

Col de la Croix Saint Robert / Mont-Dore (63)

- 2 Grands Corbeaux (*Corvus corax*) [François Guélin]
Ma remarque : seule espèce d'oiseau sur 5 km de montée ! malgré un temps splendide et peu de neige.

la Guîze / Chambon-sur-Lac (63)

- 1 Mésange huppée (*Lophophanes cristatus*) [François Guélin]

vendredi 4 janvier 2013

Bessay-sur-Allier (rue de la gare) / Bessay-sur-Allier (03)

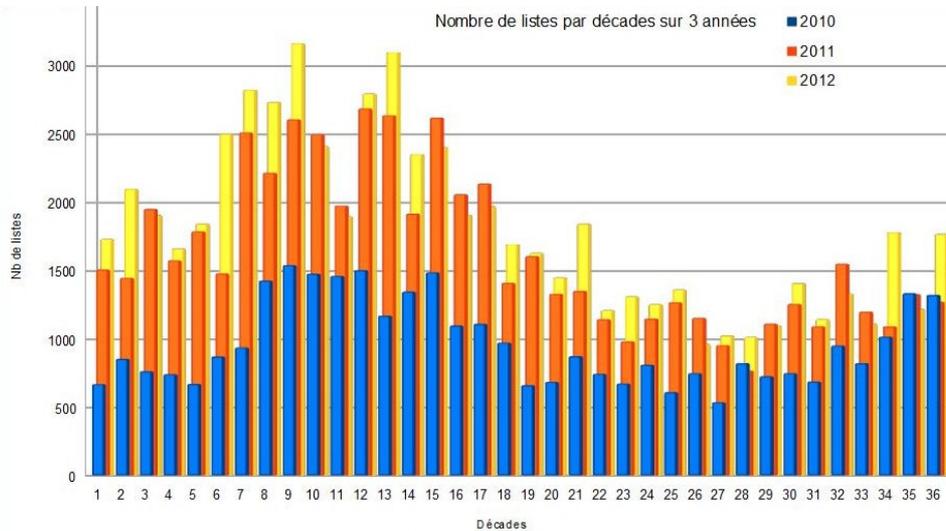
- 200 Corbeaux freux (*Corvus frugilegus*) [François Guélin]
Ma remarque : sur la colonie

le Pacage (O) (RNVA) / Ferté-Hauterive (La) (03)

- 800 Bernaches du Canada (*Branta canadensis*) [François Guélin]
- 1 Bouvreuil pivoine (*Pyrrhula pyrrhula*) [François Guélin]
- 5 Bruants des roseaux (*Emberiza schoeniclus*) [François Guélin]
- 1 Buse variable (*Buteo buteo*) [François Guélin]
- 50 Canards colverts (*Anas platyrhynchos*) [François Guélin]
- 20 Chardonnerets élégants (*Carduelis carduelis*) [François Guélin]
- 45 Corneilles noires (*Corvus corone corone*) [François Guélin]

La fiche technique de ces différentes étapes informatiques est fournie en annexe pour ceux et celles qui voudraient s'amuser intellectuellement quelques dizaines d'heures... (il faut profiter d'une immobilisation forcée, type gastro-entérite ou grippe....).

Le RESULTAT sous forme graphique montre l'activité ornithologique annuelle :



Ces courbes offrent une perspective intéressante de travail ethnologique (ou sociologique ?) sur l'activité naturaliste en Auvergne. Bien sûr, il y a des variations de l'activité de prospection tout à fait prévisibles : le printemps voit presque doubler l'effort de terrain. Les deux dernières décades de l'année sont aussi les témoins d'une prospection accrue, liée probablement à un temps libre plus important mais surtout au fait que la dernière décade n'en est pas une car elle inclut les cinq derniers jours de l'année en plus (la 37ème décade) : au lieu de 10 jours, elle en fait donc quinze !

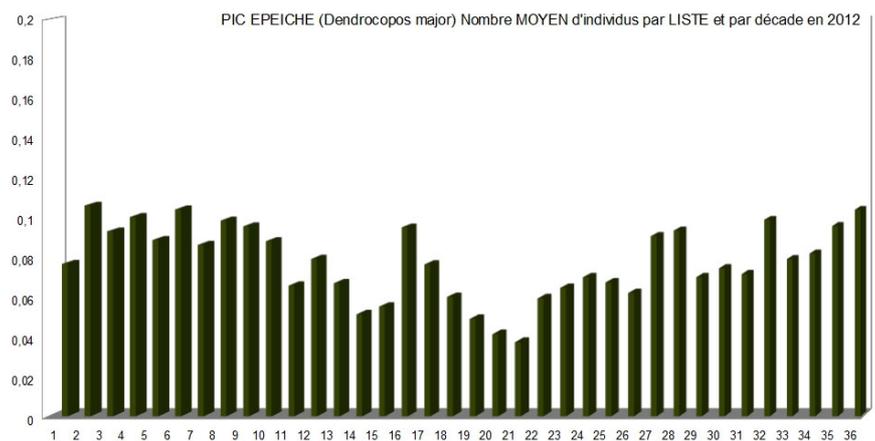
Nous proposons donc une nouvelle unité de mesure de l'intensité de prospection : la LISTE.

Les graphes (nombre cumulé d'individus par décade) fournis par FA, doivent donc être pondérés par le nombre de listes par décades : l'unité d'ordonnée de ces nouvelles courbes est donc le **nombre moyen d'individus par Liste**, unité assez concrète.

Un exemple avec la courbe du Pic épeiche (à comparer avec la courbe précédente en données brutes).

Cette courbe montre donc qu'en moyenne, on note 1 individu de Pic épeiche dans une liste sur 10 (seulement). La pondération par le nombre de Listes effectuée donc un « lissage » des données.

Mais la notion de liste pose problème: doit-on mettre au même niveau toutes les listes, même les listes très courtes, qui sont en fait des listes d'observation effectuées lors de prospections rapides et non exhaustives, par exemple en ouvrant simplement la fenêtre de sa salle à manger un 31 décembre pour noter un seul individu d'une seule espèce ?



Discussion sur la notion de « Liste »

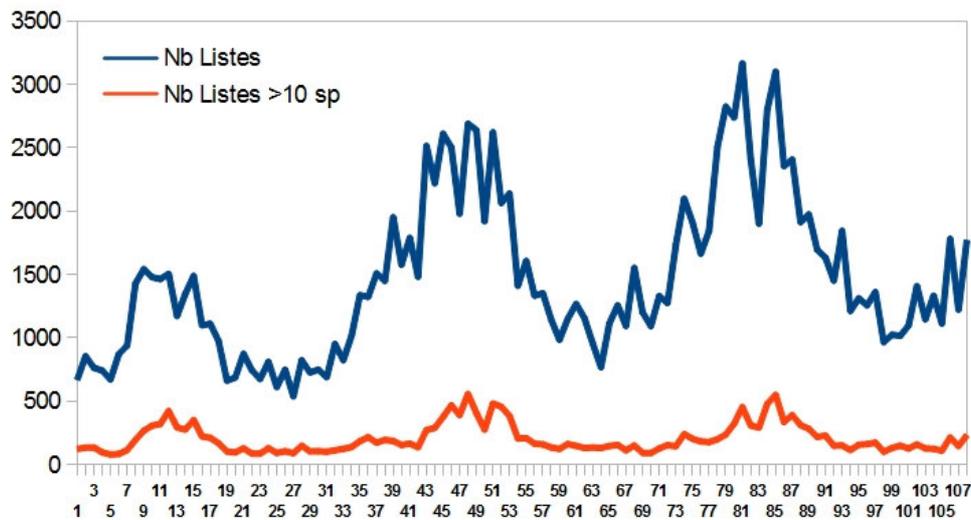
L'observation d'un écran Faune-Auvergne montre la prépondérance des données recueillies à la « billebaude », c'est à dire hors contexte d'une prospection exhaustive : devant sa maison, en allant acheter le pain, le long de l'autoroute, sur son lieu de travail (pour ces derniers cas, les observations souvent masquées :).

Effectivement ces listes n'en sont pas vraiment. La notion de Liste suppose un minimum de prospection en temps (et en surface), avec notation systématique de TOUTES les espèces. On est encore assez loin de cela en général, et profitons de ces lignes pour encourager les naturalistes à noter le plus souvent possible toutes les espèces d'un site de manière la plus exhaustive possible.

Pour trier les Listes représentatives d'une véritable « prospection » et celles obtenues ponctuellement, la seule possibilité simple qui s'offre est de choisir un seuil minimal au nombre de données, avec toutes les conséquences que cela implique : si cela permet en général d'éliminer les observations de bord d'autoroute, cela élimine aussi les données obtenues dans des milieux très pauvres (surtout en montagne par exemple où des listes en dessous de 10 espèces sont courantes), mais aussi laisse de côté les observations de rapaces nocturnes, quasiment toutes constituant l'unique donnée de la liste. Le seuil de 10 espèces est donc discutable, mais il faut bien en choisir un pour tester la méthode.

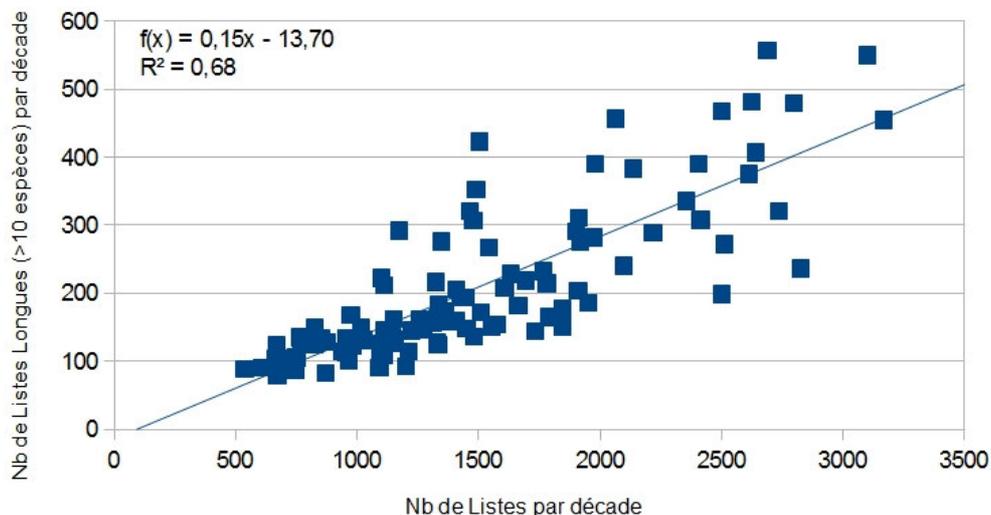
En ayant cela à l'esprit, voici les résultats de ces traitements de données traduits en graphe.

Le nombre total de Listes sur 3 années successives (2010, 2011, 2012, soit 108 décades) sur la courbe bleue. Le nombre de Listes « longues » avec **au moins 10 espèces** sur la courbe en rouge.



Évidemment et heureusement la corrélation entre les séries est bonne. Cela signifie qu'à la limite, on doit pouvoir utiliser indistinctement l'une ou l'autre de ces séries de données pour minorer le biais de la prospection sur les courbes phénologiques. La logique serait alors, quitte à choisir, d'utiliser les valeurs concernant les Listes « longues » (ici > 10 espèces), plus proches de la réalité d'un recensement de terrain tendant à une certaine exhaustivité. Cependant, il conviendrait alors de ne retenir que les seules données de ces listes longues, en supprimant celles obtenues sur les listes courtes, ce qui devient très long à réaliser techniquement.

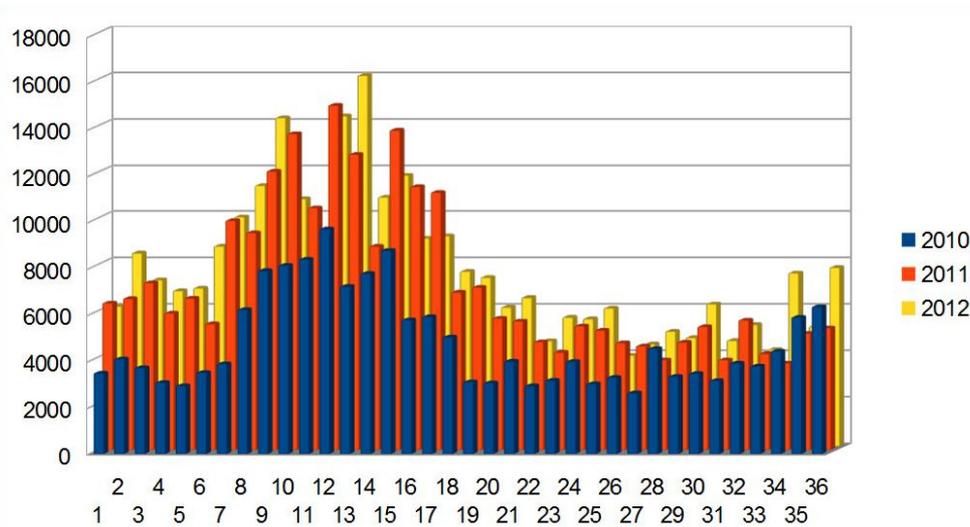
La corrélation de ces deux séries de valeurs est également visible quand on présente les résultats sous la forme d'un graphe avec 108 points (1 pour chaque décade des 3 années 2010, 2011, 2012) ayant pour abscisse le nombre de Listes, et en ordonnée le nombre de Listes « longues ».



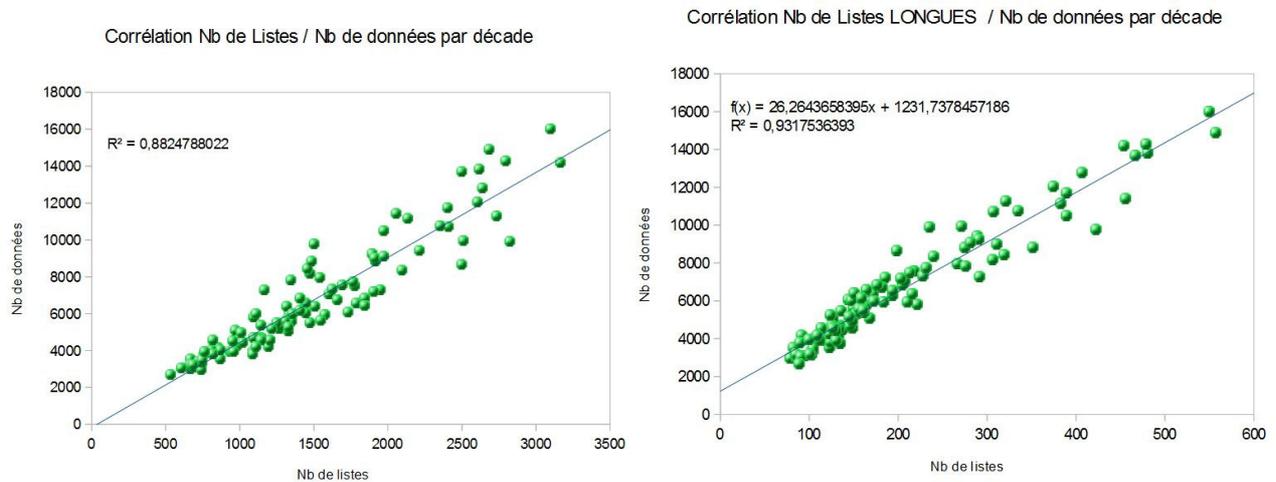
La corrélation est très bonne pour les décades qui comportent un nombre de Listes limité – en dessous de 1500 listes par décade - (en hiver par exemple). Au printemps et en été, le nombre de listes est quasiment doublé (leur longueur aussi, d’ailleurs), et la corrélation est beaucoup plus faible, les points sont plus dispersées. Mais globalement le lien est évidemment fort.

Corrélation entre le nombre de listes et le nombre total de données par décade

Le nombre total de données par décade, paramètre assez simple à utiliser, est-il lui aussi un bon indicateur de la pression d'observation ? Le graphe suivant donne le nombre décadaire de **données brutes**, sur les trois années 2010, 2011, 2012.



Nous pouvons mettre en corrélation le nombre de listes (longues ou pas) avec ce nombre de données :



Nous pouvons conclure que la corrélation est excellente entre le nombre de données par décade, et le nombre de Listes normales ou de Listes Longues. Le nombre total de données par décade est donc lui aussi un outil fiable d'évaluation de la prospection. Cette information est en plus très simple à obtenir.

Graphes tests sur quelques espèces

Nous allons comparer quatre sortes de graphes

- graphe 1 du nombre cumulé d'oiseaux par décade, non corrigé – Courbe témoin issue de FA
- graphe 2 du nombre cumulé d'oiseaux par décade, corrigé par le nombre de Listes
- graphe 3 du nombre cumulé d'oiseaux par décade, corrigé par le nombre de Listes LONGUES
- graphe 4 du nombre cumulé d'oiseaux par décade, corrigé par le nombre de données

Cinq espèces seront testées :

- le Merle noir, espèce très commune
- le Pic épeiche, considéré comme sédentaire
- le Pouillot véloce, commun mais migrateur
- La Mésange noire, parfois migratrice « invasive »
- Le Geai des chênes, avec lui aussi des mouvements migratoires invasifs.

MERLE NOIR (*Turdus merula*)

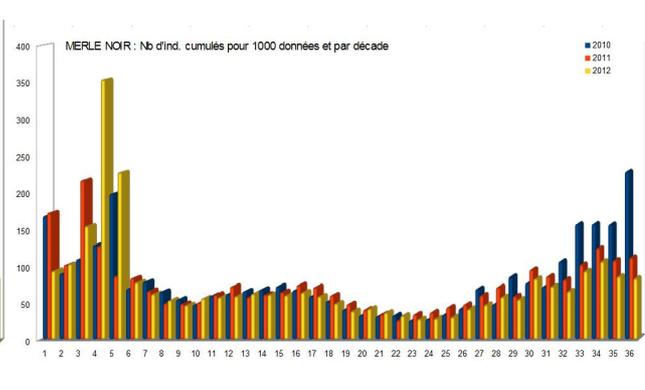
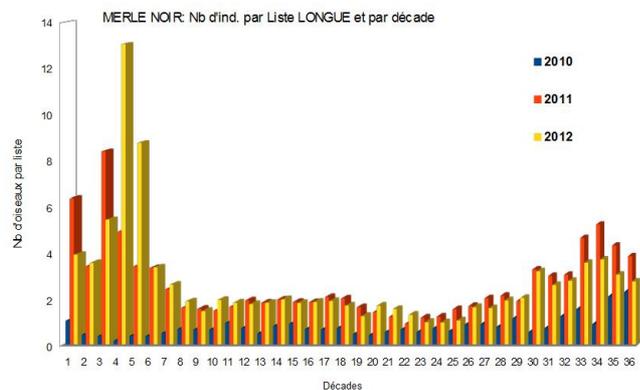
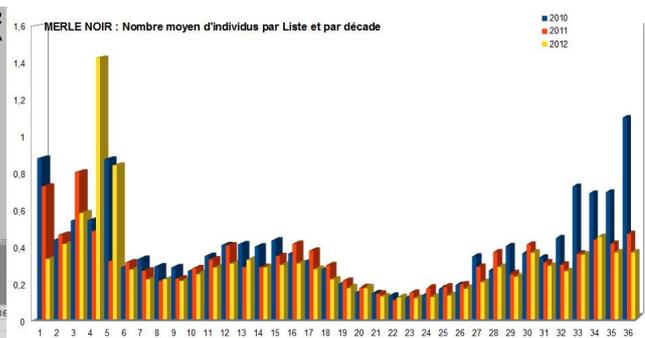
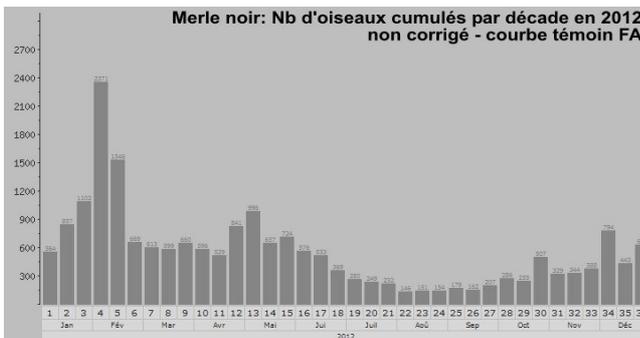
Pour le Merle noir, les valeurs hivernales élevées sont dues, après vérification, à une détection visuelle plus facile (d'autant plus que les merles peuvent se regrouper, ainsi en 2012 - 5^{ème} et 6^{ème} décades - par grand froid), mais peut-être aussi à une meilleure détection sonore déjà évoquée.

Pour cette espèce, l'amélioration des graphes en prenant en compte l'intensité de prospection amène peu de précisions supplémentaires.

Peut-être la pondération par les listes LONGUES montre-t-elle un peu mieux la réalité, notamment en novembre-décembre.



© LPO Auvergne - JM Frenoux



PIC EPEICHE (*Dendrocopos major*)

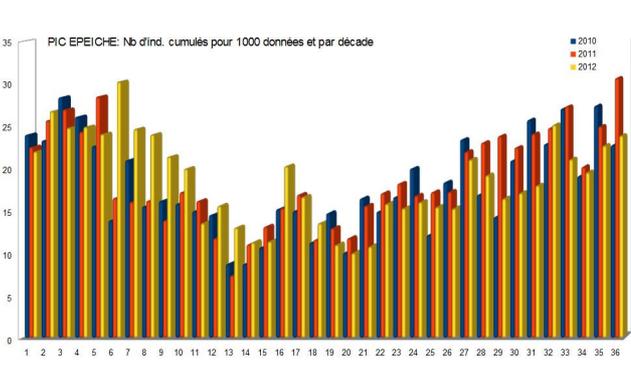
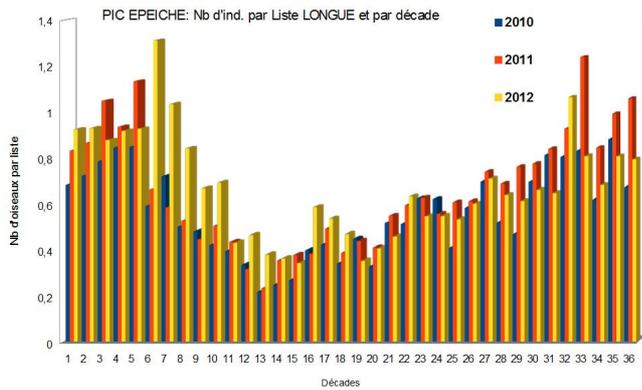
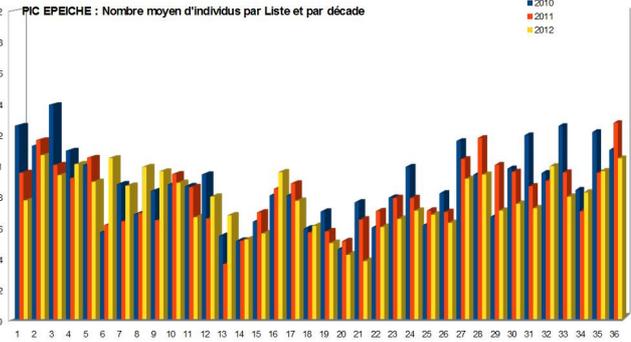
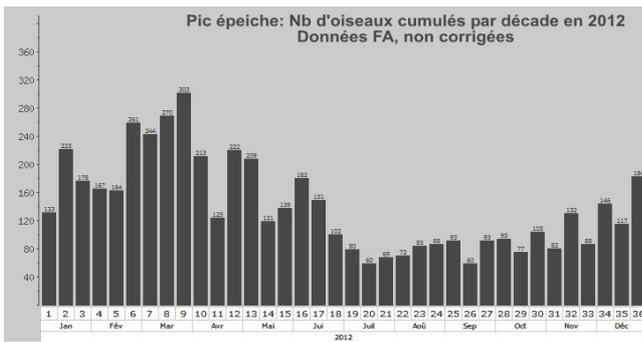
Pour le Pic épeiche, il est tout aussi difficile de trancher sur le type de graphe le plus adapté: la correction des trois types de graphes est correcte.

Le « creux » du nombre d'individus au printemps (avril-mai-juin) est probablement un problème de détection (?) dans le concert ambiant des autres espèces...

Comme expliqué précédemment, c'est une espèce typiquement sédentaire (peut-on trouver plus sédentaire en Auvergne ?) et les variations ne peuvent donc pas être très liées au nombre réel d'individus (sauf après l'envol des jeunes).



© LPO Auvergne - R Riols



POUILLOT VELOCE (*Phylloscopus collybita*)

Pour le Pouillot véloce, le graphe 1 (FA) sous-estime d'un facteur 2 les effectifs automnaux.

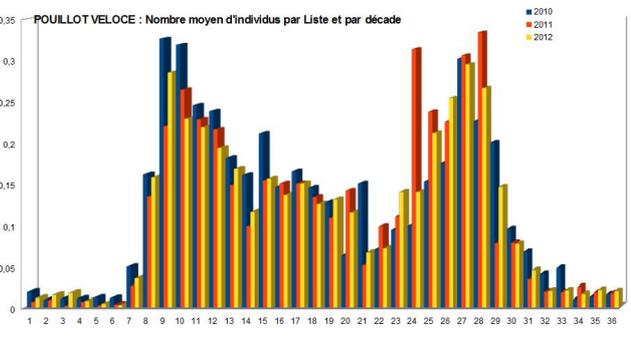
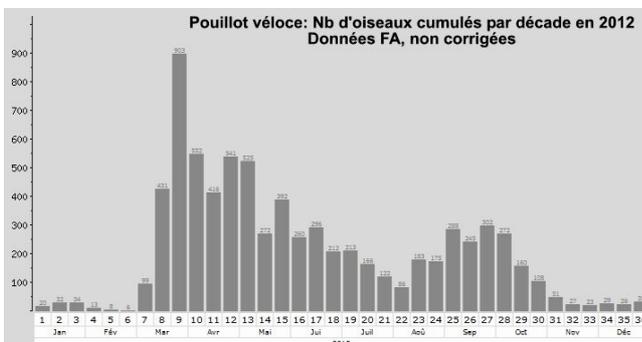
Les 3 autres graphes sont plus réalistes et montrent que le nombre d'individus par décade est au moins aussi important en automne qu'en migration de printemps.

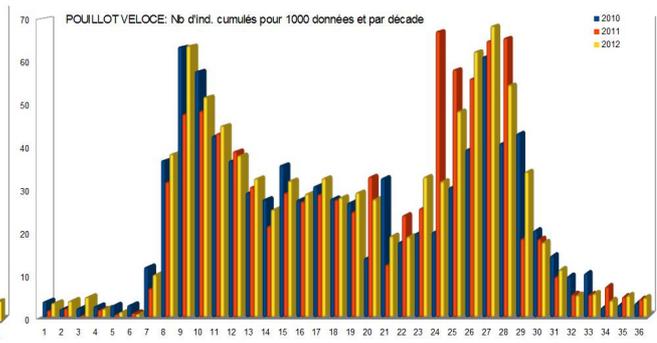
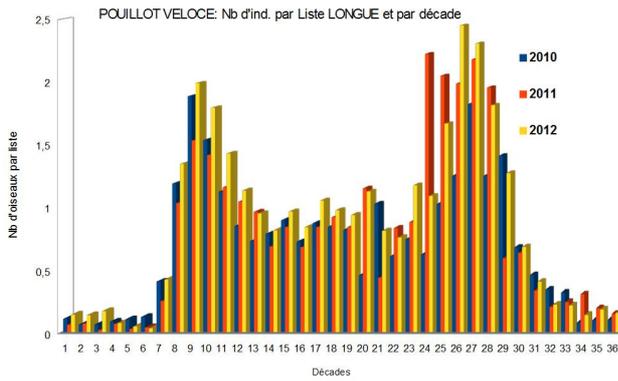
Le graphe de type 3 ou 4 (très corrélés, rappelons-le) montrent même qu'en automne, il y a plus de véloces par décades qu'au printemps, ce qui pourrait paraître logique.

Cela nous amène à choisir le type de graphe 3 ou 4, ce dernier étant le plus rapide à mettre au point (correction par le nombre total de données).



© LPO Auvergne - R Riols





MESANGE NOIRE (*Periparus ater*)

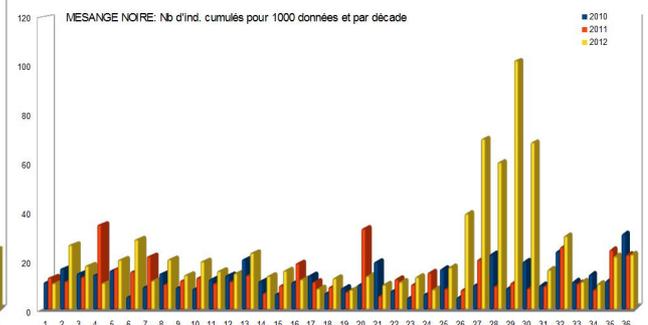
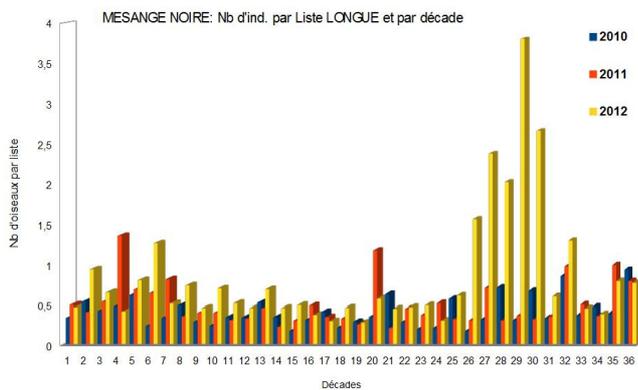
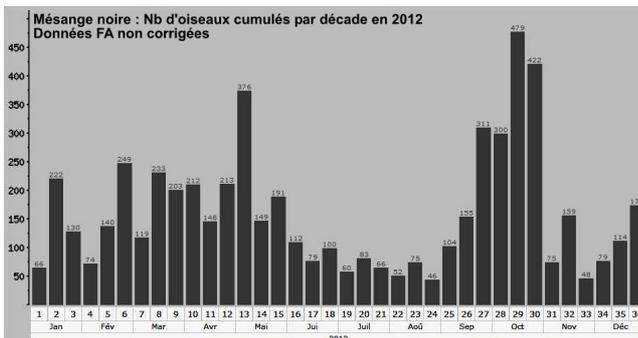
En 2012, le pic d'octobre posait évidemment la question de la réalité d'une augmentation du nombre de mésanges noires.

La courbe 2 montre bien, par comparaison des 3 années, que la phénologie de la Mésange noire est assez stable en 2010 et 2011, mais qu'en 2012, une augmentation spectaculaire (x3 ou 4) est la réalité.

Les trois types de graphes montrent bien ce phénomène d'invasion migratoire, qui avait été confirmé par les observations directes sur les cols de migration, par exemple dans le sancy.



© LPO Auvergne - JM Frenoux



GEAI DES CHENES (*Garrulus glandarius*)

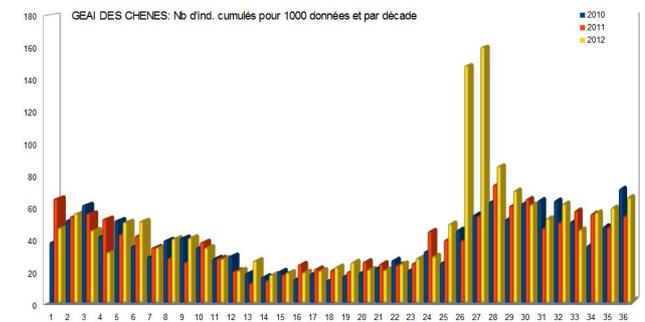
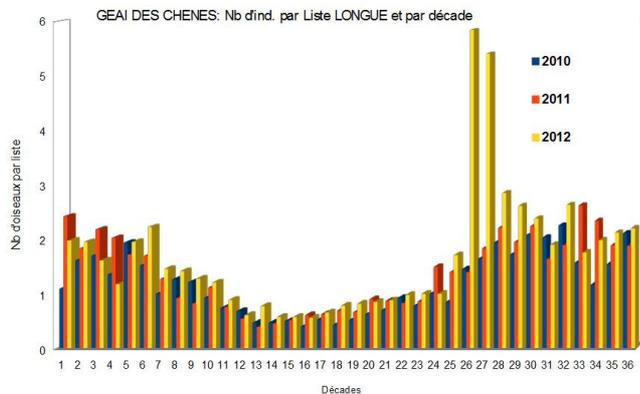
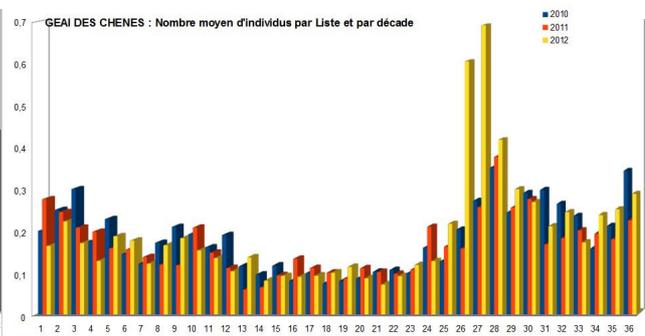
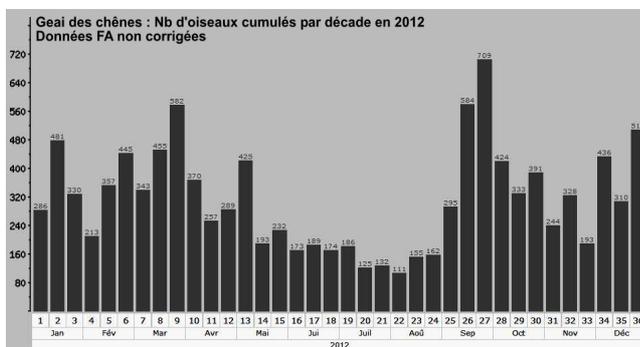
Le Geai des chênes a connu le même phénomène en 2012 que la Mésange noire : accompagné de quelques Cassenoix mouchetés, de nombreux geais ont été notés en septembre/octobre 2012.

La pondération des données brutes accentue bien le pic invasif ou migratoire, même si, chaque année, une petite augmentation est notée à l'automne, signe de mouvements migratoires discrets, ou bien d'une augmentation de la détection (le geai crie beaucoup à ces périodes, dans un silence assez général...).

Les trois types de graphes pondérés sont bien adaptés à faire ressortir une augmentation des données qui n'est pas liée à la pression d'observation.



© LPO Auvergne - R Riols



Conclusion

En conclusion de cette première partie, on peut donc dire que les trois types de graphes corrigés donnent des résultats satisfaisants. Même si l'idéal semble être la pondération par le nombre de liste longues, le pragmatisme poussera à utiliser le quatrième type de graphe : celui du nombre cumulé d'individus par le nombre total de données de la décade.

Il semble être un outil intéressant pour obtenir des courbes phénologiques moins biaisées, en particulier pour les oiseaux migrateurs communs.

ANNEXE: METHODE DE TRAITEMENT DES DONNEES

Partie 1 : Sur faune-auvergne.org et Access

→ extraire les données de l'année XXXX : par le tri multicritères, choix d'export format TXT- droits d'administrateur nécessaires.

Attention, cette première étape est longue car l'import de l'année complète n'est parfois pas possible (trop de données). Il faut importer mois par mois, ou par bimestre, ou trimestre... et faire 4, 6 ou 12 imports...

→ ouvrir une BDD Access Microsoft

→ IMPORTER la table en format TXT (nouvelle table/ importer/type de fichier : txt)

→ Faire « 1^{ère} ligne = nom des champs » puis choisir la clé primaire = champ Ref

→ Après l'import du premier fichier txt de l'année XXXX (ex : janvier + février), pour les autres, faire « ajouter à une table existante » (celle de l'année XXXX). En quatre, six à douze étapes, on constitue le fichier de l'ensemble des données de l'année XXXX...

Attention : deux problèmes à l'import : chaque table importée comporte un enregistrement vide (en N°1), et un très faible taux d'erreurs d'importation est détecté (<0,5 %) ... Penser aussi à enlever les données non validées (colonne VERIFICATION, marquées OUI).

Partie 2 : Sur Access

Il va falloir faire une suite de requêtes pour traiter progressivement cette table énorme et obtenir une liste de décades avec, en face, le nombre de listes (1 liste = 1 liste d'espèces pour un site, un jour, un ornithologue).

La programmation des requêtes est basée sur le nom de table de départ = ANNEE

→ REQUETE 1 : à partir de la table ANNEE, simplification et mise en forme en conservant 6 champs et en rajoutant un champ calculé : Décade, Jour de l'année, ID Lieu-dit, Nom, Prénom, champ calculé : concaténation NOM+Prénom, nom d'espèce. Conserve le nb total de lignes de la table initiale.

→ REQUETE 2 : elle simplifie la requête 1 en ne conservant que 5 champs sur 7 : un jour, une décade, un lieu, un « Nom+prénom=Expr1 de la requête 1 », une espèce (donc on enlève le champ nom, le champ prénom, inutiles maintenant)

→ REQUETE 3 : elle comptabilise le nb d'espèces différentes par « Nom+prénom= Expr1 », chaque ligne correspond donc à une métadonnée c.à.d. aux conditions d'obtention d'UNE liste selon la définition, et donne le nb d'espèces de chaque liste.

C'est cette requête qui est modifiable si le seuil de données minimal doit être relevé (changer manuellement via le menu « modifier la requête », la valeur numérique après le signe >).

→ REQUETE 4 : elle compte le nb d'observateurs par jour et lieu. Elle comporte donc moins de lignes que la requête précédente puisque plusieurs observateurs ont pu passer le même jour sur le même lieu.

→ REQUETE 5 : Elle totalise le nb de listes par JOUR de l'année. Elle donne donc un tableau de 365 ou 366 lignes.

→ REQUETE 6 : Totalise enfin le nb de listes par DECADE. Elle a donc 36 lignes.

Ce fichier est alors exportable en format Excel pour être utilisé dans un tableur grapheur.

Au final, seule la requête 6 est donc à utiliser, et éventuellement la requête 3 pour modifier le seuil du nb de données par liste.

Traitement des données ESPECES :

→ Extraire les données décadaires d'une année pour une espèce donnée (total du nombre d'individus par décade).

→ Reprendre le fichier ANNEE.

→ Requête 1 ESPECE : simplifie la table en ne gardant que l'espèce, son ID, la décade et le nombre d'individus de l'observation.

→ Requête 2 ESPECE ANALYSE CROISEE

Crée un tableau avec en ligne les espèces, en colonne les 36 décades (on ne peut pas faire l'inverse) et permet d'attribuer à l'intersection de chaque ligne/colonne la somme des champs NOMBRE : on obtient donc un tableau général avec la somme des individus de chaque espèce pour chaque décade. Il faut l'exporter, puis le TRANSPOSER sur tableur pour inverser espèces et colonnes, et se retrouver avec les données d'une espèce dans une colonne.

TOUS CES FICHIERS INFORMATIQUES sont à la disposition des naturalistes qui en auraient besoin.