

RÈGLES DE VALIDATION AUTOMATIQUE DES DONNÉES D'OBSERVATION D'OISEAUX

NOTICE TECHNIQUE

25 mai 2020
Version 3

PRODUCTION DU DOCUMENT :
GROUPE ORNITHOLOGIQUE NORMAND



Sommaire

I.	Présentation de l'arbre décisionnel.....	3
II.	Description des champs du fichier Source	4
III.	Utilisation de l'arbre décisionnel avec utilisation du fichier source.....	5
	Étape 1. L'espèce observée est-elle dans la liste normande ?	5
	Étape 2. L'espèce est-elle soumise à homologation régionale ou nationale ?.....	5
	Étape 3. La date d'observation est-elle conforme à la période de présence de l'espèce en Normandie ?.....	5
	Étape 4. L'espèce a-t-elle déjà été observée sur ce site ?	5
	Étape 5. Quelle est la valeur du champ O_NBRT ?.....	6
	5.1. O_NBRT = présence/absence	6
	5.2. O_NBRT = nombre de colonie.....	6
	5.3. O_NBRT = nombre d'individus, ou contient « estimé » ou « calculé »	6
	5.4. O_NBRT = nombre de couples ou nombre de nids.....	6
	5.5. O_NBRT = autres (différent des valeurs citées en 5.1, 5.2, 5.3 et 5.4).....	6
	Étape 6. Pour tester les effectifs, quel est le type de grégarisme de l'espèce ?.....	6
	6.1. TYPE_GREGAIRE = jamais.....	7
	6.2. TYPE_GREGAIRE = toujours.....	7
	6.3. TYPE_GREGAIRE = partiel_date	7
	6.4. TYPE_GREGAIRE = partiel_site.....	7
	6.5. TYPE_GREGAIRE = partiel_date&site.....	7
	Étape 7. O_NBRT indique une donnée de reproduction (nombre de couples ou nombre de nids), l'espèce est-elle nicheuse ?	8
	Étape 8. La date d'observation correspond-elle à la période de reproduction de l'espèce ?	8
	Étape 9. L'espèce a-t-elle déjà été observée en reproduction sur ce site ?.....	8
	Étape 10. Pour tester les effectifs, quel est le type de grégarisme de l'espèce ?	9
	10.1 TYPE_GREGAIRE = jamais.....	9
	10.2 TYPE_GREGAIRE = toujours.....	9
	10.3 TYPE_GREGAIRE = partiel_date.....	9
	10.4 TYPE_GREGAIRE = partiel_site	9
	10.5 TYPE_GREGAIRE = partiel_date&site.....	10
IV.	CAS DES EXCEPTIONS.....	10

I. PRÉSENTATION DE L'ARBRE DÉCISIONNEL

L'objectif de cette note est de présenter une méthodologie permettant d'attribuer aux données un des cinq niveaux de validation conformément au Cahier des Clauses Techniques Particulières produit par l'Observatoire de la Biodiversité de Normandie / Région Normandie (OBN) :

- à valider (lorsqu'aucune information ne permet de juger la fiabilité de la donnée) ;
- non plausible (espèce non présente dans la région ou les départements limitrophes ainsi que les tortues marines dans l'Orne) ;
- doute (a/ les espèces naturalisées ; b/ les espèces présentent dans les départements limitrophes mais pas en Normandie ; c/ les tortues marines, espèces exceptionnelles ; d/ les espèces assez rares à très rares ou inconnues (absentes) dans le département, difficiles à identifier et dont la localisation de l'observation se situe à moins de 10 km d'une station assez récente (moins de 11 ans) connue de la même espèce ; e/ les espèces assez rares à très rares ou inconnues et dont il n'existe pas de stations assez récentes (moins de 11 ans) et assez proches (moins de 10 km) ;
- plausible (a/ les espèces rares à inconnues (absentes), facile à identifier et dont la localisation de l'observation se situe à moins de 10 km d'une station connue de la même espèce ; b/ les espèces assez communes à très communes dans le département, difficiles à identifier et dont la localisation de l'observation se situe à moins de 10 km d'une station connue de la même espèce) ;
- validée (a/ données d'absence O_NBRE=0 si la technique d'acquisition permet de vérifier l'absence d'espèce b/ les espèces assez communes à très communes dans le département et faciles à identifier ; c/ les espèces assez rares mais faciles à reconnaître et dont la localisation de l'observation se situe à moins de 10 km d'une station connue de la même espèce ; d/ ou confirmé par un spécialiste dans le cadre d'une validation manuelle.

Cette note doit être utilisée conjointement avec le tableur (format excel) source contenant les valeurs de référence pour chaque espèce (voir chapitre II).

Les données non-plausibles, douteuses et plausibles sont validées manuellement par un spécialiste.

II. DESCRIPTION DES CHAMPS DU FICHIER SOURCE

A. TAXREF 11 (numérique)

Code d'identifiant unique de l'espèce basé sur le Référentiel taxonomique national TAXREF v11.

B. N_LATIN (texte)

Nom latin de l'espèce.

C. N_VERN (texte)

Nom vernaculaire de l'espèce.

D. CHR_CHN (oui/non)

Distingue les espèces soumises à homologation régionale ou nationale.

E. DEB_PRESENCE (date)

Date de début d'observation possible.

F. FIN_PRESENCE (date)

Date de fin d'observation possible.

G. NICHEUSE (oui/non)

Distingue les espèces nicheuses en Normandie. Si la valeur est Oui les champs H. DEB_NIDIF et I. DEB_NIDIF sont remplis.

H. DEB_NIDIF (date)

Date de début d'observation possible de données de nidification.

I. FIN_NIDIF (date)

Date de fin d'observation pour les données de nidification.

J. SITE_NIDIF (texte)

Nom du shapefile représentant les zones de nidification pour les espèces nicheuses.

K. EFF_NIDIF_MAX (numérique)

Pour les espèces non grégaires en période de nidification, indique l'effectif maximal pouvant être observé en période de reproduction.

L. SITE_PRESENCE (texte)

Nom du shapefile représentant les zones de présence des espèces normandes.

M. TYPE_GREGAIRE (texte)

Décrit le type de grégarisme de l'espèce. Valeurs possibles : jamais, total, partiel_date, partiel_site, partiel_date&site.

N. DEB_GREG (date)

Pour les espèces grégaires une partie de l'année (colonne M. TYPE_GREGAIRE = partiel_date), date de début de la période de grégarisme.

O. FIN_GREG (date)

Pour les espèces grégaires une partie de l'année (colonne M. TYPE_GREGAIRE = partiel_date), date de fin de la période de grégarisme.

P. SITE_CONCENT (texte)

Pour les espèces grégaires uniquement sur certains sites (colonne M. TYPE_GREGAIRE = partiel_site), nom du shapefile représentant les zones où des effectifs élevés peuvent être rencontrés.

Q. EFF_MAX (numérique)

Effectif maximal observable pour l'espèce.

III. UTILISATION DE L'ARBRE DÉCISIONNEL AVEC UTILISATION DU FICHIER SOURCE

Étape 1. L'espèce observée est-elle dans la liste normande ?

Le fichier Excel liste les espèces observées en Normandie depuis 1960. Sont exclues de cette liste les espèces exotiques, les échappés de captivité, les hybrides et les espèces très rarement observés en Normandie.

- OUI : l'espèce est dans la liste normande, nous passons à l'étape 2.
- NON : l'espèce n'est pas dans la liste normande, la donnée est douteuse et passe en validation manuelle.

Étape 2. L'espèce est-elle soumise à homologation régionale ou nationale ?

[Colonne D. CHR_CHN]

Certaines espèces, rares ou difficilement identifiables, sont validées uniquement par un Comité d'homologation régionale ou nationale (CHR, CHN), après renseignement d'une fiche d'homologation par l'observateur. Ces espèces ne doivent en aucun être validées automatiquement ou manuellement par un valideur non membre du CHR ou CHN.

- OUI : l'espèce est soumise à homologation, la donnée sort en validation manuelle par le CHR/CHN.
- NON : l'espèce n'est pas soumise à homologation, nous passons à l'étape 3.

Étape 3. La date d'observation est-elle conforme à la période de présence de l'espèce en Normandie ?

[Colonne E.DEB_PRESENCE et F. FIN_PRESENCE]

Les oiseaux sont caractérisés par leur période de présence en France et en Normandie qui permet de définir leur statut biologique. Nous distinguons ainsi :

- les espèces sédentaires : observables toute l'année, la période d'observation s'étale donc du 1er janvier au 31 décembre, ces espèces peuvent être nicheuses ou non nicheuses en Normandie ;
- les espèces hivernantes, observables uniquement en hiver ;
- les espèces migratrices strictes, observables pendant la migration pré-nuptiale et/ou post-nuptiale ;
- les espèces estivantes nicheuses, observables pendant la période de reproduction.

Chaque espèce a donc une période de présence en Normandie. La date est-elle conforme ?

- OUI : nous passons à l'étape 4.
- NON : la donnée passe en validation manuelle

Étape 4. L'espèce a-t-elle déjà été observée sur ce site ?

[Colonne L. SITE_PRESENCE]

Certaines espèces sont localisées géographiquement. La colonne L. SITE_PRESENCE indique pour chaque espèce les sites connus d'observation de ces espèces.

- OUI : l'espèce a déjà été observée sur le site, nous passons à l'étape 5.

- NON : c'est une nouvelle donnée de l'espèce sur le site. La donnée est taguée « plausible » mais sort en validation manuelle. Si la donnée est validée manuellement pour ce site, toute nouvelle donnée sur le même site sera validée pour cette étape et passera à l'étape 5.

La validation d'une observation sur un nouveau site nécessite une mise à jour des fichiers de localisation (shapefiles). La mise à jour se fera périodiquement, sur un pas de temps qu'il faudra définir ultérieurement, après un premier retour d'expérience sur l'utilisation du module de validation automatique. Après un ou plusieurs processus de validation, les données validées sur un nouveau site seront envoyées aux validateurs qui décideront de l'ajout de ces nouveaux sites aux shapefiles.

Étape 5. Quelle est la valeur du champ O_NBRT ?

Le format standard d'échange de données de la plateforme ODIN impose 2 champs caractérisant les effectifs d'une donnée : O_NBRE et O_NBRT.

O_NBRE est de type numérique indiquant la quantité du taxon observé, O_NBRT est un champ de type texte correspondant au type d'information indiquée à l'attribut O_NBRE.

Le test des effectifs sera différent selon la valeur indiquée dans le champ O_NBRT.

5.1.O_NBRT = présence/absence

Le taxon n'a pas été compté, le champ O_NBRE prend la valeur 1 si l'espèce est présente, 0 si l'espèce est absente (utilisé uniquement dans les cas de recherche spécifique d'espèce).

La donnée est validée automatiquement.

5.2.O_NBRT = nombre de colonie

Le nombre de colonies est une information de nidification difficile à valider automatiquement, nécessitant de définir pour chaque espèce nichant en colonie un seuil maximal de nombre de colonies possibles. Nous passons toutes les données ayant cette valeur en validation manuelle avec le tag « plausible ». La procédure pourra être modifiée en fonction du nombre de données sortantes.

5.3.O_NBRT = nombre d'individus, ou contient « estimé » ou « calculé »

Dans ce cas, le nombre indiqué dans le champ O_NBRE est une quantité d'individus, qu'elle soit précise ou estimée (ou calculée), nous testons la conformité des effectifs en passant à l'étape 6.

5.4.O_NBRT = nombre de couples ou nombre de nids

La donnée indique dans ce cas une information de reproduction. Il faut tester la conformité des conditions de reproduction avant de tester les effectifs en passant à l'étape 7.

5.5.O_NBRT = autres (différent des valeurs citées en 5.1, 5.2, 5.3 et 5.4)

Les autres valeurs possibles du champ O_NBRT ne concerne pas les données ornithologiques. Dans ce cas, il y a une erreur de saisie et la donnée est taguée « douteuse » et passe en validation manuelle.

Étape 6. Pour tester les effectifs, quel est le type de grégarisme de l'espèce ?

[Colonne M. TYPE_GREGAIRE]

Certaines espèces peuvent être observées en très grand nombre en fonction de la période d'observation et en fonction des sites d'observation. Nous distinguons donc les espèces :

- jamais grégaires ;
- toujours grégaires ;
- grégaires en fonction des sites ;
- grégaires en fonction de la période ;
- grégaire sur certains sites et à une certaine période.

6.1.TYPE_GREGAIRE = jamais

L'espèce n'est jamais grégaire, est-ce que l'effectif indiqué est inférieur ou égal à l'effectif maximal observable :

O_NBRE <= colonne Q. EFF_MAX ?

- OUI : la donnée est validée.
- NON : la donnée est douteuse et passe en validation manuelle.

6.2.TYPE_GREGAIRE = toujours

L'espèce est toujours observée en grand nombre, nous ne testons pas les effectifs, la donnée est validée automatiquement.

6.3.TYPE_GREGAIRE = partiel_date

L'espèce est observée en grand nombre une partie de l'année.

La date d'observation correspond-elle à la période de grégarisme ?

[Colonne N. DEB_GREG et O. FIN_GREG]

- OUI : la donnée est validée.
- NON : nous testons l'effectif en passant à l'étape 6.3.1

6.3.1 (La date d'observation ne correspond pas à une période de grégarisme.)

Est-ce que l'effectif indiqué est inférieur ou égal à l'effectif maximal observable :

O_NBRE <= colonne Q. EFF_MAX ?

- OUI : la donnée est validée.
- NON : la donnée est douteuse et passe en validation manuelle.

6.4.TYPE_GREGAIRE = partiel_site

L'espèce est observée en grand nombre sur certains sites.

Le site d'observation correspond-il à un site de forte concentration ?

[Colonne P. SITE_CONCENT]

- OUI : la donnée est validée.
- NON : nous testons l'effectif en passant à l'étape 6.4.1

6.4.1 (Le site d'observation ne correspond pas à une période de grégarisme.)

Est-ce que l'effectif indiqué est inférieur ou égal à l'effectif maximal observable :

O_NBRE <= colonne Q. EFF_MAX ?

- OUI : la donnée est validée.
- NON : la donnée est douteuse et passe en validation manuelle.

6.5.TYPE_GREGAIRE = partiel_date&site

L'espèce est observée en grand nombre à une certaine période de l'année mais uniquement sur certains sites. Il faut tester les dates de grégarisme mais aussi les sites de grégarisme.

La date d'observation correspond-elle à la période de grégarisme ?

[Colonne N. DEB_GREG et O. FIN_GREG]

- OUI : il faut tester maintenant le site de grégarisme en allant à l'étape 6.5.1
- NON : nous testons l'effectif en passant à l'étape 6.5.2

6.5.1 (La date d'observation correspond à une période de grégarisme.)

Le site d'observation correspond-il à un site de forte concentration ?

[Colonne P. SITE_CONCENT]

- OUI : la donnée est validée.
- NON : nous testons l'effectif en passant à l'étape 6.5.1.1

6.5.1.1 (La date d'observation correspond à une période de grégarisme, mais pas le site.)

Est-ce que l'effectif indiqué est inférieur ou égal à l'effectif maximal observable :

$O_NBRE \leq$ colonne Q. EFF_MAX ?

- OUI : la donnée est validée.
- NON : la donnée est douteuse et passe en validation manuelle.

6.5.2 (La date d'observation ne correspond pas à une période de grégarisme.)

Est-ce que l'effectif indiqué est inférieur ou égal à l'effectif maximal observable :

$O_NBRE \leq$ colonne Q. EFF_MAX ?

- OUI : la donnée est validée.
- NON : la donnée est douteuse et passe en validation manuelle.

Étape 7. O_NBRT indique une donnée de reproduction (nombre de couples ou nombre de nids), l'espèce est-elle nicheuse ?

[Colonne G. NICHEUSE]

- OUI : l'espèce est nicheuse, il faudra tester les conditions de date et de site en allant à l'étape 8.
- NON : l'espèce n'est pas nicheuse, la donnée est douteuse et passe en validation manuelle

Étape 8. La date d'observation correspond-elle à la période de reproduction de l'espèce ?

[Colonne H. DEB_NIDIF et I. FIN_NIDIF]

- OUI : la date est conforme, il faut tester si le site de nidification est conforme en allant à l'étape 9.
- NON : la date d'observation est en dehors de la période de reproduction, la donnée est taguée douteuse et passe en validation manuelle

Étape 9. L'espèce a-t-elle déjà été observée en reproduction sur ce site ?

[Colonne J. SITE_NIDIF]

- OUI : le site de reproduction est connu, nous passons à l'étape 10 pour tester les effectifs.

- NON : la donnée est taguée « plausible » mais sort en validation manuelle. Si la donnée est validée manuellement pour ce site, toute nouvelle donnée de reproduction sur le même site sera validée pour cette étape et passera à l'étape 10.

La validation d'une observation sur un nouveau site nécessite une mise à jour des fichiers de localisation (shapefiles). La mise à jour se fera périodiquement, sur un pas de temps qu'il faudra définir ultérieurement, après un premier retour d'expérience sur l'utilisation du module de validation automatique. Après un ou plusieurs processus de validation, les données validées sur un nouveau site seront envoyées aux validateurs qui décideront de l'ajout de ces nouveaux sites aux shapefiles.

Étape 10. Pour tester les effectifs, quel est le type de grégarisme de l'espèce ?

[Colonne M. TYPE_GREGAIRE]

Certaines espèces peuvent être observées en période de reproduction en très grand nombre en fonction de la période d'observation et en fonction des sites d'observation. L'étape 10 est similaire à l'étape 6, le test des effectifs se fait en fonction du type de grégarisme de l'espèce. Pour obtenir l'effectif maximal nicheur en nombre d'individus, la valeur du champ O_NBR est multiplié par 2, puis comparé à l'effectif maximal en période de reproduction

[Colonne K. EFF_NIDIF_MAX]

10.1 TYPE_GREGAIRE = jamais

L'espèce n'est jamais grégaire, est-ce que l'effectif indiqué est inférieur ou égal à l'effectif maximal observable en période de reproduction : $O_NBR \times 2 \leq \text{colonne K. EFF_NIDIF_MAX}$?

- OUI : la donnée est validée.
- NON : la donnée est douteuse et passe en validation manuelle.

10.2 TYPE_GREGAIRE = toujours

L'espèce est toujours observée en grand nombre, nous ne testons pas les effectifs, la donnée est validée automatiquement.

10.3 TYPE_GREGAIRE = partiel_date

L'espèce est observée en grand nombre une partie de l'année. La date d'observation correspond-elle à la période de grégarisme ? [Colonne N. DEB_GREG et O. FIN_GREG]

- OUI : la donnée est validée.
- NON : nous testons l'effectif en passant à l'étape 10.3.1

10.3.1 (La date d'observation ne correspond pas à une période de grégarisme.)

Est-ce que l'effectif indiqué est inférieur ou égal à l'effectif maximal observable : $O_NBR \times 2 \leq \text{colonne K. EFF_NIDIF_MAX}$?

- OUI : la donnée est validée.
- NON : la donnée est douteuse et passe en validation manuelle.

10.4 TYPE_GREGAIRE = partiel_site

L'espèce est observée en grand nombre sur certains sites. Le site d'observation correspond-il à un site de forte concentration ? [Colonne P. SITE_CONCENT]

- OUI : la donnée est validée.
- NON : nous testons l'effectif en passant à l'étape 10.4.1

10.4.1 (Le site d'observation ne correspond pas à un site de grégarisme.)

Est-ce que l'effectif indiqué est inférieur ou égal à l'effectif maximal observable :
 $O_NBRE \times 2 \leq \text{colonne K. EFF_NIDIF_MAX}$?

- OUI : la donnée est validée.
- NON : la donnée est douteuse et passe en validation manuelle.

10.5 TYPE_GREGAIRE = partiel_date&site

L'espèce est observée en grand nombre à une certaine période de l'année mais uniquement sur certains sites. Il faut tester les dates de grégarisme mais aussi les sites de grégarisme.

La date d'observation correspond-elle à la période de grégarisme ?

[Colonne N. DEB_GREG et O. FIN_GREG]

- OUI : il faut tester maintenant le site de grégarisme en allant à l'étape 10.5.1
- NON : nous testons l'effectif en passant à l'étape 10.5.2

10.5.1 (La date d'observation correspond à une période de grégarisme.)

Le site d'observation correspond-il à un site de forte concentration ?

[Colonne P. SITE_CONCENT]

- OUI : la donnée est validée.
- NON : nous testons l'effectif en passant à l'étape 10.5.1.1

10.5.1.1 (Le site d'observation ne correspond pas à un site de grégarisme.)

Est-ce que l'effectif indiqué est inférieur ou égal à l'effectif maximal observable :
 $O_NBRE \times 2 \leq \text{colonne Q. EFF_MAX}$?

- OUI : la donnée est validée.
- NON : la donnée est douteuse et passe en validation manuelle.

10.5.2 (La date d'observation ne correspond pas à une période de grégarisme.)

Est-ce que l'effectif indiqué est inférieur ou égal à l'effectif maximal observable :
 $O_NBRE \times 2 \leq \text{colonne Q. EFF_MAX}$?

- OUI : la donnée est validée.
- NON : la donnée est douteuse et passe en validation manuelle.

IV. CAS DES EXCEPTIONS

Après un processus de validation automatique, si certaines données ne correspondent à aucun cas de l'arbre décisionnel actuel, ces données seront taguées « non traitée ». Ces cas seront analysés par le comité de validation manuelle afin de mettre à jour au besoin l'arbre décisionnel.

STRUCTURE PRODUCTRICE DE CE PROTOCOLE :



Groupe Ornithologique Normand (GONm)

Le GONm est une association reconnue d'utilité publique, agréée au titre de la protection de l'environnement et comme association éducative

complémentaire de l'enseignement public.

Elle a pour principales missions l'étude, la protection des oiseaux et de leurs milieux sur les 5 départements de la région Normandie.

Elle mène également des actions de sensibilisation et d'information vers le grand-public via des animations, des stages et l'édition de livres scientifiques ou de vulgarisation.

L'ANBDD :



L'Agence Normande de la Biodiversité et du Développement Durable

a pour ambition de contribuer

à la reconquête de la biodiversité normande. Pour cela, elle se positionne en facilitateur et mobilise des acteurs régionaux aux profils divers (collectivités, entreprises, gestionnaires d'espaces naturels, etc.).

Pour répondre à cette mission, l'agence normande de la biodiversité est structurée en 3 pôles :

- **Connaissance**, dont le but est de développer et partager la connaissance sur la biodiversité normande.
- **Reconquête**, en animant des réseaux d'acteurs et en favorisant l'émergence de projets.
- **Valorisation**, en produisant des médias permettant la généralisation des bonnes pratiques régionales.

PARTENAIRES ET FINANCEURS :



ANBDD.FR