
PROJET SCO FRANCE 2021 :

EO4DROUGHTMONITORING

RAPPORT FINAL

Date	Version	Auteurs	Commentaires
29/06/2022	1.0	INSIGHT / Météo-France NC	-
29/07/2022	2.0	INSIGHT / Météo-France NC	Mise en page finalisée



Table des matières

1. Indicateurs prévisionnels : Prédiction des épisodes de sécheresse végétale.....	7
1.1. Quelques mots sur SPI-3, la variable « proxy ».....	8
1.2. Préparation des données et régionalisation de l'indice.....	11
1.3. Formulation du modèle statistique	12
1.4. GDO, le prédicteur de grande échelle	13
1.5. Modalités de mise en œuvre du modèle à 1, 2 et 3 mois d'échéance.....	18
1.6. Méthode d'apprentissage statistique	21
1.6.1. Protocole de sélection.....	21
1.6.2. Les scores de Brier	22
1.7. Mise en exploitation.....	23
1.8. Prise de décision et probabilité de sécheresse	26
2. Interface Homme Machine	29
2.1. Présentation de l'interface.....	29
2.1.1. Vue générale	29
2.1.2. Cartes actuelles et passées.....	31
2.1.3. Profil temporel	33
2.1.4. Cartes prévisionnelles	34
2.1.5. Note d'utilisation de la plateforme.....	35
2.2. Description technique.....	36
2.2.1. Fonctionnement général.....	36
2.2.1.1.1. Types de données	36
2.2.1.1.2. Flux WMS Geoserver	36
2.2.1.1.3. Description des flux de données Geoserver	37
2.2.2. Guide pour l'insertion de nouvelles données	37
2.2.2.1.1. GDI.....	37
2.2.2.1.2. LDI.....	38
2.2.2.1.3. Alert	38
2.2.2.1.4. Prevision.....	39
3. Conclusion et Externalités	40
3.1. Déroulement de l'étude	40



3.2.	Recettage solution.....	42
3.3.	Externalités pour valorisations	43
3.3.1.	Industrialisation et régionalisation	43
3.3.2.	Valorisations thématiques	43
4.	Annexes	44
	Annexe 1	44
5.	Références bibliographiques.....	46

Liste des figures

Figure 1 : Classes d'intensité des anomalies de pluies définies par SPI-3 pour juin 2017. D0, D1, D2 et D3 correspondent aux niveaux de déficits pluvieux de durée de retour 2, 6, 10 et 20 ans respectivement). W0, W1, W2 et W3 désignent les niveaux de retour équivalents en terme d'excédent pluvieux.	8
Figure 2 : Partitionnement optimal des 38 stations en 4 classes obtenu en appliquant l'algorithme PAM aux SPI-3 mois sur la période 1981-2021.	10
Figure 3 : Évolutions par décennie des moyennes de l'indice SPI- 3 mois calculées sur les 4 zones climatiquement homogènes définis pour la Nouvelle-Calédonie.	10
Figure 4 : Évolutions par décennie du SPI-3 moyen pour la commune de Bourail entre janvier 1981 et décembre 2021.	12
Figure 5 : Évolutions des anomalies de températures prévues pour les 7 prochains mois à la surface de l'océan Pacifique équatorial dans la région NINO3.4 par les 51 membres de SEAS5. Prévisions initialisées le 01/06/2022 (<i>ECMWF Charts, 2022</i>).	13
Figure 6 : Probabilités que les pluies soient supérieures à la médiane pour le trimestre juillet-août-septembre 2022 d'après SEAS5. prévisions initialisées le 01/06/2022 (<i>ECMWF Charts, 2022</i>).	14
Figure 7 : Variations de l'indicateur de prévision GDO-1 sur le domaine de la Nouvelle-Calédonie pour l'année 2015. Le pas de la grille régulière est de 1°.	16
Figure 8 : Partitionnement des points de grille obtenu en appliquant la méthode PAM aux valeurs prises par l'indicateur de prévision GDO-1 sur la période janvier 1981- décembre 2016.	17
Figure 9 : Schéma de mise en œuvre du modèle de prévision pour les échéances T+1, T+2 et T+3. Les flèches figurent la période couverte par la variable. Les flèches bleues désignent les prédicteurs du modèle, tandis que les rouges symbolisent les prédicands.	18
Figure 10 : Dispersion des tendances prévues GDO1(T) (en haut) et des conditions antérieures de sécheresse SPI2(T-1) (en bas) en fonction des catégories de sécheresse observées, SPI3(T), sur la période 1981-2021 pour la commune de Bourail.	20
Figure 11 : Distributions des Brier Skill Score (BSS) obtenues en validation croisée « Monte Carlo » après 500 itérations pour les 3 combinaisons de prédicteurs possibles : à gauche GDO1(T), au milieu GDO1(T) et SPI2(T-1), à droite SPI2(T). Probabilité de sécheresse de niveau D1, D2 ou D3.	23
Figure 12 : Brier Skill Score (BSS) obtenus en validation croisée « Monte Carlo » après 500 itérations pour les modèles de régression logistique ordinaire suivant : SPI2(T-1) + GDO1(T) (en abscisse), SPI2(T-1) (en ordonnée).	24
Figure 13 : Probabilités de sécheresse prévues sur la commune de Bourail étant données les conditions antérieures de sécheresse SPI2(T-1) et la tendance prévue GDO1(T). Le modèle de régression logistique ordinaire est ajusté sur la base de données complète (janv. 1981-décembre 2021). Les observations qui ont servi à calibrer le modèle sont repérées par des points ou des triangles, ces derniers symbolisant les occurrences de sécheresse avérées.	25

Figure 14 : Scores estimés par échantillonnage aléatoire stratifié sur la période d'apprentissage 1981-2021 pour un ensemble de seuils compris entre 5 % et 85 % : BIAS (en haut), MCC (en bas) . Prévion des occurrences de sécheresse sur la commune de Bourail à 1 mois d'échéance.	28
Figure 15 : Vue générale de la plateforme (IHM).....	30
Figure 16 : Sélection et affichage des cartes d'alertes globales	31
Figure 17 : Sélection et affichage des indicateurs actuels/passés aux échelles globales/locales.....	32
Figure 18 : Affichage du profil temporel de l'indicateur de sécheresse végétale	33
Figure 19 : Affichage et sélection des cartes prévisionnelles de sécheresse végétale.....	34
Figure 20 : Gantt réel de l'étude, comparé au planning initial	41

Liste des tableaux

Tableau 1 : Niveaux d'alerte et légende des couleurs utilisés dans le présent document pour illustrer les variations des indicateurs de prévision des conditions inhabituellement humides et sèches GDO-1, 3 et 6 mois. Les niveaux de retour sont estimés à partir des données prévues sur la période 1981-2016.	15
Tableau 2 : Matrice de confusion également appelée tableau de contingence.	26
Tableau 3 : Bilan sur la finalisation de la solution.....	42

Ce document présente les développements réalisés pour les parties Indicateurs Prévisionnels et IHM (Interface Homme Machine) dans le cadre du projet EO4DroughtMonitoring (EO4DM), labélisé par le Space Climate Observatory (SCO) en mars 2021. Une conclusion générale sera également faite à la fin du document faisant un bilan sur le déroulement global du projet ainsi que sur les valorisations à venir des produits développés. Nous invitons le lecteur à se référer au *Rapport de mi-parcours* pour plus de détails sur le contexte et les partenaires de l'étude, ainsi que sur les développements relatifs à la Bancarisation de la donnée et aux Indicateurs Actuels.

1. INDICATEURS PREVISIONNELS : PREVISION DES EPISODES DE SECHERESSE VEGETALE

Grâce à la modélisation statistique nous allons définir des outils appropriés pour prévoir l'occurrence et l'intensité des épisodes de sécheresse à l'échelle communale. La modélisation statistique confère un cadre formel permettant (1) d'industrialiser des prévisions, (2) de bâtir le « meilleur » modèle à partir de critères sur la qualité de prévision et (3) d'estimer une erreur de prévision ou un risque lié à la prise de décision.

Comme les dimensions du problème sont restreintes (quelques variables agrégées à l'échelle communale et mesurées chaque mois pendant 30 ans), nous avons privilégié les techniques statistiques rudimentaires fondées sur le modèle linéaire généralisé et l'estimation des paramètres via la méthode du maximum de vraisemblance.

Parmi tous les indicateurs de sécheresse météorologique, le SPI calculé sur 3 mois (noté ci-après SPI-3) est celui qui présente en moyenne le niveau de corrélation le plus élevé avec les indicateurs satellites utilisés pour détecter la sécheresse végétale sur les surfaces agricoles (voir section 4.5 du *Rapport de mi-parcours*). Étant donné le caractère prévisible des précipitations aux échelles infra et intra-saisonnière en Nouvelle-Calédonie (Leroy, 2007 ; Specq, 2020), SPI-3 constitue une variable dite « proxy », que l'on peut avantageusement substituer aux variables mesurées par télédétection pour prédire les occurrences de sécheresse végétale. En d'autres termes, les variations de SPI-3 représentent la partie prévisible des variations du niveau de sécheresse végétale.

Les SPI-3 mesurés chaque mois en 38 points de la Nouvelle-Calédonie entre mars 1981 et décembre 2021 constituent notre échantillon de base pour :

- la caractérisation des épisodes de sécheresse sur les 33 communes ;
- la recherche d'une meilleure méthode de prévision en minimisant une estimation de l'erreur.

L'analyse exploratoire des données constitue une étape primordiale en vue de leur prévision par apprentissage statistique. Tout d'abord, leur visualisation graphique facilite la détection des valeurs aberrantes. Par la représentation des structures de corrélation spatiale et temporelle, l'analyse oriente le modélisateur dans ses choix de prévision (sélection d'une première liste de variables prédictives, adéquation entre les échelles d'observation et de prévision, etc.) ou en vue d'estimer des données manquantes dans l'échantillon (complétion par la dernière valeur connue ou par la moyenne des plus proches voisins, etc.)

1.1. Quelques mots sur SPI-3, la variable « proxy »

Le climat de la Nouvelle-Calédonie est soumis aux variations de l'atmosphère et de l'océan induites par les phénomènes El Niño ou La Niña, et c'est usuellement l'indice SPI-3 qui est utilisé par les météorologues calédoniens pour suivre leurs retentissements sur le régime des pluies. Que nous révèlent les évolutions passées de cet indice au cours des 40 dernières années et quelles informations utiles peut-on en tirer en vue de prévoir chaque mois sa tendance ?

On le rappelle, lorsque l'indice SPI-3 est négatif, cela signifie que le cumul des pluies est inférieur à la médiane. On parle alors de situation de déficit. Les déficits pluvieux sont classés en 4 catégories. Les seuils de gravité D1, D2 et D3 correspondent aux niveaux de retour 6, 10 et 20 ans. La classe D0 désigne des anomalies légèrement négatives (cumul inférieur à la médiane). A l'opposé, les 4 niveaux d'intensité pour les périodes de pluies abondantes sont notés W0, W1, W2 et W3.

En guise d'exemple, la carte de la figure 1 présente le niveau des anomalies de pluies relevées entre avril et juin 2017. Les déficits sont relevés sur la moitié nord de la Grande Terre ; ils sont marqués sur le quart nord-ouest de la Grande Terre où les niveaux de retour 10 et 20 ans sont atteints. Ailleurs sur la moitié sud de la Grande Terre et aux îles les écarts sont proches de zéro.

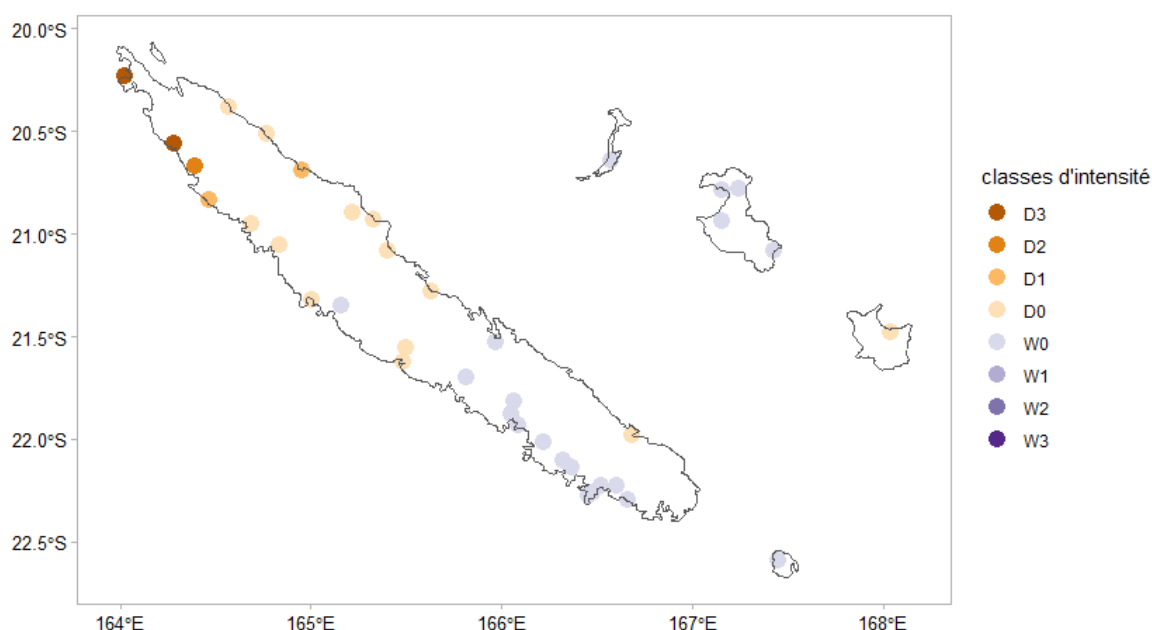


Figure 1 : Classes d'intensité des anomalies de pluies définies par SPI-3 pour juin 2017. D0, D1, D2 et D3 correspondent aux niveaux de déficits pluvieux de durée de retour 2, 6, 10 et 20 ans respectivement). W0, W1, W2 et W3 désignent les niveaux de retour équivalents en terme d'excédent pluvieux.

En général, comment varient les anomalies de précipitations sur la Nouvelle-Calédonie ? Existe-t-il des zones géographiques bien différenciées et au sein desquelles les pluies évoluent de façon homogène sous l'effet des phénomènes climatiques de grande échelle comme « El Niño Southern Oscillation » (Météo France NC , s. d.) ?

Pour identifier ces regroupements, nous avons recours à une technique statistique de partitionnement appelée Partitionnement Autour des Médoïdes (PAM ; *Kaufman and Rousseeuw, 1990*) munie de la distance euclidienne. Le nombre de classes k pour lequel les sous-ensembles sont à la fois bien compacts et bien différenciés les uns des autres est choisi de manière à maximiser la « silhouette moyenne » $\sigma(n)$, un indicateur global de la qualité d'un partitionnement en n (*Rousseeuw, 1987*), et à minimiser le nombre de stations mal classées (silhouette négative). Le partitionnement est optimal pour $k = 4$ classes. Les 4 partitions sont géographiquement homogènes : les îles Loyauté, le quart sud-ouest de la Grande Terre (de Mont-Dore à Tontouta) avec l'île des Pins, la côte ouest (de Boulouparis à Poum), la côte est (de Pouébo à Yaté). Au sein de ces 4 zones, d'un poste à l'autre, les anomalies de précipitations sur 3 mois évoluent de façon similaire en première approximation. Dit autrement, la Nouvelle-Calédonie est assujettie à un signal climatique qui, du point de vue des pluies, ne se manifeste pas de façon différenciée d'un poste à l'autre au sein d'une même zone. En conséquence, le modèle statistique ne pourra pas intégrer une tendance climatique dont l'échelle spatiale est plus fine que ces regroupements sous peine d'introduire du bruit préjudiciable à la qualité des prévisions.

On choisit la moyenne arithmétique des indices comme la variable représentative de l'évolution des anomalies de précipitations de chaque partition. Pour ces 4 nouvelles variables agrégées, les classes d'intensité sont définies à partir des niveaux de retour 6, 10 et 20 ans. Ces 6 quantiles ainsi que la médiane sont estimés par bootstrap non-paramétrique (*Davison et Hinkley, 1997*) en effectuant 1000 tirages.

Le graphique de la figure 3 met en évidence les variations des anomalies de précipitations par décennie et pour les 4 zones climatiques homogènes. Outre les deux derniers épisodes de sécheresse de 2017 et 2019, on remarque que la décennie 2000-2009 n'a pas connu d'épisode majeur. A l'opposé, la décennie 1990-1999 totalise, toute zone confondue, le plus grand nombre d'occurrences de sécheresse, notamment entre 1992 et 1995. Par endroits ce sont plus de 30 mois modérément secs (D1) à très secs (D3) qui se sont succédés. Si pour les épisodes majeurs, toutes les zones du pays sont accablées, le graphique révèle aussi des disparités d'une zone à l'autre en termes d'amplitude, mais aussi en termes de phasage, à l'entrée ou en sortie de sécheresse. Au jeu des différences, remarquons par exemple que l'épisode de 1992-1995 est le plus long et le plus intense qu'ait connu la côte ouest (zone 2) depuis 1981. S'il a bien concerné les autres zones du pays, il s'est néanmoins avéré moins intense sur la côte est (zone 3). Quant au quart sud-ouest (zone 1) et aux Loyauté (zone 4), elles ont connu une période d'accalmie au cours de l'année 1994.

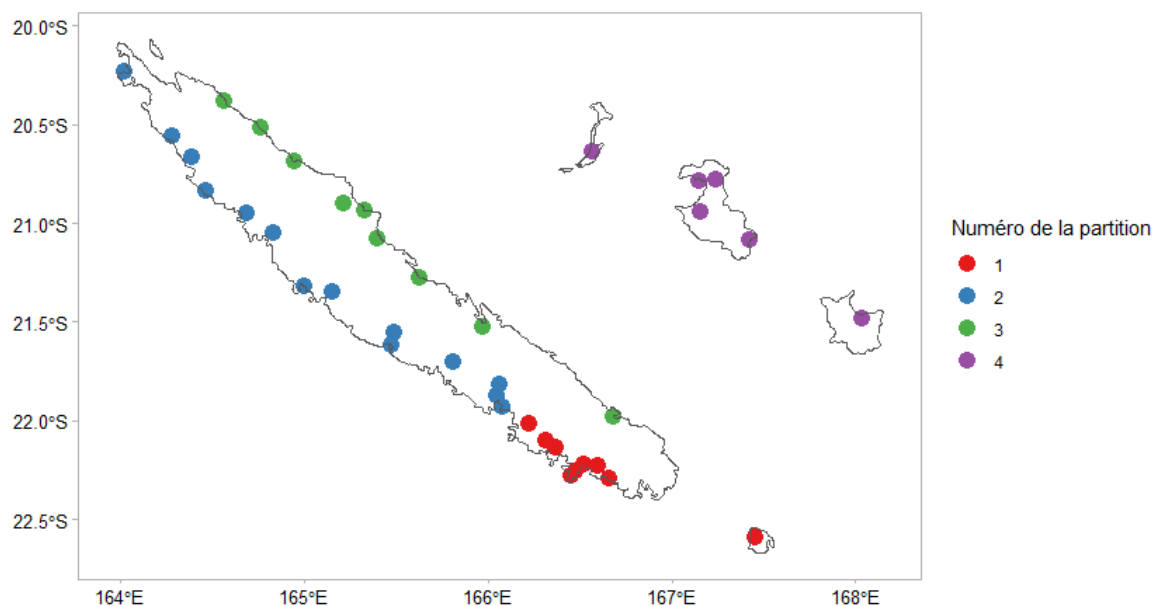


Figure 2 : Partitionnement optimal des 38 stations en 4 classes obtenu en appliquant l'algorithme PAM aux SPI-3 mois sur la période 1981-2021.

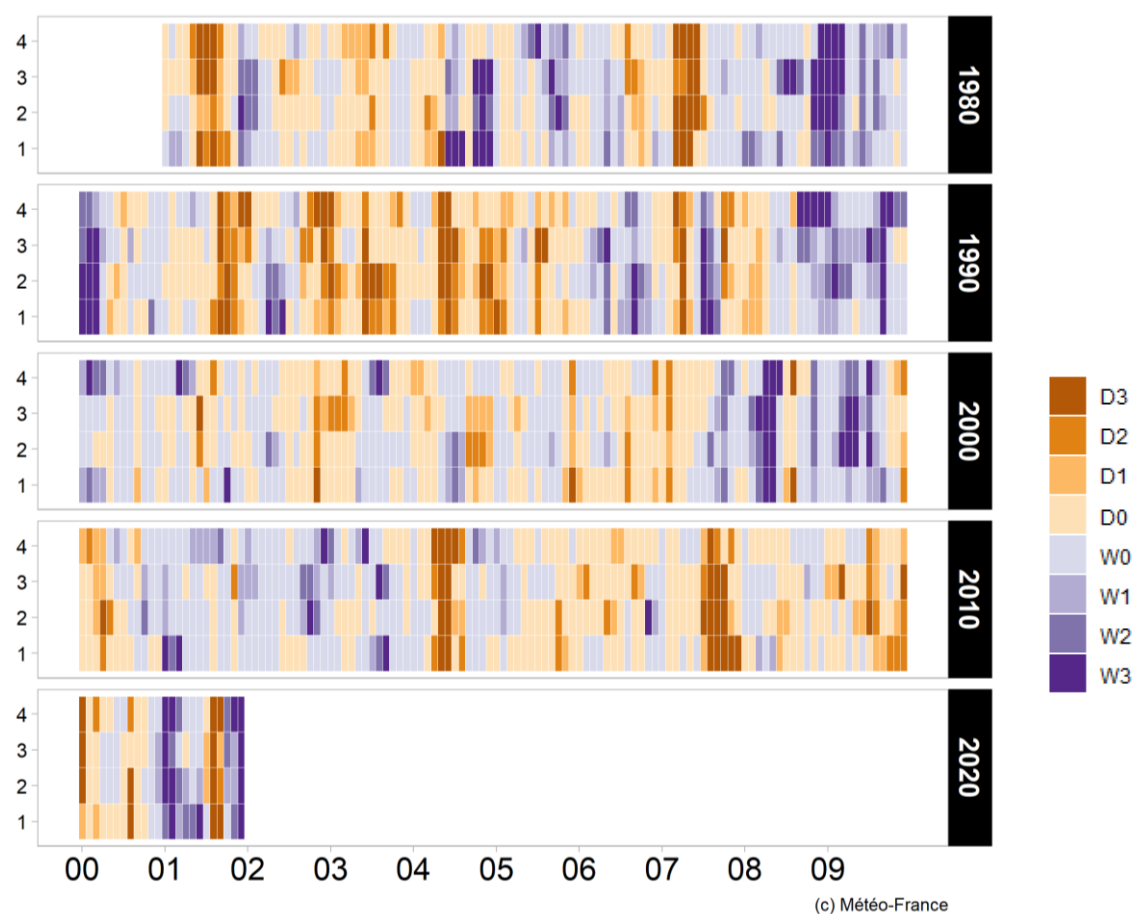


Figure 3 : Évolutions par décennie des moyennes de l'indice SPI- 3 mois calculées sur les 4 zones climatiquement homogènes définies pour la Nouvelle-Calédonie.

1.2. Préparation des données et régionalisation de l'indice

Comme l'objectif est de prévoir les occurrences de sécheresse sur chaque commune, l'estimation des valeurs de l'indice sur tout le territoire, y compris sur les communes dépourvues de stations de mesures, est une étape incontournable de la préparation des données.

Du point de vue de la géostatistique, un mois donné, l'indice de sécheresse météorologique SPI-3 est une variable régionalisée échantillonnée en 38 points. Pour reconstruire les valeurs sur tout le domaine d'intérêt, nous avons besoin de définir :

- une grille régulière couvrant tous les territoires communaux
- une technique d'interpolation spatiale.

Le pas de discrétisation est fixé à $0,1^\circ$; le maillage est positionné pour que toutes les communes du pays, même les plus petites, soient représentées par au moins un élément.

Parmi toutes les méthodes d'interpolation rudimentaires testées, c'est la pondération inverse à la distance au carré qui est retenue car c'est celle qui présente l'erreur quadratique moyenne la plus faible (estimation sur les 38 postes en validation-croisée et 492 dates). Bien qu'en théorie plus fiable (*Cressy, 2015*), la technique du krigeage n'a pas été testée, car sa mise en œuvre est fastidieuse (ajustement d'un variogramme à chaque pas de temps).

Cette reconstruction fournit les représentations spatiales de l'indice SPI-3 pour tous les mois compris entre janvier 1981 et décembre 2021.

Sur chaque commune, on définit la valeur caractéristique de l'indice. Il s'agit de la moyenne arithmétique de SPI-3 régionalisée sur l'ensemble des points de grille appartenant au territoire communal. On notera par la suite Y cette nouvelle variable aléatoire continue définie pour une commune donnée. $Y(T)$ désigne sa valeur relevée pour le mois T .

A l'instar de ce qui se fait usuellement à partir des relevés pluviométriques in situ, on définit une échelle d'intensité comportant 7 graduations. Il s'agit des quantiles de fréquence de non-dépassement $1/20$, $1/10$, $1/6$, $1/2$, $1-1/6$, $1-1/10$ et $1-1/20$. Ils sont estimés par bootstrap non-paramétrique (*Davison et Hinkley, 1997*) en effectuant 1000 tirages à chaque estimation.

A partir des 8 intervalles ainsi définis, on crée une **variable aléatoire ordinale** $y(T)$ à 8 modalités ordonnées $\{D_3 < D_2 < D_1 < D_0 < W_0 < W_1 < W_2 < W_3\}$, permettant de qualifier le niveau des anomalies de pluie à échelle de la commune.

$$y(T) = \begin{cases} D_3, & \text{si } Y(T) \leq q_{1/20} \\ D_2, & \text{si } q_{1/20} < Y(T) \leq q_{1/10} \\ \dots & \\ W_3, & \text{si } q_{1-1/20} < Y(T) \end{cases}$$

Par commodité, ces modalités sont également notées y_i : $y_1=D_3$, $y_2=D_2$, ..., $y_8=W_3$.

La figure 4 illustre les résultats obtenus pour la commune de Bourail.

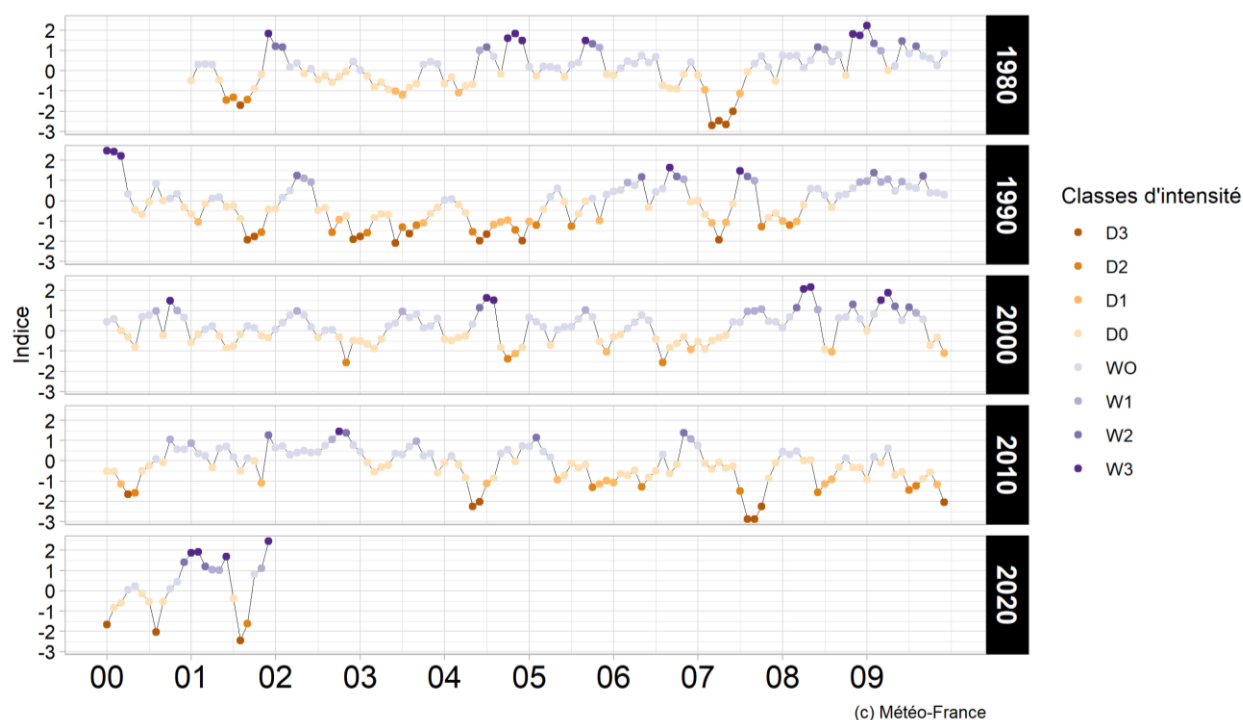


Figure 4 : Évolutions par décennie du SPI-3 moyen pour la commune de Bourail entre janvier 1981 et décembre 2021.

1.3. Formulation du modèle statistique

« A partir des données disponibles, à combien estime-t-on le risque que les cumuls de pluies sur 3 mois soient déficitaires ou que les déficits atteignent un niveau D1, D2, voire D3 ? » En langage statistique, on peut reformuler la question de la manière suivante : comment prédire la probabilité conditionnelle que le prédicand $y(T)$ tombe dans la catégorie y_i ou inférieure, étant donné un vecteur de variables prédictors $x(T)$, i.e. $P(y(T) \leq y_i(x))$?

La régression logistique ordinaire est un modèle de régression populaire en statistique et en économétrie pour les données ordinales (Greene et Hensher, 2010). Également appelée modèle cumulatif ou encore, modèle à rapports des chances proportionnels, il s'agit d'une extension de la régression logistique standard pour les prédicands multicatégoriels et ordonnés. La régression logistique ordinaire repose sur l'estimation d'une même pente pour tous les niveaux. Mathématiquement, dans notre cas, cela s'écrit comme suit :

$$\text{logit } P(y(T) \leq y_i | x(T)) = a_i - b_1 x_1(T) - \dots - b_p x_p(T), i \in \llbracket 1, 7 \rrbracket$$

$$P(y(T) \leq y_i | x(T)) = \frac{\exp(a_i - b_1 x_1(T) - \dots - b_p x_p(T))}{1 + \exp(a_i - b_1 x_1(T) - \dots - b_p x_p(T))}$$

Les a_i diffèrent d'un niveau à l'autre, contrairement aux coefficients $b_1, \dots, b_p$ associés aux p variables explicatives. Cette propriété est la garantie que $P(y \leq y_i) < P(y \leq y_{i+1})$. Il est bien évident que la probabilité que le déficit atteigne un niveau D3 ne peut pas être supérieure à la probabilité que le niveau D2 soit franchi !

Les paramètres sont estimés par maximum de vraisemblance à partir de l'échantillon $(x_1, Y_1) \dots (x_n, Y_n)$ et, sous des hypothèses similaires au cas binaire, on obtient la normalité asymptotique des estimateurs (on en déduit des intervalles de confiance et des procédures de test).

1.4. GDO, le prédicteur de grande échelle

Pour prévoir les tendances saisonnières, les météorologues utilisent des systèmes de prévision constitués d'un modèle de circulation générale atmosphérique couplé à un modèle océanique. Par rapport aux prévisions météorologiques classiques du jour pour le lendemain destinées au grand public, l'une des spécificités des prévisions saisonnières est la réalisation d'un ensemble de prévisions, appelé aussi « membres ». Le système effectue plusieurs fois la même prévision avec des états initiaux du système océan-terre-atmosphère légèrement différents. Après post-traitement, cet ensemble de simulations est exploité de manière « probabiliste », reflétant ainsi les incertitudes importantes à cette échéance (Déqué, 2003).

Parmi la quinzaine de systèmes globaux mises à disposition de la communauté internationale (WMO-LRF, s.d.), SEAS5 est la dernière génération du système conçu par le Centre européen pour les prévisions météorologiques à moyen terme (CEPMMT). Ce système repose sur un ensemble de 51 membres ; les simulations, actualisées une fois par mois, couvrent une période continue de 7 mois.

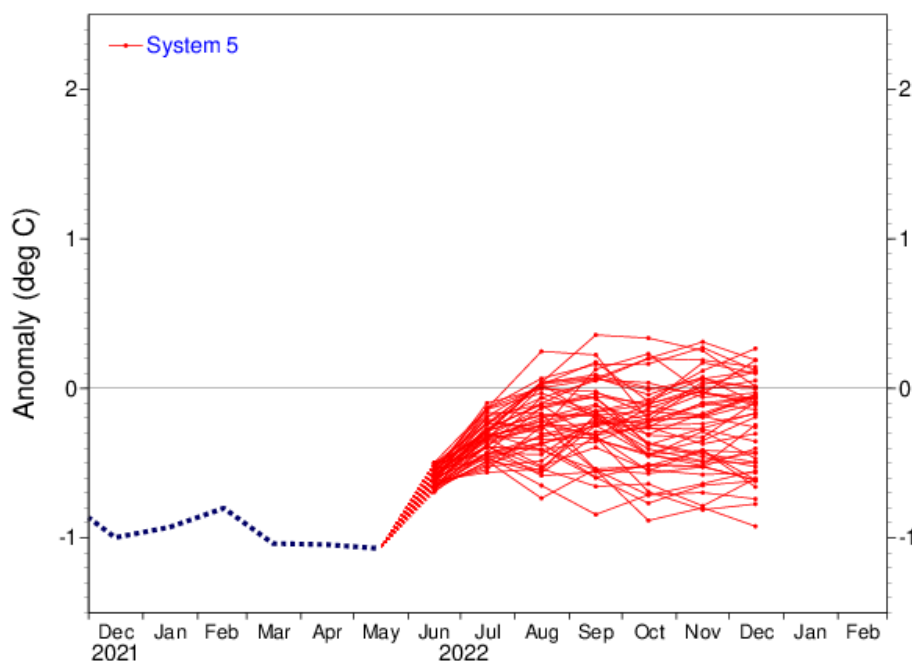


Figure 5 : Évolutions des anomalies de températures prévues pour les 7 prochains mois à la surface de l'océan Pacifique équatorial dans la région NINO3.4 par les 51 membres de SEAS5. Prévisions initialisées le 01/06/2022 (ECMWF Charts, 2022).

Le graphique ci-avant illustre les résultats obtenus pour la température à la surface de l'océan dans une région du Pacifique équatorial. Ces évolutions lentes de l'océan s'accompagnent d'une réponse spécifique de l'atmosphère, mise en évidence sur la carte de la figure 6.

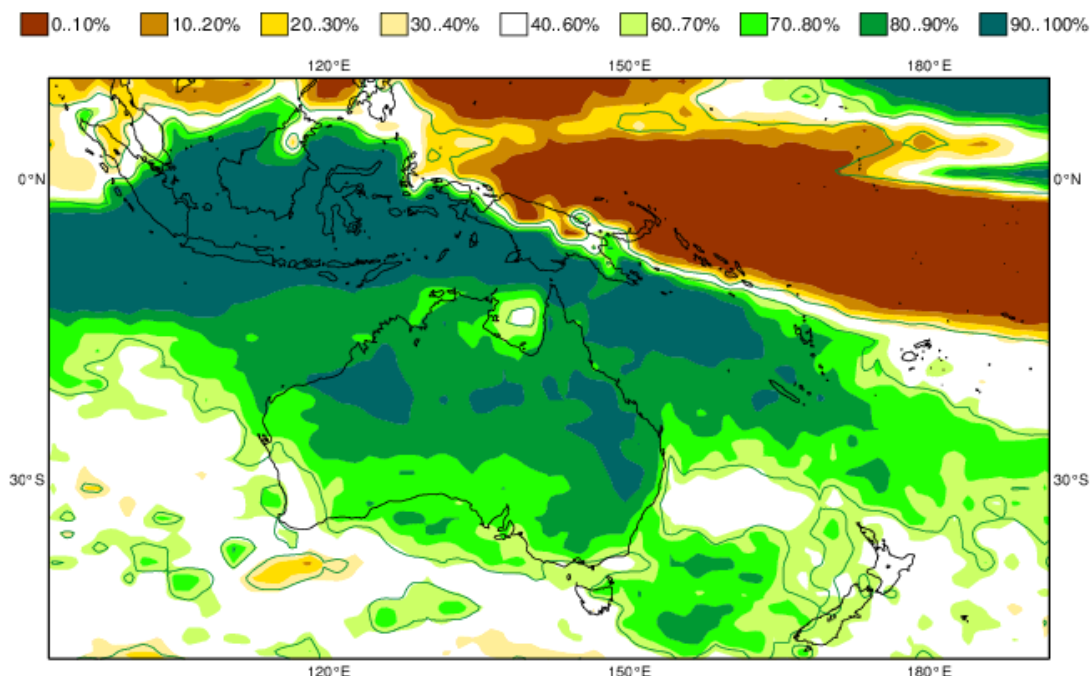


Figure 6 : Probabilités que les pluies soient supérieures à la médiane pour le trimestre juillet-août-septembre 2022 d'après SEAS5. prévisions initialisées le 01/06/2022 (ECMWF Charts, 2022).

Depuis 2018 les précipitations produites par SEAS5 sont utilisées mensuellement par le laboratoire de recherche scientifique et technique de l'Union européenne dans le cadre du Global Drought Observatory (GDO), pour élaborer des avertissements précoces de conditions de précipitations inhabituellement humides et sèches à l'échelle mondiale sur 1, 3 et 6 mois (GDO, 2019).

Ces 3 indicateurs de prévision, notés GDO-1, 3 et 6 sont calculés à partir des valeurs prévues de SPI pour la durée considérée (1, 3 ou 6 mois), elles-mêmes dérivées des cumuls de pluies prévus par SEAS5. À partir des SPI prévus, des périodes de précipitations inhabituelles sont identifiées ; elles sont prises en compte si et seulement si :

- la prévision est considérée comme robuste (c'est-à-dire avec au moins 40% des membres dénotant des prévisions intenses),
- et si les valeurs SPI correspondantes sont significatives (c'est-à-dire inférieures à -1 pour des conditions sèches, ou supérieures à +1 pour des conditions humides).

Différents niveaux d'alerte sont ensuite calculés en fonction de la période de retour de l'intensité de la prévision et de la cohérence des membres de la prévision d'ensemble. Pour de plus amples informations sur la méthodologie et les performances du système d'alerte, le lecteur est invité à lire la publication de *Lavaysse et al. (2020)*.

Les données sont produites sur une grille régulière de 1°. Disponibles en ligne à partir de janvier 2017 sur le site de l'observatoire (*GDO, 2019*), les indicateurs ont également fait l'objet d'un calcul rétrospectif sur la période 1981-2016. Ces données sont disponibles hors-ligne sur demande (*Lavaysse et al., 2020*).

Statistiquement parlant, les indicateurs GDO-1, 3 et 6 sont des variables ordinales à 9 modalités dont le sens est détaillé dans le tableau ci-dessous.

Tableau 1 : Niveaux d'alerte et légende des couleurs utilisés dans le présent document pour illustrer les variations des indicateurs de prévision des conditions inhabituellement humides et sèches GDO-1, 3 et 6 mois. Les niveaux de retour sont estimés à partir des données prévues sur la période 1981-2016.

Niveaux de l'alerte		Signification
Humide	Sec	
0		Pas d'alerte
vigilance	vigilance	Un événement humide ou sec inhabituel est prévu, avec un niveau modéré pour la fiabilité du modèle ou l'intensité des anomalies.
1	1	Un événement humide ou sec inhabituel est prévu, avec un niveau élevé pour la fiabilité du modèle ou l'intensité des anomalies. Cette prévision a une période de retour de 6 à 10 ans.
2	2	Un événement humide ou sec inhabituel est prévu, avec un niveau très élevé pour la fiabilité du modèle ou l'intensité des anomalies. Cette prévision a une période de retour de 10 à 20 ans.
3	3	Un événement humide ou sec inhabituel est prévu, avec une fiabilité du modèle et une intensité des anomalies à leur maximum. Cette prévision a une période de retour supérieure à 20 ans.

L'illustration ci-après (Fig.7) présente les variations de l'indicateur GDO-1 pour l'année 2015 sur un vaste domaine encadrant la Nouvelle-Calédonie. Visuellement, dans cet exemple, le comportement de l'indicateur sur la Nouvelle-Calédonie semble assez homogène. En janvier, mai, juin et août GDO-1 rend compte de tendances de grande échelle plus ou moins marquées. Hormis ces 4 mois, aucune tendance ne se dessine sur le domaine. L'examen sommaire de l'année 2015 ne fait ressortir aucune disparité, aucun contraste flagrant entre l'ouest de la Grande Terre, l'est ou les îles Loyauté pour ce qui concerne l'impact des phénomènes de grande échelle sur l'occurrence des événements humides ou secs en Nouvelle-Calédonie.

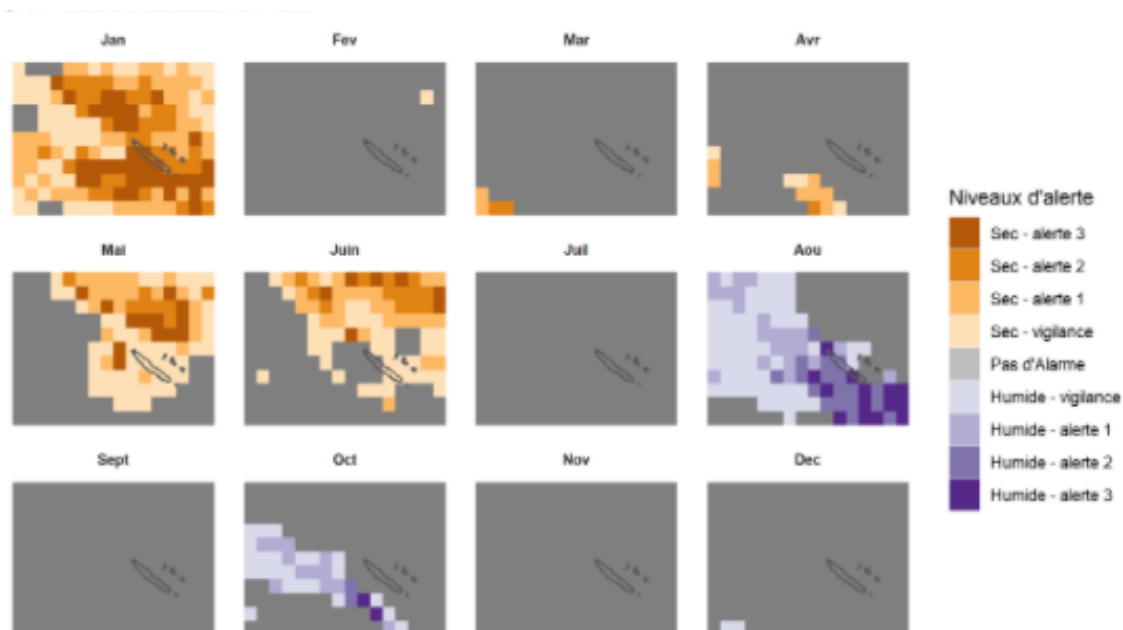


Figure 7 : Variations de l'indicateur de prévision GDO-1 sur le domaine de la Nouvelle-Calédonie pour l'année 2015. Le pas de la grille régulière est de 1°.

Les indicateurs de prévision GDO sont a priori éligibles pour figurer dans la liste des variables prédicteurs de notre modèle statistique. L'exemple précédent tend à indiquer que le signal prévu sur le territoire calédonien est assez bruité et qu'il n'y a pas lieu de différencier la valeur de l'indicateur selon la commune visée. Pour étayer cette hypothèse, le recours à une méthode statistique de partitionnement s'impose. Il s'agit de classer les différents points de grille suivant les valeurs prises par la variable ordinaire à 9 modalités. Nous devons d'abord définir une certaine notion de (dis)similarité entre les séries temporelles. La distance euclidienne n'est valable que pour les variables continues, et n'est donc pas applicable ici. C'est la distance de *Gower (1971)*, conçue pour gérer des types de données mixtes, qui est choisie. Bien qu'il existe de nombreux algorithmes capables de gérer une matrice de distance personnalisée, le partitionnement autour des médoïdes (PAM) sera utilisé ici. Quant aux nombres de classe, il existe de nombreux critères pour nous aider à choisir le nombre optimal (*Kaufman et Rousseeuw, 1990*). Nous utiliserons la largeur de la silhouette ; elle mesure la similarité d'une observation avec sa propre partition par rapport à la partition voisine la plus proche (*Rousseeuw, 1987*). Nous choisirons le nombre de classes qui réalise le meilleur compromis entre une silhouette moyenne élevée et un nombre limité de silhouettes négatives.

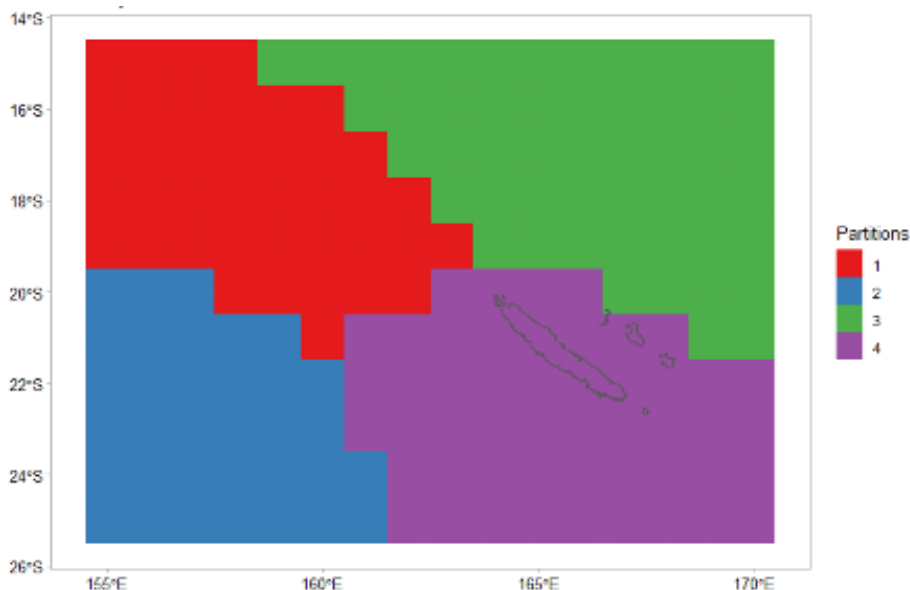


Figure 8 : Partitionnement des points de grille obtenu en appliquant la méthode PAM aux valeurs prises par l'indicateur de prévision GDO-1 sur la période janvier 1981- décembre 2016.

Pour GDO-1, c'est le partitionnement en 4 classes, présenté sur la figure 8, qui est obtenu. On s'aperçoit que les points de grille couvrant la Nouvelle-Calédonie font tous partie de la même partition. Cette homogénéité est également remarquée pour GDO-3 et 6 dont le nombre optimal de partition est estimée à 4 et 2 respectivement (résultats non représentés).

Quelle que soit la durée considérée (1, 3 ou 6 mois), la variable GDO à notre disposition est une variable ordinaire régionalisée, qualifiant l'évolution prévue des pluies en chaque point de grille (*latitude, longitude*). Pour l'intégrer à notre modèle et faciliter son agrégation à l'échelle de la Nouvelle-Calédonie, on la transforme d'abord en une variable continue $\widetilde{GDO} = g(GDO)$, puis on calcule la moyenne de la transformée sur un sous-domaine couvrant la Nouvelle-Calédonie. Ainsi définie, GDO_{NC} est une variable continue quantifiant l'intensité moyenne et la cohérence des anomalies de pluie prévues sur la Nouvelle-Calédonie.

$$\widetilde{GDO}(\text{latitude}, \text{longitude}, T) = g(GDO(\text{latitude}, \text{longitude}, T))$$

$$g(\text{alerte}) = \begin{cases} -2,06, & \text{si alerte} = \text{sec 3} \\ -1,44, & \text{si alerte} = \text{sec 2} \\ -1,11, & \text{si alerte} = \text{sec 1} \\ -0,45, & \text{si alerte} = \text{sec vigi.} \\ 0, & \text{si alerte} = \text{r.a.s.} \\ \dots \\ +2,06, & \text{si alerte} = \text{humide 3} \end{cases}$$

$$GDO_{NC}(T) = \langle \widetilde{GDO}(\text{latitude}, \text{longitude}, T) \rangle$$

Chaque niveau d'alerte i de la variable ordinaire GDO est défini par deux fréquences de non-dépassement consécutives f_i et f_{i+1} choisies parmi $\{-\infty, 1/20, 1/10, 1/6, 1/2, 1-1/6, 1-1/10, 1-1/20, +\infty\}$. La transformation g est définie pour faire correspondre à un niveau

d'alerte i la valeur moyenne de la loi normale sur l'intervalle borné par les quantiles correspondant aux fréquences de non dépassement f_i et f_{i+1} .

Par la suite, on notera GDO1, GDO3 et GDO6 les variables latentes continues pour les prévisions à 1, 3 et 6 mois d'échéance.

1.5. Modalités de mise en œuvre du modèle à 1, 2 et 3 mois d'échéance

On se place en conditions opérationnelles au début du mois T . Par convention T désigne le début du mois et l'intervalle $[T ; T+1[$ désigne les jours du mois T . Du point de vue de la prévision, c'est aussi la période de projection à 1 mois d'échéance.

Les occurrences de sécheresse qui nous intéressent sont celles définies par l'indice SPI3, soit des anomalies sur 3 mois consécutifs. Nous devons donc élaborer des modèles capables de prévoir les anomalies de pluies pour les périodes $[T-2 ; T+1[$, $[T-1 ; T+2[$ et $[T ; T+3[$ (Fig.9).

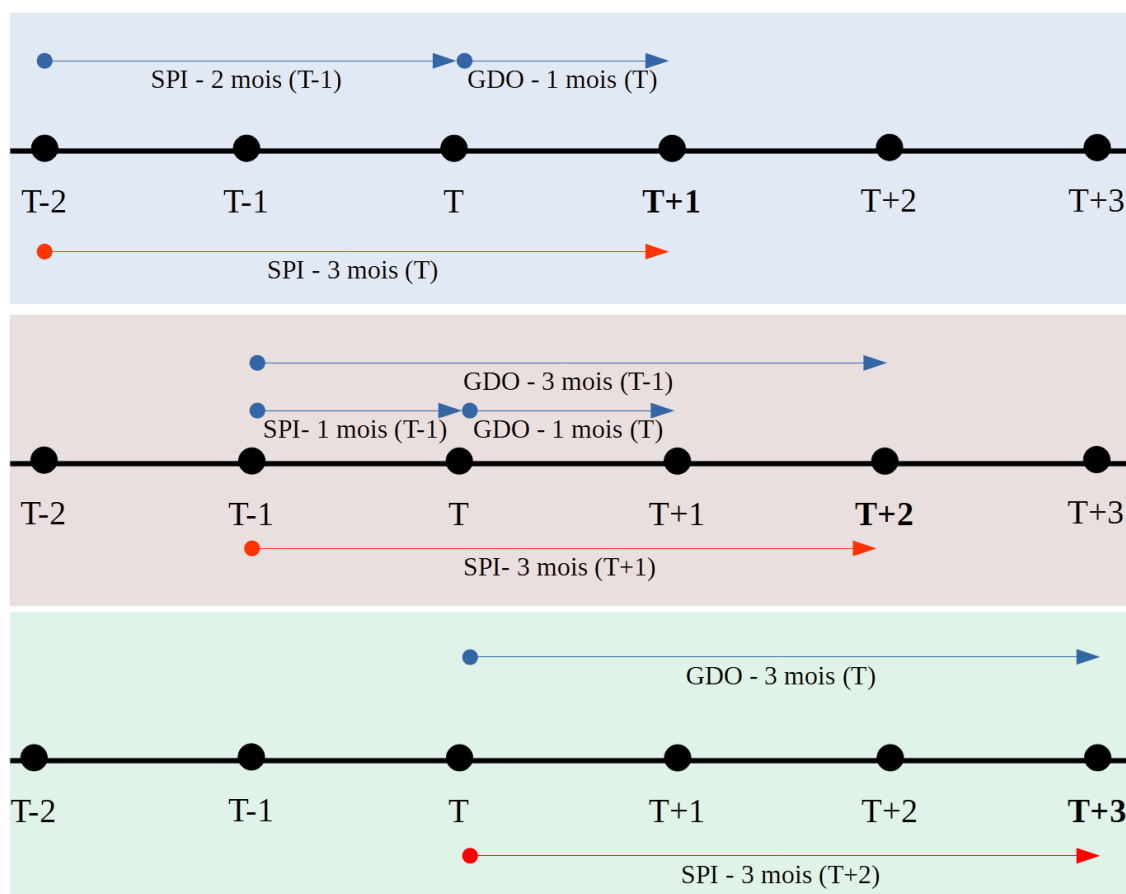


Figure 9 : Schéma de mise en œuvre du modèle de prévision pour les échéances $T+1$, $T+2$ et $T+3$. Les flèches figurent la période couverte par la variable. Les flèches bleues désignent les prédicteurs du modèle, tandis que les rouges symbolisent les prédicands.

Nous disposons des tendances prévues $GDO1(T)$ et $GDO3(T)$ qui couvrent 2 périodes futures intéressantes : $[T, T+1[$ et $[T ; T+3[$. La période $[T ; T+2[$ est largement couverte par la tendance $GDO3(T-1)$, également disponible.

Pour prévoir les anomalies sur $[T-2 ; T+1[$, on utilisera la tendance prévue par $GDO1(T)$ sur $[T, T+1[$ ainsi que le prédicteur $SPI2(T-1)$ qui quantifie les pluies sur $[T-2; T[$. Pour cette échéance, l'apprentissage statistique consistera à confronter les valeurs prévues par le modèle à l'indice $SPI3(T)$.

Pour prévoir les anomalies sur $[T-1 ; T+2[$, on utilisera la tendance prévue par $GDO3(T-1)$ sur $[T-1, T+2[$ ainsi que le prédicteur $SPI1(T-1)$ qui quantifie les pluies sur $[T-1; T[$. $GDO3(T)$ qui va au-delà de l'échéance recherchée peut également figurer dans la liste des prédicteurs candidats. Pour cette échéance, l'apprentissage statistique consistera à confronter les valeurs prévues par le modèle à l'indice $SPI3(T+1)$.

Pour prévoir les anomalies sur $[T ; T+3[$, on utilisera la tendance prévue par $GDO3(T)$ sur $[T, T+3[$. L'apprentissage statistique consistera ici à confronter les valeurs prévues par le modèle à l'indice $SPI3(T+2)$.

Pour illustrer la pertinence de ces choix, on représente la dispersion des valeurs des prédicteurs en fonction du prédicand sous la forme de diagramme à boîte. C'est le cas de la commune de Bourail pour la prévision à 1 mois d'échéance qui est choisi (Fig.10). Les diagrammes à boîte ci-après tendent à indiquer que :

- la variable $SPI2(T-1)$ et l'indice prévu $GDO1(T)$ sont liées à $SPI3(T)$;
- $SPI2(T-1)$ présente un pouvoir discriminant plus prononcé que $GDO1(T)$ vis-à-vis de la variable à prédire $SPI3(T)$.

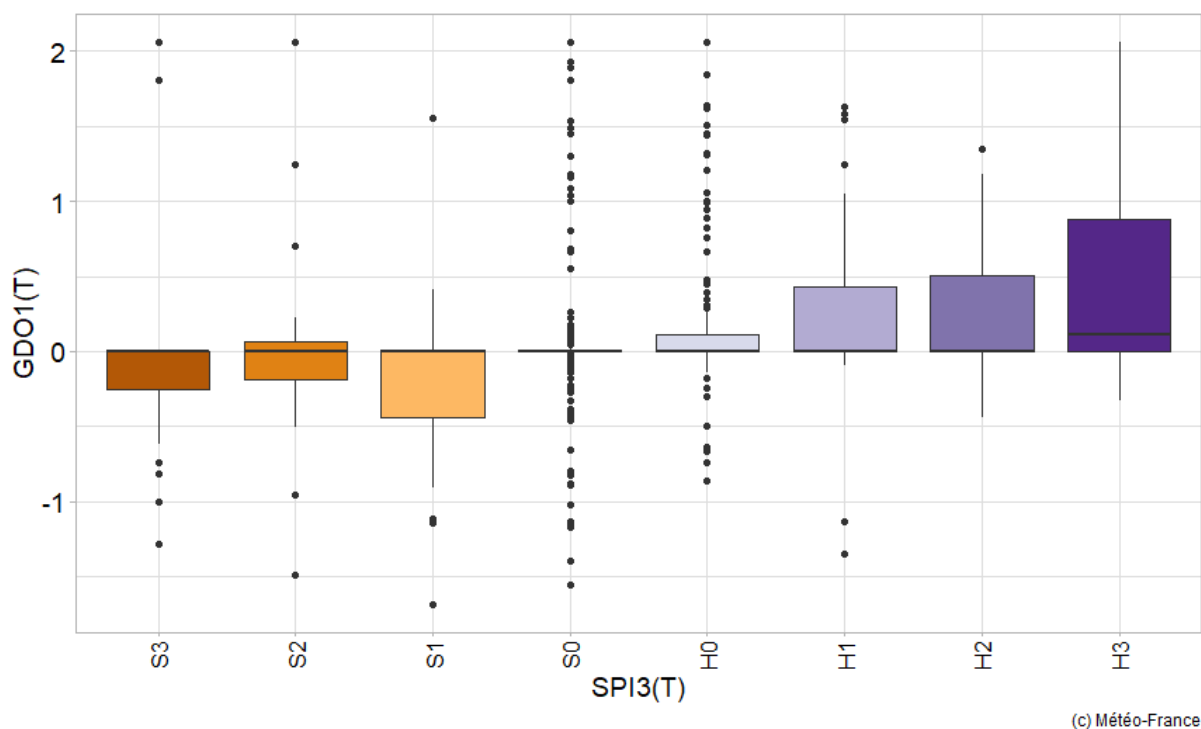


Figure 10 : Dispersion des tendances prévues GDO1(T) (en haut) et des conditions antérieures de sécheresse SPI2(T-1) (en bas) en fonction des catégories de sécheresse observées, SPI3(T), sur la période 1981-2021 pour la commune de Bourail.

1.6. Méthode d'apprentissage statistique

1.6.1. *Protocole de sélection*

$y(T)$ est la variable aléatoire ordinaire à 8 modalités ordonnées $\{y_i, i \leq 8\} = \{D_3 < D_2 < D_1 < D_0 < W_0 < W_1 < W_2 < W_3\}$, reflétant la classe des anomalies de pluie sur 3 mois à l'échelle d'une commune donnée.

Nous voulons utiliser au mieux les prévisions de grande échelle, GDO1(T) ou GDO3(T), et les conditions antérieures de pluies, comme SPI2(T-1) et SPI1(T-1), pour prévoir la probabilité conditionnelle que le prédicand $y(T)$ tombe dans la catégorie y_i ou inférieure. Pour cela, nous faisons l'hypothèse du modèle de régression logistique ordinaire, autrement dit que $\text{logit } P(y(T) \leq y_i | x(T)) \simeq a_i - b_1 x_1(T) - \dots - b_p x_p(T), i \in [1, 7]$, où les x_p sont les variables prédictives à notre disposition. Parmi ces dernières, on peut également considérer la variable qualitative à deux modalités, rendant compte de la saison en cours, ou bien encore celle spécifiant le mois de l'année.

$a_i, b_1, \dots, b_p$ sont les coefficients supposés constants de la régression estimés par la méthode du maximum de vraisemblance.

L'éventail des prédicteurs potentiels étant limité, la tentation est grande de tous les intégrer et d'estimer les paramètres b_p de la régression sur tout l'échantillon de manière à ce que le modèle s'ajuste le plus possible à notre échantillon d'observations.

Comme nous souhaitons in fine disposer d'un modèle qui réalise des prévisions à partir de données inédites, qui ne sont pas forcément présentes dans l'échantillon initial, nous ne privilégierons pas le jeu de prédicteurs complet, celui qui fait que le modèle s'ajuste le mieux aux données observées. Les performances du modèle complet sont en effet trompeuses ; elles reflètent une estimation par principe trop optimiste, de l'erreur de prévision. Celle-ci est liée aux données qui ont servi à l'ajustement du modèle et est d'autant plus faible que le nombre de prédicteurs est grand. Nous privilégierons le jeu de prédicteurs qui garantit la meilleure fiabilité, autrement dit le modèle qui réalise le meilleur compromis entre justesse et robustesse. Cela conduit à rechercher un modèle parcimonieux, c'est-à-dire avec un nombre volontairement restreint de variables prédictives.

La recherche d'un modèle fiable nous conduit logiquement à utiliser une technique dite de « validation croisée » consistant à scinder l'échantillon initial, constitué de 491 mois d'observations, en deux sous-ensembles distincts :

- Le premier, constitué de 246 mois tirés au hasard, est appelé échantillon d'apprentissage. Sa taille équivaut à 50 % de l'échantillon complet. Ces données sont utilisées pour estimer les paramètres de la régression. On envisage toutes les combinaisons de prédicteurs possibles. Tous les modèles font l'objet d'un ajustement.
- Le second sous-ensemble, constitué des 245 mois restants, est appelé échantillon de test. On mesure les écarts entre les observations qui le composent et les valeurs

prévues par les différents modèles ajustés sur l'échantillon d'apprentissage. L'évaluation des scores de prévision conclut cette première phase de test.

La variabilité de SPI3 étant marquée d'une décennie à l'autre (section 1.1), il semble présomptueux de statuer sur le choix du modèle le plus performant après seulement un découpage. On ne pourrait alors pas exclure la possibilité que la prééminence d'un modèle sur les autres ne soit le fruit du hasard du découpage. Pour accroître notre confiance dans le choix final et estimer au mieux les erreurs de prévision, on itère plusieurs fois la division de l'échantillon initial. On calcule la moyenne des scores et on trace les graphes de leur distribution (diagramme boîte). Cette technique d'estimation est également appelée « validation-croisée Monte Carlo ».

Les découpages apprentissage/test sont réalisés aléatoirement avec la seule contrainte que la distribution globale des catégories d'anomalies de pluies soit préservée dans les deux sous-ensembles. Cette disposition apporte la garantie que les catégories S3 et H3 les plus extrêmes ne sont pas mises de côté : cela améliore la représentativité des échantillons d'apprentissage et réduit la variance des estimations des scores de prévision.

1.6.2. Les scores de Brier

Le score de Brier (*Brier, 1950*) est adapté pour évaluer des prévisions exprimées sous forme de probabilité. Il est défini par :

$$BS = 1/N \sum_{k=1}^N (f_i - o_i)^2$$

où N est le nombre de prévisions, f_i et o_i sont respectivement la prévision et l'observation correspondante (qui vaut soit 0, soit 1).

Usuellement, on lui préfère sa version relative, appelée « skill score » de Brier, qui mesure les performances du modèle par rapport à un modèle trivial de référence. C'est généralement le modèle qui consiste à prévoir systématiquement la probabilité climatologique du phénomène qui est choisi à titre de comparaison. Le skill score de Brier (BSS) s'exprime comme suit :

$$BSS = 1 - \frac{BS}{BS_{clim}}$$

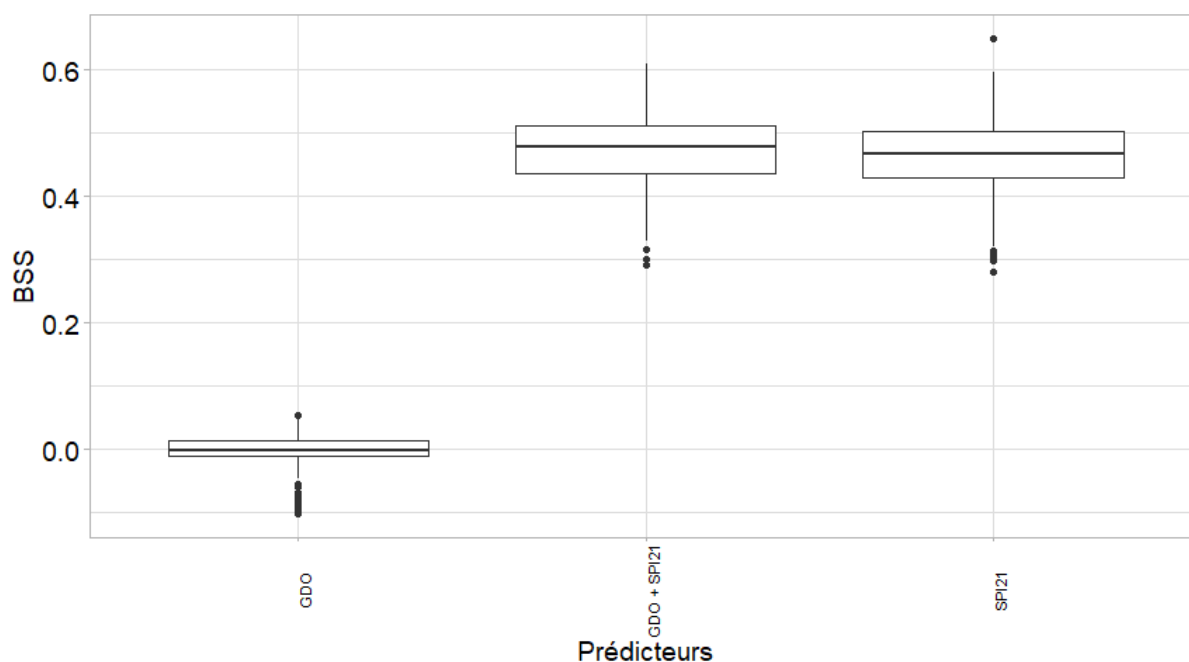
Le BSS s'exprime couramment en pourcentage. Un BSS positif identifie les prévisions meilleures que la climatologie. Le BSS atteint un maximum de 100 % avec une prévision parfaite, qui prévoit toujours et sans erreur une probabilité d'occurrence de l'événement de 100 % ou de 0 %.

1.7. Mise en exploitation

Sur la base de la liste optimale des prédicteurs établie lors de la phase d'apprentissage, on procède à la ré-estimation du modèle sur la base de données complète. Ainsi calibré, le jeu de paramètres est considéré comme arrêté jusqu'à la prochaine réactualisation du modèle (par ex. 1 fois par an). Dans le laps de temps, le modèle est mis en exploitation ; chaque mois il sert au calcul des probabilités prévues pour les 8 catégories d'anomalies de pluies.

La méthode d'apprentissage statistique est mise en œuvre pour toutes les communes calédonniennes et les 3 horizons de prévision. L'ensemble des résultats obtenus sera mis à disposition sur le site du projet.

En guise d'exemple, nous présentons ci-après les prédicteurs retenus, le modèle optimisé et les prévisions obtenues à un mois d'échéance sur la commune de Bourail. Les scores présentés sont relatifs aux prévisions de probabilité des catégories D1 ou inférieures (c'est-à-dire D1, D2 ou D3).



(c) Météo-France

Figure 11 : Distributions des Brier Skill Score (BSS) obtenues en validation croisée « Monte Carlo » après 500 itérations pour les 3 combinaisons de prédicteurs possibles : à gauche GDO1(T), au milieu GDO1(T) et SPI2(T-1), à droite SPI2(T). Probabilité de sécheresse de niveau D1, D2 ou D3.

Sur la figure 11, les Skill Score de Brier montrent clairement que les modèles incluant les conditions de pluies sur 2 mois (SPI2) font de bien meilleures prévisions que le modèle fondé uniquement sur la tendance à 1 mois (GDO1). Ce résultat n'a rien d'étonnant étant donné les poids respectifs des cumuls sur 1 et 2 mois dans la variabilité des pluies sur 3 mois ! Pour autant, cela ne remet pas en cause la pertinence de GDO1 dans la modélisation.

La figure 11 ne permet pas de départager les deux modèles à droite du graphique. Pour analyser les bénéfices éventuels de GDO1, on compare deux à deux les scores obtenus par les deux modèles à chaque itération. Les résultats sont présentés sous la forme d'un nuage de points sur la figure 12. Il apparaît qu'une majorité de points du nuage est située sous la bissectrice, ce qui tend à indiquer que les prévisions du modèle intégrant SPI2(T-1) et GDO1(T) sont de meilleure qualité que celles résultant de SPI2(T-1) seul. On peut montrer par le calcul que le BSS augmente de manière significative de 1,6 à 2,2 % (estimation au seuil de confiance 95 %).

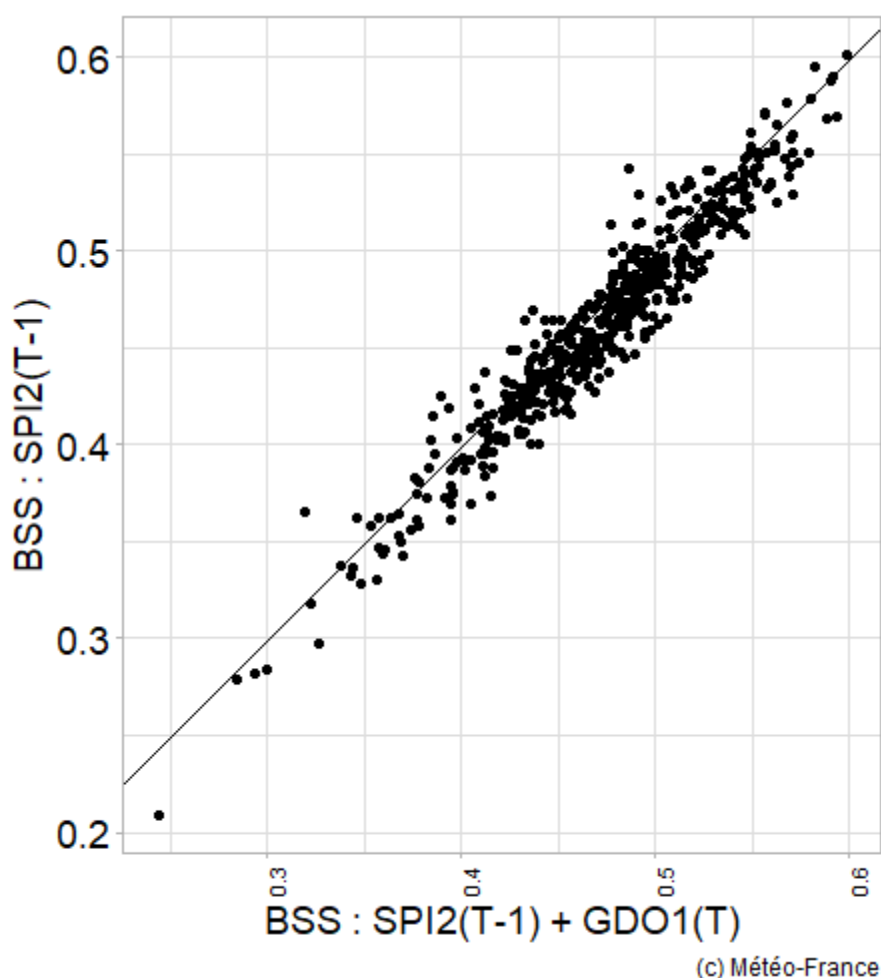


Figure 12 : Brier Skill Score (BSS) obtenus en validation croisée « Monte Carlo » après 500 itérations pour les modèles de régression logistique ordinaire suivant : SPI2(T-1) + GDO1(T) (en abscisse), SPI2(T-1) (en ordonnée).

C'est finalement le modèle à 2 variables qui est retenu pour prévoir la probabilité de sécheresse sur la commune de Bourail à 1 mois d'échéance. Ainsi défini, le modèle de régression logistique ordinaire est ajusté sur la totalité de l'échantillon. À partir des coefficients estimés et de l'équation de régression logistique ordinaire, nous sommes désormais en mesure d'extrapoler les probabilités conditionnelles $P(y(T) \leq D_1 | x)$ pour tous les couples $(SPI\ 2, GDO\ 1)$. Le diagramme 2D de la figure 13 représente les différents niveaux de probabilités obtenus pour $\{-2.0, -1.8, \dots, 1.8, 2.0\} \times \{-2.0, -1.8, \dots, 1.8, 2.0\}$.

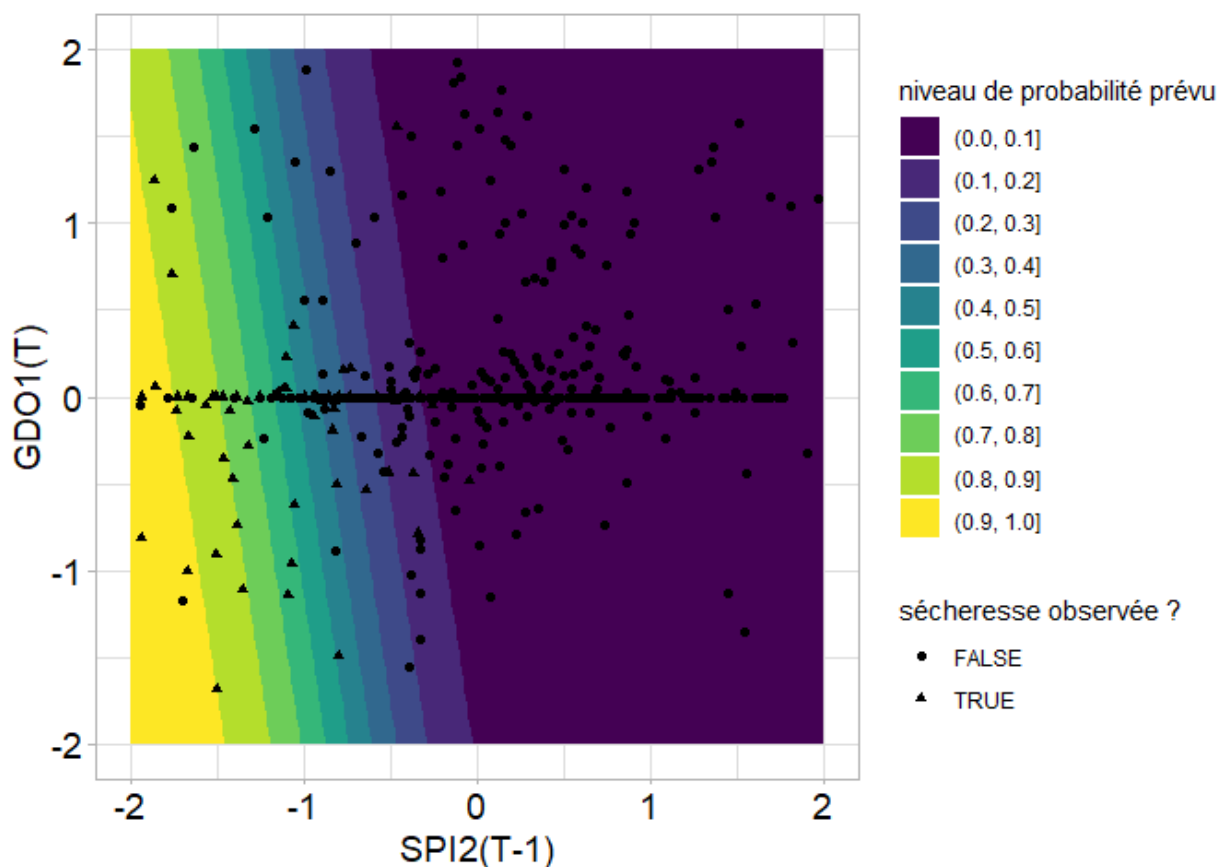


Figure 13 : Probabilités de sécheresse prévues sur la commune de Bourail étant données les conditions antérieures de sécheresse $SPI2(T-1)$ et la tendance prévue $GDO1(T)$. Le modèle de régression logistique ordinaire est ajusté sur la base de données complète (janv. 1981-décembre 2021). Les observations qui ont servi à calibrer le modèle sont repérées par des points ou des triangles, ces derniers symbolisant les occurrences de sécheresse avérées.

1.8. Prise de décision et probabilité de sécheresse

Le modèle fournit une prévision sous forme de probabilité. Dans la pratique, il est important pour le décideur que la prévision de probabilité soit assortie d'une règle de prévision $p(T)$ munie d'un seuil de décision τ à partir de cette estimation.

$$\rho(T) = \begin{cases} 1 & \text{si } P(y(T) \leq S_1) \geq \tau \\ 0 & \text{sinon} \end{cases}$$

$\tau \in [0,1]$ est un seuil fixé par l'utilisateur

Le seuil de décision τ est choisi de manière à optimiser une fonction $f : \tau \rightarrow f(\tau)$ où f est :

- soit un score qui mesure de manière équitable l'adéquation entre $p(T)$ et les observations ;
- soit un rapport coût/perte qui traduit la valeur économique des prévisions (Jolliffe et Stephenson, 2012)

La deuxième option met en œuvre une démarche d'optimisation qui tient compte du besoin exprimé par l'utilisateur final. Elle présente l'avantage de rendre sensible la qualité des prévisions, mais sa spécificité la rend inapplicable dans le cadre de notre étude. Elle nécessite d'avoir au préalable étudié la vulnérabilité de son activité à l'aléa sécheresse, en précisant notamment le coût des actions à mettre en œuvre en prévision de la sécheresse, ainsi que les pertes financières occasionnées si aucune mesure protectrice n'est prise.

Dans le domaine plus prosaïque de la statistique, la vérification des prévisions d'évènement, comme le dépassement du seuil de sécheresse, commence toujours par un tableau de contingence. Celui-ci représente la distribution conjointe des observations et des prévisions avec 4 combinaisons possibles (tableau 2).

Tableau 2 : Matrice de confusion également appelée tableau de contingence.

		Évènement observé	
		oui	non
Évènement prévu	oui	TP vrais positifs	FP faux positifs
	non	FN faux négatifs	TN vrais négatifs

$$BIAS = \frac{TP + FP}{TP + FN}$$

BIAS est le biais de fréquence. Il mesure le rapport entre la fréquence des événements prévus et la fréquence des événements observés. Il indique si le système de prévision a tendance à sous-estimer ($BIAS < 1$) ou à surestimer ($BIAS > 1$) les événements. Une prévision parfaite conduit à $BIAS = 1$. BIAS ne mesure pas la correspondance entre la prévision et les observations, mais seulement les fréquences relatives.

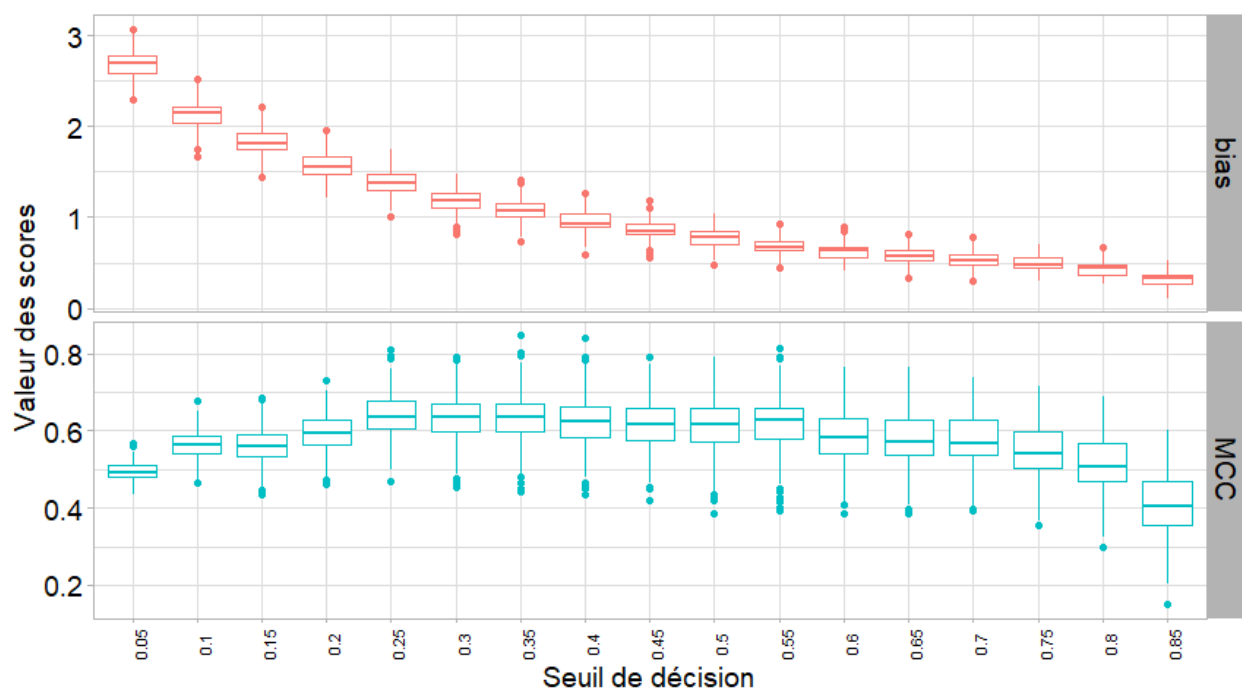
$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}}$$

Le coefficient de corrélation de Matthews (MCC ; *Matthews 1975*) est considéré comme un score adapté pour mesurer la qualité des prévisions d'occurrence même lorsque les deux classes sont de tailles très différentes (par exemple $TP+FN \ll FP+TN$). MCC varie entre -1 et 1. Un coefficient de 1 représente une prévision parfaite, 0 est équivalent à une prédiction aléatoire (par exemple, telle qu'elle serait produite par un modèle qui génère des prévisions uniformément aléatoires, ou par un modèle respectant la distribution des classes de l'ensemble d'apprentissage), et -1 indique un désaccord total entre la prévision et l'observation.

Le choix du seuil de décision se porte sur celui qui garantit le meilleur compromis entre BIAS proche de 1 et MCC élevé. Là encore, pour garantir la robustesse du choix, la sélection repose sur une technique de type « Monte-Carlo » (*Esposito et al. 2021*) :

1. Le modèle est ajusté sur l'échantillon d'apprentissage, par exemple 1981-2020 ;
2. On évalue les probabilités de sécheresse prévues $p(y \leq D_1 | x)$ sur l'ensemble d'apprentissage lui-même ;
3. Une gamme de seuils (par ex. de 5 à 95 % par pas de 5 %) est utilisée pour identifier les occurrences prévues sur l'échantillon d'apprentissage. Plus le seuil est bas, plus le nombre d'instances prédites comme sèche augmente ;
4. N sous-échantillons aléatoires sont tirés de l'ensemble d'apprentissage en veillant à respecter la proportion climatologique d'épisodes secs D_1 ; la taille relative des sous-échantillons est fixée à 33 %.
5. BIAS et MCC sont calculés pour chaque sous-ensemble et pour chacun des seuils ;
6. Pour chaque seuil, BIAS et MCC sont agrégés sur les N sous-ensembles en utilisant la médiane ;
7. Le seuil optimal de décision est choisi de telle sorte que la médiane de BIAS correspondante soit proche de 1 et que la médiane de MCC soit la plus élevée possible.

Cette méthode d'optimisation est mise en œuvre pour le seuil des prévisions de sécheresse sur la commune de Bourail (figure 14). Le coefficient de corrélation MCC croît avec le seuil pour atteindre un plateau lorsque les probabilités de décision sont comprises entre 25 et 55 %. Quant au biais de fréquence, il est proche de 1 pour les seuils compris entre 35 et 40 %. Ces deux intervalles se recoupent et on en déduit qu'un seuil compris entre 35 et 40 % fournira des prévisions d'occurrence de sécheresse fiables et bien calibrées.



(c) Météo-France

Figure 14 : Scores estimés par échantillonnage aléatoire stratifié sur la période d'apprentissage 1981-2021 pour un ensemble de seuils compris entre 5 % et 85 % : BIAS (en haut), MCC (en bas) . Prédiction des occurrences de sécheresse sur la commune de Bourail à 1 mois d'échéance.

2. INTERFACE HOMME MACHINE

Une plateforme utilisateur, ou Interface Homme Machine (IHM), a été développée afin de permettre une diffusion et une visualisation claire et simplifiée de l'information sécheresse. Les différents indicateurs développés dans le cadre de ce projet (Actuels et Prévisionnels) sont intégrés et affichés via cette plateforme. La finalité ici est de fournir un outil objectif et spatialisé pour le suivi et l'anticipation de la sécheresse végétale, servant d'aide à la décision aux institutionnels pour une prise de décision éclairée.

L'IHM se présente sous la forme d'une page web de visualisation, accessible par l'utilisateur au lien suivant : eo4dmweb. L'accès se fait en ayant généré préalablement une connexion sécurisée de type VPN (*Virtual Private Network*) au serveur d'INSIGHT sur lequel est hébergée la plateforme.

Dans une première section, nous présentons l'interface d'un point de vue utilisateur, c'est-à-dire en s'attardant particulièrement sur la description des différents onglets et outils de visualisation, tout en expliquant le contenu affiché. Une deuxième section présente le fonctionnement de l'IHM et s'attarde particulièrement sur l'utilisation et l'alimentation du serveur de données qui est basé sur *Geoserver*.

2.1. Présentation de l'interface

2.1.1. *Vue générale*

L'interface se compose de trois fenêtres principales présentant l'état passé et futur de la sécheresse en Nouvelle-Calédonie :

- 1) Carte de sécheresse actuelle ou passée
- 2) Profil temporel
- 3) Carte de sécheresse prévisionnelle

La disposition et l'aspect de ces trois fenêtres sont présentées dans la figure 15 ci-après.

1

3

2

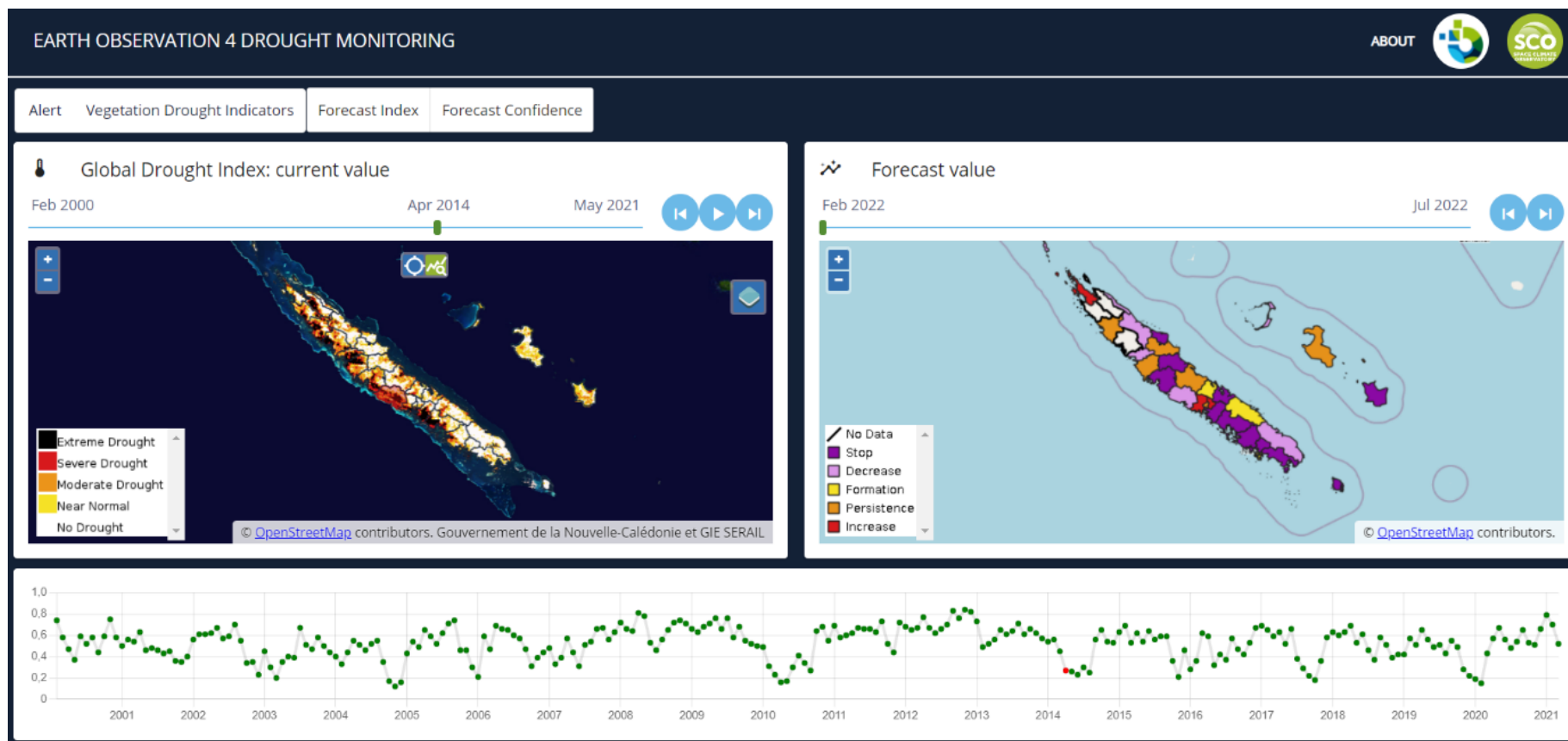


Figure 15 : Vue générale de la plateforme (IHM)

2.1.2. Cartes actuelles et passées

L'utilisateur peut sélectionner plusieurs produits via les onglets "Alert" et "Vegetation Drought Indicators". La date à afficher est choisie en déplaçant le curseur sur la barre chronologique (cf. Fig.16).

L'onglet "**Alert**" permet l'affichage de cartes mensuelles à l'échelle communale issues de l'observation de plusieurs produits (précipitation, humidité, évapotranspiration, végétation). Il s'agit du niveau le plus complet et global permettant de localiser les communes affectées ou qui seraient potentiellement affectées par la sécheresse agricole. L'utilisateur peut sélectionner :

- "Combined Indicator" qui affiche la carte d'alerte globale, avec 4 catégories d'alertes : No Alert, Watch (déficit pluviométrique), Warning (déficit d'humidité et/ou d'évapotranspiration), Alert (stress hydrique de la végétation)
- "Detailed products" qui détaille chaque sous-produit

Onglets
sécheresse
actuelle/passée

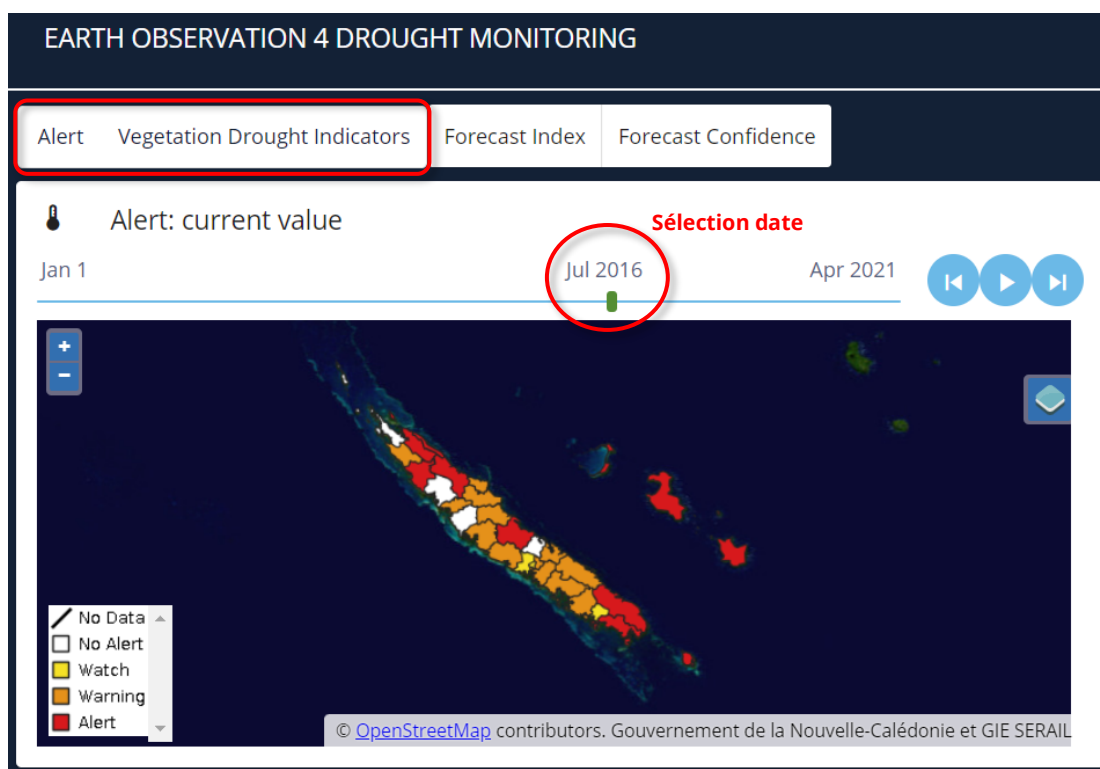


Figure 16 : Sélection et affichage des cartes d'alertes globales

L'onglet **"Vegetation Drought Indicators"** permet d'accéder à une échelle intra-communale en se focalisant sur la sécheresse végétale, c'est-à-dire uniquement sur l'état de stress hydrique des plantes (Fig.17). Deux niveaux de précisions sont disponibles :

- "Global scale" informant à une résolution de 1 km et à un pas de temps mensuel sur le niveau d'intensité de la sécheresse (onglet "Drought Indicator") et le niveau de confiance ("Confidence", fourni à la commune) estimé selon la qualité des images (nuages...) et la corrélation par rapport aux données de référence (précipitations)
- "Local scale" informant sur la sécheresse à une résolution de 10 m et un pas de temps de 10 jours. Ici, l'indice de confiance est fourni au pixel et selon la qualité de la donnée uniquement: 0 qualité minimale (non affichée) et 1 qualité maximale.

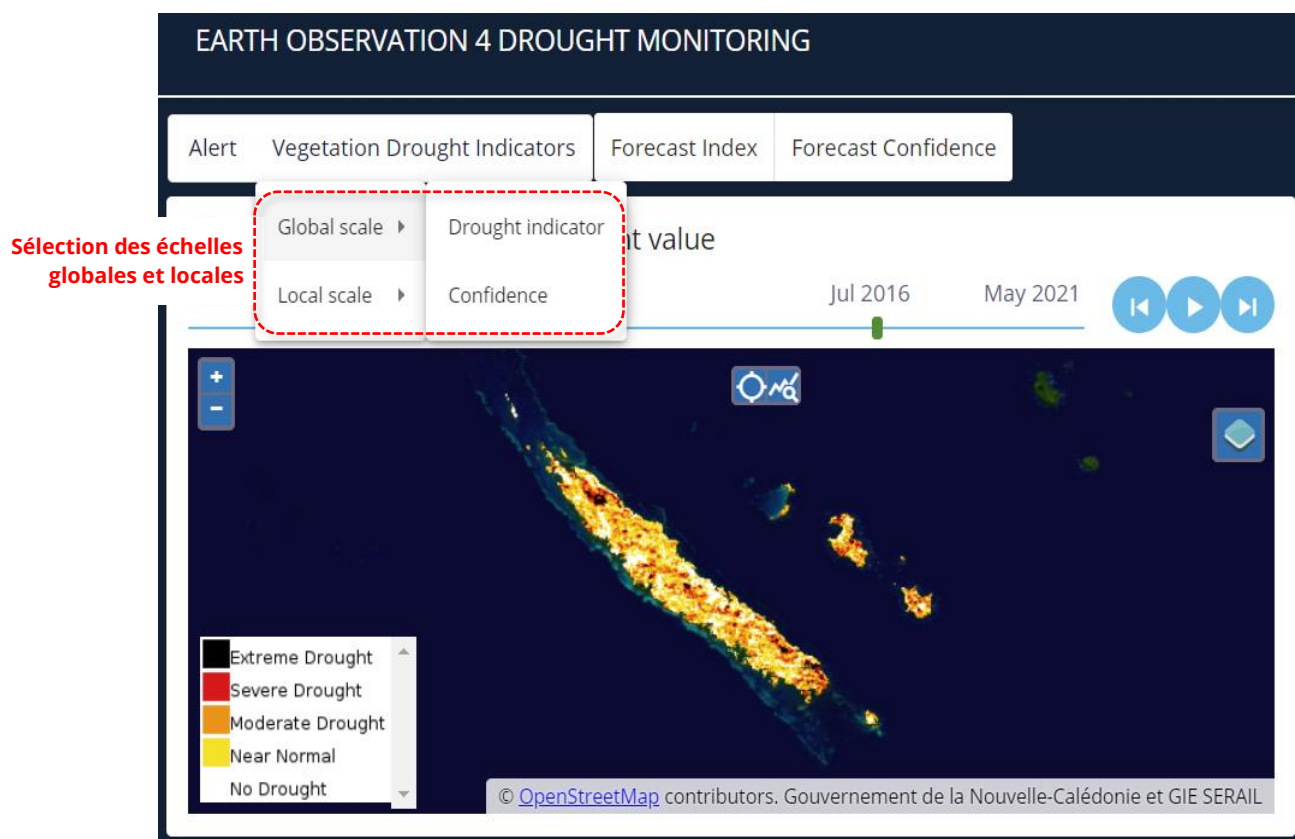


Figure 17 : Sélection et affichage des indicateurs actuels/passés aux échelles globales/locales

2.1.3. Profil temporel

L'interface permet l'affichage d'un graphique présentant le profil temporel de la sécheresse végétale au cours de la période traitée (2000 – 2021) et correspondant à un point géographique choisi sur la carte (Fig.18). Pour cela, l'utilisateur doit sélectionner :

- 1) Une carte de niveau d'intensité de la sécheresse végétale via les onglets dédiés (cf. section précédente)
- 2) Un des deux outils fournis :  pour afficher le profil temporel d'un seul pixel,  pour afficher le profil temporel de l'ensemble d'une commune (moyenne de plusieurs pixels)
- 3) La zone voulue en cliquant sur la carte (fenêtre de gauche de l'interface). Si besoin, l'utilisateur peut utiliser l'outil  afin de sélectionner les couches à afficher (indicateurs, fonds de carte)

A noter : Les valeurs minimales du profil correspondent au cas le plus défavorable (cas de sécheresse extrême), et les valeurs maximales au cas le plus favorable (bonne santé). Du fait de l'utilisation de produits différents aux échelles globales et locales, les plages de valeurs des indicateurs ne sont pas identiques entre ces deux échelles, mais présentent néanmoins la même information.

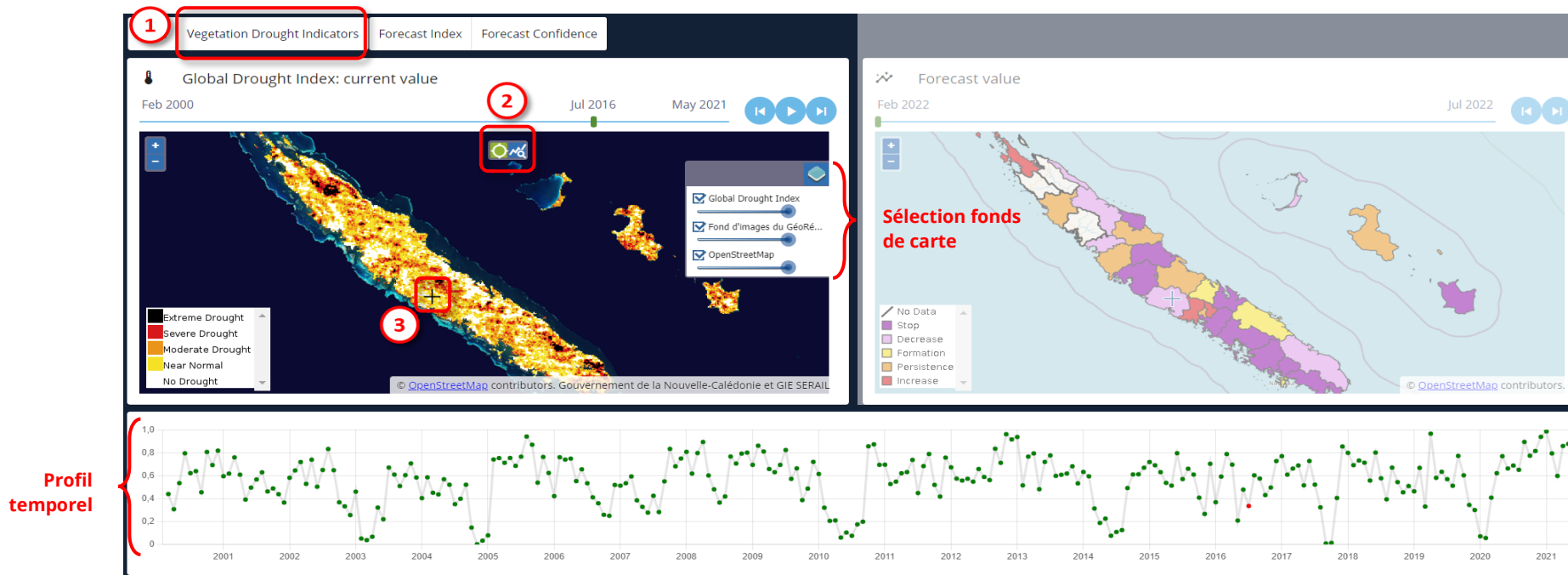


Figure 18 : Affichage du profil temporel de l'indicateur de sécheresse végétale

2.1.4. Cartes prévisionnelles

Ici, l'utilisateur peut avoir accès aux tendances d'évolution de la sécheresse végétale, qui se présentent sous la forme de cartes à l'échelle communale. Ces tendances sont faites à une prévisibilité de +1 mois, +2 mois et +3 mois (à partir du mois en cours). L'utilisateur peut sélectionner la date de prévision via la barre chronologique. Deux informations peuvent être affichées via les onglets suivant :

- **"Forecast Index"** qui indique les tendances d'évolution de la sécheresse, classées selon 5 catégories : Stop (arrêt de la sécheresse végétale), Decrease (réduction), Formation (apparition), Persistence (persistance), Increase (aggravation).
Ces évolutions sont obtenues à partir des prévisions de précipitations obtenues selon la méthodologie décrite en section 1.
- **"Forecast Confidence"** qui fournit un indice sur la confiance de la prévision. Il s'agit là du score de corrélation entre la l'indicateur satellite de sécheresse végétale (VHI) et le SPI-3 mois.

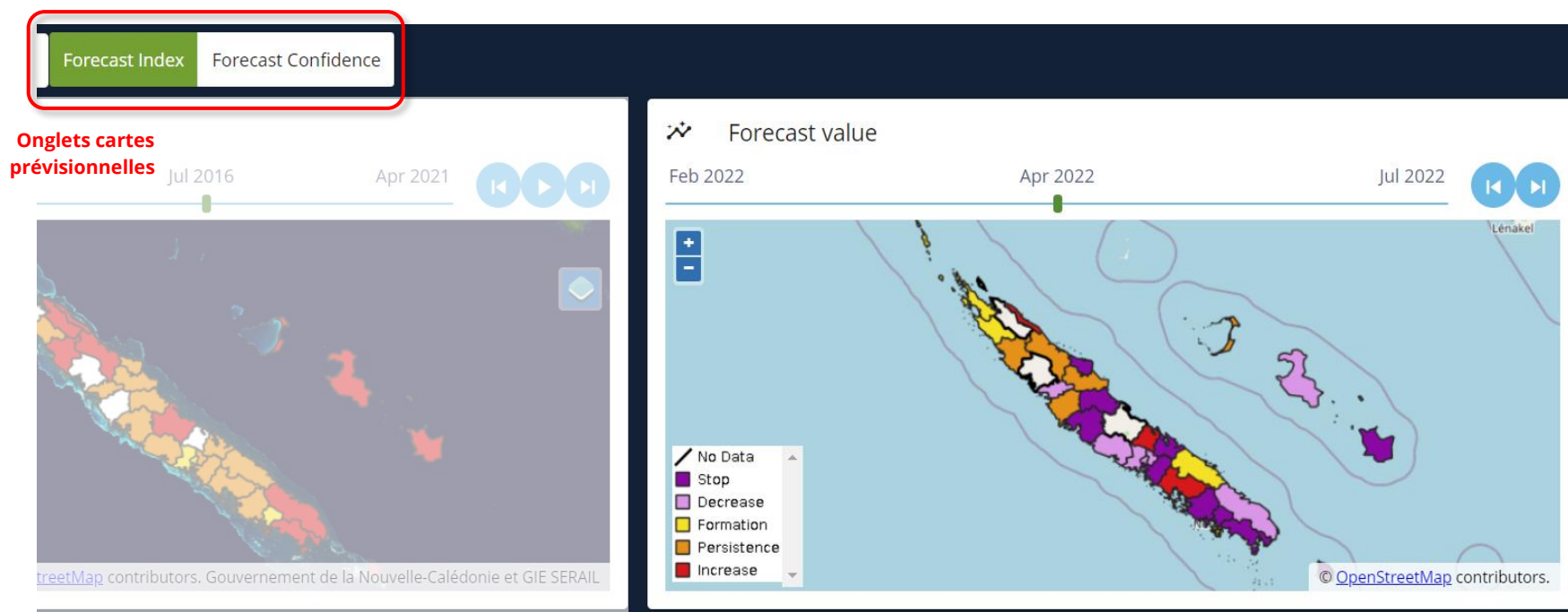


Figure 19 : Affichage et sélection des cartes prévisionnelles de sécheresse végétale

2.1.5. *Note d'utilisation de la plateforme*

La plateforme telle que développée dans ce projet permet d'accéder à plusieurs informations sur la sécheresse en Nouvelle-Calédonie. Ainsi, pour un utilisateur qui souhaiterait connaître l'état de la sécheresse végétale au cours du dernier mois traité (mois m), et les mois futurs ($m+1$, $m+2$, $m+3$), voici la démarche recommandée :

1. Observer tout d'abord le *niveau d'alerte par commune* pour le mois m permettant d'avoir un premier niveau d'information générale sur l'impact de la sécheresse au sein de chaque commune. Il s'agit d'un niveau adapté à une étude territoriale des risques de sécheresse.
2. Pour une étude plus fine (intra-communale), accéder ensuite à *l'indicateur de sécheresse végétale à l'échelle kilométrique d'une part, puis décamétrique*. Ces indicateurs apporteront une information objective plus résolue et focalisée sur le stress hydrique de la végétation.
3. Observer le *profil temporel* pour une zone particulière. Cela permettra d'avoir une information sur la durée de sécheresse végétale détectée au cours des derniers mois et sera une première information pour la tendance d'évolution.
4. Connaissant l'état actuel du niveau de sécheresse végétale, l'utilisateur peut ensuite prendre connaissance des tendances d'évolution fournies via l'indicateur de prévisions de précipitation. En parallèle, nous recommandons également à l'utilisateur de prendre connaissance du niveau de confiance pour la commune désirée, permettant d'indiquer à quel point les prévisions reflèteront l'évolution de la sécheresse végétale.

2.2. Description technique

2.2.1. *Fonctionnement général*

La plateforme est une application web développée en Angular, avec Openlayers comme librairie cartographique. Cette application consomme des flux de données fournis par une **instance de Geoserver** qui utilise :

- une base de données pour stocker les informations vectorielles qu'elle consomme
- un répertoire local contenant des geotiffs pour stocker les images raster qu'elle consomme.

2.2.1.1.1. Types de données

Le projet produit différents types de données :

- Des séries temporelles d'images à deux niveaux de résolution local et global calculés respectivement à partir d'images Sentinel2/Landsat et MODIS.
- Des séries temporelles d'indicateurs agrégés à la commune pour l'historique mais également la prévision.

Chaque type de donnée fait l'objet d'une procédure pour être insérée dans *Geoserver* et transformée en flux WMS (*Warehouse Management System*) consommable par l'application web.

2.2.1.1.2. Flux WMS Geoserver

Les séries temporelles d'images sont diffusées sous le standard WMS-T grâce au plugin *ImageMosaic* de *Geoserver*.

Les images de la série temporelle sont toutes stockées dans un répertoire avec une nomenclature simple qui fait apparaître la date de l'image dans le nom. Voici par exemple ce que contient le dossier geoserver contenant les images du flux LDI (Local Drought Indicator) :

LANDSAT SENTINEL2 NDWI 19990901.tif
LANDSAT SENTINEL2 NDWI 19990910.tif
LANDSAT SENTINEL2 NDWI 19990920.tif
LANDSAT SENTINEL2 NDWI 19991101.tif
LANDSAT SENTINEL2 NDWI 19991110.tif
LANDSAT SENTINEL2 NDWI 19991120.tif
LANDSAT SENTINEL2 NDWI 20000101.tif
LANDSAT SENTINEL2 NDWI 20000110.tif
LANDSAT SENTINEL2 NDWI 20000120.tif
LANDSAT SENTINEL2 NDWI 20000201.tif
...
LANDSAT SENTINEL2 NDWI 20210220.tif

Dans le même répertoire doivent être présents des fichiers de configuration permettant à *Geoserver* d'enregistrer dans une base de données les références des images de la série temporelle :

`datastore.properties`

`indexer.properties`

`timeregex.properties`

Une fois que tous ces fichiers sont bien en place, il suffit dans *Geoserver* de créer un nouvel entrepôt de type *ImageMosaic* et de sélectionner le dossier. *Geoserver* scanne le contenu du dossier et indexe l'ensemble des images de la série temporelle. On peut ensuite ajouter un flux WMS basé sur cet entrepôt qui dispose d'une dimension temporelle que l'on active dans l'onglet Dimensions.

L'application utilise également un flux WFS diffusant les emprises des images de la série temporelles. Ce flux est notamment utilisé par l'application web pour lister l'ensemble des dates disponibles dans la série temporelle. Il faut alors ajouter ce flux à partir de l'entrepôt connecté à la base de données (*EO4DM_vector*). Normalement, après la création de la couche liée à la série temporelle, une nouvelle table apparaît dans la base de données et l'on peut publier une couche à partir de cette table (en sélectionnant l'entrepôt *EO4DM_vector*).

Les séries temporelles d'indicateurs à la commune sont livrées sous la forme de fichiers .csv contenant pour chaque date la valeur de l'indicateur pour l'ensemble des communes. Une table est créée dans la base de données pour recevoir les informations contenues dans ces fichiers .csv.

Une couche est ensuite publiée dans l'interface de *Geoserver* à partir de l'entrepôt *EO4DM_vector* et de la table nouvellement créée.

2.2.1.1.3. Description des flux de données Geoserver

L'application web utilise 4 flux de données WMS pour fonctionner :

- GDI (*Global Drought Indicator*) : série temporelle d'images mensuelles avec une résolution de 1km
- LDI (*Local Drought Indicator*) : série temporelle d'images décadaires (tous les 10 jours) avec une résolution de 10m
- Alert : série temporelle d'indicateurs communaux mensuels
- Prévision : série temporelle d'indicateurs communaux prévisionnels avec 3 dates +1mois, +2mois et +3mois.

2.2.2. Guide pour l'insertion de nouvelles données

2.2.2.1.1. GDI

Pour insérer une nouvelle image dans la série temporelle, il faut (après avoir rajouté le fichier image au dossier) supprimer dans un premier temps l'index associé.

Pour supprimer l'index cela se passe en 2 étapes :

- Supprimer le fichier `/home/geouser/geoserver/geoserver_data/data/GLOBAL/GLOBAL.properties` sur le serveur
- Supprimer la table `GLOBAL` dans la base de données pour cela on peut utiliser la commande suivante depuis le serveur :

```
docker exec -it geoserver_geoserver_1 psql -h db -U insight  
-d eo4dm -c "DROP table \"GLOBAL\""
```

Il faut ensuite relancer l'indexation de l'entrepôt `GLOBAL` en cliquant sur le bouton « Sauvegarder » dans la page de l'entrepôt GDI de l'interface de *Geoserver*.

2.2.2.1.2. [LDI](#)

Les images brutes du niveau local sont générées par tuile Landsat. Il faut dans un premier temps **fusionner les images des différentes tuiles** pour ensuite les déposer dans le dossier de *Geoserver*. Cela se passe avec l'outil *mosaicLocal.py*. Cet outil utilise la librairie *gdal* et s'utilise comme suit :

```
python3 mosaicLocal.py -i /path/to/folder -o /output/folder
```

En entrée, les images de toutes les tuiles doivent être dans le même répertoire et avoir la nomenclature comme suit :
`LANDSAT_SENTINEL2_082075_202103_NDWI_ANOD3_POSTPROC.tif`.

L'outil crée en sortie une image avec le nom `LANDSAT_SENTINEL2_NDWI_20210320.tif`

Pour insérer une nouvelle image dans la série temporelle, il faut (après avoir rajouté le fichier image au dossier) supprimer dans un premier temps l'index associé.

Pour supprimer l'index cela se passe en 2 étapes :

- Supprimer le fichier `/home/geouser/geoserver/geoserver_data/data/LOCAL/LOCAL.properties` sur le serveur
- Supprimer la table `LOCAL` dans la base de données pour cela on peut utiliser la commande suivante depuis le serveur :

```
docker exec -it geoserver_geoserver_1 psql -h db -U insight  
-d eo4dm -c "DROP table \"LOCAL\""
```

Il faut ensuite relancer l'indexation de l'entrepôt `LOCAL` en cliquant sur le bouton « Sauvegarder » dans la page de l'entrepôt LDI de l'interface de *Geoserver*.

2.2.2.1.3. [Alert](#)

Il suffit ici **d'insérer dans la base de données une nouvelle ligne dans la table Alert**. Deux scripts bash sont fournis :

- `create_alert.bash` : permet de générer un fichier sql pour créer la table

```
create_alert.bash create_alert.sql
```

- *insert_alert.bash* : permet de générer un fichier sql pour remplir la table Alert avec les éléments des fichiers .csv

```
insert_alert.bash Alert_*.csv alerts.sql
```

Il faut **ensuite exécuter le fichier SQL** sur le serveur de la base de données. Pour cela il faut utiliser un client *psql* avec par exemple la commande suivante :

```
psql -h 192.168.13.54 -U insight eo4dm -f alert.sql
```

2.2.2.1.4. [Prevision](#)

La même procédure s'applique ici en utilisant les scripts *create_forecast.bash/insert_forecast.bash*.

3. CONCLUSION ET EXTERNALITES

3.1. Déroulement de l'étude

Pour rappel, le déroulement global de l'étude s'est décomposé en **six « Work Packages » (WP)** aboutissant à la mise au point de l'outil :

- WP1 Bancarisation de la donnée : stockage et organisation des données satellites disponibles sur le territoire de la Nouvelle-Calédonie. Un lien a également fait avec les différentes données de Météo-France Nouvelle Calédonie utiles à EO4DM
- WP2 Indicateurs Actuels : réalisation d'une chaîne de traitements des images satellites pour la production d'indicateurs actuels aux échelles (1) locales et (2) globales
- WP3 Indicateurs Prévisionnels : mise en place du calcul de l'indice SPEI (*Standardized Precipitation Evapotranspiration Index*), évaluation de la prévisibilité et tendances d'évolution de la sécheresse végétale
- WP4 IHM Actuel : développement d'une maquette de l'IHM pour la diffusion des indicateurs actuels
- WP5 IHM Prévisionnelle : diffusion des indicateurs prévisionnels et finalisation de l'IHM
- WP6 Clôture : recettage de l'IHM, retour d'expérience et valorisation des produits

L'organisation des WP tel qu'ils se sont déroulés est précisée dans le **planning réel** présenté ci-après (Fig.20). Ce dernier présente également les différents points collaborateurs et utilisateurs réalisés cours du projet ainsi que l'augmentation ou la diminution des délais prévus initialement. Pour rappel, ce projet a été estimé sur une période d'un an avec une fin prévue fin mars 2022. La fin réelle et opérationnelle du projet est définie au **29 juin 2022**, incluant la finalisation des différents WP ainsi que la livraison des livrables attendus à savoir : (1) le rapport de mi-parcours (version 1 et version 2), ainsi que le rapport final (ce document).

Un retard global de 3 mois est constaté par rapport au planning initial. Les raisons à cela sont d'une part le délai entre le lancement opérationnel du projet et le lancement administratif, engendrant notamment un retard dans l'achat et la mise en place du serveur de stockage/calcul et donc rallongeant le temps dédié à la bancarisation des données satellites. De plus, un délai dans le développement des indicateurs prévisionnels a été généré en raison des difficultés côté Météo-France à entrer en contact avec certains partenaires identifiés en métropole, et dû également à des soucis de disponibilités des référents vis-à-vis de leurs responsabilités dans la région (fortes pluies en cette année de *Niña*). Ce délai a été mis à profit pour approfondir la documentation sur la méthodologie d'évaluation des indicateurs satellites (version 2 du rapport de mi-parcours) et pour organiser une suite à ce projet (voir section 3.3).

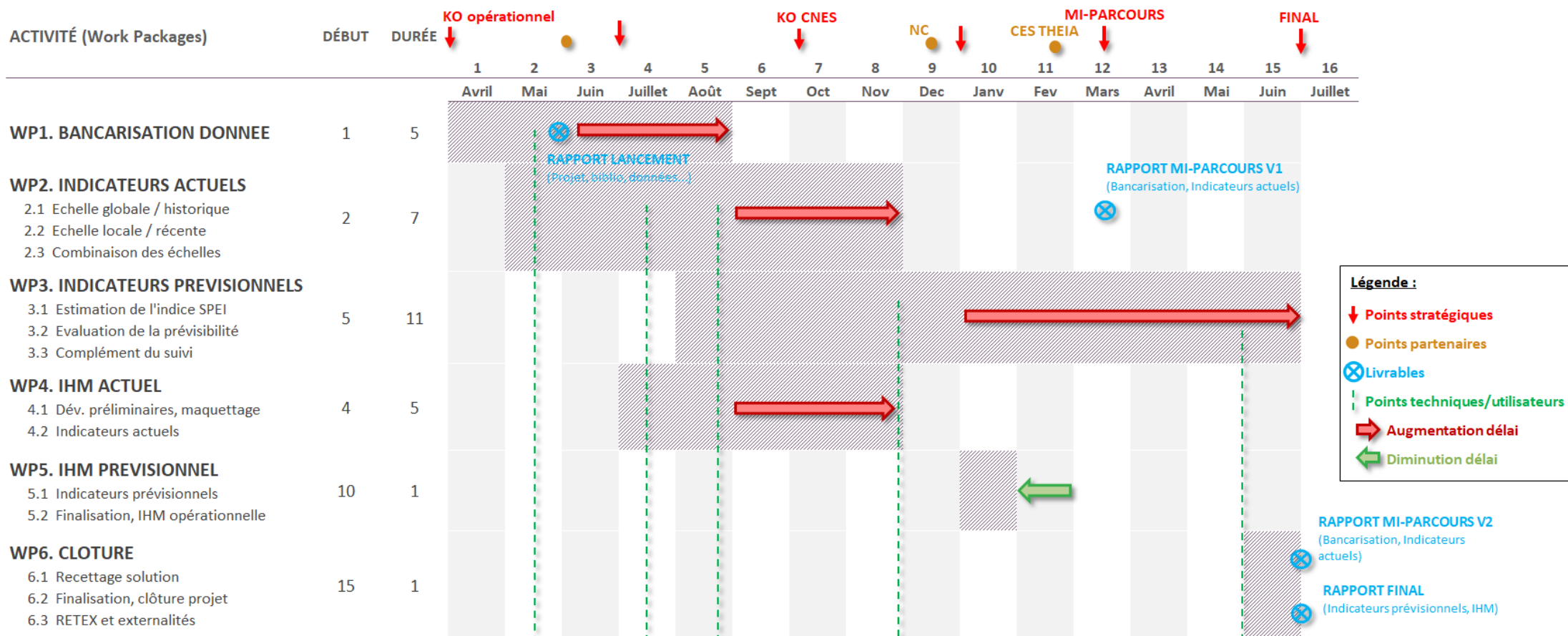


Figure 20 : Gantt réel de l'étude, comparé au planning initial

3.2. Recettage solution

L'objectif de ce projet était de **développer un outil de suivi des épisodes de sécheresse** en s'appuyant sur des indicateurs éprouvés et objectifs issus des données complémentaires d'observation de la terre. Les indicateurs mis en place devaient être diffusés via une **plateforme utilisateur de visualisation (IHM)** afin d'apporter une réponse objective, actualisée, spatialisée et temporelle aux institutionnels pour une prise de décision éclairée.

A l'issue de ce projet, nous pouvons dire que l'outil développé répond au cahier des charges initial. Le tableau ci-dessous fait un récapitulatif des différentes phases d'avancement et de finalisation de chaque tâche (Work Packages). Les différentes tâches des Work Packages ont bien été réalisées, incluant à chaque fois une finalisation des phases de développement, de validation (évaluation scientifique) et de retours utilisateurs (tests par l'Agence Rurale).

Tableau 3 : Bilan sur la finalisation de la solution

TACHES (WP)	Développements	Validation	Retours utilisateur
Indicateur ACTUEL	Finalisés	Finalisée	FAIT
Indicateur PREVISIONNEL	Finalisés	Finalisée	<i>En attente de conditions opérationnelles</i>
PLATEFORME	Finalisés	Finalisée	FAIT

En raison de l'absence de sécheresse lors du déroulement du projet (année 2021 en période de *Niña*), le test de l'outil par l'utilisateur n'a pu être réalisé en conditions opérationnelles. Cependant, l'Agence Rurale a tout de même pu tester l'outil sur des années antérieures marquées par de fortes sécheresses (ex.: fin 2019) et faire un retour utilisateur sur la plateforme (son accessibilité, son ergonomie) et sur l'indicateur de sécheresse diffusé. L'annexe 1 détaille le retour utilisateur fait par l'Agence Rurale, qui confirme donc l'utilité et l'apport de l'outil pour ses besoins de suivi et d'anticipation de la sécheresse végétale notamment dans les phases d'indemnisation des agriculteurs (meilleure qualification et spatialisation de la sécheresse, information objective pour prise de décision et argumentation face aux producteurs, etc.).

Certains développements complémentaires sont tout de même recommandés par l'Agence Rurale afin d'améliorer l'outil, notamment en augmentant la précision temporelle de l'indicateur (pas de temps plus précis) ou en donnant la possibilité à l'utilisateur d'extraire certaines cartes ou séries temporelles de sécheresse. L'Agence Rurale ajoute également qu'il serait intéressant d'afficher un indicateur "inverse" à celui de la sécheresse reflétant le niveau "d'humidité" des parcelles.

3.3. Externalités pour valorisations

Le travail mené et produit dans ce projet a permis de confirmer la pertinence et l'utilité d'un tel outil pour le suivi de la sécheresse végétale sur la Nouvelle-Calédonie. Plusieurs valorisations de ce travail sont possibles et certaines sont d'ores et déjà en place dans le cadre de projets en phase de contractualisation notamment.

3.3.1. *Industrialisation et régionalisation*

Un premier objectif à l'issue de ce projet est de permettre **la pérennisation de l'outil** développé dans ce projet SCO. INSIGHT s'est alors associé à l'Observatoire de l'Environnement pour la Nouvelle-Calédonie ([OEIL](#)) afin dans un premier temps de mettre en œuvre la production et la diffusion opérationnelle de l'indicateur de sécheresse végétale sur la Nouvelle-Calédonie, et dans un second temps d'adapter et de développer l'indicateur en question sur **d'autres territoires insulaires de la région** tels que Wallis-et-Futuna, Tahiti ou encore le Vanuatu. Cette collaboration a déjà été initiée, notamment dans le cadre d'un projet Fonds Pacifique en phase de lancement, dans lequel interviennent l'OEIL, INSIGHT, l'IRD (UMR Espace-Dev, pour le soutien scientifique) et le [CSIRO](#) (organisme gouvernemental australien pour la recherche scientifique et intervenant sur la partie régionalisation).

3.3.2. *Valorisations thématiques*

Un second objectif est de valoriser ces données via la mise en place de produits et de services permettant de les croiser à d'autres informations en réponse à des besoins thématiques dont, par exemple : l'agriculture (aide à la décision au producteur ainsi qu'aux administrations en charge du soutien à la filière), la gestion des risques (feux de forêts et aléas de glissement de terrain notamment), gestion de la ressource en eau, etc.

Dans le cas particulier d'une valorisation "agricole" de l'indicateur sécheresse, des démarches ont déjà été menées pour rassembler des acteurs locaux du domaine (Agence Rurale, Chambre d'Agriculture, agriculteurs, Gouvernement) et métropolitains ([MEOSS](#), start-up toulousaine en traitement d'images satellites) afin de mettre en place un projet dédié à la **qualification du besoin en eau et en intrant des cultures**. Il s'agira de coupler l'indicateur de sécheresse à des indices complémentaires (biomasse, rendement, intrants) afin de produire des cartes de besoins utilisables par les producteurs. L'objectif final est de fournir un **outil d'aide à la décision aux agriculteurs** permettant une optimisation de la ressource en eau et de l'utilisation d'intrants.

4. ANNEXES

Annexe 1

REPUBLIQUE
FRANCAISE
NOUVELLE-CALÉDONIE



Retours utilisateurs sur l'outil créé par le projet

EARTH OBSERVATION 4 DROUGHT MONITORING

L'Agence rurale est depuis plusieurs années demandeuse d'un outil de pilotage permettant de mesurer plus finement le niveau de sécheresse des différentes zones agricoles afin d'être plus réactif et plus précis dans la gestion de ses différents dispositifs. L'indicateur utilisé jusqu'à présent prend uniquement en compte le déficit pluviométrique. L'intérêt de ce projet est d'une part de disposer d'un indicateur reflétant plus précisément l'état de stress hydrique du végétal et de pouvoir observer certaines zones géographiques de façon plus précise, selon les zones d'élevage et de production de fourrage.

L'Agence rurale a été associée dès le départ au projet EARTH OBSERVATION 4 DROUGHT MONITORING et identifiée comme l'utilisateur principal de l'outil. Associée tout au long de la démarche de création de l'indicateur et de l'interface, l'Agence rurale a pu orienter les choix et confirmer la pertinence de l'outil. L'interface est précise et permet d'avoir une approche aussi bien globale (indicateur au niveau communal) que beaucoup plus localisée avec une résolution de 10 m qui permet une analyse au niveau parcellaire.

L'historique des données a pu être pris en compte dans l'outil et permettront de mesurer l'évolution et les tendances passées afin de les comparer à l'état actuel. La fréquence d'actualisation des images doit permettre un suivi réactif de l'état de sécheresse.

Cet outil a été complété par un outil prévisionnel de sécheresse dont les applications seront précieuses pour l'Agence rurale comme pour les agriculteurs, afin de planifier budgets et cycles culturaux.

Accessibilité IHM

- Connexion à l'interface simple
- Prise en main facile de l'IHM, ergonomie adaptée

Utilité visée pour l'Agence rurale

- Meilleure qualification de l'état de sécheresse par zone géographique ;
- Possibilité de découpage plus précis des zones ;
- Possibilité d'argumenter auprès des agriculteurs avec indicateurs et cartes à l'appui ;
- Possibilité de « corrélérer » les consommations de fourrages à l'état de sécheresse de la zone ;
- Possibilité de « corrélérer » les rendements des parcelles à l'état de sécheresse de la zone ;
- Le stress hydrique permet de donner des indications sur la qualité du fourrage produit et de préconiser des compléments alimentaires (dont les prix sont subventionnés par l'Agence rurale) ;

- Facilite la prise de décision sur les dispositifs de lutte ou d'indemnisation de la sécheresse ;
- Permet d'anticiper les budgets à prévoir pour dispositifs sécheresse selon prévisions de sécheresse.
- L'outil prévisionnel de sécheresse pourrait également devenir un outil d'aide à la décision pour la planification des productions et notamment le choix des cultures à semer.

Développement complémentaires envisagés

- Possibilité de définir un pas de temps plus précis (dates début/fin) sur le profil temporel ;
- Possibilité d'extraire des cartes (format image) et les données du profil temporel (format Excel) ;
- Possibilité d'ajouter des strates sous communales (Tomo, Nessadiou, Bouraké,...) dont le microclimat est parfois caractéristique ;
- Possibilité d'ajouter les couches en provenance de Géorep afin notamment de consulter le cadastre ;
- Possibilité de créer un indicateur inverse à celui de la sécheresse reflétant le niveau « d'humidité » des parcelles.

Analyse de l'indicateur de sécheresse sur les parcelles de la Ouenghi

- Les zones cultivées et irriguées en période de sécheresse (Déc 2019) présentent un indicateur ne mentionnant aucune sécheresse ou un indicateur proche de la normale
- On peut constater au niveau parcellaire une aggravation du niveau de stress hydrique (Janv 20)

A ce niveau de détail l'information proposée est précise et semble conforme avec les observations de terrain



5. REFERENCES BIBLIOGRAPHIQUES

Cressie, N. (2015). Statistics for spatial data. John Wiley & Sons.

Davison, A.C. and Hinkley, D.V. (1997) Bootstrap Methods and Their Application. Cambridge University Press.

Déqué M., 2003 : « La prévision numérique à l'échelle saisonnière : que sait-on faire et que peut-on espérer ? », La Météorologie, N°41-mai 2003

ECMWF Charts, 2022 : long range forecast charts, consulté le 1 juin 2022, <<https://www.ecmwf.int/en/forecasts/charts>>

Esposito C., Landrum G.A., Schneider N., Stiefl N. & Riniker S.. GHOST: Adjusting the Decision Threshold to Handle Imbalanced Data in Machine Learning. *Journal of Chemical Information and Modeling* **2021** 61 (6), 2623-2640 DOI: 10.1021/acs.jcim.1c00160

GDO, 2019 : Global Drought Observatory Mapviewer, European Commission – Joint Research Center, consulté le 1 juin 2022, <https://edo.jrc.ec.europa.eu/gdo>

Gower (1971) A general coefficient of similarity and some of its properties. *Biometrics* **27** 857-874.

Greene, W. H., & Hensher, D. A. (2010). Modeling ordered choices: A primer. Cambridge University Press.

Jolliffe, I. T., & Stephenson, D. B. (Eds.). (2012). Forecast verification: a practitioner's guide in atmospheric science. John Wiley & Sons.

Kaufman L, Rousseeuw P. 1990. Finding Groups in Data: An Introduction to Cluster Analysis. Wiley: New York, USA.

Lavaysse, C., Stockdale, T., McCormick, N., & Vogt, J. (2020). Evaluation of a New Precipitation-Based Index for Global Seasonal Forecasting of Unusually Wet and Dry Periods, *Weather and Forecasting*, 35(4), 1189-1202.

Leroy, Anne ; Céron, Jean-Pierre ; Un défi de la prévision saisonnière : la descente d'échelle - La Nouvelle-Calédonie, un exemple à suivre, La Météorologie, 2007, N° 58 ; p. 36-44

MF-NC s. d.. Météo-France Nouvelle-Calédonie, *En savoir plus – ENSO*, consulté le 1 juin 2022, <<https://www.meteo.nc/en-savoir-plus/comprendre-la-meteo/phenomenes-particuliers/ENSO>>

Rousseeuw, P.J. (1987) : Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Comput. Appl. Math.* 20, 53-65.

Specq, D., L. Batté (2020). Improving subseasonal precipitation forecasts through a statistical-dynamical approach : application to the southwest tropical Pacific. *Climate Dynamics*.

WMO-LRF s.d., World Meteorological Organization -Long Range Forecast, consulté le 1 juin 2022, <<https://www.wmolc.org/>>