

Apport de la cartographie numérique des sols pour prédire l'hydromorphie et l'extension des zones humides potentielles à l'échelle régionale

B. Lemerrier^(1,2,3), M. Lacoste⁽⁴⁾, M. Loum^(1,2,3), L. Berthier^(1,2,3), A.L. Le Bris^(1,2,3) et C. Walter^(1,2,3)

- 1) AGROCAMPUS OUEST, UMR 1069, Sol Agro et hydrosystème Spatialisation, F-35000 Rennes, France
- 2) INRA, UMR 1069, Sol Agro et hydrosystème Spatialisation, F-35000 Rennes, France
- 3) Université Européenne de Bretagne, France
- 4) INRA, US 1106, InfoSol, F-45000 Orléans, France

* : Auteur correspondant : blandine.lemerrier@agrocampus-ouest.fr

RÉSUMÉ

Les outils de la Cartographie Numérique des Sols (CNS) permettent de prédire des propriétés pédologiques sur de vastes étendues à partir d'informations limitées sur les sols et de variables exhaustives traduisant les facteurs de formation des sols. L'objectif de cette étude était d'établir, à l'échelle de la région Bretagne, des cartes de prédiction de l'hydromorphie et des zones humides par apprentissage automatique, et de les valider avec des données indépendantes. La première étape a consisté à établir un modèle basé sur les relations sol-environnement physique pour prédire l'hydromorphie selon 4 classes. La méthode appliquée est une classification supervisée par arbre stochastique optimisée (algorithme AdaBoost-SAMME implémenté dans la boîte à outil 'adabag' du logiciel R). Le modèle a été calibré à partir de 1652 points où l'hydromorphie était connue et de 14 variables environnementales exhaustives sur la région, puis extrapolé à l'ensemble de la zone d'étude (27 360 km²). Dans une seconde étape, l'extension des zones humides potentielles a été dérivée de celle de l'hydromorphie par reclassement des valeurs prédites selon les critères pédologiques de l'arrêté du 1er octobre 2009 relatif à l'identification et à la délimitation des zones humides. Les prédictions ont été validées à partir d'informations pédologiques (données ponctuelles, cartes pédologiques précises numérisées et référentiel régional pédologique de Bretagne à 1/250 000) et d'inventaires de zones humides considérés comme fiables disponibles dans le Finistère.

Les variables qui contribuent le plus au modèle de prédiction de l'hydromorphie sont l'occupation du sol, le matériau parental et la pluviométrie, suivies de la courbure verticale et de la dénivellée au cours d'eau. La précision interne du modèle est satisfaisante : le taux de prédictions correctes calculé à partir des données de calibration est de 77 %, l'indice de Kappa de 70 % et les 4 classes d'hydromorphie sont correctement prédites. La distribution spatiale des sols hydromorphes et des zones humides est très cohérente avec les connaissances préexistantes bien que l'hydromorphie et les zones humides semblent surestimées. Cette surestimation est confirmée par la validation des résultats du modèle par comparaison aux observations ponctuelles indépendantes (n = 3 348) et aux cartes des

sols précises : les taux de prédictions correctes sont de 55 et 36 % respectivement, et les sols non hydromorphes sont sous-estimés. Cette étude a permis d'établir une procédure complète de prédiction de l'hydromorphie et de l'extension spatiale des zones humides potentielles à l'échelle d'une région, à une résolution fine et assortie une incertitude. La procédure est répétable et adaptable, ce qui permet d'envisager d'améliorer les performances du modèle, voire de la transposer à d'autres secteurs qui disposeraient de suffisamment de données pédologiques et de variables prédictives exhaustives.

Mots clés

cartographie numérique des sols, classification supervisée par arbre stochastique optimisée, hydromorphie, zones humides, Bretagne.

SUMMARY

PREDICTION OF SOIL REDOXIMORPHIC FEATURES AND WETLANDS EXTENT AT REGIONAL SCALE USING A DIGITAL SOIL MAPPING APPROACH

Digital Soil Mapping (DSM) approaches enable the prediction of soil properties over wide extents, using sparse pedological data and exhaustive information about soil forming factors. The aim of this study was to establish predictive maps of soil redoximorphic features and wetland extent at the scale of Bretagne (a northwestern France region) using a machine learning procedure, and validate the predictions with independent data. Firstly, a soil-landscape model was defined to predict soil redoximorphic features clustered into four classes. The applied method was a tree-based supervised classification using a stochastic gradient boosting procedure (AdaBoost-SAMME algorithm implemented in the R package 'adabag'). The model was fitted on 1652 points of known soil redoximorphic features and 14 environmental variables of regional extent, and extrapolated to the whole study area (27,360 km²). Secondly, potential wetlands extent was obtained by reclassification of redoximorphic features prediction, according to pedological criteria described in the French regulation on wetland identification and delineation. Predictions were validated with pedological information (point-specific soil observations, digital precise soil maps, 1 :250,000 soil map) and accurate terrain wetlands inventories available in Finistère (western part of the study area). The most influential variables in redoximorphic features prediction were soil use, soil parent material and rainfall, and then profile (vertical) curvature and elevation above the nearest stream. The internal accuracy of the model was correct : the overall accuracy was 77 %, the Kappa index was 70 % and all of the 4 classes of redoximorphic features were well predicted. The spatial distribution of redoximorphic soils and wetlands were consistent with previous knowledge, even if redoximorphic features and wetlands seemed over-estimated. This over-estimation was confirmed by the comparison of the predictions with independent point observations (n =3 348) and precise soil maps : overall accuracies were 55 and 36 % respectively, and naturally well-drained soils were under-estimated.

With this study, a complete procedure was established to predict soil redoximorphic features and potential wetlands over a region, with fine resolution and with information about uncertainty. This procedure is repeatable, and adjustable, allows future model performance improvement, and transposition to other territories having enough pedological information and exhaustive predictive variables.

Key-words

Digital soil mapping, stochastic gradient boosting classification tree, soil redoxymorphic features, wetlands, Bretagne (France)

RESUMEN

APORTE DE LA CARTOGRAFÍA NUMÉRICA DE LOS SUELOS PARA PREDECIR LA HIDROMORFÍA Y LA EXTENSIÓN DE LAS ZONAS HÚMEDAS POTENCIALES A ESCALA REGIONAL

Las herramientas de la Cartografía Numérica de Suelos (CNS) permiten predecir propiedades pedológicas sobre amplias superficies a partir de informaciones limitadas sobre los suelos y de variables exhaustivas que traducen los factores de formación de los suelos. El objetivo de este estudio estaba establecer, a la escala de la región Breñaña, mapas de predicción de la hidromorfía y de las zonas húmedas por aprendizaje automático y validarlos con datos independientes. La primera etapa consistió establecer un modelo basado sobre las relaciones suelo - medio ambiente físico para predecir la hidromorfía según 4 clases. El método aplicado está una clasificación supervisada por árbol estocástico optimizada (algoritmo AdaBoost-SAMME implementado en la caja de herramienta "adabag" del programa R). El modelo se calibra a partir de 1652 puntos donde la hidromorfía estaba conocida y de 14 variables ambientales exhaustivas sobre la región, después extrapolado a toda la zona de estudio (27 360 km²). En una segunda etapa, la extensión de las zonas húmedas potenciales se deriva de la de la hidromorfía por una reclasificación de valores predichos según los criterios pedológicos que se encuentran en el orden del 1 de octubre de 2009 relativo a la identificación y delimitación de las zonas húmedas. Las predicciones se validaron a partir de informaciones pedológicas (datos puntuales, mapas pedológicos precisos numerizados y referencial pedológico de Breñaña a 1/250 000) y de inventarios de las zonas húmedas consideradas como fiables disponibles en el Finisterre.

Las variables que contribuyan lo más al modelo de predicción de la hidromorfía están la ocupación del suelo, el material parental y la pluviometría seguidos por la curvatura vertical y del desnivel al curso de agua. La precisión interna del modelo está satisfactoria : la

taza de predicciones correctas calculada a partir de los datos de calibración está de 77 %, el índice de Kappa de 70 % y las 4 clases de hidromorfía están correctamente predichas. La distribución espacial de los suelos hidromórficos y de las zonas húmedas está muy coherente con los conocimientos preexistentes aunque la hidromorfía y las zonas húmedas parecen sobrevalorizadas. Esta sobrestimación se confirmó por la validación de los resultados del modelo por comparación a las observaciones puntuales independientes ($n = 3348$) y a los mapas de suelos precisos : la tasa de predicciones correctas están de 55 y 36 % respectivamente, y los suelos no hidromórficos están infravalorizados.

Este estudio permitió establecer un procedimiento completo de predicción de la hidromorfía y de la extensión espacial de las zonas húmedas potenciales a la escala de una región, a una resolución fina y aderezada con una incertidumbre. El procedimiento está repetible y adaptable, lo que permite planear de mejorar las prestaciones del modelo, incluso transponerlo a otros sectores que dispondrían suficientemente de datos pedológicos y de variables predictivas exhaustivas.

Palabras clave

Cartografía numérica de suelos, clasificación supervisada por árbol estocástico optimizada, hidromorfía, zonas húmedas, Bretaña.

Les démarches classiques de cartographie des sols sont basées sur des observations directes des sols regroupées en unités cartographiques et/ou unités typologiques de sol par expertise (Legros, 1996). Les méthodes classiques de cartographie se heurtent au coût de mise en œuvre quand il s'agit de couvrir de vastes territoires comme les départements ou les régions, elles sont peu reproductibles, et l'exploitation des cartes reste liée aux choix du pédologue, souvent guidés par les objectifs initiaux des levés (drainage, épandage, irrigation, etc.). Depuis quelques années, la Cartographie Numérique des Sols (CNS) complète de façon opérationnelle la cartographie classique des sols. Sans remplacer les observations de terrain, la CNS permet de produire, quelle que soit l'échelle de travail, une information homogène assortie d'une estimation des incertitudes, de façon documentée et répétable. La CNS peut être définie comme « la création et le renseignement de systèmes d'information spatiale sur les sols par des modèles numériques traduisant les variations spatiales et temporelles des sols et de leurs propriétés à partir d'observation des sols, de connaissances des sols préalables et de variables environnementales » (Lagacherie et McBratney, 2007). Les techniques de CNS sont nombreuses et basées sur l'interpolation entre des observations ponctuelles du sol (statistiques, géostatistiques) ou les relations sol-paysage (Dobos *et al.*, 2006).

Les relations entre le sol et les facteurs de formation du sol sont connues de longue date et ont été très largement prises en compte pour cartographier les sols. Tout d'abord formalisées par Jenny (1941), ces relations ont été généralisées par l'approche scorpan (McBratney *et al.*, 2003) qui fournit un cadre conceptuel aux études de CNS. Scorpan fait référence aux 7 facteurs de formation des sols : s : sol (mesures précédentes, autres paramètres pédologiques que celui qui est à

prédire) ; c : climat ; o : organismes vivants (faune, végétation, activités humaines), r : relief ; p : matériau parental ; a : facteur temps ; n : position dans l'espace ; qui permettent de prédire une classe de sol ou un paramètre pédologique.

L'apprentissage automatique est une technique de CNS qui a pour objectif d'extraire et d'exploiter automatiquement l'information présente dans un jeu de données pour résoudre l'équation du modèle sol-paysage. Dans le cas d'une classification supervisée, il s'agit de définir un modèle d'organisation spatiale des sols à partir d'un jeu de données d'apprentissage, qui correspond à des situations connues de sol et de paysage, et ensuite d'appliquer ce modèle pour prédire la variable d'intérêt dans les zones où elle est inconnue.

La mise en œuvre d'une classification pédologique supervisée suppose d'être en mesure de construire un modèle de prédiction, c'est-à-dire de disposer d'informations pédologiques représentatives des situations observables sur la zone d'étude (variable à prédire) et d'informations sur les facteurs de différenciation des sols (variables prédictives) exhaustives sur la zone d'étude. La prédiction repose sur des algorithmes statistiques qui formalisent les lois de distribution des sols utilisées implicitement par le pédologue cartographe à partir des données disponibles (Laroche *et al.*, 2011), puis extrapolent ces lois à l'ensemble de la zone d'étude. La résolution de la carte de prédiction dépend de la résolution des variables d'entrée du modèle. Les travaux de CNS par classification supervisée menés en France à l'échelle départementale ou régionale ont permis de prédire des unités cartographiques de sol (Grinand *et al.*, 2008, Laroche *et al.*, 2011) ou des propriétés des sols (Lacoste *et al.*, 2011 ; Lemerrier *et al.*, 2012).

Dans cette étude, l'hydromorphie des sols a été prédite à l'échelle régionale à partir d'observations pédologiques ponctuelles et de variables environnementales exhaustives. La

distribution spatiale du caractère hydromorphe des sols n'est pas aisée à prédire, du fait de l'interaction entre les déterminants connus avec précision à l'échelle régionale (topographie, hydrologie) et les déterminants connus avec une précision moindre à l'échelle d'étude (pédologie, géologie, tectonique récente, aménagements). L'hydromorphie est donc un paramètre pertinent à étudier du point de vue strictement méthodologique, mais aussi thématique. Manifestation morphologique résultant de l'engorgement prolongé en eau dans un horizon pédologique ou un solum, l'hydromorphie s'exprime sous forme de taches, de concentrations, de colorations spécifiques ou de décolorations, résultant de la dynamique, en milieu réducteur ou alternativement réducteur puis réoxydé, de deux éléments colorés : le fer et le manganèse (Baize et Jabiol, 2011). Elle peut aussi se manifester par une augmentation de la teneur en matière organique et une différenciation de sa composition, dues au ralentissement de l'activité biologique en conditions d'anoxie. Les causes de l'hydromorphie impactent directement les potentialités agronomiques du sol, et régulent les flux d'eau dans les paysages, la genèse des processus de ruissellement et d'érosion et la qualité des eaux. La régulation de la qualité des eaux se fait par la rétention des matières en suspension et des éléments associés, et les transformations et dégradations spécifiques dues aux conditions d'anaérobiose (dénitrification, dégradation des micropolluants organiques) (Barneaud et Fustec, 2007). L'hydromorphie est donc un paramètre pédologique de première importance, impliqué dans les fonctions de production, environnementales et sociétales des sols. Ces travaux s'inscrivent dans un contexte de prise en compte croissante des zones humides dans la gestion des territoires. Les zones humides sont cruciales pour la biodiversité et rendent des services écosystémiques (Millennium Ecosystem Assessment, 2005). L'inventaire des zones humides est un outil d'aide à la décision pour la planification, l'aménagement, la gestion de la ressource en eau et la gestion de la biodiversité. L'inventaire des zones humides est une étape préalable indispensable à la préservation et à la gestion durable de ces milieux sensibles. Depuis la parution de l'arrêté du 24 juin 2008 modifié par l'arrêté du 1^{er} octobre 2009 et de la circulaire du 18 janvier 2010 relatifs à l'application des articles L214-7-1 et R211-108 du code de l'environnement, la description de l'hydromorphie du sol et/ou de la végétation est un critère de définition et de délimitation des zones humides. D'après l'arrêté du 1^{er} octobre 2009, les sols caractéristiques de zones humides sont définis par la profondeur d'apparition de l'hydromorphie et son intensité. Ils correspondent, à tous les HISTOSOLS, tous les REDUCTISOLS et aux autres sols présentant des traits rédoxyques débutant à moins de 25 cm de profondeur et se prolongeant ou s'intensifiant en profondeur, ou des traits rédoxyques débutant à moins de 50 cm de profondeur, se prolongeant ou s'intensifiant en profondeur, et de traits rédoxyques apparaissant entre 80 et 120 cm de profondeur. Dans certains contextes pédologiques

particuliers (FLUVIOSOLS, PODZOSOLS humiques et humoduriques), l'excès d'eau prolongé ne se traduit pas par des traits morphologiques habituels, et une expertise des conditions hydrogéomorphologiques peut être nécessaire pour identifier les zones humides.

Le besoin d'inventaire des zones humides s'exprime à différentes échelles et correspond à des besoins spécifiques. Au niveau régional ou des Schémas Directeurs d'Aménagement et de Gestion des Eaux (SDAGE), une estimation homogène de la superficie des zones humides et de leur répartition spatiale est recherchée pour la mise en place des Schémas Régionaux de Cohérence Ecologique (SRCE), comparer et optimiser les inventaires de terrain, sensibiliser et informer les gestionnaires de l'espace, aménageurs et citoyens. A l'échelle des Schémas d'Aménagement et de gestion des Eaux (SAGE), et des entités administratives locales, l'inventaire des zones humides est un outil de planification territoriale et d'urbanisme réalisé sur la base de campagnes de terrain et d'échanges avec les acteurs locaux. Les travaux à une échelle fine s'appuient également sur des données disponibles sur des territoires plus larges : photographies aériennes, topographie, hydrologie, etc., et parfois sur des enveloppes de zones humides potentielles délimitant les zones où la probabilité de rencontrer des zones humides est forte (Mérot *et al.*, 1995 ; Mérot *et al.*, 2003). La mise en œuvre des enveloppes de zones humides potentielles est relativement rapide et peu coûteuse. Elle repose sur le calcul d'indices topographiques ou topo-climatiques. Cependant, cette approche est insuffisante pour prédire l'hydromorphie des sols, qui dépend de facteurs autres que la topographie et le climat, tels que le substrat rocheux ou la tectonique récente (Chaplot et Walter, 2007), qui impactent le relief et la qualité des altérites (Bonnet *et al.*, 1998), ou le sol. En effet, l'eau s'accumule dans les versants par ruissellement de surface et de subsurface contrôlé par la gravité (topographie) et/ou la faible infiltration des sols, saprolites ou substrats (Curmi *et al.*, 1998). La présence de discontinuités locales des matériaux géologiques (fracturation, variation du faciès des altérites...) et des sols ainsi que l'existence d'aménagements hydrauliques modifient le fonctionnement hydrologique effectif des sols et induisent des différences entre les zones humides prédites par les indices topographiques et les zones humides observées sur le terrain. Les indices topographiques rendent compte uniquement des zones humides liées au modelé du terrain et non au matériau géologique ou au sol, comme peuvent l'être certaines zones humides de versant ou de plateau. Dans certaines régions, ces zones humides peu connectées au réseau hydrographique représentent des superficies importantes.

L'objectif de ce travail a été de mettre en œuvre des outils de cartographie numérique pour prédire l'hydromorphie des sols à l'échelle de la région Bretagne. Pour cela, une technique d'apprentissage automatique et plus précisément une classification supervisée basée sur des arbres de classification a été

appliquée. La probabilité d'occurrence associée aux classes prédites et la comparaison avec des cartes pédologiques précises (1/25 000) permettent de valider l'approche. Dans un second temps, l'hydromorphie modélisée a été valorisée pour prédire l'extension spatiale des zones humides, elle-même validée avec des données indépendantes.

MATERIELS ET METHODES

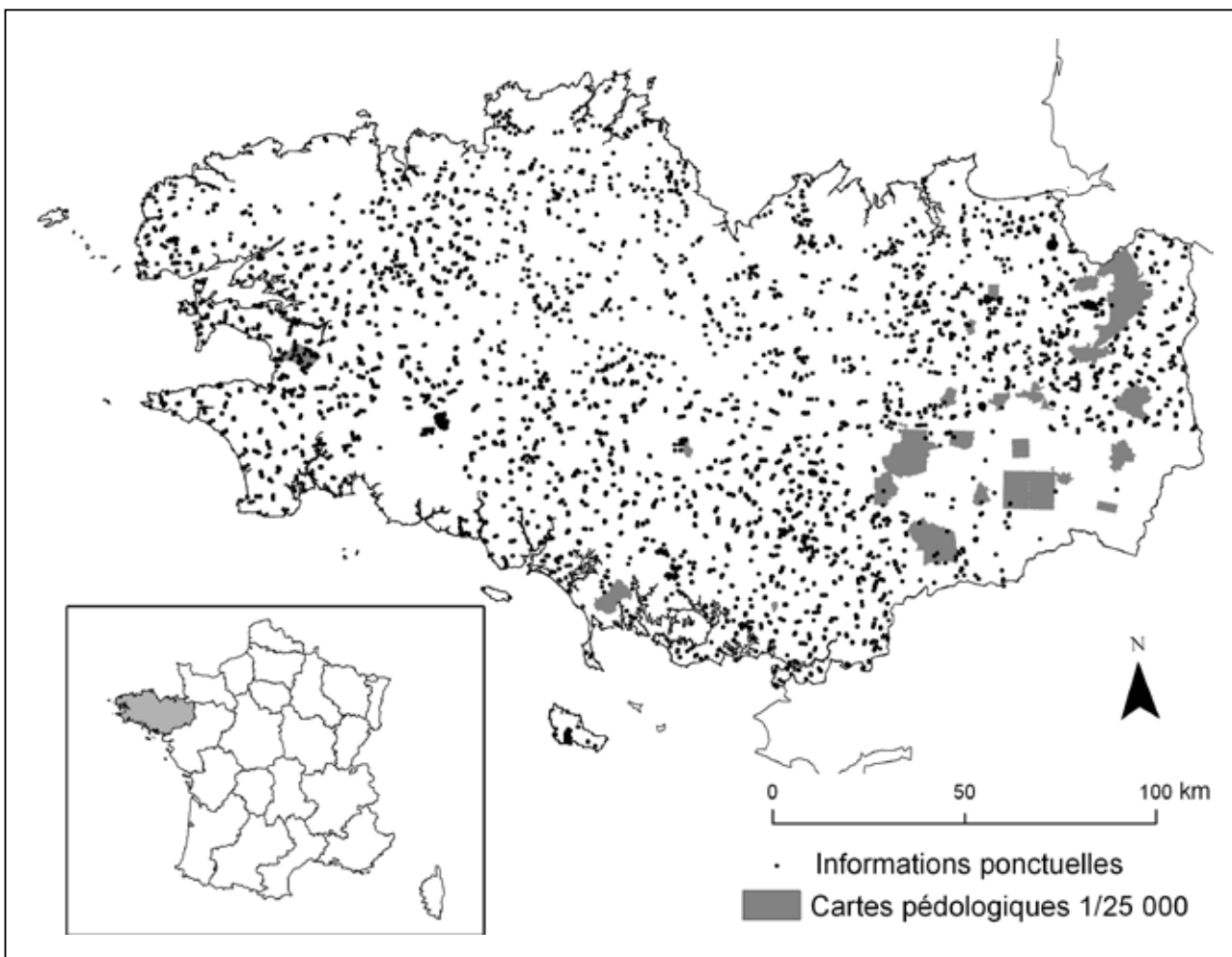
Zone d'étude

La zone d'étude (*figure 1*) est la région administrative de Bretagne, péninsule du nord-ouest de la France dont la superficie est d'environ 27 360 km². Elle appartient au Massif armo-

ricain constitué essentiellement d'un socle antécambrien et d'une couverture paléozoïque (Lardeux, 2002). Les matériaux sont pour la plupart métamorphiques et granitiques (schiste, granite, micaschiste, gneiss). Au Quaternaire, des dépôts de limons éoliens ont recouvert le socle rocheux dans le nord de la région. Il en résulte une couverture limoneuse en place ou remaniée, d'épaisseur décroissante du nord vers le sud, présente jusque dans le bassin de Rennes à l'est. Les sols ont en majorité des BRUNISOLS, des BRINSOLS-REDOXISOLS, des RANKOSOLS, des NEOLUVISOLS et des FLUVIOSOLS-REDOXISOLS (Lemercier *et al.*, 2010). Le climat est de type océanique tempéré, avec des températures moyennes annuelles comprises entre 10,5 et 13 °C et des précipitations moyennes annuelles de 970 mm.an⁻¹, variant de 600 mm.an⁻¹ à 1500 mm.an⁻¹ le long d'un gradient principal est-ouest. Le

Figure 1 - Zone d'étude et répartition spatiale des informations pédologiques disponibles, ponctuelles et surfaciques à l'échelle 1/25 000.

Figure 1 - Punctual observations and 1:25.000 resolution soil maps available in digital soil databases for the study area.



relief, généralement modéré, est fortement corrélé aux formations géologiques. L'altitude varie de 0 à 390m et les reliefs les plus accentués se trouvent dans la partie centre-ouest de la Bretagne. Le réseau hydrographique est dense (890m de cours d'eau par km²) et caractérisé par l'absence de grand fleuve hormis la Vilaine. Les petits cours d'eau d'ordres 1 et 2 selon la classification de Stralher (1952) représentent 70% du linéaire. Le réseau hydrographique dense, le relief mollement vallonné et le substrat géologique peu perméable ont pour conséquence la présence de nombreuses zones humides de fond de vallées d'une superficie allant de quelques ares à quelques hectares, plus ou moins reliées entre elles et au réseau hydrographique (Durand *et al.*, 2000). Ces zones sont en interaction avec le versant (du fait du flux hydrique provenant de l'ensemble du versant) et le réseau hydrographique, surtout pour les ordres faibles (Mourier *et al.*, 2008). L'indice topographique de Beven et Kirky (1979) modifié par Mérot *et al.* (1995) estime les zones humides potentielles à 22% de la superficie de la Bretagne et représente bien les zones humides de fond de vallées, dont le fonctionnement hydrologique a été décrit (Durand *et al.*, 2000).

Prédiction de l'hydromorphie

Données d'apprentissage : informations ponctuelles

La classe d'hydromorphie est disponible pour un total de 5000 sondages à la tarière ou profils pédologiques réalisés en majorité entre 2006 et 2009, parmi lesquels 1652 ont été utilisés pour calibrer le modèle et 3 348 pour le valider. En chacun de ces points, l'hydromorphie est décrite en 10 classes numérotées de 0 (absence d'indice d'hydromorphie) à 9 (hydromorphie très intense affectant l'ensemble du profil) selon la méthodologie de Rivière *et al.* (1992). La nature des traits d'hydromorphie, leur profondeur d'apparition, leur intensité et leur succession verticale permettent de définir la classe d'hydromorphie. Pour les besoins de l'étude, ces 10 classes ont été regroupées en 4 groupes :

1. Hydromorphie absente ou débutant à plus de 80cm, correspondant aux classes 0, 1 et 2 définies par Rivière *et al.* (1992) et I et II du GEPPA (1981) ;
2. Hydromorphie débutant entre 40 et 80cm de profondeur et se poursuivant ou s'intensifiant en profondeur, correspondant aux classes 3 et 4 définies par Rivière *et al.* (1992) et à la classe III du GEPPA (1981) ;
3. Hydromorphie débutant à moins de 40 cm de profondeur et se poursuivant ou s'intensifiant en profondeur, correspondant aux classes 5 et 6 définies par Rivière *et al.* (1992) et IVb, IVc, IVd, Vb et Vc du GEPPA (1981) ;
4. Hydromorphie manifeste sur toute l'épaisseur du sol avec présence de traits réductiques, correspondant aux classes 7, 8 et 9 définies par Rivière *et al.* (1992) et Vd, VI et H du GEPPA (1981).

Variables prédictives

Les variables prédictives ont été sélectionnées ou dérivées à partir des informations disponibles de façon exhaustive et homogène sur la zone d'étude, et décrivant un des facteurs de formation des sols. Ainsi, 14 variables environnementales ont été utilisées pour bâtir le modèle puis l'extrapoler à l'ensemble de la zone d'étude (*tableau 1*). Elles sont de type :

- topographique : altitude, pente locale, courbures longitudinale et transversale, longueur des versants, dénivelée au cours d'eau, unités topographiques ;
- pédologique : matériau parental prédit par une méthode similaire à celle présentée ici ;
- géophysique (spectrométrie gamma) : rapport potassium/thorium, écart à la moyenne de l'unité lithologique correspondantes des émissions de potassium ;
- hydrologique : distance hydrologique au cours d'eau, indice topographique modifié ;
- paysagique : unités paysagères définies par télédétection ;
- climatique : moyenne des précipitations annuelles.

Préalablement à la modélisation, les variables prédictives ont été converties en données matricielles d'une résolution de 50m et leur superposition exacte a été vérifiée. La transformation des données matricielles a consisté en un ré-échantillonnage par affectation par le voisin le plus proche, réalisée avec le logiciel ArcGIS 10.0®.

Algorithme de prédiction

Les techniques d'apprentissage automatique utilisent un algorithme pour déterminer ou expliciter les relations entre une variable à prédire et des variables prédictives. La procédure de prédiction de l'hydromorphie comporte deux étapes : l'apprentissage (élaboration des règles de classification) à partir de situations où la variable à prédire et les variables prédictives sont connues puis l'extrapolation à l'ensemble de la zone d'étude où seules les variables prédictives sont connues (*figure 2*). La méthode d'apprentissage automatique choisie est une classification par arbres stochastique optimisée. Elle consiste en la partition récursive binaire du jeu de données d'apprentissage en fonction des variables prédictives de façon à minimiser les erreurs de prédiction et à produire des groupes finaux homogènes. Une procédure d'optimisation stochastique (Friedman, 2001 ; Friedman et Meulman, 2003) améliore la qualité du modèle produit en affectant un poids plus important aux individus mal classés. Des arbres élémentaires courts sont construits à partir d'une fraction du jeu de données et le modèle final combine l'ensemble des arbres simples. La classification par arbres stochastique optimisée présente de nombreux avantages : elle est non paramétrique, c'est-à-dire qu'elle ne nécessite pas d'hypothèse sur la distribution des variables ni de transformation de celles-ci, elle n'est pas sensible aux valeurs manquantes ni aux variables peu corrélées à la variable à prédire, elle prend en compte

Tableau 1- Variables environnementales utilisées pour prédire l'hydromorphie.**Table 1 - Environmental predictors used to predict soil redoximorphic features class.**

Nom	Description (unités)	Moyenne (étendue)
Topographiques (dérivées du modèle numérique de terrain à 50m)		
alti	Altitude (m)	97 (0-382)
pent	Pente locale (%)	5,63 (0-146)
courv	Courbure verticale (m.100 m ⁻¹)	0 (-9-9)
courh	Courbure horizontale (m.100 m ⁻¹)	0 (-8-6)
lgvers	Longueur du versant (m)	571 (0-2430)
deniv	Altitude relative par rapport au cours d'eau le plus proche (m)	16 (0-302)
topo	Unités topographiques définies à partir des dérivées du MNT par analyse des correspondances multiples suivie d'une classification ascendante hiérarchique (Lemerrier, non publié)	9 classes nominales
Pédologique		
mpp	Matériau parental prédit par arbre de classification (Lacoste <i>et al.</i> , 2011)	20 classes nominales
Géophysique aéroporté, spectrométrie gamma (Bonijoly <i>et al.</i> , 1999)		
ktratio	Rapport K/Th (Wilford <i>et al.</i> , 1997)	0,19 (-1,1-3,06)
devk	Ecart des émissions de K à la moyenne de l'unité lithologique correspondante (Chantraine <i>et al.</i> , 2002)	0 (-1,94-2,11)
Hydrologiques (dérivées du modèle numérique de terrain à 50m)		
dist	Distance hydrologique au cours d'eau le plus proche (m)	353 (0-2483)
mcti	Indice topographique modifié (Mérot <i>et al.</i> , 1995, après Beven et Kirkby, 1979)	4 (0-27)
Paysagique (résolution 250m)		
occup	Unités paysagères définies par télédétection à partir d'images du capteur Modis (Le Dû-Blayo <i>et al.</i> , 2008)	19 classes nominales
Climatique (moyenne de 2000 à 2009, AURELHY, météo-France)		
pluv	Pluviométrie totale annuelle (mm)	968 (579-1589)

l'interaction entre les variables sans qu'elle soit connue *a priori*, et elle peut gérer à la fois des variables quantitatives et qualitatives (Friedman et Meulman, 2003). L'algorithme utilisé est SAMME – Stagewise Additive Modeling using a Multi-class Exponential loss function - (Zhu *et al.*, 2009), qui généralise l'algorithme de classification AdaBoost (Freund et Schapire, 1997). AdaBoost-SAMME est implémenté dans la boîte à outils 'adabag' du logiciel R (Alfaro-Cortes *et al.*, 2012). Cet algorithme a été choisi car il était le seul outil libre en mesure d'effectuer une classification multi-classe optimisée au moment des travaux.

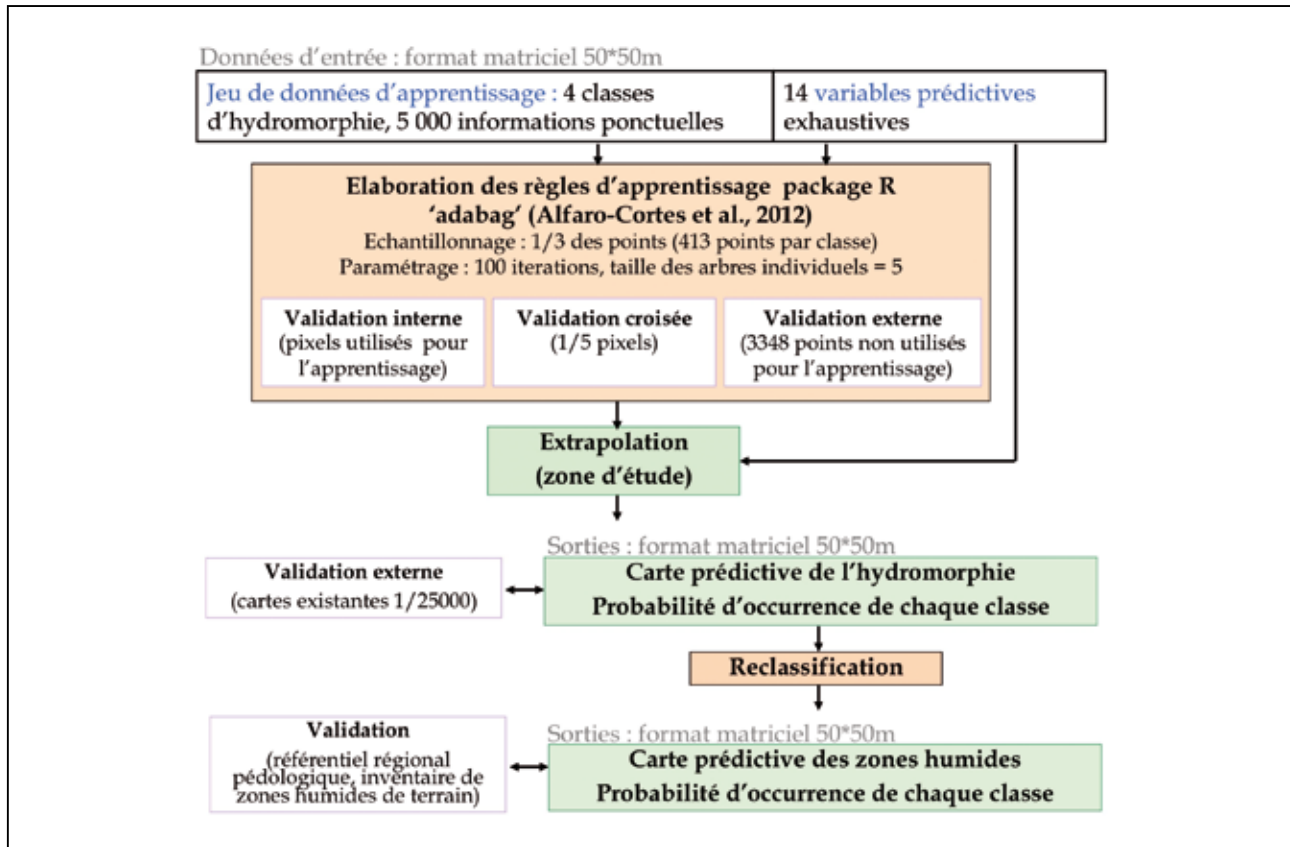
Interprétation des résultats du modèle et validation

Pour chaque pixel, le modèle fournit la probabilité d'appartenance à chaque classe. La classe prédite est celle dont la probabilité d'occurrence est la plus élevée.

La précision du modèle est estimée par quatre procédures de validation distinctes :

- validation interne : estimée sur l'ensemble des pixels utilisés pour construire le modèle, elle permet d'estimer l'erreur intrinsèque du modèle ;
- validation croisée : l'échantillon est divisé en 5, puis chaque sous-échantillon est utilisé successivement comme jeu de validation, les 4 autres échantillons constituent le jeu d'apprentissage ; la moyenne des 5 erreurs moyennes fournit l'erreur de prédiction ;
- validation externe ponctuelle : réalisée sur les 3 348 données ponctuelles disponibles et non utilisées pour construire le modèle ; ces informations ont été assimilées à un pixel de 50 m ;
- validation externe par cartes : effectuée par comparaison avec des cartes précises (1/25 000) préexistantes, couvrant 1 031 km², soit 3 % de la zone d'étude, localisées essentiellement à l'est de la zone d'étude (*figure 1*). Sur ces cartes, l'hydromorphie a été représentée selon la méthode Rivière *et al.* (1992). L'hypothèse que les unités cartographiques sont pures a été faite.

Figure 2 - Méthodologie générale de prédiction de l'hydromorphie et de l'extension spatiale des zones humides.
Figure 2 - Methodological overview for soil redoximorphic features and wetlands spatial extent prediction.



L'estimation de la précision des résultats repose sur plusieurs indicateurs de performance calculés à partir des matrices de confusion. La performance globale est estimée par l'indice de précision globale qui est le rapport entre le nombre de pixels correctement prédits et le nombre total de pixels, et l'indice de kappa qui est un indicateur de concordance corrigé par la probabilité que la prédiction correcte soit due à la chance (Cohen, 1960). Des indices par classe de réponse sont également calculés : la précision du producteur, qui correspond à la probabilité pour un pixel d'une classe donnée sur le terrain d'être bien classé sur la carte (complément à 100 % de l'erreur de commission), et la précision de l'utilisateur, qui correspond à la probabilité pour un pixel classé en dans une classe donnée sur la carte d'appartenir à la même classe sur le terrain (Story et Congalton, 1986). D'après Schmidt *et al.* (2008), pour une classe prédite donnée, la précision du producteur exprime la sur-estimation de la classe et la précision de l'utilisateur sa sous-estimation.

Lacoste *et al.* (2011) ont appliqué une procédure similaire sur la même zone d'étude pour prédire le matériau parental et

ont mis en évidence une corrélation très forte entre la probabilité d'occurrence (toutes classes confondues) et la précision globale de la validation externe, ce qui montre que la probabilité d'occurrence de la classe prédite est un bon indicateur de la précision du résultat.

Prédiction des zones humides

Evaluation numérique des sols

L'hydromorphie étant un critère réglementaire pour l'identification et la délimitation des zones humides, la prédiction de l'hydromorphie à l'échelle régionale peut être utilisée pour évaluer l'extension des zones humides et aider aux inventaires de terrain. Dans le cas de la Bretagne, les sols des cas particuliers prévus dans la réglementation sont rares : les FLUVIOSOLS TYPIQUES sont inféodés à des vallées alluviales d'ordre important très peu fréquentes, de même que les PODZOSOLS humiques ou humoduriques. La règle générale est alors appliquée pour identifier les zones humides. Il s'agit d'interpréter les valeurs d'un paramètre pédologique pour

estimer un processus pouvant affecter le fonctionnement de l'écosystème, ce qui fait appel au concept d'Évaluation Numérique des Sols (ENS). L'ENS a été définie par Carré *et al.* (2007) comme étant la prédiction quantitative de paramètres pédologiques difficiles à mesurer, nécessaires à l'évaluation des risques pesant sur le sol et des fonctions du sol, utilisant des sorties de CNS.

Selon la réglementation en cours (arrêté du 1^{er} octobre 2009), les classes d'hydromorphie prédites 1 et 2 ne concernent pas les sols caractéristiques de zones humides, et l'intégralité des sols de la classe 4 caractérisent les zones humides. La classe 3 regroupe à la fois des sols caractéristiques de zones humides (classes IVd, Vb et Vc) et des sols non caractéristiques (classes IVb, IVc). Une classification en deux étapes a été réalisée :

- Identification des Zones Non Humides ('ZNH') : elles correspondent aux pixels pour lesquels la somme des probabilités d'occurrence des classes 1 et 2 est supérieure à la somme des probabilités d'occurrence des classes 3 et 4.
- Au sein des zones non classées comme non humides, si la probabilité de la classe 3 est supérieure à celle de la classe 4, le pixel est classé en Zone Intermédiaire ('ZI') ; à l'inverse, si la probabilité de la classe 4 est supérieure à celle de la classe 3, le pixel est classé en Zone Humide ('ZH').

Ainsi, pour chaque individu I de coordonnées i et j ,

Si $p(1)_{i,j} + p(2)_{i,j} > p(3)_{i,j} + p(4)_{i,j}$ alors $I_{i,j}$ est 'ZNH'

Si $p(3)_{i,j} + p(4)_{i,j} > p(1)_{i,j} + p(2)_{i,j}$ et $p(3)_{i,j} > p(4)_{i,j}$ alors $I_{i,j}$ est 'ZI'

Si $p(3)_{i,j} + p(4)_{i,j} > p(1)_{i,j} + p(2)_{i,j}$ et $p(4)_{i,j} > p(3)_{i,j}$ alors $I_{i,j}$ est 'ZH'

Avec $p(X) =$ probabilité d'occurrence associée à la classe X .

Une prédiction des zones humides est ainsi produite, assortie d'une incertitude représentée par :

- la somme des probabilités d'occurrence des classes 1 et 2, pour la classe 'ZNH' ;

$$p(\text{ZNH})_{i,j} = p(1)_{i,j} + p(2)_{i,j}$$

- la probabilité d'occurrence de la classe 3 rapportée à la somme des probabilités des classes 3 et 4 pour la classe 'ZI' ;

$$\forall I_{i,j} \notin \text{ZNH}, p(\text{ZI})_{i,j} = \frac{p(3)_{i,j}}{p(3)_{i,j} + p(4)_{i,j}}$$

- la probabilité d'occurrence de la classe 4 rapportée à la somme des probabilités des classes 3 et 4 pour la classe 'ZH'.

$$\forall I_{i,j} \notin \text{ZNH}, p(\text{ZH})_{i,j} = \frac{p(4)_{i,j}}{p(3)_{i,j} + p(4)_{i,j}}$$

Une probabilité proche de 1 associée à la classe 'ZI' ou 'ZH' ne sera pas interprétée comme une quasi-certitude d'appartenance à cette classe, mais comme une probabilité élevée

d'y appartenir, sachant que l'on n'est probablement pas en zone non humide ('ZNH').

Validation

La validation de la prédiction des zones humides a été effectuée en comparant les résultats de la prédiction avec des bases de données indépendantes : le Référentiel Régional Pédologique (RRP) de Bretagne (Lemerrier *et al.*, 2010) labellisé en 2012 par le Ministère en charge de l'agriculture, et des délimitations de zones humides réalisées à l'échelle parcellaire dans le Finistère et regroupés au sein d'une base de données unique : l'Inventaire Permanent des Zones Humides du Finistère, dit IPZH29 (Forum des marais atlantiques et Conseil général du Finistère).

Pour la comparaison avec le RRP, les classes 'ZI' et 'ZH' prédites ont été agrégées puis leur proportion a été calculée par unité cartographique de sol du RRP. Parallèlement, la base sémantique du RRP a été utilisée pour représenter, pour chaque unité cartographique de sol, la part de la superficie occupée par des sols caractéristiques de zones humides identifiés selon les critères pédologiques de l'arrêté du 1^{er} octobre 2009.

Concernant l'IPZH29, un indice de fiabilité est accordé aux inventaires, selon la méthodologie employée pour les établir. Pour cette étude, seuls les inventaires dont l'Indice de Fiabilité (IF) est maximal (classe 6) sont considérés. Ces inventaires couvrent 84 communes correspondant à 2 063 km², soit un tiers du territoire départemental (*figure 3*). Les inventaires dont l'IF est 6 correspondent aux zones humides délimitées par une étude de terrain répondant aux critères botaniques et/ou pédologiques d'identification des zones humides (arrêté du 1^{er} octobre 2009) et suivant la procédure du cahier des charges départemental. L'IPZH29 juge leur précision et leur calage géométriques bons.

RESULTATS

Modélisation de l'hydromorphie

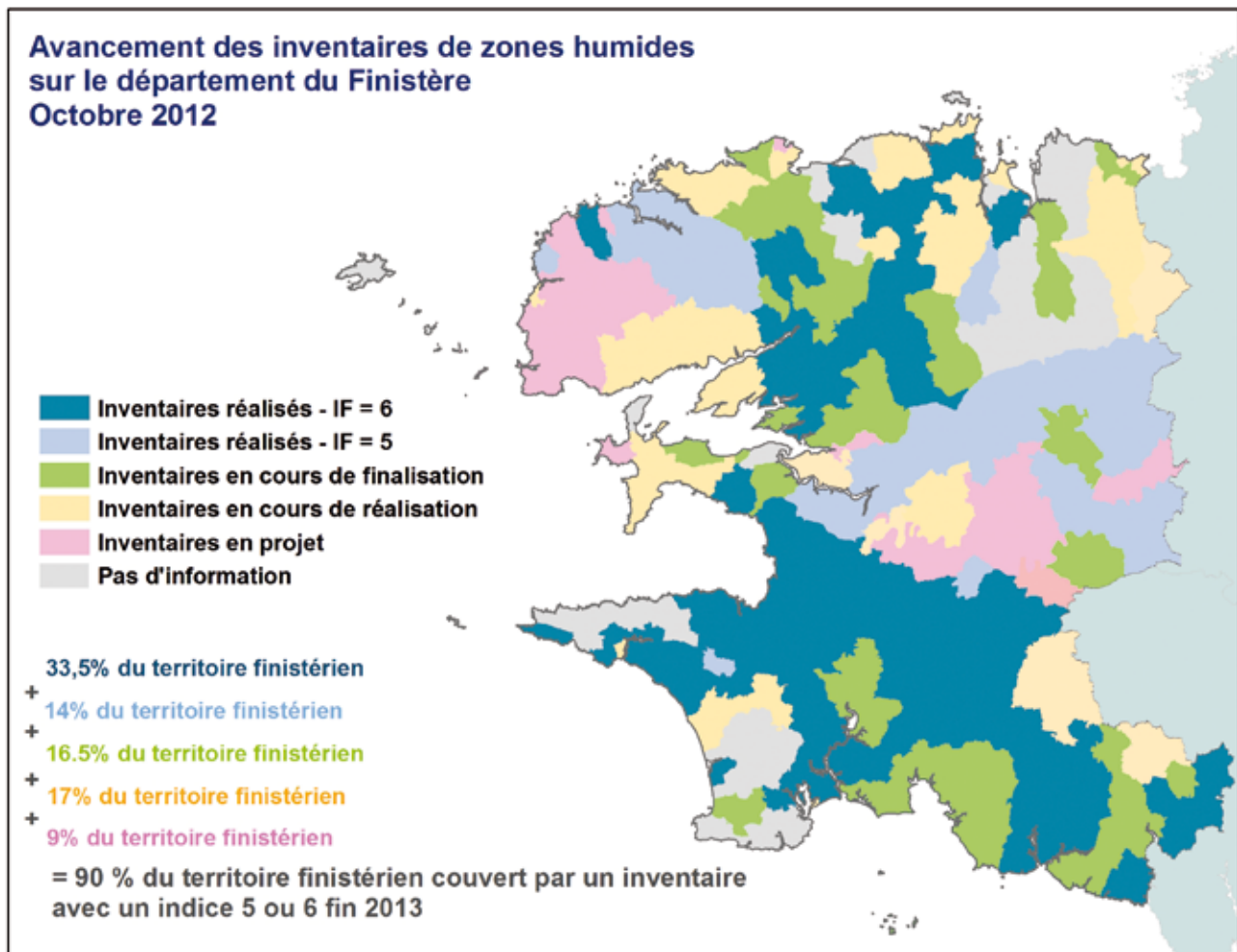
Caractéristiques du modèle

Le modèle de prédiction de l'hydromorphie élaboré se caractérise par la contribution relative des variables dans la classification et sa performance intrinsèque, estimée à partir des données utilisées pour le calibrage.

Les 14 variables environnementales ont toutes une importance relative supérieure à 23 % (*figure 4*), ce qui montre la pertinence du jeu de variables prédictives choisi pour établir le modèle. Les trois variables les plus discriminantes sont l'occupation du sol, le matériau parental et la pluviométrie. Elles ont

Figure 3 - Inventaires de terrain réalisés dans le Finistère (www.zoneshumides29.fr) utilisés pour valider la prédiction des zones humides.

Figure 3 - Terrain-based inventories available in Finistère (www.zoneshumides29.fr) used to validate wetlands prediction.



chacune une contribution relative supérieure à 75 %. Les trois variables qui participent le moins à l'élaboration du modèle ont une importance relative comprise entre 23 et 28 %. Il s'agit de la distance au cours d'eau, de la pente et de la longueur du versant.

La validation interne, basée sur les données utilisées pour calibrer le modèle, caractérise la précision intrinsèque du modèle. La précision globale est de 77 % et l'indice de Kappa de 70 %, illustrant la bonne performance intrinsèque du modèle (tableau 2). La variabilité des résultats entre les précisions du producteur et de l'utilisateur et entre les classes est faible. La validation croisée montre un taux de précision global et un indice de Kappa plus faibles : 47 % et 29 % respectivement, et une variabilité limitée des indices pour une même classe. La classe 3 est un peu moins bien prédite que les autres classes. L'examen de la matrice de confusion de la validation croisée

(non montrée) met en évidence que la valeur prédite diffère le plus souvent d'une seule classe par rapport la valeur observée. Les résultats obtenus sur le jeu de données de calibration montrent que le modèle permet de restituer les situations « apprises » sans surestimer ni sous-estimer certaines classes par rapport aux autres.

Carte de l'hydromorphie et validation externe

L'extrapolation du modèle de classification fournit une cartographie exhaustive de l'hydromorphie à l'échelle régionale (figure 5). La classe 1 représente 38 % du territoire régional, la classe 2 26 %, la classe 3 20 % et la classe 4 16 %. La carte montre une nette démarcation est-ouest, avec des sols nettement plus hydromorphes en Ille-et-Vilaine que dans les autres départements. Le sud du Morbihan et dans une moindre me-

Tableau 2 - Indices de précision de la prédiction de l'hydromorphie calculés sur les données d'apprentissage et externes.**Table 2** - Accuracy of soil redoximorphic features prediction computed for training and external datasets.

		Classes d'hydromorphie				Précision globale	Indice de Kappa
		1	2	3	4		
Jeu de données d'apprentissage							
Validation interne	Nombre de pixels prédits	447	388	378	439	0,77	0,70
	Précision du producteur	0,88	0,71	0,64	0,86		
	Précision de l'utilisateur	0,81	0,76	0,70	0,81		
Validation croisée	Nombre de pixels prédits	389	418	403	442	0,47	0,29
	Précision du producteur	0,56	0,43	0,32	0,57		
	Précision de l'utilisateur	0,60	0,43	0,33	0,53		
Jeux de données externes							
Validation ponctuelle	Nombre de pixels prédits	1 757	688	494	409	0,55	0,15
	Précision du producteur	0,58	0,35	0,34	0,69		
	Précision de l'utilisateur	0,95	0,07	0,20	0,09		
Validation par cartes	Nombre de pixels prédits	70 793	156 528	132 898	52 274	0,36	0,16
	Précision du producteur	0,29	0,47	0,42	0,33		
	Précision de l'utilisateur	0,73	0,28	0,30	0,30		

sure le sud du Finistère et le littoral nord portent également une proportion plus importante de sols hydromorphes. Dans la partie centre-ouest de la Bretagne, les sols peu et moyennement hydromorphes (classes 2 et 3) sont peu représentés. Dans ce secteur, les sols hydromorphes (classe 4) sont très liés au réseau hydrographique qui incise un relief plus accidenté que dans l'est de la région. La probabilité d'occurrence associée à la classe d'hydromorphie prédite est comprise entre 0,25 et 0,55, la moyenne étant de 0,34. En comparaison avec des études similaires (Lacoste *et al.*, 2011 ; Lemercier *et al.*, 2012), le niveau de probabilité d'occurrence est faible, ce qui traduit un niveau d'incertitude élevé du résultat de la classification. Les probabilités d'occurrence les plus faibles sont associées aux classes 2 et 3.

L'estimation régionale de l'hydromorphie est validée par comparaison avec les résultats de 3 348 observations ponctuelles d'une part et les cartes pédologiques à 1/25 000 vectorisées existantes en Bretagne d'autre part. La validation ponctuelle montre que la classe prédite correspond à la classe observée dans 55 % des cas, et un indice de Kappa de 0,15 (tableau 2). Cependant, les disparités entre classes sont importantes : la classe 1 est la mieux prédite tout en étant sous-estimée par le modèle. Les situations de sols très hydromorphes (classe 4) sont correctement représentées par le modèle, mais souvent surestimées. Les classes 2 et 3 sont moins bien pré-

dites et la classe 2 surtout est surestimée. Les résultats de la validation externe à partir de cartes pédologiques montrent également une sous-estimation de la classe 1 et une surestimation des classes 2 et 3. En comparaison avec les cartes pédologiques, la classe 4 n'est ni surestimée, ni sous-estimée par le modèle en termes d'effectif global. Les indices du producteur et de l'utilisateur de même ordre de grandeur (33 et 30 %) mais faibles observés pour la classe 4 sont liés au fait que cette classe est sous-estimée dans la partie ouest de la région et surestimée dans la partie est. De façon générale, le modèle de prédiction établi rend bien compte des grands ensembles de sols hydromorphes connus en Bretagne, mais surestime l'hydromorphie par rapport aux observations de terrain.

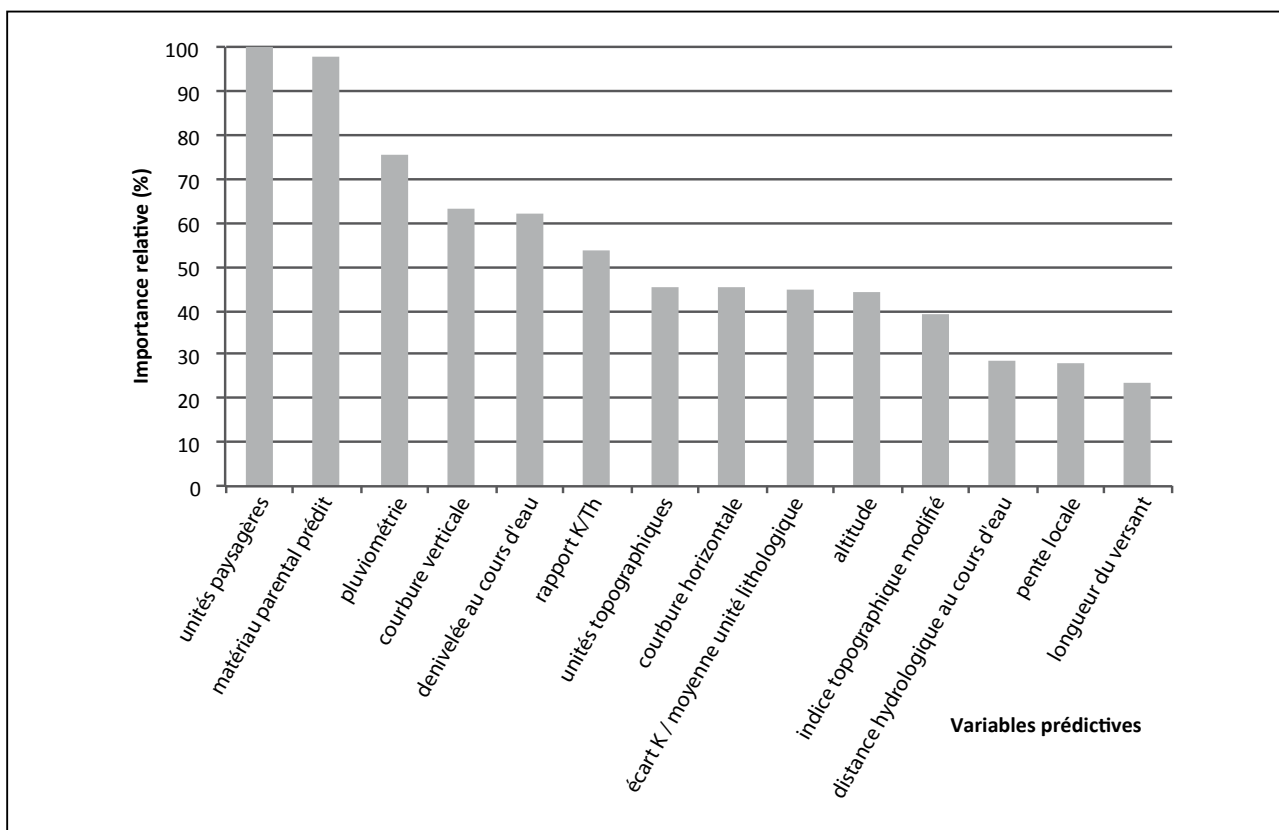
Estimation des zones humides

Cartographie prédictive des zones humides

L'hydromorphie prédite à l'échelle régionale a été utilisée pour affecter une classe de zone humide à chaque pixel : 'ZNH' (hors zone humide), 'ZI' (zone intermédiaire) ou 'ZH' (zone humide). La classe 'ZH' couvre 17 % de la superficie régionale, 65 % du territoire ne se trouvent pas en zone humide et 18 % se situent dans la zone intermédiaire où peuvent se situer des zones humides, mais à l'humidité moins intense ou en abon-

Figure 4 - Contribution relative des variables environnementales au modèle de prédiction de l'hydromorphie (explicitation des sigles dans le tableau 1).

Figure 4 - Relative contribution of environmental variables to the model of soil redoximorphic features prediction (see table 1 for code definition).



dance plus faible que dans la classe 'ZH'. La représentation cartographique des résultats (figure 6) met en évidence une distinction claire entre l'est et l'ouest de la région : à l'est la classe intermédiaire est largement représentée, alors qu'elle l'est beaucoup moins à l'ouest. D'une façon générale, la probabilité d'occurrence de la classe prédite est faible, mettant en évidence une incertitude forte sur le résultat prédit.

Validation de la prédiction des zones humides

Les zones humides prédites ont été comparées à une information pédologique disponible à l'échelle régionale, le Référentiel Régional Pédologique (RRP), et des inventaires communaux de zones humides issus de travaux de terrain localisés dans le Finistère, regroupés au sein de l'Inventaire Permanent des Zones Humides du Finistère (IPZH29).

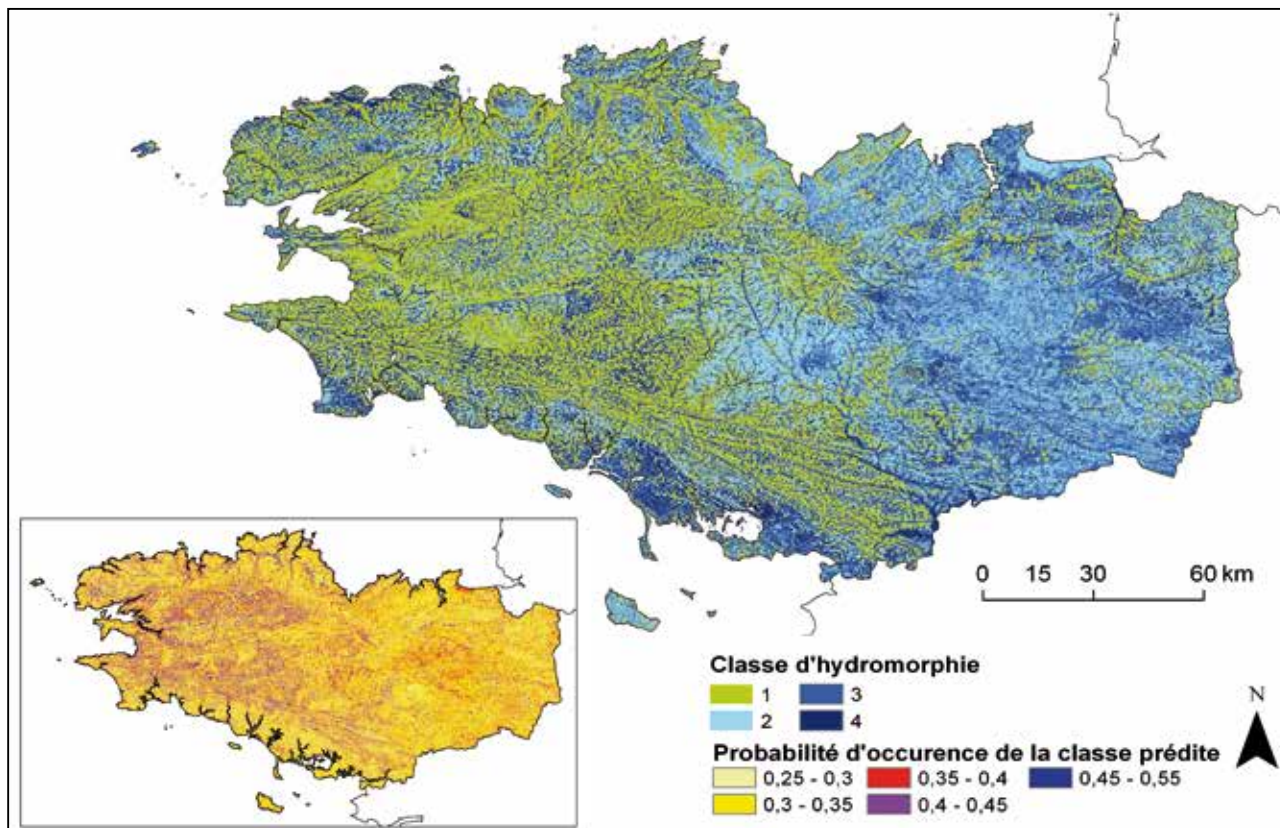
La figure 7 représente l'estimation de la superficie des zones humide par UCS du RRP, selon les zones humides modélisées (classes 'ZI' et 'ZH' agrégées) et les données du RRP. Les deux cartes présentent de fortes similitudes : leur structure

générale est semblable, la différenciation est-ouest est commune aux deux cartes et dans les deux cas la frange côtière est plus hydromorphe que l'intérieur des terres. De nouveau, la modélisation tend à surestimer l'extension des zones humides. Le fait que les classes 'ZI' (dont tous les sols ne caractérisent pas des zones humides) et 'ZH' aient été regroupées pour établir la figure 7A explique en partie cette surestimation.

Les 84 communes dont les inventaires de zones humides ont été versés à l'IPZH29 et réalisés avec un degré de fiabilité maximal représentent une superficie de 2 063 km² dont 189 km² (9,2 %) de zones humides. Sur la même zone d'étude, la classe 'ZH' de la prédiction représente 21,3 % du territoire et la classe 'ZI' 9,3%. 59,7% des zones humides inventoriées sont prédites comme telles (classes 'ZH') et 25,8 % des pixels appartenant à la classe 'ZH' entrent effectivement dans l'inventaire des zones humides. Ces résultats montrent une fois de plus que la prédiction surestime l'extension effective des zones humides, mais de façon très hétérogène selon les communes. Pour certaines, où une proportion importante du territoire a

Figure 5 - Prédiction de la classe d'hydromorphie à l'échelle de la Bretagne obtenue à partir d'un modèle par arbre de classification optimisé, et probabilité d'occurrence de la classe prédite.

Figure 5 - Predicted soil redoximorphic classes at regional scale based on a boosted classification tree model, and probability of occurrence of the predicted class.



été inventoriée comme appartenant à une zone humide, l'adéquation entre l'inventaire et la prédiction est bonne, alors que pour les communes dont une faible proportion du territoire est inventoriée en zone humide, le modèle surestime largement les zones humides. Ces deux cas sont présentés par la *figure 8* : les communes de Plomodiern, Ploeven, Cast et Quemeneven (*figure 8A*) montrent une bonne correspondance entre le modèle et les inventaires, alors que sur la commune de Nevez (*figure 8B*), les zones humides inventoriées se limitent généralement à une bande de quelques mètres de part et d'autre du cours d'eau et les zones humides prédites s'étendent plus largement. Les zones humides inventoriées et prédites ont le plus souvent des localisations similaires. Les incohérences portent sur leur extension spatiale.

DISCUSSION

La mise en œuvre de techniques de cartographie numérique des sols a permis d'établir une procédure complète de prédiction de l'hydromorphie, puis des zones humides. Les grandes étapes ont été l'élaboration du modèle de prédiction de l'hydromorphie basé sur une classification par arbres de classification optimisée (algorithme implémenté dans le package R 'adabag', Alfaro-Cortes *et al.*, 2012), l'extrapolation du modèle à l'ensemble de la zone d'étude à partir des variables environnementales exhaustives, la reclassification du résultat pour prédire les zones humides, et enfin la validation des résultats par comparaison avec des données indépendantes ou non. Les techniques de CNS ont permis de répondre à un besoin de connaissance des sols qui combine une résolution fine et une étendue régionale de l'information produite. La méthodologie adoptée a été renseignée précisément, ce qui permet d'assurer sa traçabilité et sa répétabilité. Il sera notamment possible et relativement simple de mettre à jour et préciser les prédictions

Figure 6 - Prédiction des zones humides de Bretagne obtenue par reclassification de la prédiction de l'hydromorphie à partir d'un modèle par arbre de classification optimisé, et probabilité d'occurrence de la classe prédite (ZNH : zone non humide ; ZI : zone intermédiaire ; ZH : zone humide). Avertissement : du fait de l'échelle de restitution, les classes ZI et ZH peuvent sembler surestimées. Elles représentent respectivement 18 et 17 % de la superficie de la zone d'étude.

Figure 6 - Predicted wetlands in Bretagne (France) by reclassifying soil redoximorphic features prediction based on a boosted classification tree model, and probability of occurrence of the predicted class (ZNH : no wetland ; ZI : Intermediate land ; ZH : wetland). Warning : due to the poor resolution of maps, ZI and ZH classes seemed to be overestimated. They respectively represent 18 and 17% of the study area.

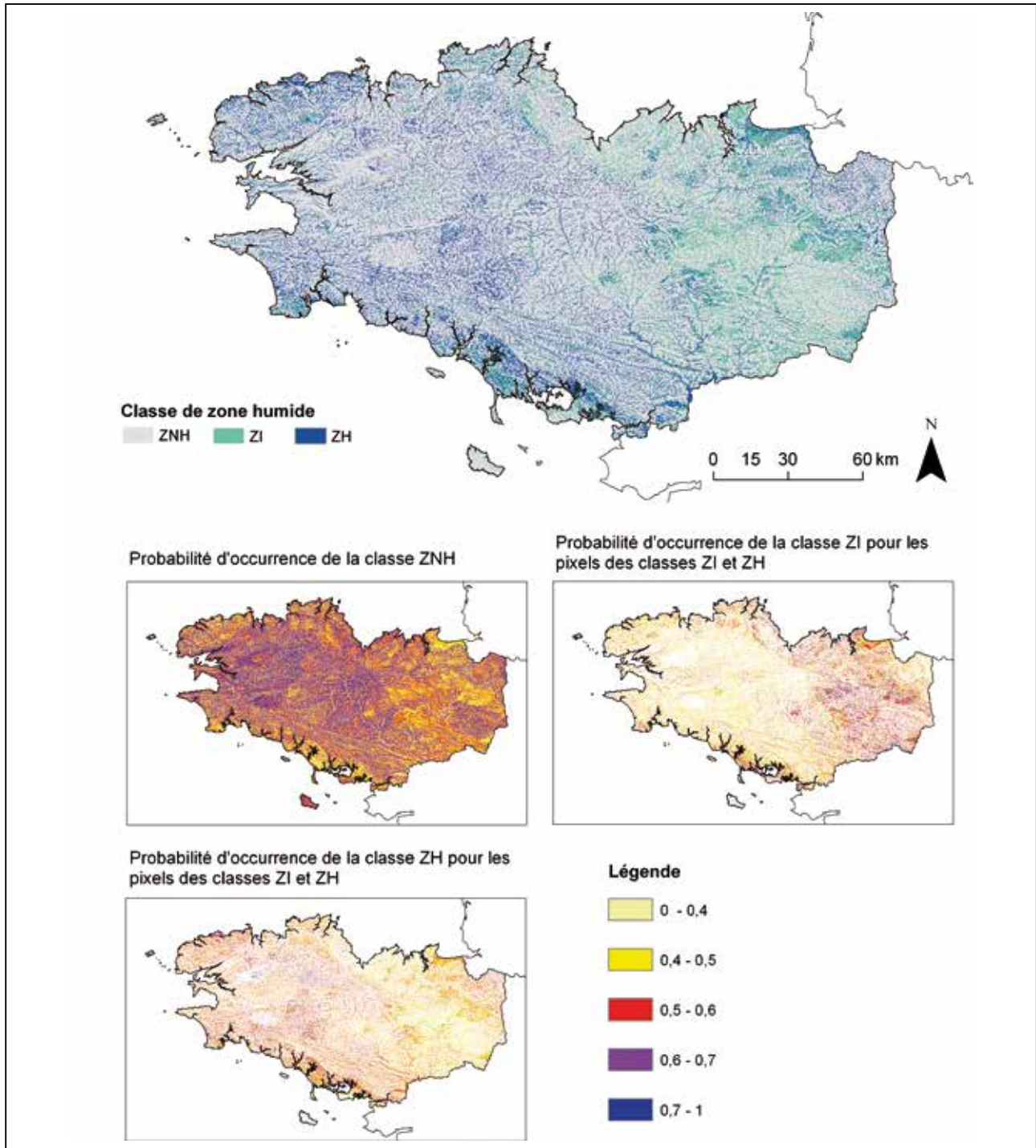
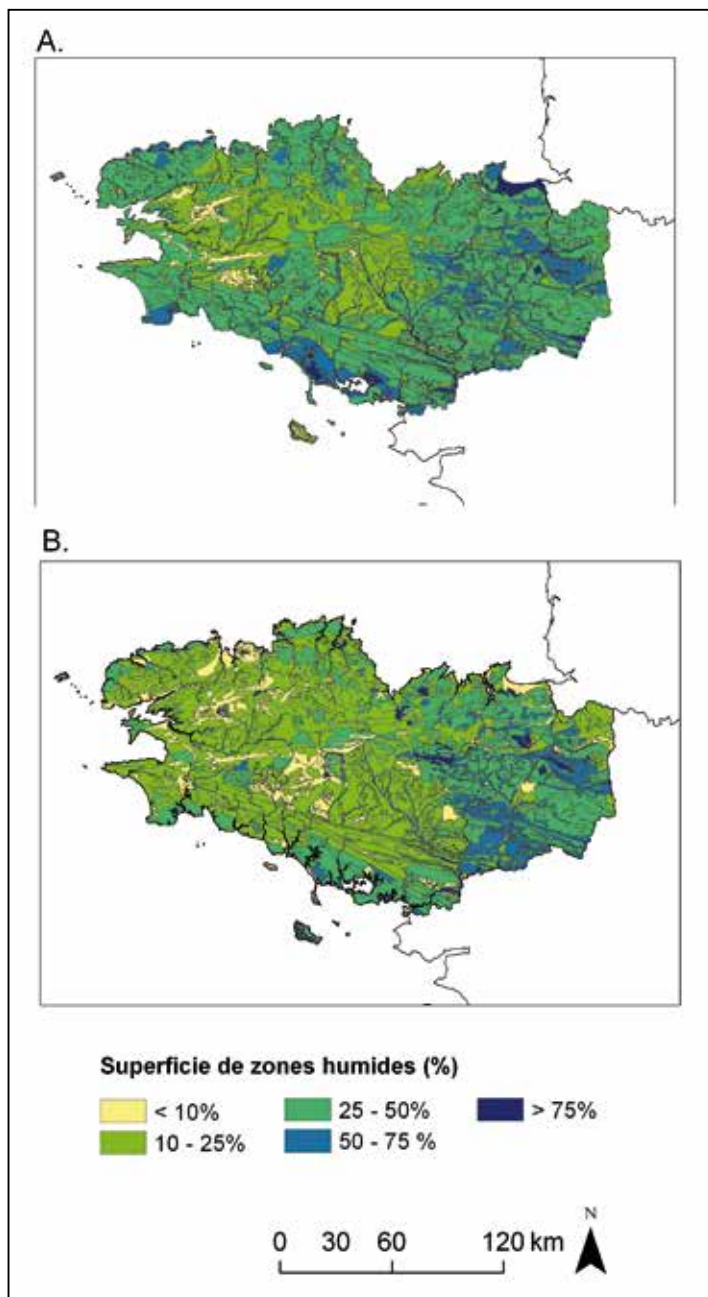


Figure 7 - Représentation des zones humides (classes ZI et ZH) intégrées par unités cartographique de sol du référentiel régional pédologique de Bretagne, à partir des zones humides prédites (A) et de la superficie occupée par des types de sols caractéristiques de zones humides décrits dans le référentiel régional pédologique.

Figure 7 - Wetlands (intermediate and wetland classes) aggregated by soil mapping units area of the 1 : 250 000 soil map of Brittany derived from predicted wetlands (A,) and from proportion of soil types characterizing wetlands described in the 1 : 250 000 soil map.



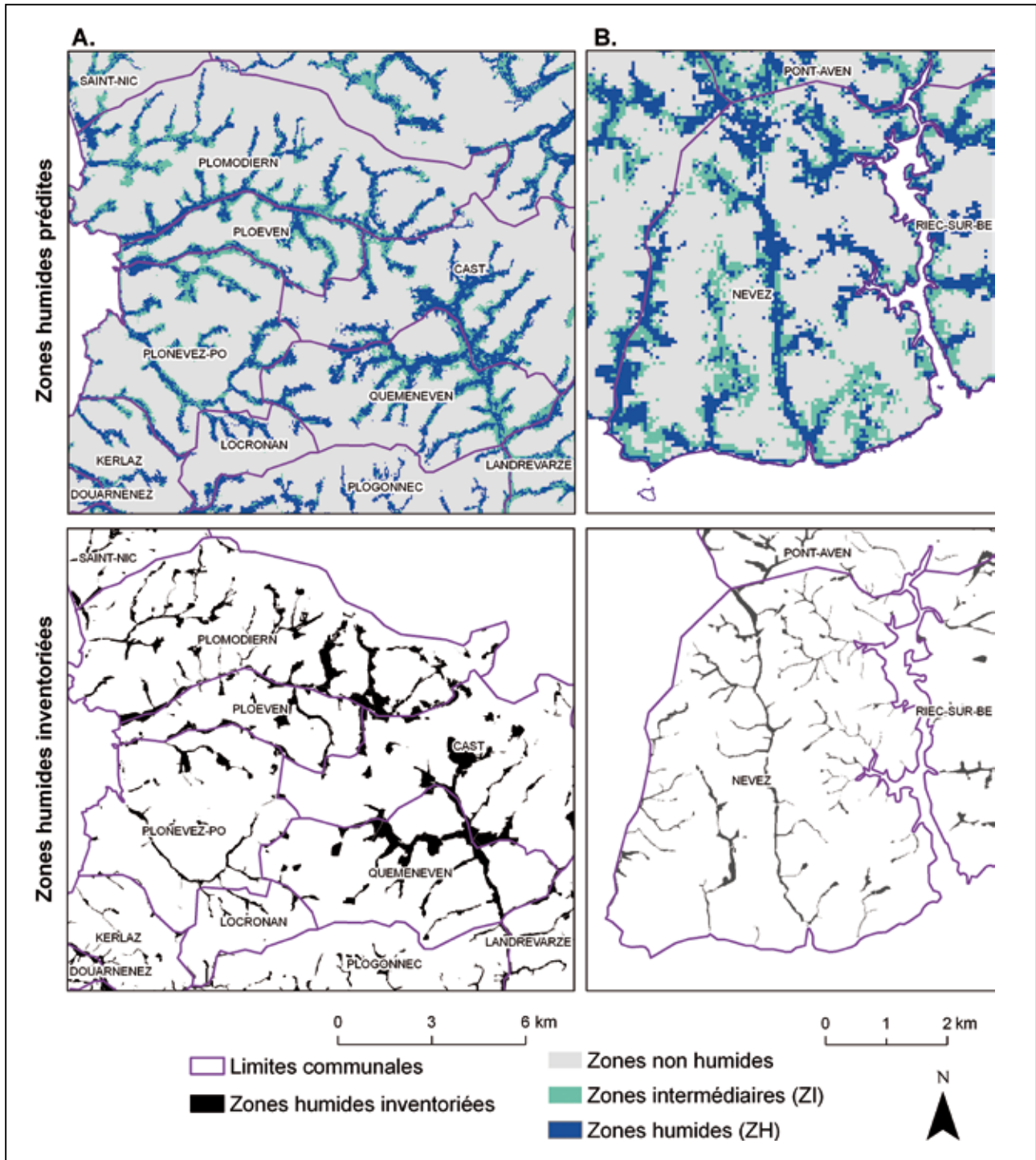
si de nouvelles données pédologiques ou variables environnementales prédictives sont disponibles.

L'hydromorphie est un paramètre pédologique assez simple à caractériser et restituer sous forme de classes en fonction de la profondeur d'apparition et l'intensité des traits morphologiques spécifiques, mais qui dépend de nombreux facteurs pédologiques, topographiques, hydrologiques, géologiques, climatiques et anthropiques qui interagissent à des échelles spatiales variables. La prédiction de l'hydromorphie n'est donc pas aisée du fait de l'interaction entre les facteurs. En Bretagne, la classification des pixels selon 4 classes d'hydromorphie montre un taux d'accords bruts, ou précision globale, de 55 % (validation externe ponctuelle). Des résultats du même ordre de grandeur avaient été obtenus avec des techniques similaires : 44 % pour la prédiction d'unités cartographiques de sol dans le sud de la Corrèze (Grinand *et al.*, 2008), 54 % pour la prédiction du matériau parental en Bretagne (Lacoste *et al.*, 2011) et 52 % pour la prédiction de l'hydromorphie sur la moitié nord de l'Ille-et-Vilaine (Lemerrier *et al.*, 2012). La grande majorité des erreurs est liée à une surestimation de l'hydromorphie par le modèle. D'un point de vue cartographique, les résultats sont conformes aux connaissances préexistantes sur la répartition régionale des sols hydromorphes.

Concernant les zones humides, trois classes ont été prédites : hors zone humide ('ZNH'), zone humide ('ZH') et zone intermédiaire ('ZI'). De même que l'hydromorphie, l'extension des zones humides est surestimée par rapport aux surfaces estimées grâce au RRP et aux inventaires de terrain effectués dans le Finistère et regroupés dans l'IPZH29. Cependant, dans une optique de préservation des ressources, il vaut mieux que les zones humides soient surestimées que sous-estimées. Les sols associés à la classe 'ZI' présentent une hydromorphie relativement peu marquée (absence d'horizons réductiques ou histiques) débutant à moins de 40 cm de profondeur, se prolongeant ou s'intensifiant en profondeur. Selon l'arrêté du 1^{er} octobre 2009 relatif aux critères de définition et de délimitation des zones humides, les sols de la classe 'ZI' caractérisent une zone humide si et seulement si leur hydromorphie débute à moins de 25 cm. Or les informations ponctuelles utilisées pour établir le modèle ne permettent pas de distinguer cette situation (classes Va, Vb et Vc) des cas non caractéristiques de zones humides (IVc et IVd). Dans la partie est de la zone d'étude, la classe 'ZI' est nettement plus représentée que la classe

Figure 8 - Comparaison de l'extension de zones humides prédites et inventoriées (IPZH29) dans le cas de communes où le ratio entre les zones humides inventoriées et prédites est compris entre 0,6 et 0,8 (A) et est de 0,2 (B). Les zones humides des communes situées au nord de Plomodiern et à l'ouest de Nevez n'ont pas été inventoriées.

Figure 8 - Comparison of spatial extent of predicted and inventoried wetlands (IPZH29) for municipalities where inventoried versus predicted wetlands ratio is between 0.6 and 0.8 (A) and is 0.2 (B). Wetlands were not inventoried in Municipalities located in the north of Plomodiern and in the western of Nevez.



'ZH', et inversement dans la partie ouest. La comparaison avec les inventaires existants et les enveloppes de zones humides potentielles basées sur le calcul d'un indice topographique (Mérot *et al.*, 2003) suggère que les résultats de la classe 'ZI' doivent être analysés différemment selon la zone considérée : dans l'est cette classe regroupe en grande partie des zones humides, alors que dans la partie ouest, la presque totalité des zones humides inventoriées se distribuent presque qu'exclusivement dans la classe 'ZH', et pas dans la classe 'ZI'.

Des éléments de compréhension de ces résultats sont à rechercher également dans l'élaboration du modèle lui-même, et notamment l'importance relative des variables prédictives et les caractéristiques de l'échantillonnage. Les variables prédictives les plus importantes (occupation du sol, matériau parental, pluviométrie) traduisent des facteurs de distribution de l'hydromorphie qui interviennent à l'échelle de grandes entités physiographiques : le matériau parental caractérise la perméabilité du substrat et des sols, la pluviométrie caractérise les apports d'eau et l'occupation du sol est un facteur intégrateur des conditions du milieu et de l'organisation spatiale des activités humaines qui en découlent. Les variables topographiques qui expliquent la variabilité de l'hydromorphie à l'échelle des versants, plus locale, ont une importance secondaire. Parmi les variables topographiques, la courbure verticale de la pente, qui traduit la convexité ou la concavité de la pente, et la dénivellée par rapport au cours d'eau le plus proche contribuent le plus à la classification. Moran et Bui (2002), Grinand *et al.* (2008), Behrens *et al.* (2010) et Lehmann *et al.* (2013) ont montré l'intérêt d'utiliser à la fois des données topographiques brutes et ces mêmes données lissées par l'application de filtres de convolution comme variables prédictives de classifications supervisées. L'intégration de variables lissées selon des rayons variables de convolution permet de prendre en compte différentes échelles de structuration spatiale des informations topographiques, mais implique un nombre de variables explicatives plus important.

La performance interne du modèle est correcte : le taux d'accords bruts entre la prédiction et l'observation est de 77 %, et toutes les classes ont une qualité de prédiction du même ordre de grandeur, ce qui signifie que le modèle restitue correctement l'information dans les situations sur lesquelles les relations variables prédictives – hydromorphie ont été explicitées. Par contre, suite à l'extrapolation à l'ensemble de la région, l'hydromorphie et l'extension des zones humides sont surestimées. Cela peut être lié au fait que, sur la zone d'extrapolation, un certain nombre de combinaisons des variables environnementales n'étaient pas bien représentées dans l'échantillonnage de calibration, et l'hydromorphie correspondant à ces situations inédites pour le modèle n'est donc pas correctement prédite. Chaplot *et al.* (2003) avaient déjà montré la difficulté de modéliser correctement l'hydromorphie sur la base des relations sol-topographie lorsque

les modèles sont appliqués dans des contextes différents de ceux dans lesquels ils ont été élaborés. La qualité des prédictions pourrait être améliorée par l'adaptation de la stratégie d'échantillonnage : la stratification peut être envisagée non seulement en fonction de la classe d'hydromorphie observée, mais aussi en fonction de la localisation géographique, pour mieux couvrir la diversité des contextes. Les ensembles géographiques issus des travaux sur la tectonique récente (Bonnet *et al.*, 1998), qui sont au nombre de 5 en Bretagne, pourraient être un support intéressant de la stratification de l'échantillonnage étant donné que la tectonique récente influence la répartition des sols hydromorphes (Chaplot *et al.*, 1999 ; Chaplot et Walter, 2007),.

Enfin, l'étape de validation externe à partir des cartes pédologiques existantes permet de comparer des restitutions cartographiques entre elles. Cependant, les cartes à grande échelle valorisables sont limitées en nombre, et ne se répartissent pas de façon homogène dans la zone d'étude. La validation par comparaison à des cartes pédologiques n'a pu se faire que pour les contextes pédo-géo-topographiques caractérisés par les cartes. De plus, les unités cartographiques des cartes de validation n'ont pas été délimitées selon le seul critère d'hydromorphie. Elles sont donc probablement impures, sans qu'on puisse estimer leur degré de pureté vis-à-vis de l'hydromorphie. L'utilisation des inventaires de zones humides pour valider l'extension prédite des zones humides potentielles doit se faire avec précaution, même dans le cas où les inventaires sont réalisés selon les critères de l'arrêté du 1er octobre 2009. En effet, les zones humides inventoriées ne sont pas forcément les zones humides effectives selon les critères pédologiques retenus dans l'arrêté. Dans les faits, il n'y a pas de toujours de consensus autour de la définition des zones humides, notamment dans le cas des zones cultivées qui ne présentent pas ou plus de végétation caractéristique des zones humides. Sur les 84 inventaires communaux de l'IPZH29 ayant un indice de fiabilité maximal, seuls 7 % de la superficie des zones humides inventoriées sont des parcelles cultivées. L'inventaire des zones humides effectives à grande échelle résulte de travaux de terrain qui restent indispensables, fréquemment associés à un processus de concertation entre les acteurs du territoire. Ainsi, la validation de travaux scientifiques par des inventaires communaux n'est pas toujours objective, les inventaires intégrant souvent des enjeux économiques, agricoles et environnementaux locaux, qui se superposent aux critères pédologiques et botaniques fixés par la réglementation.

CONCLUSION

Les informations spatialisées, les méthodes quantitatives pour décrire l'organisation des sols et des processus pédologiques ainsi que les outils pour gérer et visualiser les don-

nées sont maintenant largement disponibles, et ouvrent de nouvelles opportunités pour prédire les sols et leurs propriétés (Grunwald, 2009 ; Walter et al., 2011). L'hydromorphie est un paramètre pédologique de première importance qui influence les fonctions de production, environnementales et sociétales des sols, et les zones humides sont des milieux sensibles qu'il convient de connaître en vue de leur gestion durable et de la préservation de leurs fonctions écosystémiques. La mise en œuvre d'approches de cartographie numérique des sols a permis de dresser des cartes de prédiction de l'hydromorphie par apprentissage à partir de données pédologiques ponctuelles et de variables environnementales exhaustives puis extrapolation à l'ensemble de la zone d'étude. Cette approche permet d'avoir une prédiction de l'hydromorphie en 4 classes, à une résolution fine sur une région entière, avec une incertitude quantifiée, basée sur l'observation des sols. Les résultats montrent que les grands ensembles de sols hydromorphes sont retrouvés et précisés à l'échelle régionale et que le modèle tend à surestimer l'hydromorphie par rapport aux observations de terrain. Ce résultat a ensuite été utilisé dans une démarche d'évaluation numérique des sols pour prédire l'extension des zones humides. Les résultats sont intéressants, même si les zones humides prédites couvrent une surface plus importante que les zones humides inventoriées. La validation de ces cartes prédictives est une étape essentielle mais difficile à mener par manque d'inventaires de terrains fiables en nombre suffisant. Les conclusions de ce travail sont très encourageantes et des ajustements permettront d'optimiser les résultats. Il s'agira notamment d'adapter l'échantillonnage, et le choix des variables explicatives, voire l'algorithme d'apprentissage. Les cartes de prédiction de l'hydromorphie et des zones humides sont complémentaires des travaux de terrain et permettront d'optimiser les campagnes de terrain ultérieures, que ce soit pour réaliser des cartes pédologiques ou des inventaires de zones humides. Pour une caractérisation à une échelle fine, la probabilité d'occurrence associée aux classes prédites peut aider à adapter l'effort d'échantillonnage en densifiant les observations quand la probabilité est faible ou hétérogène.

Une des perspectives de ce travail est de transposer la méthode à d'autres territoires, la procédure complète de prédiction de l'hydromorphie et des zones humides étant établie. Cela implique de disposer d'informations ponctuelles et/ou de cartes précises pour calibrer le modèle et valider les résultats, et de variables environnementales exhaustives. Quoiqu'il en soit, l'interprétation correcte des résultats nécessitera toujours une expertise pédologique locale.

REMERCIEMENTS

Cette étude a été menée dans le cadre du Réseau Mixte Technologique (RMT) Sols et Territoires, dont les auteurs remercient les membres. La comparaison de la prédiction des zones humides potentielles avec les zones humides inventoriées a pu être réalisée grâce au forum des marais atlantiques et au conseil général du Finistère qui ont mis à disposition les données de l'inventaire permanent des zones humides du Finistère.

BIBLIOGRAPHIE

- Alfaro-Cortes E., Gamez-Martinez M., Garcia-Rubio N., 2012 - Package 'adabag'. *Appl. multiclass AdaBoost.M1, AdaBoost-SAMME and Bagging*. R package version 3.1., 21 p.
- Baize D., Jabiol B., 2011 - Guide pour la description des sols. Editions Quae, Versailles (France), 429 p.
- Barnaud G., Fustec E., 2007 - Conserver les zones humides : pourquoi ? comment ? Editions Educagri, Dijon (France) et Quae, Versailles (France), 296 p.
- Behrens, T., Schmidt, K., Zhu, A. X., Scholten, T., 2010 - The ConMap approach for terrain-based digital soil mapping. *European Journal of Soil Science*, 61, 1, pp. 133 – 143.
- Beven K.J., Kirkby M.J., 1979 - A physically based variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24, pp. 43-69.
- Bonijoly D., Perrin J., Truffet C., Asfirane F., 1999 - Couverture géophysique aéroportée du Massif armoricain. Rapport BRGM R-40471, 75 p.
- Bonnet S., Guillocheau F., Brun J.P., 1998 - Relative uplift measured using river incisions : the case of the Armorican basement (France). *Compte-Rendus de l'Académie des Sciences Paris*, 327, série IIa, pp. 245-251.
- Moran, C.J., Bui, E.N., 2002 - Spatial data mining for enhanced soil map modelling. *International Journal of Geographical Information Science*, 16, 6, pp. 533-549.
- Carré F., McBratney A.B., Mayr T., Montanarella L., 2007 - Digital soil assessments : Beyond DSM. *Geoderma*, 142, pp. 69-79.
- Chantraine J., Rabu D., Béchenec F., 2002 - Carte géologique numérique à 1/250 000 du Massif armoricain, version 1.1., BRGM Editions.
- Chaplot V., Walter C., Curmi P., 1999 - Arguments cartographiques en faveur du rôle de la tectonique récente sur la distribution régionale des sols du Massif armoricain. *Compte-Rendus de l'Académie des Sciences Paris*, 329, pp. 487-493.
- Chaplot V., Walter C., Curmi P., 2003 - Testing quantitative soil-landscape models for predicting the soil hydromorphic index at a regional scale. *Soil Science*, 168, 6, pp. 445-454.
- Chaplot V., Walter C., 2007 - Improving the spatial prediction of soil s at local and regional levels through a better understanding of soil-landscape relationships : soil hydromorphy in the Armorican Massif of Western France. *In* : P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), *Digital Soil Mapping, an introductory perspective*. *Developments in soil science*, vol. 31. Elsevier Science, Amsterdam (The Netherlands), pp. 507-520.
- Cohen J., 1960 - A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, 20, pp. 27-46.
- Conseil général du Finistère, Forum des Marais Atlantiques, 2010 - Guide méthodologique. Inventaire des zones humides à l'échelle locale sur le département du Finistère - Édition octobre 2010. 44 p.
- Curmi P., Durand P., Gascuel-Odoux C., Mérot P., Walter C., Taha A., 1998 - Hydromorphic soils, hydrology and water quality : spatial distribution and functional modeling at different scales. *Nutrient Cycling in Agroecosystems*, 50, pp. 127-142.

- Dobos E., Carré F., Hengl T., Reuter H.I., Tóth G., 2006 - Digital Soil Mapping as a support to production of functional maps. EUR 22123 EN. Office for Official Publications of the European Communities, Luxembourg, 68 p.
- Durand P., Gascuel-Oudou C., Kao C., Merot P., 2000 - Une typologie hydrologique des petites zones humides ripariennes. *Etude et Gestion des Sols*, 7, 3, pp. 207-218.
- Freund Y., Schapire R., 1997 - A decision theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55, pp. 119-139.
- Friedman, J.H., 2001 - Greedy function approximation : a gradient boosting machine. *The annals of statistics*, 29, 5, pp. 1189-1232.
- Friedman, J.H., Meulman, J.J., 2003 - Multiple additive regression trees with application in epidemiology. *Statistics in Medicine*, 22, pp. 1365-1381.
- Groupe d'Étude des Problèmes de Pédologie Appliquée (GEPPA), 1981. Synthèse des travaux de la commission de cartographie 1970-1981. INRA-CESRPF. 20 p.
- Grinand C., Arrouays D., Laroche B., Martin M.P., 2008 - Extrapolating regional landscapes from an existing soil map : sampling intensity, validation procedures, and integration of spatial context. *Geoderma*, 143, pp. 180-190.
- Grunwald S., 2009 - Multi-criteria characterization of recent digital soil mapping modeling approaches. *Geoderma*, 152, pp. 195-207.
- Jenny, H., 1941. *Factors of Soil Formation, A System of Quantitative Pedology*. McGraw-Hill, New York. 281 p.
- Lacoste M., Lemerrier B., Walter C., 2011 - Regional mapping of soil parent material by machine learning based on punctual training data. *Geomorphology*, 133, 1-2, pp. 90-99.
- Lagacherie P., McBratney A.B., 2007 - Spatial soil information systems and spatial soil inference systems : perspectives for Digital Soil Mapping. In : P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), *Digital Soil Mapping, an introductory perspective*. *Developments in soil science*, vol. 31. Elsevier, Amsterdam (The Netherlands), pp. 3-24.
- Lardeux H. (Coord.), 2002 - *Guide géologique de la Bretagne*. 3^e édition. Collection Guides géologiques régionaux, Editions Dunod, Paris (France), 221 p.
- Laroche B., Arrouays D., Olivier D., Lecerf N., 2011 - Essais de cartographie numérique des pédopaysages à 1/250 000 dans les départements de l'Aube et de la Marne. Enseignements, limites et perspectives. *Etude et Gestion des Sols*, 18, 1, pp. 17-31.
- Le Dû-Blayo L., Bourget E., Corpetti T., Gouery P., Michel K., 2008 - Esquisse cartographique des pédopaysages de Bretagne par télédétection. Rapport final, 82 p + annexes.
- Legros J.P., 1996 - *Cartographie des sols*. Editions Presses polytechniques et universitaires romandes, Lausanne (Suisse), 321 p.
- Lehmann S., Eimberck M., Martin M.P., Arrouays D., 2013 - Cartographie numérique d'une carte pédologique au 1/50 000 dans le Doubs, France. *Etude et Gestion des Sols*, 20, 1, pp. 27-46.
- Lemerrier B., Berthier L., Buquen G., Dutin G., Guichard S., Le Bris A.L., Renault-Gardan S., Quidu O., Walter C., Cluzeau D., Guernion M., Mercier V., Pérès G., Piron D., Rougé L., Le Dû-Blayo L., Michel K., Gouéry P., Bourget E., Corpetti T., Tico S., Dupont C., Giteau J.L., Guéillot C., Hanocq D., 2010 - Programme Sols de Bretagne (2005-2010) - Rapport final. Document UMR 1069 SAS, 111 p. + annexes
- Lemerrier B., Lacoste M., Loum M., Walter C., 2012 - Using boosted classification tree-based models to extrapolate at regional scale local knowledge embedded in detailed soil maps : a two-step approach. *Geoderma*, 171-172, pp. 75-84.
- McBratney A.B., Mendonça Santos M.L., Minasny B., 2003 - On digital soil mapping. *Geoderma*, 117, pp. 3-52.
- Mérot P., Ezzahar B., Walter C., Arousseau P., 1995 - Mapping waterlogging of soils using digital terrain models. *Hydrological Processes*, 9, pp. 27-34.
- Mérot P., Squidant H., Arousseau P., Hefting M., Burt T., Maitre V., Kruke M., Butturini A., Thenail C., Viaud V., 2003 - Testing a climato-topographic index for predicting wetlands distribution along an European climate gradient. *Ecological Modelling*, 163, pp. 51-71.
- Mourier B., Walter C., Mérot P., 2008. Soil distribution in valleys according to stream order. *Catena* 72, 3, pp. 395-404.
- Millennium Ecosystem Assessment, 2005 - *Ecosystems and Human Well-being : Biodiversity Synthesis*. World Resources Institute, Washington, DC. 86 p.
- Rivière J.M., Dupont C., Tico S., 1992 - Méthode tarière Massif armoricain, Editions INRA, 24 p.
- Story, M., Congalton, R.G., 1986 - Accuracy assessment : A user's perspective. *Photogrammetric Engineering and Remote Sensing*, 52, 3, pp. 397-399.
- Schmidt K., Behrens T., Scholten T., 2008 - Instance selection and classification tree analysis for large spatial datasets in digital soil mapping. *Geoderma*, 146, pp. 138-146.
- Strahler A.N., 1952 - Hypsometric (area-altitude) analysis of erosional topography. *Geological Society of America Bulletin*, 63, pp. 1117-1142.
- Walter C., King D., Lagacherie P., Robbez-Masson J.M., 2011 - L'analyse spatiale des sols : description, modélisation et représentation de la variabilité spatiale des sols. In : Girard M.C., Walter C., Rémy J.C., Berthelin J., Morel J.L. (Dir.), *Sol et environnement*, 2^e édition. Editions Dunod, Paris (France), pp. 427-453.
- Wilford J.R., Bierwirth P.N., Craig M.A., 1997 - Application of airborne gamma-ray spectrometry in soil/regolith mapping and applied geomorphology. *Journal of Australian Geology and Geophysics*, 17, 2, pp. 201-216.
- Zhu J., Zou H., Rosset S., Hastie T., 2009 - Multi-class AdaBoost. *Statistics and Its Interface*, 2, pp. 349-360.

RÉFÉRENCE INTERNET :

- Forum des Marais Atlantiques et Conseil Général du Finistère. Inventaire Permanent des Zones Humides du Finistère, serveur de cartographie dynamique. <http://zoneshumides29.fr:8080/cartoweb3/htdocs/ipzh.php> (consulté le 22 juillet 2013).

