Nowcasting agro-climatique par fusion satellite et régimes spatio-temporels : cadre de recherche institutionnel multi-cultures et multi-régions

Résumé

Ce document construit un indicateur d'offre agricole à partir d'imagerie satellite gratuite (Sentinel-1, Sentinel-2), en traitant explicitement les sources de bruit instrumental propres à chaque capteur, en modélisant le stress hydrique par un processus de régime spatio-temporel à plusieurs états plutôt qu'un simple seuillage, et en le confrontant à des modèles d'apprentissage profond comme alternatives de robustesse. Le signal est validé en dehors de son propre cadre de construction (corrélation avec les rendements agricoles USDA), testé sur plusieurs cultures et plusieurs régions productrices, et évalué par un protocole de backtest incluant coûts d'exécution réalistes, contraintes de capacité et tests de robustesse sur épisodes extrêmes. L'ambition est un signal simultanément scalable, robuste, faiblement corrélé aux primes de risque standard, persistant, et compatible avec une mise en production.

1 Positionnement et originalité

L'usage le plus répandu de l'imagerie satellite en finance agricole consiste à calculer un indice de végétation brut et à le régresser directement sur un rendement ou un prix. Cette approche pose trois problèmes rarement traités ensemble dans la littérature académique ou dans les notes de recherche buy-side accessibles publiquement : le bruit instrumental des capteurs est ignoré ou traité de façon uniforme malgré sa forte hétérogénéité, le stress hydrique est réduit à un état binaire alors qu'il évolue par paliers de sévérité distincts, et la structure spatiale du phénomène (une sécheresse ne s'arrête pas aux limites administratives d'un comté) est ignorée par des modèles comté par comté indépendants.

Ce projet apporte quatre éléments différenciants. Premièrement, une quantification explicite du bruit radar (dépendant de l'angle d'incidence et de la rugosité de surface) et du bruit optique (dépendant de la correction atmosphérique), intégrée directement dans la covariance d'observation du filtre de fusion plutôt que traitée comme une constante. Deuxièmement, un modèle de régime à quatre états avec dépendance spatiale explicite entre comtés voisins. Troisièmement, une couche de comparaison avec des modèles séquentiels et spatio-temporels modernes, utilisée comme test de robustesse plutôt que comme unique brique de production. Quatrièmement, un protocole de validation et de backtest construit selon les standards d'une équipe de recherche institutionnelle : validation externe indépendante, généralisation multi-cultures et multi-régions, coûts d'exécution réalistes, tests de stress et correction pour le biais de sélection de spécification.

2 Données et prétraitement

2.1 Sources

Sentinel-2 fournit une image optique multispectrale à 10 mètres de résolution, cycle de revisite de cinq jours. Sentinel-1 fournit une image radar en bande C à 5 mètres de résolution, cycle de six jours, insensible aux conditions météorologiques. Les deux missions sont gratuites et accessibles sans restriction pour un usage de recherche via le Copernicus Data Space Ecosystem.

2.2 Correction et quantification du bruit radar

Le coefficient de rétrodiffusion Sentinel-1 dépend fortement de l'angle d'incidence local et de la rugosité de surface, ce qui rend une utilisation brute du signal instable dans le temps et dans l'espace. Le prétraitement comprend un aplatissement du terrain (*terrain flattening*) fondé sur un modèle numérique de terrain, qui normalise le signal en $\gamma^0 = \sigma^0 / \cos(\theta_{\text{loc}})$ où θ_{loc} est l'angle d'incidence local calculé à partir de la géométrie réelle du sol plutôt que de l'angle satellite nominal, ainsi qu'un filtrage multitemporel du chatoiement (*speckle*), qui exploite la redondance des passages successifs plutôt qu'un filtrage spatial isolé qui dégraderait la résolution.

Le bruit résiduel n'est cependant pas éliminé et doit être quantifié plutôt qu'ignoré. On estime, sur des cibles de calibration stables (sol nu post-récolte, réflecteurs urbains), la variance résiduelle du signal en fonction de classes d'angle d'incidence et d'un proxy de rugosité dérivé du modèle numérique de terrain, en s'appuyant sur les modèles semi-empiriques de rétrodiffusion du type Oh ou Dubois qui relient σ^0 à l'humidité du sol, la rugosité de surface et l'angle d'incidence. La variance ainsi estimée $\hat{\sigma}^2_{\text{sar}}(\theta_t, \rho_i)$ alimente directement la covariance d'observation du filtre de fusion (section 4.1), qui devient hétéroscédastique : une observation radar prise sous un angle défavorable ou sur un terrain rugueux pèse mécaniquement moins dans la mise à jour de l'état filtré.

2.3 Correction atmosphérique et normalisation du signal optique

Le signal Sentinel-2 brut (réflectance au sommet de l'atmosphère) n'est pas directement exploitable : il doit être corrigé de l'aérosol, de la vapeur d'eau et des cirrus pour obtenir une réflectance de surface, via une correction atmosphérique standard (type Sen2Cor), et masqué des pixels nuageux, d'ombre de nuage et de neige à l'aide de la couche de classification de scène associée plutôt qu'un seuil de nébulosité arbitraire. Une normalisation BRDF (méthode dite du *c-factor*) ramène chaque acquisition à une géométrie de visée et d'illumination de référence, pour éviter qu'une variation d'angle de prise de vue ne soit confondue avec une variation réelle de l'état de la végétation.

Comme pour le radar, l'incertitude résiduelle de la correction atmosphérique (propagée depuis l'incertitude d'estimation de l'épaisseur optique des aérosols jusqu'à l'indice de végétation par une approximation delta) fournit une variance d'observation $\hat{\sigma}^2_{\text{opt}}(\text{AOD}_t, \theta_{\text{vue},t})$, également hétéroscédastique, utilisée dans le filtre de fusion.

2.4 Couverture multi-cultures et multi-régions

Le calendrier phénologique est indexé non pas en jours calendaires mais en degrés-jours de croissance cumulés (*growing degree days*), ce qui aligne les stades phénologiques entre cultures et latitudes différentes. Une courbe de référence $g_c(t)$ est calibrée séparément par culture c (maïs, soja, blé, coton), avec une fenêtre de sensibilité au stress hydrique propre à chaque culture (floraison femelle pour le maïs, remplissage des gousses pour le soja, épiaison pour le blé, formation des capsules pour le coton). Pour l'extension hors des États-Unis, le masque de parcelles agricoles USDA Cropland Data Layer est remplacé par des produits harmonisés à couverture mondiale ou régionale (ESA WorldCereal, MapBiomas pour le Brésil), et l'instrument de trading associé est choisi selon la liquidité disponible plutôt

que selon la localisation géographique de la production : l'exposition à la production ukrainienne, par exemple, est prise via les futures CBOT plutôt qu'un contrat local, et les zones sous occupation ou affectées par le conflit sont explicitement exclues du signal agrégé pour éviter de confondre abandon de parcelles lié au conflit et stress hydrique réel.

3 Cadre méthodologique

3.1 Fusion des observations par modèle d'état

Pour chaque comté, l'état latent de santé végétale suit une marche aléatoire $x_t = x_{t-1} + w_t$, $w_t \sim \mathcal{N}(0, Q)$, observée par les deux capteurs :

$$y_t^{\text{opt}} = H_{\text{opt}}\, x_t + v_t^{\text{opt}}, \quad v_t^{\text{opt}} \sim \mathcal{N}(0, R_t^{\text{opt}}), \qquad R_t^{\text{opt}} = \hat{\sigma}^2_{\text{opt}}(\text{AOD}_t, \theta_{\text{vue},t})$$

$$y_t^{\text{sar}} = H_{\text{sar}}\, x_t + v_t^{\text{sar}}, \quad v_t^{\text{sar}} \sim \mathcal{N}(0, R_t^{\text{sar}}), \qquad R_t^{\text{sar}} = \hat{\sigma}^2_{\text{sar}}(\theta_t, \rho_i)$$

Un filtre de Kalman avec lissage de Rauch-Tung-Striebel fournit l'estimation $\hat{x}_{t|T}$. Contrairement à une fusion à bruit constant, les covariances R_t^{opt} et R_t^{sar} varient ici acquisition par acquisition selon la géométrie et les conditions atmosphériques réelles, ce qui évite qu'une observation dégradée ne soit prise au même titre qu'une observation de bonne qualité.

3.2 Détection de régime spatio-temporelle

Le résidu $\varepsilon_i(t) = \hat{x}_{i,t|T} - g_c(t)$, calculé après retrait de la courbe de référence phénologique de la culture c, est modélisé par une chaîne de Markov cachée à quatre états $s_i(t) \in \{\text{sain, stress léger, stress sévère, reprise}\}$, avec des émissions gaussiennes $\varepsilon_i(t) \mid s_i(t) = j \sim \mathcal{N}(\mu_j, \sigma_j^2)$ ordonnées $\mu_{\text{stress sévère}} < \mu_{\text{stress léger}} < 0 \leq \mu_{\text{sain}}$, l'état de reprise étant caractérisé par un excès de croissance positif transitoire distinct de l'état sain.

État	Interprétation	Caractéristique
Sain	croissance conforme à la référence	moyenne proche de 0, variance faible
Stress léger	déficit hydrique modéré	moyenne négative, variance modérée
Stress sévère	déficit hydrique marqué	moyenne fortement négative, variance élevée
Reprise	rebond après stress	moyenne positive transitoire

Table 1: Régimes de la chaîne de Markov cachée

Un HMM indépendant par comté ignore la structure géographique du phénomène : une sécheresse affecte typiquement un ensemble de comtés contigus de façon corrélée. On couple donc les états à travers un champ de Markov caché spatial (*Hidden Markov Random Field*) : la probabilité jointe des états à la date t combine la transition temporelle propre à chaque comté et un terme d'interaction spatiale de type Potts sur le graphe de contiguïté des comtés,

$$P\big(s_1(t), \ldots, s_n(t) \mid s.(t-1)\big) \;\propto\; \prod_i P\big(s_i(t) \mid s_i(t-1)\big) \cdot \exp\Big(\eta \sum_{i \sim j} \mathbb{1}[s_i(t) = s_j(t)]\Big)$$

L'inférence exacte par forward-backward n'étant plus praticable en présence du couplage spatial, elle est approchée par une procédure EM à champ moyen ou par maximisation conditionnelle itérée. Le modèle standard suppose implicitement une durée de séjour géométrique dans chaque état, ce qui est peu réaliste pour un épisode de sécheresse ayant une durée caractéristique propre ; un modèle semi-markovien caché (HSMM), avec une distribution de durée explicite par état plutôt qu'implicite, est intégré à la couche de comparaison ci-dessous pour tester si cette hypothèse de durée coûte en pouvoir explicatif.

3.3 Couche de comparaison et test de stabilité

Le modèle spatio-temporel ci-dessus est confronté à quatre alternatives, utilisées comme test de robustesse plutôt que comme remplacement direct : un réseau récurrent (GRU ou LSTM) entraîné en séquence-vers-label pour classer l'état de stress à partir des séries brutes fusionnées NDVI/SAR, sans hypothèse de structure markovienne ; un transformeur spatio-temporel, où les poids de dépendance spatiale entre comtés sont appris par un mécanisme d'attention plutôt que fixés par un graphe de contiguïté, permettant de capturer des dépendances de plus longue portée liées à des systèmes météorologiques partagés ; un modèle de changement de régime bayésien complet, où l'état-espace et les paramètres de régime sont estimés conjointement avec incertitude a posteriori par MCMC particulaire, fournissant un intervalle de crédibilité sur chaque probabilité de régime plutôt qu'une seule estimation ponctuelle ; et le modèle semi-markovien caché mentionné plus haut.

Modèle	Rôle dans la comparaison
GRU / LSTM	robustesse à l'hypothèse markovienne
Transformeur spatio-temporel	robustesse à l'hypothèse de voisinage géographique fixe
Changement de régime bayésien	quantification de l'incertitude d'estimation
HSMM	robustesse à l'hypothèse de durée géométrique

Table 2: Modèles de la couche de comparaison

Chaque modèle est évalué sur les mêmes plis hors-échantillon selon deux critères : la vraisemblance ou l'exactitude de classification du régime par rapport à un label de stress dérivé des rendements USDA (section 5.1), et le ratio de Sharpe obtenu lorsque sa probabilité de régime remplace celle du modèle de référence dans l'indice agrégé. Le modèle retenu pour la production est le plus simple qui n'est pas dominé sur ces deux critères, plutôt que le plus complexe, pour des raisons d'interprétabilité et de vitesse de ré-estimation.

La stabilité des paramètres du modèle de référence est testée entre les sous-périodes 2016-2019 et 2020-2023 par un test du rapport de vraisemblance opposant une estimation poolée à une estimation séparée par sous-période,

$$LR = -2\Big[\ell_{\text{poolé}}(\hat{\theta}) - \ell_{\text{2016-19}}(\hat{\theta}_1) - \ell_{\text{2020-23}}(\hat{\theta}_2)\Big] \;\sim\; \chi^2_q \text{ sous } H_0 : \theta_1 = \theta_2$$

En cas de rejet, le calendrier de ré-estimation en production est resserré ; ce test est lui-même exécuté en continu sur fenêtre glissante pour déclencher une alerte de dérive de régime plutôt que d'être une vérification ponctuelle.

3.4 Indice d'offre agrégé

L'indice national de stress d'offre pondère les probabilités de régime par les parts de production et par un poids de sévérité par état,

$$S_t = \sum_i \pi_i \sum_j w_j \, \gamma_i^j(t)$$

Les poids w_j ne sont pas fixés arbitrairement : ils sont estimés par régression hors-échantillon des écarts de rendement USDA observés en fin de saison sur les probabilités de régime (section 5.1), ce qui ancre la sévérité relative des états dans une donnée agronomique réelle plutôt que dans un choix de modélisateur.

3.5 Signal de trading par cointégration

La cointégration entre S_t et le prix des futures F_t est testée par Engle-Granger ou Johansen. L'écart $z_t = F_t - \beta S_t - \alpha$ est modélisé par un processus d'Ornstein-Uhlenbeck $dz_t = -\kappa z_t \, dt + \sigma \, dW_t$,

calibré par régression sur sa version discrétisée. La position est proportionnelle à $z_t/\hat{\sigma}_z$, dimensionnée inversement à la demi-vie $\ln(2)/\kappa$, et exprimée de préférence comme un spread calendaire entre contrats de récolte consécutifs plutôt qu'en position directionnelle pure. La stabilité du coefficient β est vérifiée séparément par culture et par région, plutôt que supposée transférable telle quelle.

4 Validation externe et généralisation

4.1 Validation contre les rendements USDA

La probabilité de régime $\gamma_i(t)$ agrégée sur une saison est corrélée aux écarts de rendement county-level publiés par l'USDA NASS par rapport à leur tendance de long terme, en dehors de tout usage dans la construction de l'indice de trading, pour vérifier que le signal capture une réalité agronomique et non un artefact interne au pipeline. Cette validation est menée à l'échelle du comté, plus fine que l'agrégat national utilisé pour le trading, pour s'assurer que l'agrégation ne masque pas un signal défaillant au niveau local.

4.2 Robustesse multi-cultures et multi-régions

Le cadre est répliqué sur quatre cultures (maïs, soja, blé, coton) avec des courbes de référence et des fenêtres de sensibilité propres à chacune, et sur quatre bassins de production (États-Unis, Brésil, Argentine, Ukraine) avec des masques de parcelles et des calendriers agricoles adaptés à chaque hémisphère et à chaque cycle de culture. La performance et la significativité du signal sont rapportées séparément par culture et par région plutôt qu'en moyenne agrégée, pour éviter qu'une région ou une culture dominante ne masque l'échec du signal ailleurs.

5 Protocole de backtest institutionnel

5.1 Coûts d'exécution et roulement

Le glissement d'exécution est modélisé par une loi d'impact en racine carrée, $\text{Impact}(Q) = \lambda\, \sigma_{\text{jour}} \sqrt{Q/ADV}$, appliquée à l'entrée et à la sortie de chaque position, plutôt qu'un coût de transaction fixe. Le roulement des contrats suit le calendrier réel d'échéance CBOT et une fenêtre de roulement en TWAP sur plusieurs jours, avec un coût de roulement égal au spread bid-ask réalisé plus l'écart de portage entre échéances, plutôt qu'un contrat continu supposé sans coût.

5.2 Contraintes de capacité

La taille de position est plafonnée à une fraction de l'open interest et du volume moyen quotidien du contrat, ce qui borne explicitement la capacité de la stratégie et alimente directement la discussion de scalabilité (section 6).

5.3 Tests de stress

La stratégie est évaluée spécifiquement sur des épisodes extrêmes identifiés ex ante plutôt que sur la seule moyenne du backtest : la sécheresse du Midwest américain de 2012, la perturbation logistique de 2020, et le choc sur les céréales consécutif au conflit russo-ukrainien de 2022. La performance, le comportement de position et le drawdown sont rapportés séparément sur chacune de ces fenêtres.

5.4 Comparaison à un référentiel interne et stabilité du Sharpe

Les rendements de la stratégie sont régressés sur un référentiel de momentum de série temporelle et de portage sur matières premières, deux styles standards dans un fonds multi-stratégies, pour quantifier

la part de performance non expliquée par ces primes de risque usuelles et documenter le caractère réellement différenciant du signal plutôt que sa simple redondance avec un facteur déjà connu. Le ratio de Sharpe est suivi par fenêtre glissante plutôt que rapporté comme un chiffre unique, et corrigé par un ratio de Sharpe déflaté (Bailey & López de Prado, 2014) qui tient compte du nombre de variantes de spécification testées durant la recherche (choix de cultures, de régions, de fenêtres), pour éviter de surestimer la significativité d'un résultat issu d'une recherche à plusieurs essais.

5.5 Sensibilité à l'horizon

La cadence informationnelle du signal est bornée par le cycle de revisite satellite, de l'ordre de cinq à six jours : un rafraîchissement intrajournalier du signal lui-même n'a pas de sens et n'est pas testé. La sensibilité de la performance à la cadence d'exécution est en revanche testée entre un rééquilibrage hebdomadaire et mensuel de la position, pour documenter l'horizon de détention le mieux adapté au rythme réel d'arrivée de l'information.

6 Propriétés visées du signal

La scalabilité est bornée explicitement par les contraintes de capacité de la section 5.2, plutôt que supposée illimitée. La robustesse est établie par la réplication multi-cultures, multi-régions et par la comparaison aux modèles de la section 3.3, plutôt que par la seule performance sur un unique couple culture-région. La non-corrélation aux primes de risque usuelles est testée explicitement en section 5.4 par régression sur des référentiels standards. La persistance du signal est caractérisée par la demi-vie du processus d'Ornstein-Uhlenbeck et par une régression des rendements futurs sur le signal à horizons croissants, pour documenter la vitesse de décroissance du pouvoir prédictif. L'exploitabilité en production suppose enfin que la latence du pipeline de traitement (acquisition, correction atmosphérique et radiométrique, filtrage, inférence de régime) reste compatible avec la cadence de rééquilibrage retenue en section 5.5, ce qui doit être mesuré empiriquement sur l'infrastructure de traitement effectivement utilisée plutôt que supposé négligeable.

7 Limites et extensions

Le principal risque conceptuel reste celui du proxy : une probabilité de régime mesure un stress de végétation, pas directement un rendement agricole, même après validation externe. L'extension multi-région introduit une hétérogénéité de qualité et de disponibilité des données (masques de parcelles, calendriers agricoles, liquidité des instruments de couverture) qui rend la comparaison entre bassins de production délicate, et le cas ukrainien impose de traiter séparément la confusion entre abandon de parcelles lié au conflit et stress hydrique réel. La couche de comparaison en deep learning, si elle documente la robustesse du modèle de référence, a un coût de calcul et un besoin de données d'entraînement nettement supérieurs, ce qui justifie de la garder en couche de validation plutôt que de la déployer telle quelle en production. Une extension naturelle consiste à répliquer le cadre sur d'autres variables hydrologiques, comme le niveau de navigabilité du Rhin ou du Mississippi, pour un signal équivalent sur des actions industrielles exposées au fret fluvial.

Références indicatives

Small (2011) pour l'aplatissement du terrain SAR ; Oh, Sarabandi & Ulaby (1992) et Dubois, van Zyl & Engman (1995) pour les modèles semi-empiriques de rétrodiffusion ; documentation Sen2Cor et Roy et al. (2016) pour la correction atmosphérique et la normalisation BRDF ; Rabiner (1989) pour l'algorithme forward-backward et Baum-Welch ; Zhang, Brady & Smith (2001) pour les champs de Markov cachés spatiaux ; Yu (2010) pour les modèles semi-markoviens cachés ; Hamilton (1989) pour

les modèles à changement de régime ; Engle & Granger (1987) pour la cointégration ; Vasicek (1977) pour le processus d'Ornstein-Uhlenbeck ; Almgren & Chriss (2001) pour le modèle d'impact de marché ; Bailey & López de Prado (2014) pour le ratio de Sharpe déflaté ; Durbin & Koopman, *Time Series Analysis by State Space Methods*, pour le filtrage et lissage de Kalman.