

## Première contribution d'une approche de *downscaling* climatique statistique sur le bassin versant de la Seine

Emmanuel Garbolino<sup>1\*</sup>, Arman Ranjbari<sup>1,2</sup>, Jasha Oosterbaan<sup>1</sup>

<sup>1</sup> Mines Paris PSL/ISIGE

<sup>2</sup> Université de la Côte d'Azur/UMR ESPACE

\* emmanuel.garbolino@minesparis.psl.eu

### Résumé

*Cette étude présente une méthode de désagrégation climatique (ou downscaling) fondée sur la régression linéaire multiple (MLR), appliquée à une portion du bassin de la Seine, plus précisément au bassin versant de l'Orgeval et à la plaine alluviale de la Bassée. L'objectif est de transformer les données issues des modèles climatiques globaux (GCM), à faible résolution, en projections locales à très haute résolution spatiale (30 mètres), adaptées aux besoins d'analyse écologique et de planification territoriale.*

*La méthode repose sur l'utilisation de variables topographiques et géographiques (altitude, pente, latitude, distance à la mer) pour modéliser les températures et précipitations locales. Elle se distingue par sa simplicité de mise en œuvre et son caractère reproductible, reposant sur des données ouvertes et conformes au concept FAIR (Findable, Accessible, Interoperable, Reusable). L'implémentation sous Python à l'aide d'outils standards (scikit-learn, GDAL, NumPy) permet une application rapide et transparente, tout en facilitant la reproductibilité scientifique.*

*Les résultats sont très encourageants : la MLR montre une excellente précision pour les températures ( $R > 0,90$ ,  $RMSE < 0,3$  °C) et des performances satisfaisantes pour les précipitations ( $R > 0,70$ ,  $RMSE < 4$  mm). L'interprétation des coefficients de régression est aisée, permettant d'identifier clairement les facteurs physiques dominants tels que l'altitude ou la distance à la mer.*

*Au-delà de ses performances techniques, cette approche offre une solution robuste et rapide pour générer des surfaces climatiques locales cohérentes. Elle ouvre des perspectives prometteuses pour des approches prospectives en écologie, notamment dans l'étude des effets du changement climatique sur les forêts riveraines du bassin de la Seine. Par sa simplicité, son respect des standards FAIR et son adaptabilité, la MLR constitue un outil efficace et durable pour le développement de services climatiques opérationnels et la modélisation écologique à haute résolution. Ce rapport discute des limites de cette approche et présente également des voies technologiques d'amélioration du modèle.*

### Points clefs

- ✓ Méthode de désagrégation climatique par régression linéaire multiple (MLR) pour produire des données climatiques à très haute résolution (30 m).
- ✓ Simple à mettre en œuvre, transparente et reproductible, la MLR utilise des données conformes aux principes FAIR, facilitant l'interprétation physique des résultats.
- ✓ Les résultats sont prometteurs pour des approches prospectives en écologie et l'évaluation des impacts du changement climatique.

## Abstract

*This study presents a climate downscaling method based on Multiple Linear Regression (MLR), applied to a section of the Seine River basin, specifically the Orgeval catchment and the Bassée alluvial plain. The objective is to transform coarse-resolution data from Global Climate Models (GCMs) into local, very high-resolution projections (30 meters), suitable for ecological analysis and territorial planning.*

*The method relies on topographic and geographic variables (altitude, slope, latitude, distance to the sea) to model local temperature and precipitation patterns. It stands out for its ease of implementation and high reproducibility, being based on open data compliant with the FAIR principles (Findable, Accessible, Interoperable, Reusable). The implementation in Python using standard libraries (scikit-learn, GDAL, NumPy) enables a fast, transparent, and scientifically reproducible workflow.*

*The results are very encouraging: the MLR approach achieves excellent accuracy for temperature ( $R > 0.90$ ,  $RMSE < 0.3$  °C) and satisfactory performance for precipitation ( $R > 0.70$ ,  $RMSE < 4$  mm). The interpretation of regression coefficients is straightforward, allowing for clear identification of dominant physical factors such as altitude or distance to the sea.*

*Beyond its technical performance, this approach provides a robust and efficient solution for generating coherent local climate surfaces. It opens up promising perspectives for prospective ecological studies, particularly regarding the potential impacts of climate change on riparian forests in the Seine basin. Thanks to its simplicity, adherence to FAIR standards, and adaptability, the MLR method constitutes an effective and sustainable tool for developing operational climate services and high-resolution ecological modeling. The report also discusses the limitations of this approach and outlines technological pathways for model improvement.*

## Key points

- ✓ *Multiple Linear Regression (MLR)–based climate downscaling method used to produce very high-resolution (30 m) climate data.*
- ✓ *Simple to implement, transparent, and reproducible, the MLR approach relies on data compliant with FAIR principles, facilitating the physical interpretation of results.*
- ✓ *The results are promising for prospective ecological studies and for assessing the impacts of climate change.*

## **Introduction : pourquoi le *downscaling* climatique est-il nécessaire ?**

Le changement climatique constitue aujourd’hui l’un des enjeux majeurs de la recherche scientifique et de la gouvernance environnementale mondiale. Les travaux du Groupe d’experts intergouvernemental sur l’évolution du climat (GIEC, 2021) confirment que l’augmentation des températures moyennes, l’intensification des extrêmes climatiques et la modification des régimes hydrologiques exercent une pression croissante sur les écosystèmes, les ressources naturelles et les sociétés humaines. Dans ce contexte, la disponibilité de données climatiques fiables, cohérentes et spatialement détaillées est devenue un impératif pour l’évaluation des impacts locaux du changement climatique, la gestion durable des territoires et la mise en œuvre de politiques d’adaptation.

Or, les modèles climatiques globaux (*Global Climate Models* – GCM), bien qu’indispensables pour comprendre les mécanismes planétaires du climat et projeter ses évolutions futures, opèrent généralement à des résolutions spatiales comprises entre 50 et 200 km (Wilby & Wigley, 1997 ; GIEC, 2021). À cette échelle, les hétérogénéités locales – telles que la topographie ou la proximité des plans d’eau – sont lissées, rendant les sorties de ces modèles inadaptées aux analyses régionales, à la planification territoriale ou à l’évaluation écologique à fine échelle (Haylock *et al.*, 2008 ; Giorgi & Mearns, 1991). Cette inadéquation entre la résolution des modèles globaux et les besoins des décideurs locaux explique l’essor des méthodes de désagrégation climatique, ou *downscaling*, qui visent à traduire les informations climatiques issues des GCM en projections locales à haute résolution. Ces approches, devenues incontournables dans les études d’impact (Garbolino *et al.*, 2021), permettent de rendre les résultats des modèles climatiques opérationnels pour la gestion des territoires (Maraun, 2018 ; Teutschbein & Seibert, 2012).

L’enjeu principal du *downscaling* n’est donc pas uniquement scientifique : il est également opérationnel et stratégique. Les politiques publiques d’adaptation au changement climatique, les stratégies de gestion des ressources naturelles ou encore les plans d’urbanisme nécessitent une information climatique localisée, cohérente et comparable dans le temps (GIEC, 2021).

Les décideurs doivent disposer de cartographies climatiques fines pour anticiper les impacts des scénarios d’émissions sur les ressources en eau, la biodiversité, les risques naturels ou la productivité agricole (Reder *et al.*, 2025). Le *downscaling* statistique permet ainsi d’alimenter ces processus décisionnels en fournissant des indicateurs à l’échelle des bassins versants, des communes ou des écosystèmes sensibles.

De plus, les approches fondées sur des données ouvertes et conformes aux principes FAIR (*Findable, Accessible, Interoperable, Reusable* ; Wilkinson *et al.*, 2016) répondent aux exigences croissantes de transparence et de reproductibilité scientifique. Elles permettent la co-production de la connaissance entre chercheurs, institutions publiques et gestionnaires territoriaux, favorisant l’intégration de la science dans la décision (Baño-Medina *et al.*, 2021).

Nos recherches s’inscrivent dans la continuité des réflexions engagées au sein du PIREN-Seine depuis la phase 8, au cours de laquelle la question de la désagrégation climatique avait déjà fait l’objet d’investigations approfondies. En particulier, les travaux de Boé et Mass, présentés en 2021, ont introduit une méthode de désagrégation « *hybride statistico-dynamique* », fondée sur le développement d’un modèle climatique régional via une approche par analogues construits. Cette méthode visait notamment à surmonter les limites de la stationnarité inhérente aux approches strictement statistiques. La méthode proposée par Boé et Mass a été appliquée à l’échelle européenne, à partir de données climatiques de résolution spatiale d’environ 150km, pour aboutir à des résultats d’une résolution spatiale de 12km.

Dans le cadre du présent rapport, nous prolongeons ces réflexions en adoptant une démarche de désagrégation explicitement statistique, appliquée à très haute résolution spatiale (30m). Si notre approche diffère de celle développée en 2021 par son positionnement méthodologique, elle s’inscrit dans le même objectif général : produire des données climatiques régionalisées robustes, adaptées aux besoins d’analyse environnementale et hydrologique du bassin de la Seine. Elle constitue ainsi une nouvelle étape dans l’effort collectif de développement et de comparaison de méthodes de *downscaling* au sein du PIREN-Seine.

Dans ce contexte, la méthode MLR se distingue par sa simplicité d'implémentation, sa compatibilité avec les standards de données ouverts et sa capacité à produire des résultats directement exploitables dans les systèmes d'aide à la décision environnementale. Elle constitue ainsi un pont entre la recherche académique et la planification opérationnelle (Maraun, 2018). La présente étude s'inscrit dans le cadre du bassin de la Seine, et plus particulièrement sur deux territoires emblématiques : l'Orgeval et la Bassée. Ces territoires constituent des laboratoires hydrologiques et écologiques de référence dans les programmes du PIREN-Seine, caractérisés par des interactions complexes entre dynamiques fluviales, forêts riveraines et pressions anthropiques.

La modélisation des variables climatiques à très haute résolution spatiale (30 mètres) est ici envisagée pour simuler les processus écologiques locaux, notamment la réponse des écosystèmes ripariens face à l'évolution du climat. Cet aspect sera présenté dans un rapport ultérieur. Ainsi, la désagrégation climatique par MLR appliquée au bassin de la Seine illustre la manière dont une méthodologie scientifique rigoureuse et accessible peut être intégrée dans une démarche de science appliquée au service des territoires (Reder *et al.*, 2025).

## 1. Méthodologie

### 1.1. Données et sources

La MLR a pour objectif de produire des grilles climatiques fines à partir de données globales (*WorldClim*, Copernicus DEM), en intégrant les déterminants topographiques, géographiques et physiques spécifiques du bassin.

Cette approche répond à plusieurs besoins :

- Fournir des données d'entrée adaptées aux modèles écologiques et hydrologiques régionaux ;
- Renforcer la capacité d'analyse prospective sur les impacts du changement climatique dans les territoires étudiés ;
- Soutenir la prise de décision territoriale en matière d'aménagement, de gestion de l'eau ou de conservation de la biodiversité.

La figure suivante présente les différentes étapes de la méthodologie, depuis l'acquisition des données climatiques et des variables explicatives jusqu'à la production des données climatiques à très haute résolution spatiale.

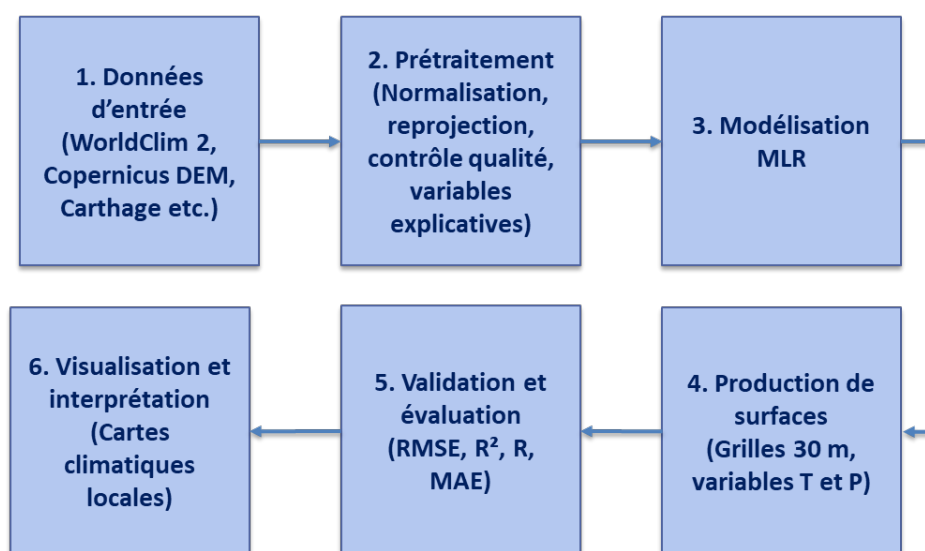


Figure 1. Ce flux de travail illustre les étapes principales du modèle MLR, depuis la collecte et le prétraitement des données jusqu'à la production et la validation des surfaces climatiques à 30 m de résolution.

Les données climatiques utilisées proviennent de la base *WorldClim 2* (Fick & Hijmans, 2017), offrant des surfaces climatiques interpolées à une résolution spatiale d'environ 1 km<sup>2</sup>, établies à partir de plus de 4.000 stations météorologiques. La production de ces données repose sur les sorties de modèles issues du Coupled Model Intercomparison Project Phase 6 (CMIP6, Eyring *et al.*, 2016) qui est associé au World Climate Research Programme (WCRP). La gouvernance du WCRP associe notamment le World Meteorological Organization (WMO), l'International Science Council (ISC, anciennement ICSU), et l'Intergovernmental Oceanographic Commission (IOC) de l'United Nations Educational, Scientific and Cultural Organization (UNESCO). Les variables climatiques exploitées incluent les températures minimales et maximales moyennes ainsi que les précipitations totales mensuelles, sur la période de référence 1970–2000, conformément aux standards du GIEC (2021). Pour les scénarios futurs, nous avons travaillé sur les scénarios SSP245 et SSP585 et les périodes suivantes : 2021-2040, 2041-2060, 2061-2080 et 2081-2100). Nous avons également sélectionné les modèles suivants pour tester leur intérêt dans le cadre du *downscaling* climatique :

- CMCC-ESM2 (Centro Euro-Mediterraneo sui Cambiamenti Climatici, Italy) ;
- EC-Earth3-Veg (EC-Earth Consortium – Sweden, Netherlands, Ireland, etc.) ;

- IPSL-CM6A-LR (Institut Pierre-Simon Laplace, France) ;
- MIROC6 (Tokyo Univ., *National Institute for Environmental Studies* etc., Japan) ;
- MPI-ESM1-2-HR (*Max Planck Institute for Meteorology*, Germany).

L'ensemble de ces données climatiques pour l'actuel et le futur constitue un jeu de données au format « *raster* » (TIF) de presque 2 To, nécessitant le développement en propre d'outils informatiques optimisés pour le traitement de données volumineuses.

Ces données globales ont été complétées par des couches spatiales de prédicteurs (variables explicatives) à très haute résolution spatiale (résolution spatiale de 30 m) issues du Copernicus DEM (*Digital Elevation Model*). Les variables dérivées comprennent l'altitude, la pente, la latitude, la longitude, la distance à la mer et les distances aux cours d'eau et lacs. Les données concernant le tracé des lacs et des cours d'eau proviennent de la base de données CarTHAgE® (CARtographie THématique des AGences de l'eau et du ministère chargé de l'environnement) produite par le SANDRE (Service d'Administration Nationale des Données et Référentiels sur l'Eau). Ces facteurs topographiques, géographiques et physiques sont reconnus pour influencer significativement la variabilité spatiale des conditions climatiques locales (Gutiérrez *et al.*, 2019).

Chaque variable a été reprojctée et harmonisée dans un système de coordonnées commun (EPSG:4326), puis normalisée afin d'assurer la comparabilité entre les grilles spatiales. Un contrôle qualité a été réalisé pour éliminer les valeurs aberrantes ou incohérentes (ex. pixels nuls, erreurs de topographie locale).

## 1.2. Variables explicatives et dépendantes

Les variables dépendantes (Y) correspondent aux paramètres climatiques à estimer à haute résolution (températures et précipitations locales). Les variables explicatives (X) sont constituées des prédicteurs topographiques, physiques et géographiques mentionnés précédemment.

La régression linéaire multiple (MLR) vise à modéliser la relation empirique entre ces variables selon la formule suivante :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

Où :

$\beta_0$  est l'ordonnée à l'origine ;

$\beta_i$  les coefficients de régression représentant l'influence respective de chaque prédicteur ;

$\varepsilon$  l'erreur résiduelle.

Cette formulation repose sur l'hypothèse que les facteurs géographiques exercent une influence additive et proportionnelle sur la variable climatique cible (voir paragraphes 1.5 et 3.3). L'objectif est d'identifier la contribution relative de chaque variable topographique à la variabilité climatique régionale (Wilby & Wigley, 1997).

L'ensemble des analyses a été conduit sous Python, en s'appuyant sur les bibliothèques *scikit-learn* pour la modélisation statistique, NumPy pour les opérations matricielles, et GDAL/Rasterio pour la gestion des données géospatiales.

Les résultats ont été exportés sous la forme de rasters de 30 m de résolution spatiale, permettant une visualisation directe des gradients climatiques locaux.

## 1.3. Validation statistique et métriques

Les performances du modèle ont été évaluées à l'aide de plusieurs indicateurs statistiques complémentaires :

- Coefficient de corrélation R : il permet d'évaluer la qualité de la relation linéaire entre les valeurs observées et simulées. Il varie entre -1 et +1, où des valeurs proches de +1 indiquent une corrélation positive forte, tandis que des valeurs proches de 0 traduisent une absence de relation linéaire et les

valeurs proches de -1 indiquent une corrélation négative qui varie de manière proportionnelle mais opposée. Cet indicateur informe sur la cohérence du modèle dans la reproduction des variations des données observées, mais ne renseigne pas sur l'amplitude des écarts entre observations et simulations ;

- Le coefficient de détermination  $R^2$  : il mesure la proportion de la variance des observations qui est expliquée par le modèle. Il s'agit du carré du coefficient de corrélation linéaire  $R$ , et sa valeur varie entre 0 et 1.  $R^2 = 1$  indique que le modèle reproduit parfaitement les observations, sans erreur résiduelle et  $R^2 = 0$  signifie que le modèle n'explique aucune part de la variance observée, les prédictions n'étant pas meilleures qu'une estimation basée sur la moyenne des observations. Le  $R^2$  permet d'évaluer la qualité de l'ajustement global du modèle : plus sa valeur est élevée, plus les simulations suivent fidèlement la variabilité des observations ;
- Erreur quadratique moyenne (RMSE) : elle quantifie la dispersion moyenne des erreurs de prédiction. Elle est calculée comme la racine carrée de la moyenne des carrés des écarts entre valeurs observées et simulées. Une faible valeur de RMSE traduit une bonne performance globale du modèle. Cet indicateur est toutefois sensible aux erreurs extrêmes, ce qui en fait un bon outil pour détecter la présence d'écarts ponctuels importants entre simulations et observations ;
- Erreur absolue moyenne (MAE) : elle évalue la magnitude moyenne des erreurs, indépendamment de leur signe. Contrairement au RMSE, la MAE est moins sensible aux valeurs extrêmes. Elle exprime la précision moyenne du modèle et fournit une mesure directe de l'erreur typique entre simulation et observation ;
- Analyse des résidus spatiaux : elle consiste à cartographier les différences entre valeurs simulées et observées afin d'identifier des biais systématiques liés à des facteurs géographiques ou environnementaux. Cette approche permet notamment de localiser des zones de sous ou de surestimation persistantes, d'évaluer la robustesse du modèle dans des contextes spécifiques et d'orienter d'éventuelles améliorations du modèle (paramétrisation, forçages, résolution spatiale, etc.).

Ces métriques ont été calculées séparément pour chaque variable et chaque zone d'étude, permettant une évaluation spatiale fine de la performance du modèle. Les analyses ont confirmé des corrélations élevées ( $R > 0,95$ ) pour les températures et des corrélations légèrement moins marquées ( $R \approx 0,80-0,90$ ) pour les précipitations.

#### **1.4. Zones d'étude**

La zone d'application de la méthodologie couvre les départements de l'Aube (10), de la Marne (51), de la Seine-et-Marne (77) et de l'Yonne (89) en raison du chevauchement des futures zones d'études des effets du changement climatique sur les structures végétales de la Bassée et de l'Orgeval.

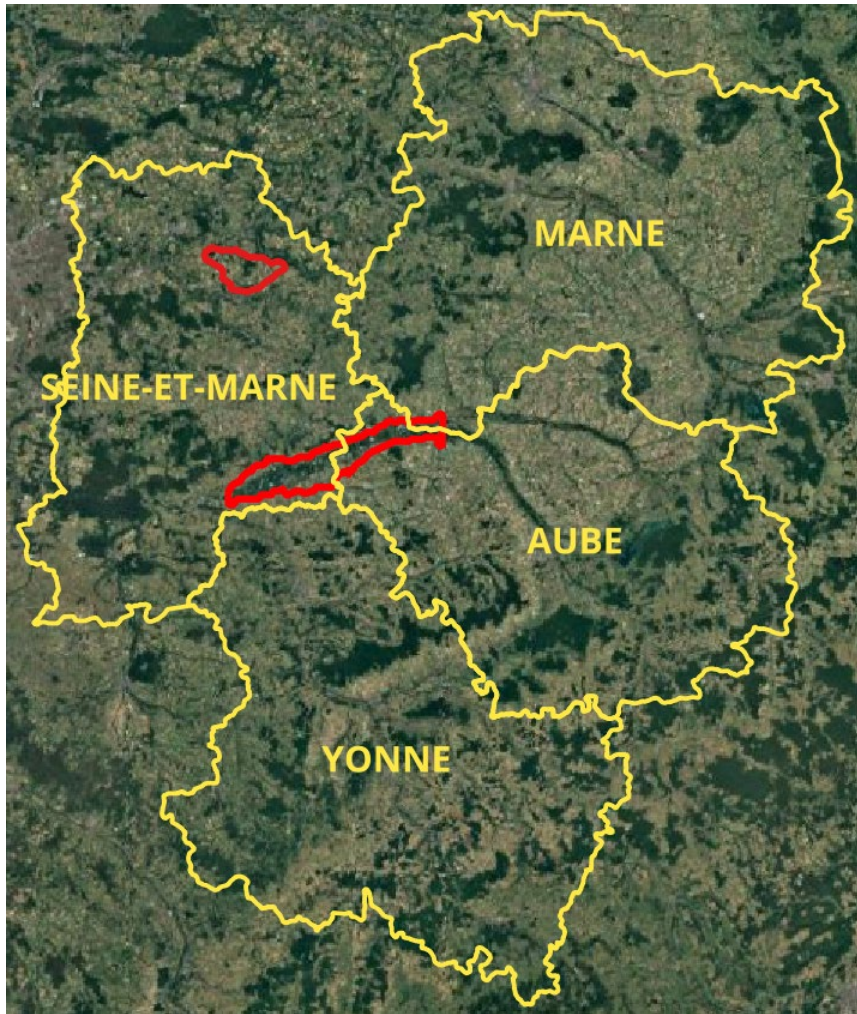


Figure 2. Départements sélectionnés pour l'application du *downscaling* climatique concernant les deux zones d'étude.

La MLR a donc été testée sur ces zones du bassin de la Seine qui couvrent des contextes géographiques variés :

- Régions de plaine (vallée de la Seine, Bassée) : climats tempérés à faible gradient altitudinal.
- Régions de transition (Orgeval, Brie) : zones soumises à des influences mixtes océanique et continentale.

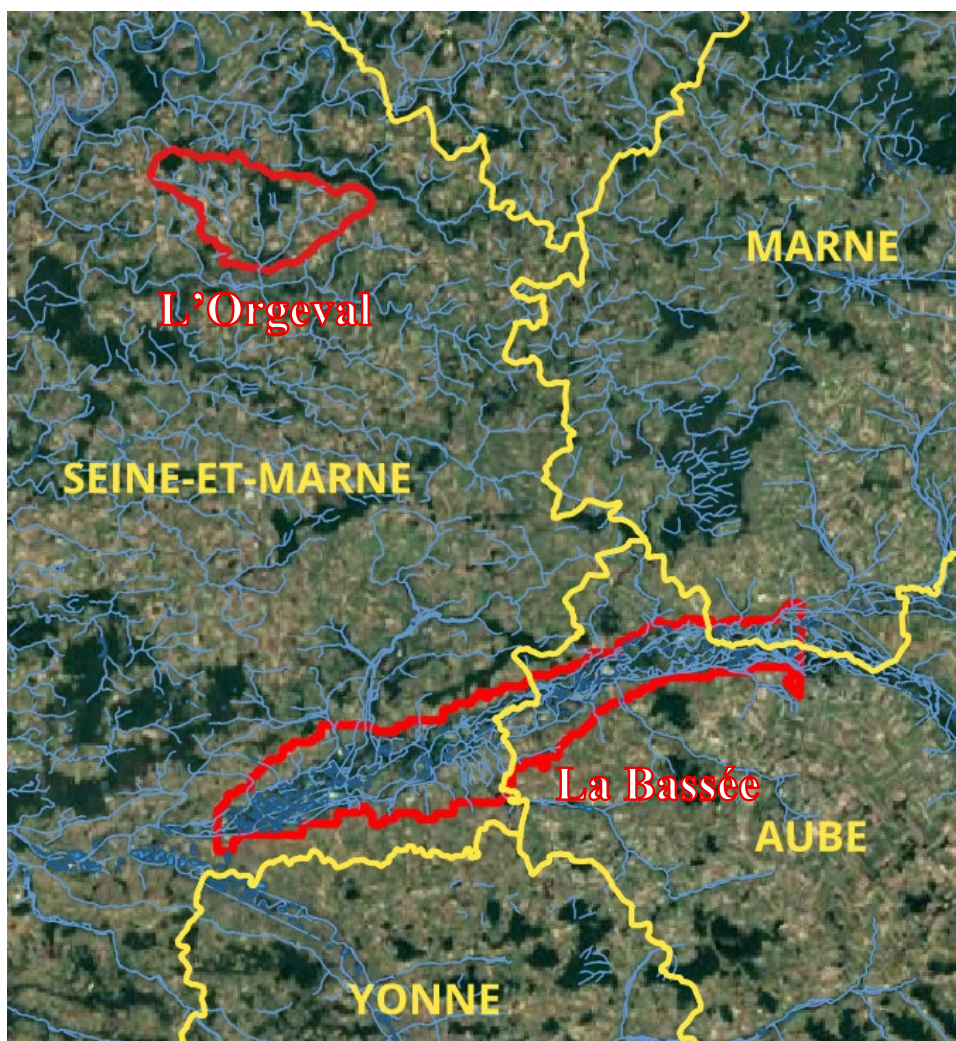


Figure 3. Zones d'étude pour l'application du *downscaling* climatique pour l'étude de la dynamique de la végétation.

Cette diversité spatiale garantit une évaluation robuste du modèle sur des gradients géographiques représentatifs du bassin.

### 1.5. Limites méthodologiques

La MLR repose sur l'hypothèse de linéarité et de stationnarité des relations entre les prédicteurs et la variable dépendante, ce qui limite sa capacité à reproduire des processus non linéaires tels que la convection ou les effets microclimatiques complexes qui peuvent influencer l'expression locale des variables climatiques.

De plus, les incertitudes liées aux données d'entrée (erreurs de mesure, interpolation des stations météorologiques, biais de MNT) peuvent affecter la précision régionale du modèle.

Enfin, sous conditions de changement climatique, la structure statistique reliant les variables peut évoluer, compromettant la transférabilité temporelle du modèle (Cannon, 2011 ; GIEC, 2021).

Ces limites justifient l'exploration de modèles hybrides combinant MLR et approches d'apprentissage automatique pour capturer des dépendances non linéaires tout en conservant la transparence et la robustesse des modèles linéaires (Baño-Medina *et al.*, 2021 ; Wang & Tian, 2022). Ces aspects constituent aujourd'hui des voies d'amélioration du modèle présenté dans ce rapport.

## 2. Résultats

### 2.1. Contribution des variables explicatives pour le *downscaling*

L’application de la méthodologie sur les variables explicatives a permis de calculer les statistiques  $R^2$ ,  $R$ , RMSE et MAE au regard des variables climatiques (Prec, Tmin, Tmax) de la période de référence (1970-2000). Pour rappel, les températures utilisées sont exprimées en degrés Celsius et concernant les moyennes mensuelles de normales climatiques (période de 30 ans) pour les températures de la nuit (Tmin) et les températures du jour (Tmax). Les précipitations sont exprimées en mm et concernent les cumuls mensuels moyens de normales climatiques.

Les tableaux suivants présentent les résultats de ces statistiques, sachant qu’il s’agit de moyennes calculées à chaque fois sur 12 mois pour chaque variable climatique. Ces résultats concernent respectivement les départements de l’Aube (10), de la Marne (51), de la Seine-et-Marne (77) et de l’Yonne (89).

Tableau 1. Statistiques relatives aux variables explicatives pour le département 10 (Aube)

Département	Variables climatiques	Variables explicatives	R <sup>2</sup>	R	RMSE	MAE
10	Prec	DEM_m	0.55	0.72	3.95	3.30
10	Prec	Lakes_km	0.03	0.13	6.14	5.07
10	Prec	Latitude	0.45	-0.61	4.36	3.69
10	Prec	Longitude	0.48	0.67	4.22	3.44
10	Prec	Rivers_km	0.03	-0.08	6.11	4.87
10	Prec	Sea_km	0.75	0.83	2.34	1.84
10	Prec	Slope_deg	0.20	0.41	5.46	4.42
10	Tmax	DEM_m	0.82	-0.90	0.21	0.17
10	Tmax	Lakes_km	0.11	-0.31	0.48	0.40
10	Tmax	Latitude	0.41	0.64	0.39	0.31
10	Tmax	Longitude	0.14	-0.35	0.47	0.39
10	Tmax	Rivers_km	0.01	-0.07	0.51	0.42
10	Tmax	Sea_km	0.39	-0.62	0.39	0.32
10	Tmax	Slope_deg	0.41	-0.64	0.39	0.30
10	Tmin	DEM_m	0.45	-0.58	0.16	0.13
10	Tmin	Lakes_km	0.25	-0.49	0.20	0.16
10	Tmin	Latitude	0.21	0.31	0.20	0.16
10	Tmin	Longitude	0.12	-0.09	0.22	0.17
10	Tmin	Rivers_km	0.16	-0.37	0.21	0.17
10	Tmin	Sea_km	0.24	-0.24	0.20	0.16
10	Tmin	Slope_deg	0.12	-0.25	0.22	0.17

Pour le département de l’Aube (10), les précipitations présentent des relations contrastées avec les variables géographiques :

- Altitude (DEM\_m) montre une corrélation positive modérée à forte ( $R = 0.72$  ;  $R^2 = 0.55$ ). Cela traduit une augmentation des précipitations avec l’altitude, phénomène classique lié à l’effet orographique. Les erreurs (RMSE = 3.95 mm ; MAE = 3.30 mm) confirment une dispersion modérée des écarts entre valeurs observées et estimées, ce qui atteste d’une très bonne performance du modèle par rapport à cette variable climatique.
- Distance à la mer (Sea\_km) est le meilleur prédicteur ( $R^2 = 0.75$  ;  $R = 0.83$ ). La corrélation positive indique que les précipitations diminuent à mesure que l’on s’éloigne de la mer (ou inversement selon le signe de référence utilisé). Cette variable explique donc une grande partie de la variabilité spatiale des précipitations dans le département.
- Latitude et longitude présentent des corrélations modérées ( $R \approx \pm 0.6$ ), suggérant des gradients spatiaux

nord-sud et est-ouest des précipitations, probablement liés à des effets régionaux de circulation atmosphérique et à la topographie du territoire.

- Proximité aux rivières (Rivers\_km) et aux lacs (Lakes\_km) n’apportent que très peu d’information ( $R^2 \approx 0.03$ ), ce qui traduit une influence marginale de ces éléments sur les précipitations à l’échelle départementale.
- Pente (Slope\_deg) montre une relation faible ( $R^2 = 0.20$ ), probablement associée à des micro-effets topographiques locaux sans tendance marquée.

Pour les températures diurnes (Tmax), les relations sont ici particulièrement marquées avec le relief :

- Altitude (DEM\_m) présente une corrélation négative très forte ( $R = -0.90$  ;  $R^2 = 0.82$ ), traduisant la diminution de la température avec l’altitude (gradient thermique vertical). Les faibles erreurs (RMSE =  $0.21^\circ\text{C}$  ; MAE =  $0.17^\circ\text{C}$ ) confirment la robustesse de cette relation.
- Pente (Slope\_deg) et distance à la mer (Sea\_km) montrent des corrélations négatives modérées à fortes ( $R \approx -0.6$ ), suggérant que les zones à fort relief ou éloignées des influences océaniques tendent à présenter des températures maximales plus faibles.
- Les autres variables (lacs, rivières, longitude) ont des  $R^2$  faibles ( $< 0.15$ ), ce qui indique une influence marginale sur la variabilité des températures maximales.

Les températures nocturnes (Tmin) présentent des relations plus diffuses et généralement plus faibles :

- Altitude (DEM\_m) conserve une corrélation négative modérée ( $R = -0.58$  ;  $R^2 = 0.45$ ), reflétant un refroidissement nocturne plus marqué en altitude.
- Proximité des plans d’eau (Lakes\_km,  $R = -0.49$  ;  $R^2 = 0.25$ ) exerce une influence non négligeable : les zones proches des lacs tendent à bénéficier d’un effet adoucissant sur les températures minimales, probablement lié à l’inertie thermique des masses d’eau.
- Les autres variables (latitude, longitude, rivières, pente, distance à la mer) présentent des corrélations faibles ( $|R| < 0.4$ ), traduisant une influence limitée et hétérogène.
- L’effet maritime est perceptible mais modéré ( $R^2 = 0.24$ ), indiquant une atténuation partielle des extrêmes nocturnes à proximité de la mer.

Les statistiques montrent ainsi que, pour les précipitations, la distance à la mer et l’altitude sont les facteurs dominants de la variabilité spatiale, tandis que les autres variables n’apportent qu’une contribution secondaire. Pour la température maximale, elle est principalement contrôlée par l’altitude, suivie par des effets secondaires de la pente et de la distance à la mer. Enfin, les températures minimales dépendent principalement de l’altitude et de la proximité des lacs, avec des influences secondaires de la mer et du relief local. Ces résultats semblent montrer la combinaison d’un effet orographique marqué et d’une influence maritime résiduelle sur la partie occidentale du département apportant une structure climatique complexe liée à la transition entre le plateau de Langres et les plaines champenoises.

Tableau 2. Statistiques relatives aux variables explicatives pour le département 51 (Marne)

Département	Variables climatiques	Variables explicatives	R2	R	RMSE	MAE
51	Prec	DEM_m	0.10	0.31	5.27	4.49
51	Prec	Lakes_km	0.05	-0.17	5.43	4.58
51	Prec	Latitude	0.07	-0.07	5.34	4.45
51	Prec	Longitude	0.52	0.71	3.76	3.11
51	Prec	Rivers_km	0.01	-0.07	5.52	4.60
51	Prec	Sea_km	0.32	0.54	4.57	3.77
51	Prec	Slope_deg	0.03	0.13	5.47	4.56
51	Tmax	DEM_m	0.63	-0.79	0.30	0.25
51	Tmax	Lakes_km	0.03	0.15	0.48	0.38

51	Tmax	Latitude	0.15	-0.38	0.44	0.36
51	Tmax	Longitude	0.13	0.34	0.45	0.36
51	Tmax	Rivers_km	0.01	0.07	0.48	0.39
51	Tmax	Sea_km	0.35	0.58	0.39	0.31
51	Tmax	Slope_deg	0.27	-0.52	0.41	0.32
51	Tmin	DEM_m	0.33	-0.56	0.19	0.15
51	Tmin	Lakes_km	0.12	-0.33	0.22	0.17
51	Tmin	Latitude	0.07	-0.23	0.22	0.18
51	Tmin	Longitude	0.22	0.41	0.20	0.16
51	Tmin	Rivers_km	0.08	-0.27	0.22	0.17
51	Tmin	Sea_km	0.29	0.49	0.19	0.15
51	Tmin	Slope_deg	0.02	-0.10	0.23	0.18

Pour le département de la Marne (51), les précipitations sont surtout liées à la longitude ( $R = 0.71$  et  $R^2 = 0.52$ ) et à la distance à la mer ( $R = 0.54$  et  $R^2 = 0.32$ ), avec un gradient est-ouest assez net. Les températures diurnes dépendent fortement de la topographie ( $R = -0.79$  et  $R^2 = 0.63$ ) et les températures nocturnes sont légèrement influencées par la topographie ( $R = -0.56$  et  $R^2 = 0.33$ ) et la distance à la mer ( $R = 0.49$  et  $R^2 = 0.29$ ). Il semble que le relief peu contrasté de ce territoire réduise l'effet orographique sur les variables climatiques : la variabilité est surtout spatiale (longitudinale), liée à la continentalité croissante vers l'est. Le département de la Marne présente ainsi des relations plus faibles que le département de l'Aube (10), confirmant une homogénéité climatique plus marquée et une influence topographique atténuée.

Tableau 3. Statistiques relatives aux variables explicatives pour le département 77 (Seine-et-Marne)

Département	Variables climatiques	Variables explicatives	R2	R	RMSE	MAE
77	Prec	DEM_m	0.30	0.43	1.56	1.23
77	Prec	Lakes_km	0.07	0.22	1.82	1.45
77	Prec	Latitude	0.12	0.00	1.77	1.40
77	Prec	Longitude	0.40	-0.04	1.33	1.04
77	Prec	Rivers_km	0.11	0.19	1.78	1.40
77	Prec	Sea_km	0.13	-0.03	1.74	1.40
77	Prec	Slope_deg	0.01	-0.09	1.87	1.49
77	Tmax	DEM_m	0.47	-0.65	0.26	0.21
77	Tmax	Lakes_km	0.05	-0.17	0.34	0.27
77	Tmax	Latitude	0.33	-0.56	0.28	0.22
77	Tmax	Longitude	0.20	-0.34	0.31	0.24
77	Tmax	Rivers_km	0.07	0.25	0.33	0.26
77	Tmax	Sea_km	0.19	0.33	0.31	0.25
77	Tmax	Slope_deg	0.02	-0.09	0.34	0.28
77	Tmin	DEM_m	0.63	-0.80	0.18	0.14
77	Tmin	Lakes_km	0.17	-0.41	0.27	0.22
77	Tmin	Latitude	0.10	-0.23	0.28	0.22
77	Tmin	Longitude	0.56	-0.75	0.20	0.15
77	Tmin	Rivers_km	0.01	0.02	0.29	0.24
77	Tmin	Sea_km	0.09	-0.21	0.28	0.23
77	Tmin	Slope_deg	0.01	0.12	0.29	0.24

Pour le département de la Seine-et-Marne (77), les relations sont globalement faibles à modérées avec des

précipitations peu structurées ( $R^2 \leq 0.4$ , souvent proches de 0.1) pour l’ensemble des variables explicatives. Pour les températures diurnes, les corrélations sont modérées avec le relief ( $R = -0.65$  et  $R^2 = 0.47$ ) et la Latitude ( $R = -0.56$  et  $R^2 = 0.33$ ). Les températures nocturnes montrent une forte corrélation avec la topographie ( $R = -0.80$  et  $R^2 = 0.63$ ) et la Longitude ( $R = -0.75$  et  $R^2 = 0.56$ ). Le département ayant globalement une topographie très peu marquée, car étant situé dans le bassin parisien, les effets du relief sont peu marqués. Les gradients latitudinaux et longitudinaux dominant, traduisant des différences de rayonnement et d’influences continentales internes.

Tableau 4. Statistiques relatives aux variables explicatives pour le département 89 (Yonne)

Département	Variables climatiques	Variables explicatives	R2	R	RMSE	MAE
89	Prec	DEM_m	0.58	0.76	3.73	3.02
89	Prec	Lakes_km	0.09	0.27	5.54	4.40
89	Prec	Latitude	0.46	-0.64	4.18	3.27
89	Prec	Longitude	0.47	0.64	3.86	3.18
89	Prec	Rivers_km	0.05	0.22	5.67	4.60
89	Prec	Sea_km	0.65	0.77	3.05	2.45
89	Prec	Slope_deg	0.04	0.20	5.68	4.61
89	Tmax	DEM_m	0.79	-0.88	0.21	0.17
89	Tmax	Lakes_km	0.06	-0.19	0.45	0.34
89	Tmax	Latitude	0.17	0.40	0.42	0.33
89	Tmax	Longitude	0.16	-0.36	0.42	0.33
89	Tmax	Rivers_km	0.06	-0.22	0.45	0.34
89	Tmax	Sea_km	0.23	-0.46	0.41	0.32
89	Tmax	Slope_deg	0.06	-0.25	0.45	0.34
89	Tmin	DEM_m	0.63	-0.77	0.14	0.11
89	Tmin	Lakes_km	0.11	-0.33	0.24	0.19
89	Tmin	Latitude	0.20	0.35	0.23	0.18
89	Tmin	Longitude	0.22	-0.36	0.22	0.18
89	Tmin	Rivers_km	0.09	-0.30	0.24	0.19
89	Tmin	Sea_km	0.29	-0.42	0.21	0.16
89	Tmin	Slope_deg	0.03	-0.11	0.25	0.20

Enfin, pour le département de l’Yonne qui a été intégré à cette étude en raison de sa proximité immédiate avec une partie du territoire de la vallée de La Bassée (moins de 4 km), on remarque que les précipitations sont corrélées à la distance à la mer ( $R = 0.77$  et  $R^2 = 0.65$ ) et à la topographie ( $R = 0.76$  et  $R^2 = 0.58$ ), ces valeurs étant très semblables à celles de l’Aube (10). Les températures diurnes montrent une forte relation avec le relief ( $R = -0.88$  et  $R^2 = 0.79$ ) et les températures nocturnes montrent quant à elles une dépendance marquée au relief ( $R = -0.77$  et  $R^2 = 0.63$ ) et à la distance à la mer ( $R = -0.42$  et  $R^2 = 0.29$ ). Ces résultats semblent converger vers une influence conjointe de la topographie (plateaux du Morvan, collines de l’Auxerrois) et d’une distance à la mer importante, entraînant un effet de continentalité marqué. Ils montrent une structure quasi identique avec le département de l’Aube (10), ce qui souligne que ces deux départements partagent des reliefs intermédiaires et des positions similaires dans la transition entre bassins atlantiques et continentaux.

Ainsi, pour les quatre départements, la structure générale des relations entre les variables climatiques et explicatives est globalement cohérente :

- Les températures maximales et minimales sont fortement corrélées à l’altitude (DEM\_m), avec un signe négatif, traduisant le gradient thermique vertical classique.
- Les précipitations sont principalement influencées par la distance à la mer (Sea\_km) et dans une moindre mesure par la longitude, indiquant des effets de continentalité et un gradient ouest–est.

- Les autres variables (lacs, rivières, pente) jouent un rôle secondaire ou local, souvent avec des coefficients de corrélation faibles ( $|R| < 0.4$ ).

Toutefois, l’intensité des relations varie d’un département à l’autre selon la morphologie du relief, la position géographique et la proximité des influences océaniques.

## 2.2. Performances de la MLR sur les températures et les précipitations pour l’actuel

Le tableau suivant présente les résultats de synthèse de l’application du *downscaling* climatique sur les quatre départements pour la période actuelle.

Tableau 5. Statistiques moyennes du *downscaling* climatique sur les 4 départements pour l’actuel.

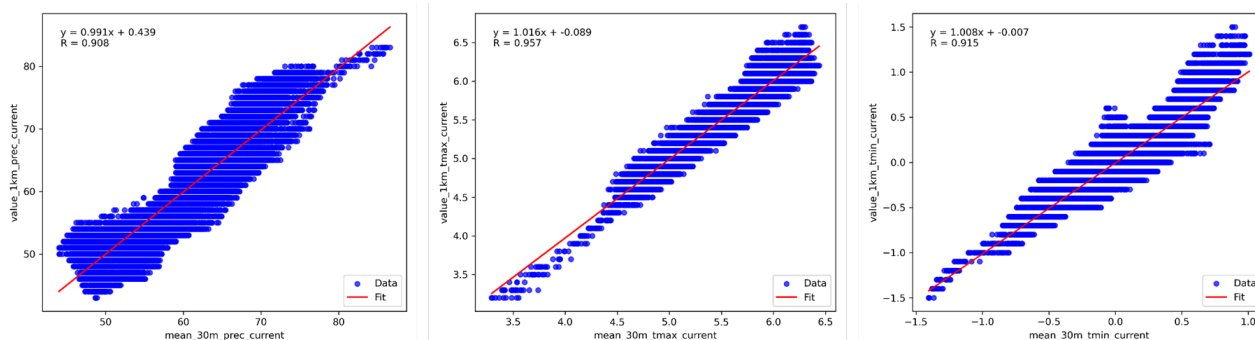
Variables climatiques	R <sup>2</sup>	R	RMSE	MAE
Précipitations	0.73	0.85	3.26	2.61
Tmax	0.83	0.91	0.21	0.16
Tmin	0.73	0.85	0.16	0.13

Les résultats obtenus pour la régression linéaire multiple (MLR) appliquée aux températures minimales et maximales sur les départements de l’Aube (10), de la Marne (51), de la Seine-et-Marne (77) et de l’Yonne (89) montrent des corrélations très élevées, avec des coefficients de détermination R compris entre 0,85 (Tmin) et 0,91 (Tmax). Les erreurs moyennes quadratiques (RMSE) varient entre 0,16 °C et 0,21 °C, ce qui traduit une excellente capacité du modèle à reproduire les gradients thermiques régionaux, y compris dans les zones de plaine et dans les vallées de la Seine et de l’Yonne, où la topographie reste modérée.

Ces résultats rejoignent les conclusions de Cannon (2011) et Teutschbein & Seibert (2012) sur la capacité des modèles linéaires à décrire efficacement les gradients thermiques moyens à l’échelle régionale.

Les précipitations constituent la variable la plus délicate à modéliser statistiquement en raison de leur forte variabilité spatiale et temporelle, amplifiée par les contrastes entre plateaux et vallées du bassin de la Seine. Sur les quatre départements étudiés, la MLR obtient des coefficients de corrélation R moyens de 0.73 avec des RMSE mensuels moyens de 3.26 mm, ce qui traduit également un très bon niveau de prédiction des précipitations dans ces territoires. Ces résultats confirment la sensibilité de la MLR à la complexité orographique et à la densité du réseau d’observation (Baño-Medina *et al.*, 2021).

Enfin, nous avons aussi procédé à la comparaison entre les valeurs des variables climatiques de chaque cellule de 1 km et la moyenne de ces variables pour les cellules de 30 m correspondantes qui se chevauchent. Cette méthode permet d’évaluer dans quelle mesure la réduction d’échelle a préservé les modèles de données d’origine. La figure suivante est un exemple pour les trois variables climatiques en janvier.



*Figure 4. Validation spatiale de la prédiction des variables climatiques par la méthode de downscaling climatique au regard des données climatiques de référence pour l’actuel.*

Chaque point bleu correspond à une paire de valeurs (moyenne 30 m, valeur 1 km), et la ligne rouge représente la régression linéaire ajustée. Les résultats montrent une forte cohérence entre les deux jeux de données, avec des coefficients de corrélation supérieurs à 0.90 pour les trois variables (R = 0.908 pour P, R = 0.957 pour Tmax et R = 0.915 pour Tmin). Les pentes de régression proches de 1 (0.99 à 1.02) et les ordonnées à l’origine proches de zéro indiquent une préservation fidèle des gradients et amplitudes spatiales après le passage à la résolution de 1 km.

Ces performances traduisent une excellente stabilité du modèle de désagrégation spatiale, capable de restituer la variabilité locale tout en maintenant la cohérence statistique des champs climatiques originaux. En d’autres termes, la réduction d’échelle opérée ne génère ni biais systématique ni distorsion majeure dans la distribution spatiale des valeurs, confirmant la robustesse du processus de validation pour la période actuelle.

### 2.3. Fiabilité pour les scénarios futurs SSP245 et SSP585

Le tableau suivant présente les résultats moyens des statistiques appliquées aux scénarios SSP245 et SSP585 pour les périodes 2021-2040, 2041-2060, 2061-2080 et 2081-2100). Ces statistiques sont appliquées aux modèles climatiques suivants :

- CMCC-ESM2 (Centro Euro-Mediterraneo sui Cambiamenti Climatici, Italy) ;
- EC-Earth3-Veg (EC-Earth Consortium – Sweden, Netherlands, Ireland etc.) ;
- IPSL-CM6A-LR (Institut Pierre-Simon Laplace, France) ;
- MIROC6 (Tokyo Univ., National Institute for Environmental Studies etc., Japan) ;
- MPI-ESM1-2-HR (Max Planck Institute for Meteorology, Germany).

*Tableau 6. Statistiques moyennes du downscaling climatique sur les 4 départements pour le futur pour l’ensemble des scénarios et des périodes.*

Modèles	R <sup>2</sup>			R			RMSE			MAE		
	P	Tmax	Tmin	P	Tmax	Tmin	P	Tmax	Tmin	P	Tmax	Tmin
<b>CMCC-ESM2</b>	0.72	0.84	0.80	0.85	0.91	0.89	3.37	0.21	0.15	2.70	0.17	0.11
<b>EC-Earth3-Veg</b>	0.74	0.83	0.73	0.86	0.91	0.85	3.34	0.21	0.17	2.67	0.16	0.13
<b>IPSL-CM6A-LR</b>	0.74	0.83	0.72	0.86	0.91	0.85	3.30	0.21	0.17	2.63	0.16	0.13
<b>MIROC6</b>	0.73	0.83	0.72	0.85	0.91	0.85	3.23	0.21	0.17	2.58	0.17	0.13
<b>MPI-ESM1-2-HR</b>	0.72	0.83	0.77	0.85	0.91	0.88	3.28	0.21	0.16	2.62	0.17	0.13

Les projections effectuées à partir des sorties GCM-WorldClim sous le scénario SSP585 (fortes émissions) prévoient, à l’horizon 2100, une augmentation moyenne de +3,0 à +3,4 °C pour la température maximale annuelle sur l’ensemble du bassin de la Seine. Cette hausse est légèrement plus marquée dans le sud du bassin (Aube et Yonne). Les précipitations annuelles devraient diminuer de 5 à 10 %, principalement en été, tandis que les hivers resteraient légèrement plus humides (GIEC, 2021).

Les cinq modèles présentent des performances très similaires, illustrant la stabilité du processus de réduction

d’échelle appliqué aux différents jeux de données. Les valeurs élevées de  $R^2$  (comprises entre 0.72 et 0.84) et de  $R$  (de 0.85 à 0.91) indiquent une forte corrélation entre les champs climatiques désagrégés à 30 m et les données de référence à 1 km. Les faibles valeurs de RMSE ( $\approx 3.3$  mm pour les précipitations,  $0.21$  °C pour Tmax,  $0.16$  °C pour Tmin) et de MAE ( $\approx 2.6$  mm,  $0.17$  °C et  $0.12$  °C respectivement) confirment une erreur moyenne très faible, témoignant de la fiabilité du *downscaling* à restituer la variabilité spatiale des variables climatiques.

Les précipitations affichent des valeurs de  $R^2$  comprises entre 0.72 et 0.74 et des  $R$  proches de 0.85, indiquant une bonne conservation de la variabilité spatiale après désagrégation. Les RMSE varient entre 3.23 et 3.37 mm, et les MAE entre 2.58 et 2.70 mm, ce qui montre que les écarts entre les valeurs désagrégées et agrégées restent faibles à l’échelle régionale. Ces résultats traduisent une capacité du modèle de *downscaling* à reproduire fidèlement les gradients pluviométriques locaux, même si les précipitations demeurent la variable la plus difficile à reproduire avec précision en raison de leur forte hétérogénéité spatiale.

Les performances sont particulièrement élevées pour la température maximale, avec des  $R^2$  entre 0.83 et 0.84 et des  $R$  constants à 0.91 pour tous les modèles. Les erreurs moyennes sont faibles (RMSE =  $0.21$  °C, MAE =  $0.16$ – $0.17$  °C), ce qui démontre que la structure spatiale du champ de température diurne est très bien conservée à haute résolution. La cohérence entre modèles suggère que les relations topographiques (altitude, pente) intégrées dans la méthode de *downscaling* ont été correctement transférées du maillage kilométrique à la maille de 30 m.

Pour les températures nocturnes, les  $R^2$  varient de 0.72 à 0.80, avec des  $R$  entre 0.85 et 0.89, confirmant une très bonne cohérence entre échelles spatiales. Les erreurs moyennes (RMSE =  $0.15$ – $0.17$  °C, MAE =  $0.11$ – $0.13$  °C) sont légèrement inférieures à celles de Tmax, ce qui semble traduire une plus grande homogénéité spatiale des températures minimales et donc une meilleure performance du modèle sur cette variable. Ces résultats confirment que la variabilité thermique nocturne, souvent liée à des phénomènes locaux (inversions thermiques, effets de vallées, inertie des surfaces), est bien captée par la désagrégation fine.

Les performances sont quasi équivalentes entre modèles, mais de légères différences peuvent être notées :

- Les modèles MIROC6 et IPSL-CM6A-LR affichent les plus faibles RMSE et MAE, indiquant une très bonne stabilité spatiale des champs simulés ;
- Les modèles CMCC-ESM2 et MPI-ESM1-2-HR montrent une légère dispersion supplémentaire sur les précipitations, mais conservent des corrélations très élevées sur les températures ;
- Le modèle EC-Earth3-Veg offre des résultats intermédiaires, très proches de ceux du modèle IPSL-CM6A-LR, confirmant la cohérence entre modèles des simulations désagrégées.

Ces différences mineures reflètent davantage les spécificités internes des modèles climatiques globaux (paramétrisations physiques et dynamiques) que le comportement du *downscaling* lui-même, dont la performance reste stable.

Les résultats montrent que le processus de *downscaling* climatique à très haute résolution spatiale (de 1 km à 30 m) conserve très efficacement la structure statistique et spatiale des données climatiques initiales, quel que soit la variable considérée ou le modèle CMIP6 utilisé.

Les corrélations élevées ( $R > 0.85$ ) et les faibles erreurs moyennes démontrent la robustesse, la précision et la transférabilité de la méthode à différentes conditions climatiques et géographiques, ces transpositions ayant été testées antérieurement sur d’autres territoires de France (Garbolino *et al.*, 2021) ainsi qu’en Europe, Afrique et Amérique Centrale.

Ainsi, cette validation intermodèle confirme que la désagrégation à 30 m permet une restitution fiable de la variabilité climatique locale, constituant une base solide pour les applications environnementales et territoriales fines (hydrologie, écologie, gestion des risques, planification territoriale, etc.).

Les figures ci-dessous présentent un exemple de comparaison de descente d’échelle sur le territoire de la vallée de La Bassée pour les températures diurnes actuelles de juin.

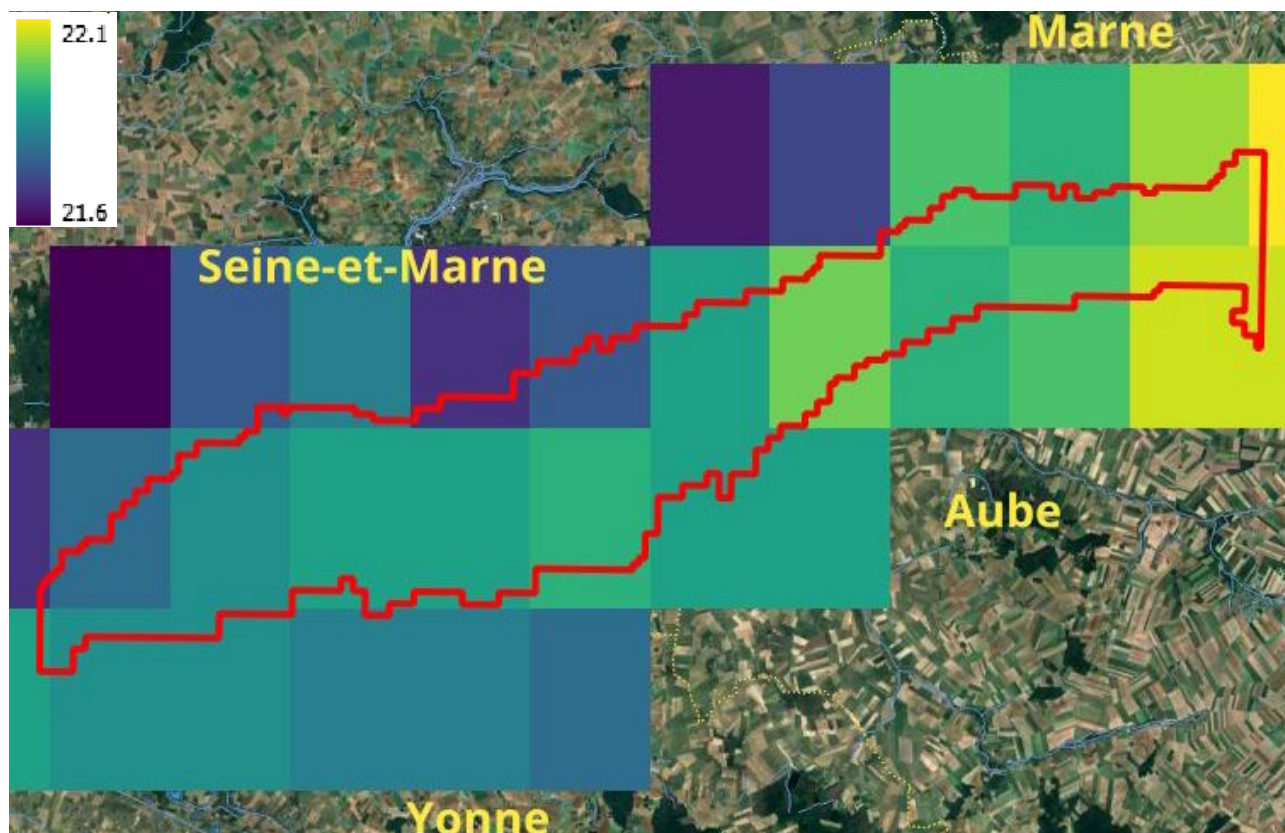


Figure 5. Températures diurnes actuelles pour le mois de juin, résolution spatiale de 8 km (Météo France).

La carte à 8 km de résolution montre la répartition générale des températures maximales de juin à l’échelle régionale.

Les valeurs sont lissées, traduisant des tendances thermiques larges : un gradient Ouest–Est est perceptible, avec des températures plus faibles vers la Seine-et-Marne et plus élevées vers la Marne et l’Aube.

Cependant, cette échelle ne permet pas de capturer les variations locales liées à la topographie (altitude, vallées etc.). Les discontinuités départementales sont visibles, ce qui traduit la limite d’un maillage trop large pour une analyse environnementale fine.

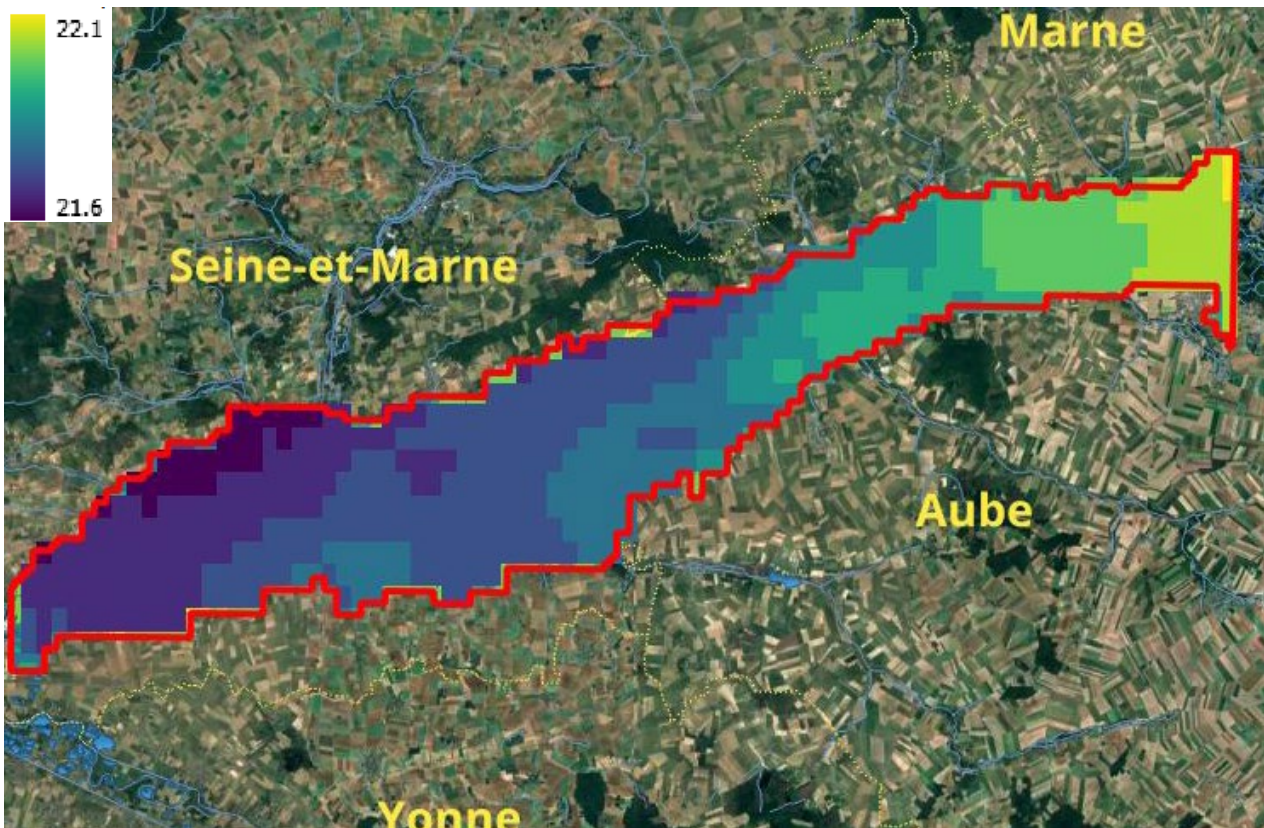


Figure 6. Températures diurnes actuelles pour le mois de juin, résolution spatiale de 1 km (WorldClim).

La carte à 1 km issue de WorldClim offre une représentation plus détaillée des gradients thermiques, tout en maintenant une cohérence régionale avec les données Météo-France.

Les zones plus fraîches de la partie occidentale (Seine-et-Marne) et les zones plus chaudes du sud-est (Aube, Marne) sont mieux différenciées.

Néanmoins, la variabilité intra-départementale reste limitée : certaines hétérogénéités locales observables dans le paysage (vallées, plateaux agricoles, zones boisées) ne sont pas encore perceptibles à cette résolution.

Cette échelle constitue un compromis régional pertinent, mais reste insuffisante pour des applications nécessitant une forte précision spatiale (hydrologie, gestion des espaces naturels, agriculture).

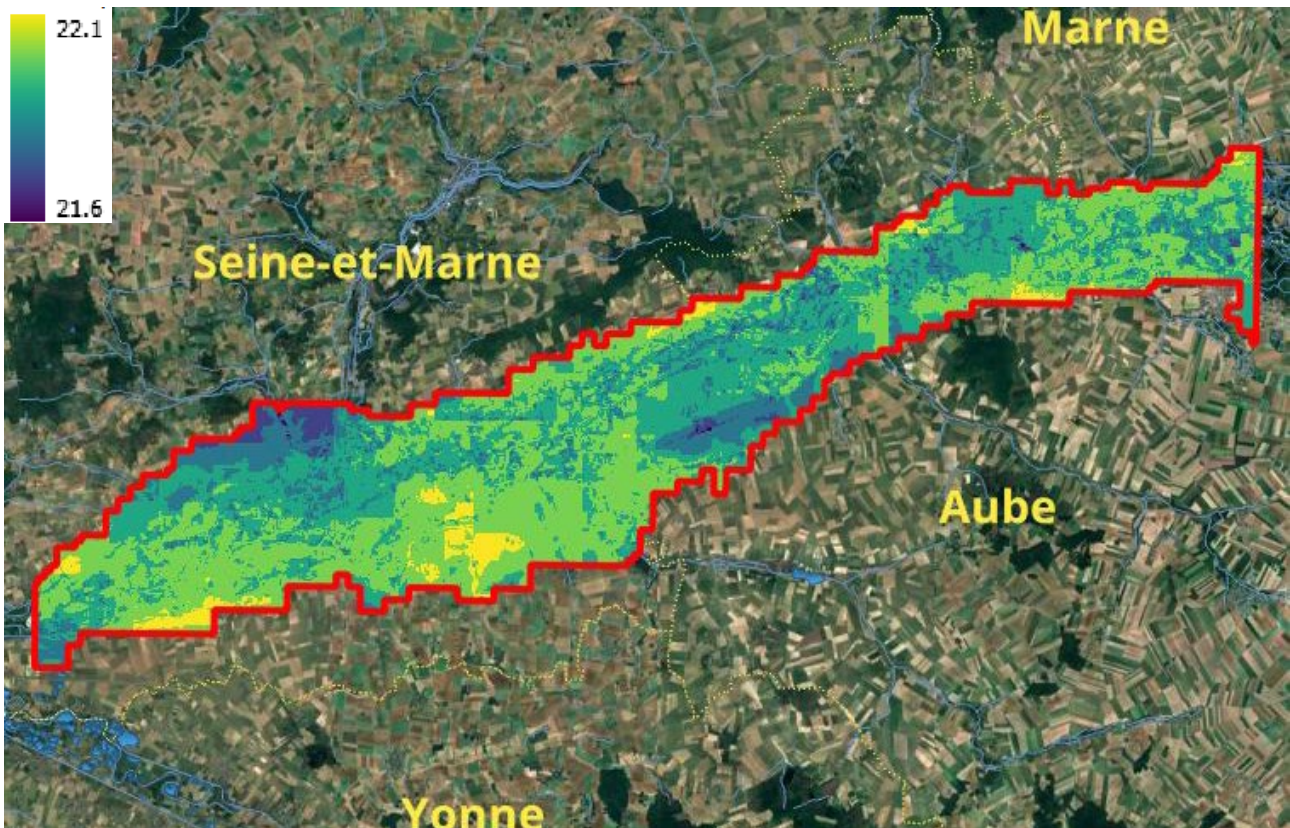


Figure 7. Températures diurnes actuelles pour le mois de juin, résolution spatiale de 30 m (Modèle de *downscaling* climatique).

La carte obtenue par la régression multilinéaire (MLR) appliquée au *downscaling* à 30 m de résolution spatiale présente une structure spatiale plus détaillée. Les gradients thermiques sont continus et finement nuancés, révélant la variabilité thermique à l’échelle intra-parcellaire. Les effets du relief, des vallées humides et des versants exposés apparaissent nettement.

On observe par exemple :

- Des zones plus fraîches le long des vallées et sur les plateaux élevés de la Seine-et-Marne et du nord de l’Yonne ;
- Des zones plus chaudes sur les versants sud et les plaines ouvertes de la Marne et de l’Aube.

Cette carte témoigne de la capacité du modèle de *downscaling* à restituer la variabilité microclimatique en s’appuyant sur des variables explicatives à fine échelle (altitude, pente, distance aux plans d’eau, etc.). La continuité spatiale obtenue et la cohérence générale avec les tendances des modèles à plus grande échelle confirment la robustesse du processus de désagrégation statistique.

### 3. Discussion

L’analyse conjointe des trois résolutions relatives aux précédents résultats met en évidence une progression nette de la précision spatiale : les données Météo-France (8 km) capturent les grands gradients régionaux, celles de WorldClim (1 km) affine la variabilité départementale et le modèle MLR à 30 m restitue la variabilité topoclimatique locale.

Cette hiérarchie démontre que la méthode de *downscaling* MLR par régression multilinéaire permet de transférer efficacement l’information climatique à très haute résolution, tout en préservant la cohérence avec les structures climatiques d’échelle supérieure.

Elle offre ainsi une base solide pour des applications locales sensibles au climat, telles que la modélisation des habitats, la viticulture, la gestion forestière ou les études d’adaptation au changement climatique.

Nous présentons ci-après quelques éléments de discussion sur les limites et l’intérêt de cette approche.

#### 3.1. Interprétation critique des résultats

Les résultats obtenus confirment la robustesse de la régression linéaire multiple (MLR) pour la modélisation des variables climatiques et son potentiel opérationnel dans un cadre régional comme celui du bassin de la Seine.

Les corrélations élevées observées pour les températures minimales et maximales traduisent la capacité du modèle à reproduire les gradients spatiaux dominants, notamment ceux liés à l’altitude et à la latitude, qui expliquent l’essentiel de la variabilité observée (Fick & Hijmans, 2017).

Les performances homogènes enregistrées sur les départements de l’Aube (10), de la Marne (51), de la Seine-et-Marne (77) et de l’Yonne (89) démontrent la stabilité du modèle dans les zones de plaine et de transition, où les gradients topographiques sont réguliers et les effets microclimatiques limités.

En revanche, la dispersion plus marquée des précipitations dans les zones de relief modéré (sud de l’Yonne, bordures du plateau champenois) souligne les limites de la MLR pour modéliser des processus atmosphériques complexes, fortement influencés par la convection, l’humidité relative et les circulations locales (Cannon, 2011).

Cette variabilité met en évidence l’intérêt d’intégrer, dans de futures itérations du modèle, des composantes non linéaires ou probabilistes, capables de mieux représenter les effets d’échelle et la dynamique orographique (Wang & Tian, 2022).

Dans l’ensemble, les résultats confirment que la MLR, malgré sa simplicité mathématique, demeure un outil de descente d’échelle statistique performant, stable et interprétable, répondant efficacement aux besoins d’analyse climatique à haute résolution.

#### 3.2. Comparaison avec la littérature récente

Les conclusions de cette étude s’inscrivent dans la continuité des travaux antérieurs démontrant la fiabilité de la MLR pour la désagrégation statistique des températures (Teutschbein & Seibert, 2012 ; Gutiérrez *et al.*, 2019 ; Garbolino *et al.*, 2021).

Plusieurs études européennes, notamment celles de Maraun (2019) et Reder *et al.* (2025), ont souligné que les modèles linéaires multi-prédicteurs offrent un compromis satisfaisant entre précision, simplicité et coût de calcul.

Toutefois, les recherches récentes en apprentissage profond appliqué à la climatologie (Baño-Medina *et al.*, 2021 ; Wang & Tian, 2022) montrent que l’intégration de réseaux neuronaux convolutionnels (CNN) ou de modèles de régression quantile (QR) améliore significativement la modélisation des extrêmes pluviométriques et thermiques. Ces approches permettent de capter les interactions spatiales complexes et les non-linéarités locales que la MLR tend à lisser.

Cependant, ces techniques avancées présentent encore des contraintes fortes :

- Un besoin en ressources computationnelles élevées,
- Une opacité des processus internes limitant l’interprétation scientifique,
- Une dépendance importante aux volumes de données d’apprentissage.

Dans le contexte de l’étude sur le bassin de la Seine, où la transparence et la reproductibilité sont essentielles pour la planification environnementale et hydrologique, la MLR reste une solution privilégiée.

Sa capacité à s’appuyer sur des données ouvertes et FAIR (Wilkinson *et al.*, 2016) en fait un modèle robuste, directement intégrable dans des démarches d’aide à la décision territoriale (GIEC, 2021).

### **3.3. Limites et incertitudes des approches statistiques**

Les principales limites de la MLR tiennent à la linéarité des relations modélisées et à l’hypothèse de stationnarité temporelle des processus climatiques.

Sous conditions de changement climatique, la structure statistique liant les variables explicatives (altitude, latitude, distance à la mer, etc.) et les variables dépendantes (températures, précipitations) peut évoluer de manière non proportionnelle (Cannon, 2011). Cela réduit la transférabilité temporelle du modèle, en particulier lorsqu’il est appliqué à des scénarios futurs.

Les incertitudes proviennent également des erreurs de mesure et d’interpolation dans les données sources, notamment celles issues de WorldClim 2 et du Copernicus DEM (Fick & Hijmans, 2017).

La représentativité spatiale des stations météorologiques constitue un autre facteur critique : dans les zones à faible densité d’observation, les biais d’interpolation peuvent influencer la calibration du modèle. Des techniques de correction de biais, telles que la méthode EQM (*Equidistant Quantile Mapping*), ont démontré leur capacité à réduire les erreurs systématiques dans les simulations régionales (Reder *et al.*, 2025). Cependant, leur efficacité reste fortement dépendante du contexte topographique et climatique local, ce qui impose une adaptation spécifique à chaque bassin versant étudié.

### **3.4. Implications pour la recherche et la modélisation climatique**

Les résultats obtenus sur le bassin de la Seine renforcent la pertinence scientifique et opérationnelle de la MLR comme méthode de référence pour la désagrégation climatique à haute résolution.

Sa simplicité, son faible coût de calcul et sa compatibilité avec des données ouvertes en font un outil adapté aux applications régionales, notamment pour la gestion hydrologique, la modélisation écologique et la planification territoriale.

À moyen terme, la convergence entre la MLR et les approches issues de l’intelligence artificielle (notamment les modèles hybrides MLR–IA) ouvre la voie à des systèmes intégrés combinant interprétabilité, performance et adaptabilité (Baño-Medina *et al.*, 2021).

Ces systèmes permettront de mieux représenter les processus non linéaires tout en préservant la transparence scientifique requise pour l’analyse des impacts climatiques.

L’intégration de la MLR dans des plateformes décisionnelles climatiques régionales constituerait une avancée majeure pour les territoires, en facilitant la mise à disposition de données climatiques locales précises, explicables et interopérables.

Elle favoriserait ainsi la création de services climatiques opérationnels, au croisement de la recherche académique et des politiques publiques d’adaptation (GIEC, 2021).

## 4. Conclusion

Cette étude confirme que la régression linéaire multiple (MLR) demeure une méthode fiable, transparente et efficace pour la descente d'échelle climatique, en particulier dans les contextes régionaux à topographie modérée comme celui du bassin de la Seine.

Appliquée aux départements de l'Aube (10), de la Marne (51), de la Seine-et-Marne (77) et de l'Yonne (89), la MLR a permis de générer des surfaces climatiques à très haute résolution spatiale (30 m) présentant une cohérence spatiale et statistique tout à fait pertinente et utile pour les zones d'étude de l'Orgeval et de La Bassée, notamment dans la perspective de leur utilisation pour la gestion des milieux naturels.

La stabilité des coefficients de régression, la robustesse des corrélations thermiques ( $R > 0,90$ ) et la faible sensibilité aux biais locaux font de la MLR un outil de référence pour les études climatiques régionales.

Cependant, la complexification croissante des phénomènes climatiques et la non-linéarité des interactions atmosphériques imposent une évolution méthodologique. A l'heure où les outils d'intelligence artificielle et les modèles de machine learning se développent dans le domaine du climat (Baño-Medina *et al.*, 2021 ; Wang & Tian, 2022), les approches comme la MLR conservent un rôle central par leur interprétabilité et leur robustesse statistique. Leur combinaison avec des méthodes d'apprentissage profond ouvre la voie à des modèles hybrides MLR-IA, capables de mieux représenter les processus non linéaires tout en maintenant la lisibilité scientifique des résultats (Baño-Medina *et al.*, 2021).

Ces premiers résultats de *downscaling* climatique à Très Haute Résolution Spatiale sont considérés comme suffisamment précis pour être utilisés dans le cadre de l'évaluation des effets du changement sur les structures végétales des zones d'étude de l'Orgeval et de la vallée de La Bassée. Ils seront alors intégrés dans une approche prospective en écologie fondée sur l'utilisation de modèles de distribution spatiale des espèces de ces territoires (Species Distribution Model – SDM) développés par l'ISIGE – Mines Paris PSL. Ils permettront de comparer les probabilités pour trouver ou non dans le futur des situations environnementales favorables au développement des espèces végétales principalement constitutives des milieux naturels actuellement observés.

## Bibliographie

- Baño-Medina, J., Manzanas, R. and Gutiérrez, J. M. (2021). On the suitability of deep convolutional neural networks for continental-wide downscaling of climate change projections. *Climate Dynamics*, 57(9), 2941–2951.
- Boé J. & Mass A. (2021). Méthode de désagrégation hybride statistico-dynamique pour les scénarios climatiques sur le bassin de la Seine. *PIREN-Seine phase 8 - Rapport 2021*.
- Cannon, A. J. (2011). Quantile regression neural networks: Implementation in R and application to precipitation downscaling. *Computers & Geosciences*, 37(9), 1277–1284.
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geosci. Model Dev.*, 9, 1937–1958, <https://doi.org/10.5194/gmd-9-1937-2016>, 2016.
- Fick, S. E., and Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302–4315.
- Garbolino, E., Hinojos Mendoza, G., Gutierrez Ramos, C.A., Mariscal Guerra, J., Jáquez Frías, L. and Heredia Corral, D.M. (2021). Integrating climate change scenarios for assessing baseline scenario trends. *International Association of Impact Assessment (IAIA) conference IAIA2021 “Smartening Impact Assessment in Challenging Times”*, 18-21 May 2021.
- GIEC, 2021 : Résumé à l’intention des décideurs. In: *Changement climatique 2021: les bases scientifiques physiques. Contribution du Groupe de travail I au sixième Rapport d’évaluation du Groupe d’experts intergouvernemental sur l’évolution du climat* [publié sous la direction de Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu et B. Zhou]. *Cambridge University Press*.
- Giorgi, F. and Mearns, L. O. (1991). Approaches to the simulation of regional climate change: A review. *Reviews of Geophysics*, 29(2), 191–216.
- Gutiérrez, J. M., Maraun, D., Widmann, M., Huth, R., & San Martín, D. (2019). An intercomparison of statistical downscaling methods over Europe: Results from the VALUE perfect predictor experiment. *International Journal of Climatology*, 39(9), 3750-3781.
- Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones and M. New (2008). A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research*, 113, D20119.
- Maraun, D. (2018). *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge University Press.
- Reder, A., Fedele, A., Manco, M. and Mercogliano, P. (2025). Estimating pros and cons of statistical downscaling based on EQM bias adjustment as a complementary method to dynamical downscaling. *Scientific Reports*, 15(1).
- Rummukainen, M. (2010). State-of-the-art with regional climate models. *Wiley Interdisciplinary Reviews: Climate Change*, 1(1), 82–96.
- Teutschbein, C. and Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies. *Journal of Hydrology*, 456, 12–29.
- Wang, F. and Tian, D. (2022). On deep learning-based bias correction and downscaling of multiple climate model simulations. *Climate Dynamics*, 59(11), 3451–3468.
- Wilby, R. L. and Wigley, T. M. L. (1997). Downscaling general circulation model output: A review of methods and limitations. *Progress in Physical Geography*, 21(4), 530–548.
- Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018.