



**HAL**  
open science

# Apport de la télédétection et des variables auxiliaires dans l'étude de l'évolution des périodes de sécheresse

Nesrine Farhani

► **To cite this version:**

Nesrine Farhani. Apport de la télédétection et des variables auxiliaires dans l'étude de l'évolution des périodes de sécheresse. Météorologie. Université Paul Sabatier - Toulouse III; Université de Carthage (Tunisie), 2022. Français. NNT : 2022TOU30022 . tel-03715302

**HAL Id: tel-03715302**

**<https://theses.hal.science/tel-03715302>**

Submitted on 6 Jul 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# THÈSE



En vue de l'obtention du  
**DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE**  
Délivré par l'Université Toulouse 3 - Paul Sabatier

Cotutelle internationale: Université de Carthage

Présentée et soutenue par  
**Nesrine FARHANI**

Le 7 février 2022

**Apport de la télédétection et des variables auxiliaires dans  
l'étude de l'évolution des périodes de sécheresse**

Ecole doctorale : **SDUZE - Sciences de l'Univers, de l'Environnement et de  
l'Espace**

Spécialité : **Surfaces et interfaces continentales, Hydrologie**

Unité de recherche :  
**CESBIO - Centre d'Etudes Spatiales de la Biosphère**

Thèse dirigée par  
**Gilles BOULET et Zohra Lili CHABAANE**

Jury

M. Yves TRAMBLAY, Rapporteur  
Mme Sihem BEN ABDALLAH, Rapporteur  
M. Etienne LEBLOIS, Rapporteur  
M. Lionel JARLAN, Examinateur  
Mme Hedia CHAKROUN, Examinatrice  
M. Gilles BOULET, Directeur de thèse  
Mme Zohra LILI CHABAANE, Co-directrice de thèse  
Mme Julie Carreau, invitée

---

## Remerciements

À l'issue de la finalisation de ce travail, je suis convaincue que la thèse est loin d'être un travail individuel. En effet, je n'aurais jamais pu réaliser ce travail doctoral sans le soutien d'un grand nombre de personnes dont la générosité, la bonne humeur et l'intérêt manifestés à l'égard de ma recherche m'ont permis de progresser dans cette phase délicate pour « l'apprenti chercheur ».

En premier lieu, je tiens à remercier mes directeurs de thèse, Gilles Boulet et Zohra Lili Chabaane pour la confiance qu'ils m'ont accordée en acceptant de diriger et d'encadrer ce travail doctoral et pour leur aide précieuse dans la réalisation de ce Projet. Qu'ils trouvent ici, le témoignage de ma profonde gratitude et de ma haute considération. Un remerciement particulier pour Gilles Boulet pour son rôle déterminant qu'il a joué à toutes les étapes de cette thèse, pour sa confiance, sa disponibilité et ses discussions très enrichissantes qui m'ont aidée à mieux développer mon esprit de chercheur.

Je souhaite également remercier ma co-encadrante de thèse Julie Carreau de toute l'aide qu'elle m'a apportée durant ces années de thèse. Julie n'a jamais hésité ni tardé à me soutenir et à m'encourager. Je la remercie de tous les échanges instructifs et pour toutes les heures qu'elle a consacrées pour m'apprendre les notions de statistiques et de programmation.

Ce travail a aussi bénéficié du soutien scientifique et amical de Zeineb Kassouk et Michel Le Page qui ont joué un rôle important dans ce travail. Merci Michel pour les heures que tu avais passées sur mes scripts python pour me débloquer.

Je tiens à remercier les membres de mon comité de pilotage, M. Jérôme Demarty, Mme. Gwladys Toulemonde, M. Sadok El Asmi et M. Habaieb Hamadi pour leurs commentaires et pistes de recherche.

J'adresse aussi mes remerciements à M. Yves Trambly, M. Etienne Leblois et Mme Sihem Ben Abdallah d'avoir accepté de relire cette thèse et d'en être rapporteurs. La version finale de ce mémoire a bénéficié de leur relecture très attentive et de leurs remarques précieuses. J'aimerais remercier M. Lionel Jarlan d'avoir accepté d'être président du jury. Je remercie également Mme Hedia Chakroun d'avoir accepté d'être examinatrice de ce travail.

Je tiens à exprimer de tout mon cœur mes remerciements ainsi que ma reconnaissance à mes parents Heni et Nabila et à mon mari Amine Ben Dhia qui n'ont pas hésité à m'encourager à mener ce projet de thèse, et qui m'ont donné, la force lors de mes moments de faiblesse et la confiance lors de mes moments de doute. Je vous remercie d'avoir toujours cru en moi et de m'avoir toujours encouragé à aller de l'avant. Je remercie ainsi ma soeur, mon beau frère et mes beaux parents pour leur soutien durant ce parcours.

Ces années de thèse n'auraient pas été agréables sans la convivialité de l'équipe et les doctorants de l'IRD à l'INAT. Je remercie toute personne qui a contribué à rendre cette expérience plus agréable. Je pense alors à tous les doctorants de notre laboratoire pour la bonne ambiance de groupe et tous les moments de partage convivial : Safa Bousbih, Wafa Chebbi, Sameh Saadi, Azza Gorrab, Amal Sebai, Hayfa Zayani, Raja Mabrouki... Je remercie également Radhouane Hamdi, Gaspard Nougaret et Bernard Mougénot pour leur présence et sympathie.

## Résumé

La surveillance des sécheresses dans les régions arides et semi-arides est cruciale car ses conséquences pour l'agriculture peuvent être dramatiques. Afin d'aider les décideurs à établir de bonnes pratiques de gestion de la ressource en eau et d'atténuation du risque des sécheresses, nous nous intéressons à l'analyse des indices de stress hydriques. À cette fin, un modèle de bilan d'énergie à double source permet, en combinant de l'information satellitaire (température de surface, NDVI, albédo et LAI) et de l'information météorologique (température de l'air, humidité relative de l'air, vitesse du vent et rayonnement global), de simuler l'évapotranspiration ainsi que le stress hydrique. Ces deux variables doivent être fournies d'une façon continue et sur une longue période temporelle pour une analyse adéquate des périodes de sécheresses. Or, les réseaux d'observations météorologiques sont parfois insuffisants (faible densité des sites instrumentés et périodes d'observation courtes et souvent non-concomitantes). Notre premier objectif est alors de simuler des scénarios de différentes variables climatiques afin de les prolonger. Nous avons adapté un générateur de conditions météorologiques "MetGen" qui permet de combler les lacunes présentes sur une période d'observation et de projeter des scénarios sur une période distincte de la période d'observation. MetGen exploite parmi ses co-variables, les données de réanalyses qui fournissent des variables à faible résolution spatiale (environ 31 km), comme source d'information importante. Nous comparons cette méthode avec des méthodes de correction de biais (univariée et multivariée) qui exploitent également les données de réanalyses. Cette approche statistique est validée selon deux volets : l'évaluation de la capacité (1) à bien reproduire les variables météorologiques et (2) à bien restituer les variables de bilan d'énergie. Les analyses, menées avec les données des stations météorologiques du système d'observations, ont permis de valider MetGen sur une période de validation (2011-2016). Nous avons utilisé alors cette méthode afin de simuler des données climatiques sur toute la période d'étude (2000-2019). Cette série ainsi que celle provenant des réanalyses brutes sont utilisées comme forçages climatiques du modèle d'énergie à double source SPARSE, afin de simuler deux indices de stress thermiques  $SI_{SWG}$  et  $SI_{ERA5}$  issus du générateur et des réanalyses ERA5 respectivement, à une échelle kilométrique. Ces deux indices sensibles aux anomalies de température de surface, sont comparés avec d'autres indices standardisés issus de différentes longueurs d'onde : le NDVI issu du visible/proche infrarouge, SWI du micro-onde et un indice standardisé de précipitations UPI qui est utilisé comme une référence pour notre analyse. Cette analyse est effectuée en termes de pertinence, de cohérence et de précocité pour la détection d'une sécheresse agronomique. Les deux indices thermiques ont montré des bonnes performances pour la détection du stress, notamment  $SI_{SWG}$  qui a montré plus de précision et de capacité à détecter le stress hydrique d'une façon précoce. Ces analyses et tous ces approches statistiques sont effectuées au niveau du bassin versant de Merguellil situé au centre de la Tunisie et qui présente un modèle typique des régions semi-arides.

**Mots clés.** Variables météorologiques, évapotranspiration, semi-arides, modèle de bilan d'énergie, ERA5, indice de stress, générateur stochastique

## Abstract

In arid and semi-arid areas, water is a major limitation factor for agricultural production. Indeed, these areas are characterized by a short rainy season and strong irregularity in time and space of precipitation events. This induces more frequent annual and intra-seasonal droughts. Evapotranspiration that characterizes plant water use and water stress are needed to better manage water resources and agrosystem health. They both can be simulated by a dual source energy balance model that relies on meteorological variables (air temperature, relative humidity, wind speed and global radiation) and satellite data (surface temperature, NDVI, albedo and LAI). These variables might be simulated for a long period in order to be adequate for drought studies purposes. However, available meteorological observations may often be insufficient to account for the temporal variability present in the study area (sparsity of gauged networks, the lack of long observation periods and the presence of numerous gaps). Our first objective is then to adapt a stochastic weather generator "MetGen" driven by large-scale reanalysis data (about 31 km of spatial resolution) to semi-arid climates and to the sub-daily resolution. MetGen serves to fill in missing data and to provide a temporal extension of multiple meteorological variables. It is compared with two state-of-the-art bias correction methods, univariate and multivariate methods, applied to large-scale reanalysis data. The surrogate series that are either produced by MetGen and the bias correction methods or taken as the un-processed reanalysis data, are evaluated in terms of their ability (1) to reproduce the statistical properties of the meteorological observations and (2) to reproduce energy balance outputs when constrained by observations series. The evaluation of these different statistical methods is performed on a validation period which included the observation period (2011-2016). Then, we used MetGen and the unprocessed reanalyses data to generate meteorological data during the whole study period (2000-2019). These surrogate series are used therefore to constrain the dual-source model Soil Plant Atmosphere and Remote Evapotranspiration (SPARSE) in order to simulate water stress indices  $SI_{SWG}$  and  $SI_{ERA5}$  from MetGen and ERA5 reanalyses successively. Stress index anomalies retrieved from SPARSE are then compared to anomalies in other wave lengths in order to assess their consistency, reliability and capacity to detect incipient water stress and early droughts at the kilometer resolution. Those are the root zone soil moisture at low resolution derived from the microwave domain, active vegetation fraction cover deduced from NDVI time series and a uniformized precipitation index UPI as a reference for these analyses. Both thermal stress indices show a good performance to detect water status, especially using  $SI_{SWG}$  which show more precision and ability to identify incipient water stress. Our analyses are carried on in the Kairouan area in central Tunisia which is subject to semi arid climate.

**Keywords.** meteorological variables, evapotranspiration, semi-arid, dual energy balance model, ERA5, stress index, weather generator

## خلاصة

ان رصد الجفاف بالمناطق القاحلة وشبه القاحلة يكتسي أهمية كبرى لأن مخلفات الجفاف على الفلاحة لا يمكن أن تكون الا وخيمة. وحتى توفر المادة اللازمة لأصحاب القرار ليتسنى لهم اتخاذ الإجراءات اللازمة على مستوى التحكم في الموارد المائية والتخفيف من وطأة الجفاف، قمنا بتحليل مؤشرات الإجهاد المائي. لذلك إستعملنا نموذج توازن الطاقة ثنائي المصدر، يجمع البيانات المتأتية من الأقمار الصناعية والبيانات المتأتية من الأرصاد الجوية ( حرارة الهواء، نسبة الرطوبة فيه، سرعة الريح والإشعاع العام) وذلك لمعرفة النتج والتبخّر

فهذان المصدران يجب أن يتوفرا بصفة مستمرة وعلى مدة زمنية طويلة حتى يكون التحليل مناسباً لفترات الجفاف. الا أن شبكات الرصد الجوية تكون في بعض الأحيان غير كافية للكثافة المنخفضة للمواقع المجهزة او لقصر وغير تزامن فترات المراقبة في أغلب الأحيان. لذلك فقد هيأنا لكل السيناريوهات المناخية للعمل على اطالته عبر وسيلة إحصائية، نستطيع من خلالها سد كل الفجوات في فترة المراقبة ببيانات إعادة التحليل التي توفر متغيرات ذات دقة مكانية منخفضة ( حوالي 31 كم) كمصدر مهم للمعلومة

وقد قارنا هذه الطريقة بطرق أخرى لتصحيح التحيز ( والتي تستعمل أيضا بيانات إعادة التحليل ، فوقفنا على صحة هذا المنهج الإحصائي وفقا لمكونين اثنين: تقييم القدرة (1) لإعادة انتاج متغيرات الأرصاد الجوية بشكل جيد. (2) لاستعادة متغيرات توازن الطاقة بشكل صحيح

فالتحليلات التي أجريت أثبتت صحة هذه الطريقة طوال الفترة الفاصلة بين (2011 و 2016). لذلك فقد استعملنا هذه الطريقة لمجمل البيانات المناخية لكل فترة الدراسة (من سنة 2000 الى سنة 2019)

إستعملنا هذه السلسلة بالإضافة الى تلك المتأتية من عمليات إعادة التحليل الأولية لإنشاء مؤشرين حراريين متأيين على التوالي من المولد المناخي وإعادة التحليل من خلال النموذج على مقياس كيلومترى

تم مقارنة هذين المؤشرين الحساسين لتغيرات حرارة السطح ببعض المؤشرات القياسية متأية من موجات مختلفة. فالمؤشران الحراريان قد أظهرتا قدرة كبيرة على توضيح ورصد الإجهاد المائي بصفة مبكرة . كل هذه التحاليل وهذه المقاربات الإحصائية قد تمت في مجرى وادي مرق الليل بالوسط التونسي ويعتبر أحسن مثال للمناطق شبه الجافة

**الكلمات الدليلية :** الجفاف، الأرصاد الجوية، المناطق القاحلة، النتج والتبخّر، نموذج توازن، الطاقة

# TABLE DES MATIÈRES

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Partie A</b>	<b>11</b>
2.1	Etat de l'art . . . . .	11
2.1.1	Indices et indicateurs usuellement utilisés pour la détection du stress hydrique . . . . .	12
2.1.1.1	Indices basés sur les données météorologiques . . . . .	12
2.1.1.2	Indices basés sur les données satellitaires . . . . .	15
2.1.2	Indices basés sur les anomalies . . . . .	24
2.1.2.1	Indices issus d'une normalisation basée sur le minimum et le maximum . . . . .	25
2.1.2.2	Indices issus d'une normalisation standard . . . . .	26
2.1.3	Indices et anomalies combinés . . . . .	26
2.1.4	Mise en regard de certains indices et anomalies appliqués aux régions semi-arides de la Méditerranée . . . . .	28
2.1.5	Modèle de bilan d'énergie pour la détermination de l'évapotranspiration . . . . .	31
2.1.5.1	Approches contextuelles . . . . .	33
2.1.5.2	Approches de single source . . . . .	34
2.1.6	Méthodes de descente d'échelle . . . . .	36
2.1.6.1	Méthodes Perfect-prog . . . . .	37
2.1.6.2	Model output statistics MOS . . . . .	39
2.1.6.3	Revue sur les travaux de comparaison des différentes méthodes . . . . .	41
2.2	Matériel et Méthodes . . . . .	44
2.2.1	Zone d'étude . . . . .	44
2.2.1.1	Contexte géographique . . . . .	44
2.2.1.2	Contexte climatique . . . . .	45
2.2.1.3	Contexte hydrique . . . . .	46
2.2.1.4	Contexte pédologique et occupation du sol . . . . .	47
2.2.2	Données climatiques . . . . .	48
2.2.2.1	Données climatiques ponctuelles . . . . .	49
2.2.2.2	Données climatiques à large échelle : la ré-analyse . . . . .	50
2.2.3	Données satellitaires . . . . .	55
2.2.4	Données d'occupation du sol . . . . .	56
2.3	Conclusion partielle et synthèse . . . . .	57

<b>3</b>	<b>Partie B</b>	<b>59</b>
3.1	Introduction . . . . .	59
3.2	Développement de la méthode statistique . . . . .	59
3.2.1	Résumé de l'article . . . . .	60
3.2.2	Publication . . . . .	61
3.3	Résultats complémentaires . . . . .	96
3.3.1	Ajustement des variables . . . . .	96
3.3.1.1	Dépendance inter-variables . . . . .	96
3.3.1.2	Lois de probabilité choisies . . . . .	97
3.3.1.3	Jeu de co-variables choisi . . . . .	97
3.3.2	Simulation des variables . . . . .	98
3.3.2.1	Extension temporelle . . . . .	98
3.3.2.2	Imputation . . . . .	102
3.4	Conclusion partielle et synthèse . . . . .	102
<b>4</b>	<b>Partie C</b>	<b>105</b>
4.1	Introduction . . . . .	105
4.2	Interpolation journalière de l'Evapotranspiration . . . . .	105
4.2.1	Interpolation basée sur le rayonnement global . . . . .	107
4.2.2	Interpolation basée sur API . . . . .	107
4.2.3	Comparaison des deux méthodes d'interpolation Rg/API . . . . .	110
4.3	Indices de sécheresse . . . . .	115
4.3.1	Résumé de l'article . . . . .	115
4.3.2	Publication . . . . .	117
4.4	Utilisation de l'indice thermique pour le stress de la végétation . . . . .	149
4.4.1	Simulation de la transpiration . . . . .	150
4.4.2	Indice de stress végétation . . . . .	151
4.5	Comparaison SI et ESI . . . . .	153
4.6	Conclusion partielle et synthèse . . . . .	155
<b>5</b>	<b>Conclusion et perspectives</b>	<b>159</b>

## TABLE DES FIGURES

1.1	Ressources en eau renouvelables totales par habitant en 2014 (m <sup>3</sup> /an) (AQUAS-TAT, 2015) . . . . .	2
1.2	Processus et facteurs de déclenchement des différents types de sécheresse . .	4
2.1	Organigramme des étapes à suivre pour la simulation de l'évapotranspiration et l'indice du stress issu du thermique . . . . .	11
2.2	Indicateurs usuellement utilisés selon les différentes longueurs d'onde. . . . .	15
2.3	Réflectance de la végétation au niveau du bande rouge (R) et proche infra-rouge (IR) en fonction de l'état de la végétation (saine, malade). . . . .	16
2.4	schématisation d'un capteur actif et passif. . . . .	17
2.5	illustration graphique pour le modèle WDI (Water Deficit Index) qui estime un facteur de stress pour le point C en fonction des températures limites A et B, identifiées par gamme de taux de couverture de la végétation. (Adaptée de Lagouarde & Boulet (2016)) . . . . .	23
2.6	Echanges radiatifs simplifiés terre-atmosphère. . . . .	32
2.7	Schémas résistifs mono- et double-source : $r_{as}$ et $r_{ac}$ forment les résistances aérodynamiques des transferts respectivement du composite sol/air et végétation/air. $T_{sol}$ et $T_C$ représentent les températures du sol et de la canopée, successivement. $H_s$ et $H_c$ forment les flux de chaleur sensible issus du sol et de la canopée, successivement. . . . .	34
2.8	Schématisation de l'application des méthodes de descente d'échelle pour fournir des données à une échelle adéquate pour les modèles d'impact. . . . .	37
2.9	Exemple illustratif de l'effet de correction de biais sur une température à faible résolution spatiale par rapport à une température observée. . . . .	40
2.10	Localisation géographique du bassin versant de Merguellil. . . . .	44
2.11	Le cumul des précipitations annuelles sur la période 1986-2021, mesurées à la station synoptique de Kairouan. . . . .	45
2.12	Les moyennes, les minimums et les maximums mensuelles des températures journalières mesurées à la station de Kairouan entre 1986 et 2021. . . . .	46
2.13	Diagramme Ombrothermique : Précipitations et températures mensuelles moyennes mesurées à la station de Kairouan de 1986 à 2021. . . . .	47
2.14	Carte d'occupation du sol de la plaine de Merguellil (hiver-printemps 2019-2020) effectuée dans le cadre des travaux de thèse de Hayfa Zayani en utilisant une série temporelle de NDVI dérivée des images Sentinel-2. . . . .	48
2.15	Localisation des stations météorologiques au niveau de la plaine de Merguellil.	49
2.16	Dépendance inter-sites des différentes variables climatiques pendant la période entre 08/12/2014 et 31/12/2016. . . . .	51

2.17	Localisation de la maille ERA5 utilisée pour extraire les variables à large échelle nécessaires pour notre étude. . . . .	52
2.18	Graphiques quantile-quantile des données climatiques observées à Ben Salem et les données climatiques issues des réanalyses. . . . .	54
2.19	Grille MODIS téléchargée. . . . .	56
2.20	Évolution annuelle de la variation des pourcentages des céréales et la portion du sol nu sur la zone d'étude. . . . .	57
2.21	schéma pour illustrer les étapes de transformation des cartes OS en cartes des pourcentages selon les pixel de MODIS. . . . .	57
3.1	Graphique complet des relations d'inter-dépendance des variables. . . . .	97
3.2	Comparaison des cycles annuels (en haut) et journaliers (en bas) des variables Pression atmosphérique et Précipitations, observées au niveau de la station de Ben Salem, simulées par le générateur et issues des ré-analyses. . . . .	100
3.3	Graphiques quantile-quantile des valeurs observées de la pression atmosphérique et des précipitations à Ben Salem (axe des x) et les valeurs simulées à partir des différents scénarios climatiques issus de MetGen, des deux méthodes de correction de biais (univariée et multivariée) et des ré-analyses brutes. . . . .	101
3.4	Schéma représentatif de la méthode de validation de l'imputation. . . . .	102
3.5	Graphiques quantile-quantiles des valeurs observées de la pression atmosphérique et des précipitations à Ben Salem et leurs valeurs gap-remplées. . . . .	103
4.1	interpolation de l'évapotranspiration (Delogu <i>et al.</i> , 2021) . . . . .	106
4.2	Variation journalière des valeurs de $ET_j$ sans interpolation, $ET_j$ issue de l'interpolation par $EF_{max}$ , suivant la méthode utilisée dans les travaux de Delogu <i>et al.</i> (2021), et les données de mesures de scintillomètre XLAS durant une période de l'année agricole 2013-2014 (forçage climatique issu du générateur stochastique). . . . .	110
4.3	Illustration de la variation de la variable fraction évaporative (EF) en fonction de la méthode de son calcul. . . . .	111
4.4	Variation journalière des valeurs de $ET_j$ sans interpolation, $ET_j$ issue de la nouvelle méthode d'interpolation par API, et les données de mesures de scintillomètre XLAS durant une période de l'année agricole 2013-2014 (forçage climatique issu du générateur stochastique). . . . .	111
4.5	Séries temporelles de l'évapotranspirations journalières et celles issues de deux méthodes d'interpolation testées, en utilisant différents forçages climatiques et satellitaires, sur l'année agricole 2013-2014. . . . .	113
4.6	Variations journalières des valeurs de $ET_j$ sans interpolation, $ET_j$ issue de deux méthodes d'interpolation par API et par Rg (forçage climatique issu du générateur stochastique) et les données de mesures de scintillomètre XLAS durant toute la période où on dispose des mesures XLAS (de mars 2013 jusqu'au Juin 2015). . . . .	114
4.7	Schématisation de l'évapotranspiration au niveau du composite "sol-plante" et la transpiration issue uniquement de la plante. . . . .	149

---

4.8	Variation décadaire de la moyenne de transpiration journalière simulée sur un pixel MODIS contenant un pourcentage élevé de céréale sec. . . . .	152
4.9	Variation décadaire de l'indice de stress calculé à partir de l'évapotranspiration et l'indice calculé seulement à partir de la composante de transpiration, tous les deux simulés sur un pixel MODIS contenant un pourcentage élevé de céréale sec durant l'année 2009-2010. . . . .	152
4.10	Comparaison de l'évapotranspiration de référence et l'évapotranspiration potentielle sur toute la période d'étude, toutes les deux simulées à partir d'un forçage climatique issu du générateur stochastique, représentant la plaine. . .	155
4.11	Variation décadaire des indices standardisés SI et ESI et les données météorologiques standardisées durant l'année agricole 2012-2013. . . . .	156

## LISTE DES TABLEAUX

2.1	Indices de sécheresses météorologiques habituellement utilisés. . . . .	14
2.2	Comparaison entre les capteurs passifs et actifs du micro-onde. . . . .	19
2.3	Résolution temporelle et spatiale de chaque satellite et capteur associé effectuant des mesures de LST au niveau de l'infra rouge thermique. . . . .	20
2.4	Indices de sécheresses utilisés pour la détection du stress au niveau des régions semi arides de la Méditerranée. . . . .	30
2.5	Avantages et limitations des méthodes de descente d'échelle. . . . .	43
2.6	Variables issues des ré-analyses ERA5 utilisées dans le cadre du travail. . . . .	53
3.1	Paramètres et propriétés des différentes lois de probabilité utilisées. . . . .	97
3.2	Covariables sélectionnées pour la pression atmosphérique et les précipitations. Les périodes d'oscillations pour définir un cycle annuel sont spécifiées en jours, et pour un cycle diurne en des heures. Pour les effets mémoires la longueur de la plus longue période de lag est identifiée. La colonne autre contient tous les covariables supplémentaires qu'on a rajoutées pour améliorer la simulation des variables . . . . .	99

## Problématique

Depuis le milieu du 20<sup>e</sup> siècle, on observe des changements dans l'intensité et la fréquence des phénomènes climatiques extrêmes : une augmentation des températures extrêmes chaudes, une élévation du niveau de la mer et des événements de précipitations intenses dans un certain nombre de régions (Min *et al.*, 2011). Les projections climatiques indiquent qu'on aura par conséquent des vagues de chaleur plus intenses, plus fréquentes et plus longues dans de nombreuses régions (Trenberth *et al.*, 2014). Les précipitations extrêmes vont devenir plus fréquentes au niveau de certaines zones, ce qui fait augmenter le risque d'inondations (Hirabayashi *et al.*, 2013). Ces changements climatiques ont un impact direct sur le bilan hydrique terrestre, sur la disponibilité de l'eau douce et également sur les tendances futures relatives aux réserves hydriques. Le stress hydrique se mesure entre autres en mettant en vis à vis l'utilisation de l'eau et les réserves disponibles. Près de 4 milliards de personnes sont touchées par la pénurie d'eau pour au moins un mois par an (Loudière & Gourbesville, 2020). En effet, la disponibilité d'eau douce par habitant est très variable selon les régions, comme montré au niveau de la figure 1.1. On observe que les pays qui disposent de moins de  $1.700 \text{ m}^3/\text{an}$  d'eau par habitant sont en situation de pénurie marginale grave, et ils sont plus vulnérables aux problèmes de sécheresse. Ainsi, cette figure montre que la plupart des pays présentant une pénurie d'eau sont concentrés en Afrique du Nord, au Proche-Orient et en Afrique subsaharienne. Également, on estime que, d'ici 2050, 685 millions de personnes vivant dans plus de 570 villes subiront un déclin supplémentaire de 10 % de la disponibilité d'eau douce, en raison des changements climatiques (Loudière & Gourbesville, 2020).

La croissance démographique et le développement économique entraînent également l'augmentation de la consommation hydrique dans certains pays (Asrar *et al.*, 2019). A l'échelle mondiale, on trouve que 69 % des prélèvements des eaux douces sont destinés à des usages agricoles, essentiellement pour l'irrigation mais aussi pour l'élevage et pour l'aquaculture. Au niveau des pays en zones semi-aride, ce taux est beaucoup plus important et peut atteindre 95 % (FAO, 2011). L'industrie utilise 19% des prélèvements de l'eau et les municipalités absorbent les 12 % restants. L'augmentation des usages aggrave la situation de stress hydrique essentiellement au niveau des régions déjà touchées par le stress. Par ailleurs, le problème de l'eau devient un enjeu majeur à régler afin de préserver la sécurité hydrique et alimentaire des sept milliards d'habitants que compte aujourd'hui notre planète, ou les neuf milliards et demi qui seront présents en 2050 (Lutz *et al.*, 2001).

Les précipitations continentales forment le seul patrimoine d'eau douce de l'humanité et la véritable ressource renouvelable sur terre. A l'échelle globale, ce capital évalué à 110 000 Milliards de  $\text{m}^3$  par an, montre une répartition très inégale : environ 64% des précipitations

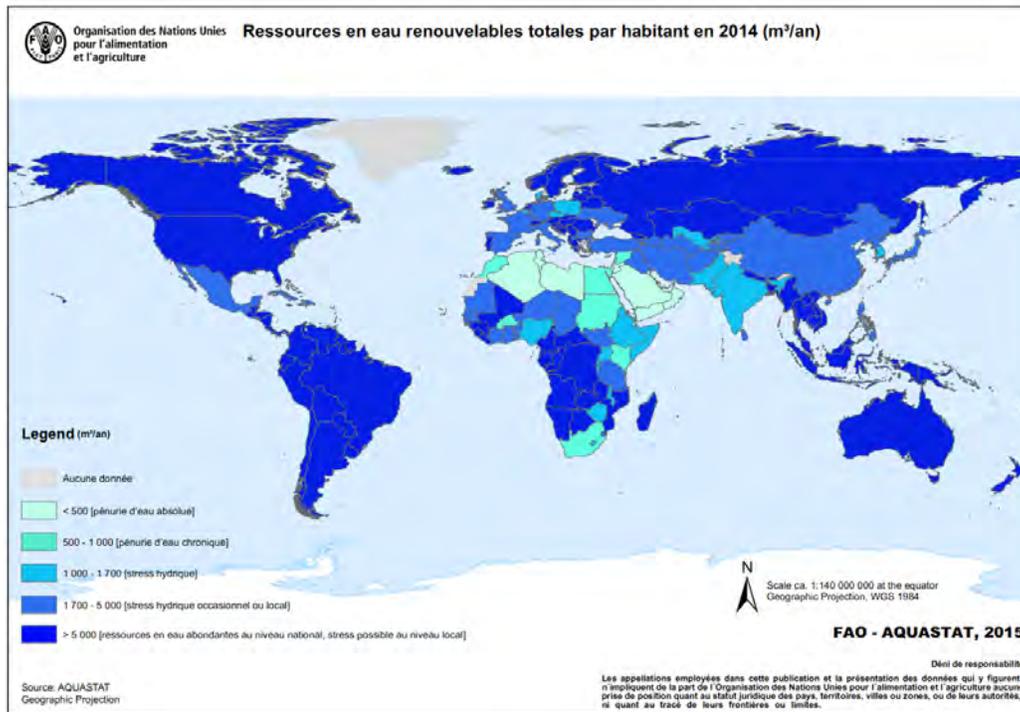


Figure 1.1 – Ressources en eau renouvelables totales par habitant en 2014 (m<sup>3</sup>/an) (AQUASTAT, 2015)

sont reprises par évapotranspiration, les 36% restants alimentent les écoulements : rivières et nappes souterraines qui approvisionnent de l'eau pour l'irrigation, et le reste est partagé entre l'eau potable et les industries (Besbes *et al.*, 2014). Face à cette disponibilité limitée en eau et à ces contraintes climatiques et socio-économiques, le bilan hydrique risque d'être déficitaire et les ressources en eau douce risquent d'être insuffisantes (Besbes *et al.*, 2014).

Une attention particulière est donnée aux pays de l'Afrique du nord, qui montrent ces dernières années, une forte variabilité climatique inter-annuelle (Schilling *et al.*, 2020). Ce fait revient à leur zone climatique (aride-subhumide) qui les rend très sensibles à l'augmentation de la température et à la diminution des précipitations. Péridy *et al.* (2012) montrent que depuis les années 1970, la température a déjà augmenté d'environ 0,5 °C et les précipitations ont diminué d'au moins 20% (IPCC, 2021). Ces changements climatiques risquent par ailleurs d'augmenter la demande évaporative, ce qui devrait avoir des conséquences sur le rendement et la durée du cycle des cultures (Alboghday & El-Hendawy, 2016). D'autre part, d'après Schmitz *et al.* (2013), la pénurie d'eau dans toutes les régions d'Afrique du nord va continuer à augmenter progressivement jusqu'en 2050. Ces conditions climatiques vont engendrer par conséquent des périodes de sécheresse plus fréquentes et plus longues (Tramblay *et al.*, 2020). Ces régions qui sont fortement dépendantes du secteur agricole en termes de contribution au PIB et aux exportations, sont particulièrement touchées par les pénuries d'eau. La question relative à l'impact du climat sur le secteur agricole dans ces pays devient de plus en plus le sujet d'intérêt pour les climatologues et les agronomes.

La Tunisie fait partie d'un groupe de pays du sud de la Méditerranée ayant un poten-

tiel hydrique très modeste. Les précipitations moyennes sur le pays estimées à  $36 \text{ km}^3/\text{an}$  sont reprises à 88% par l'évapotranspiration (Besbes *et al.*, 2013). La Tunisie est considérée comme l'un des pays méditerranéens les plus vulnérables au changement climatique et une terre de climat contrasté (Fathalli *et al.*, 2019; Schilling *et al.*, 2020). Elle est située au niveau d'une zone de transition entre un climat méditerranéen humide et un climat saharien sec. Par conséquent, le climat tunisien varie d'extrêmement aride au sud avec une importante variabilité inter-annuelle de pluies et des épisodes de sécheresse sévères, à un climat plus tempéré au nord avec des hivers relativement frais et humides et des étés chauds et secs (Fathalli *et al.*, 2019). Les précipitations moyennes annuelles varient de plus de 1600 mm en extrême nord le long des montagnes, à moins de 50 mm au niveau du Sahara (Fathalli *et al.*, 2019). Ces gradients pluviométriques intègrent une forte variabilité spatio-temporelle et multi-échelles (Slimani *et al.*, 2007). Là encore, cette fluctuation climatique est couplée à une demande en eau qui croît sans cesse (Leduc *et al.*, 2007), en raison de l'augmentation de la population, du développement industriel et principalement de l'intensification de l'agriculture qui forme le principal consommateur des ressources en eau (80%) (Baccour *et al.*, 2012). Le secteur agricole représente 8% du commerce extérieur tunisien, 10% de l'investissement intérieur et plus de 15% de l'emploi total (Zouabi & Peridy, 2015). Pratiquée sur environ 400000 hectares, l'irrigation provient à 75% des eaux souterraines (Besbes *et al.*, 2013). Bien que ces ressources soient considérées comme des ressources renouvelables, certains réservoirs (les plus profonds généralement) forment des ressources non renouvelables à l'échelle humaine vu le temps nécessaire pour leur régénération (Amigues *et al.*, 2006; Brochet, 1977). Ces réservoirs représentent la plus grande part des ressources souterraines en Tunisie. Elles sont exploitées d'une façon intense et commencent à présenter des signes de dégradation de la qualité chimique de ses eaux, d'où l'importance de la préservation de ces biens environnementaux et l'établissement des bonnes pratiques de gestion de ces ressources rares (Ministère de l'Environnement et du Développement Durable, 2005).

La sécheresse est un aléa naturel engendré par la variabilité climatique (Van Loon & Van Lanen, 2013). Elle apparaît lors d'un déficit de bilan hydrique marqué par rapport à une situation habituelle pour une période donnée et une région déterminée (Brochet, 1977; Olukayode Oladipo, 1985). Elle est déclenchée essentiellement suite à un manque de précipitations sur une période prolongée (saisonniers, annuelle, inter-annuelle) (McKee *et al.*, 1993). D'une façon générale, ce phénomène intègre un nombre de processus météorologiques, hydrologiques et biophysiques ayant des implications socio-économiques (Tramblay *et al.*, 2020). De ce fait, on distingue différents types de sécheresses : - la sécheresse météorologique qui est toujours associée à un déficit prolongé de précipitations (Mishra & Singh, 2010; Tramblay *et al.*, 2020), - la sécheresse agricole qui affecte la végétation et les animaux d'élevage indirectement (Brochet, 1977), selon Olukayode Oladipo (1985), elle apparaît lorsque la teneur en eau disponible dans le sol est insuffisante pour répondre à la demande évaporative de la végétation, et - la sécheresse hydrologique qui se traduit par une anomalie négative en approvisionnement des eaux souterraines (les rivières, les réservoirs de stockage, les lacs et les eaux souterraines) (Tramblay *et al.*, 2020). Comme montré au niveau de la figure 1.2, les facteurs de déclenchement de chaque type de sécheresse sont reliés en cascade : un manque de précipitations va provoquer une sécheresse météorologique, qui par la suite va influencer la

teneur en humidité du sol et induire une sécheresse agricole. La teneur en eau au niveau du sol est également affectée par les augmentations de l'évapotranspiration réelle exigée par la demande atmosphérique, ce qui amplifie davantage le déficit hydrique (Tramblay *et al.*, 2020; Vicente-Serrano *et al.*, 2020). Finalement, une faible recharge du sol va provoquer une sécheresse hydrologique retardée. L'incidence d'une sécheresse n'est pas seulement liée au manque de précipitations, mais dépend également de la bonne gestion de l'utilisation des ressources en eau disponibles (Olukayode Oladipo, 1985).

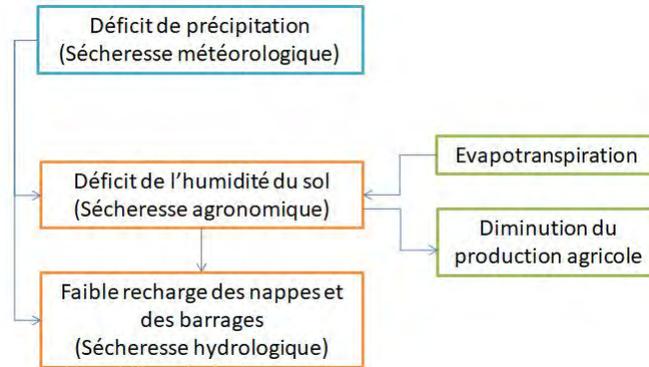


Figure 1.2 – Processus et facteurs de déclenchement des différents types de sécheresse

L'agriculture est souvent le premier secteur touché par la sécheresse (Wilhite & Glantz, 1985; Narasimhan & Srinivasan, 2005), du fait de sa forte dépendance aux réserves hydriques. La sécheresse agronomique affecte la croissance, le développement de la végétation ainsi que sa productivité due à un déficit pluviométrique (De Oliveira *et al.*, 2014). En effet, un déficit pluviométrique au cours de la saison de croissance et une insuffisance de la teneur en eau au niveau de la zone racinaire (Sur *et al.*, 2019) peuvent avoir des impacts subséquents sur l'agriculture : diminution de l'activité photosynthétique, arrêt de croissance, faible rendement et même mortalité de la végétation (Vicente-Serrano *et al.*, 2020; Fahad *et al.*, 2017). D'autre part, un stress hydrique de la végétation est enregistré lorsque l'eau disponible au niveau de la zone racinaire est insuffisante pour répondre à la demande atmosphérique (De Oliveira *et al.*, 2014; Sur *et al.*, 2019). En effet, la plante utilise l'eau disponible au niveau du sol via ses racines, afin d'assurer ses fonctionnalités phénologiques. En conditions de stress hydrique, la plante va réduire son évapotranspiration réelle ( $ET_r$ ), qui est la quantité d'eau évaporée sous les conditions climatiques et d'humidité du sol actuelles (Thornthwaite, 1948), par rapport à un taux d'évapotranspiration potentielle ( $ET_p$ ) qui devrait être effectuée sous les mêmes conditions climatiques.  $ET_p$  est la quantité d'eau maximale qui peut être évaporée si la disponibilité de l'eau n'est pas un facteur limitant (Thornthwaite, 1948).  $ET_p$  dépend énormément des facteurs climatiques, qui représentent la demande évaporative climatique (Vicente-Serrano *et al.*, 2020). Un faible rapport ( $ET_r/ET_p$ ) témoigne par ailleurs de l'existence d'un stress hydrique (UNCCD, 2017). Son intensité est ainsi estimée à partir de ce rapport (Anderson & Kustas, 2008; Boulet *et al.*, 2015; Saadi *et al.*, 2018). Ce rapport présente des valeurs comprises entre 0 qui désigne un stress extrême (pas d'eau disponible), et 1 qui désigne à son tour une condition d'humidité suffisante pour satisfaire la demande climatique (Anderson *et al.*,

2011a). Contrairement à une sécheresse, le stress hydrique peut disparaître si on amène à la végétation la quantité d'eau nécessaire pour compenser ses pertes, d'où l'importance d'identifier le stress hydrique d'une façon précoce afin de limiter ses dégâts.

Une sécheresse peut durer typiquement de quelques mois à quelques années (Vogt *et al.*, 2018). On trouve les "Megadroughts" qui persistent durant des décennies, comme les sécheresses sévères observées au centre des États-Unis au cours des *XV<sup>e</sup>* et *XIV<sup>e</sup>* siècles (Stahle *et al.*, 2007), ou bien aussi les "Flash Droughts" qui durent moins de 3 mois (Otkin *et al.*, 2013). Les impacts de la sécheresse varient par conséquent suivant le secteur considéré mais aussi selon sa durée. Une sécheresse peut avoir des répercussions graves et durables : ressources en eau en deçà des possibilités de remobilisation, denrées alimentaires rares et chères, sols érodés, bétail affaibli et des conflits d'usage entre les différents secteurs (Wilhite & Glantz, 1985). Ses dégâts qui touchent une multitude de secteurs économiques et une large population, présentent à l'échelle mondiale 20% des dégâts provoqués par les aléas naturels avec un coût annuel estimé à 80 milliards de dollars. La sécheresse est la catastrophe naturelle la plus coûteuse et la plus meurtrière de notre époque (Gerber & Mirzabaev, 2017; Kogan, 2001). Il est important alors d'identifier ce phénomène et le caractériser. De même, une identification précoce de la sécheresse peut jouer un rôle très important dans la stratégie d'atténuation de la sécheresse (Anderson *et al.*, 2016; Otkin *et al.*, 2013).

L'identification d'une période de sécheresse et la détermination de sa nature forment une tâche complexe et ambiguë. En effet, la sécheresse est caractérisée selon différentes dimensions dynamiques incluant son intensité, sa durée et sa répartition spatiale, ce qui la rend difficile à caractériser (Zargar *et al.*, 2011). On recourt alors à des indices de sécheresse basés sur des indicateurs reflétant l'état hydrique du couvert mis en question. Ces indices assurent par ailleurs le suivi de la sécheresse et la détection des différents stades de son évolution grâce à l'exploitation des différents jeux de données climatiques, hydriques et phénologiques. Les indices de sécheresse ont été développés au cours du 20<sup>me</sup> siècle et appliqués dans différents domaines de recherche : météorologie, hydrologie, agronomie, ainsi qu'au niveau des applications de gestion des ressources en eau (Niemeyer *et al.*, 2008). Un bon indice de stress doit être fiable et robuste. Selon Narasimhan & Srinivasan (2005), un bon indice doit aussi pouvoir refléter l'évolution des courtes périodes de sécheresse, ne doit pas avoir une saisonnalité (l'indice doit indiquer la sécheresse en été ou en hiver) et doit bien définir la répartition spatiale de la sécheresse indifféremment s'il s'agit d'une zone humide ou aride. Pour cela, on applique une standardisation des valeurs de l'indice pour le rendre comparable sur différentes échelles temporelles et spatiales (Mishra & Singh, 2010). Ces valeurs standardisées sont souvent calculées comme une déviation par rapport à une moyenne sur une période de référence, et sont donc appelées "anomalies". Elles peuvent être exprimées en termes de probabilité (division par l'écart type pour une distribution normale par exemple). Par ailleurs, chaque valeur d'anomalie est associée à une probabilité d'occurrence de la sécheresse (Mega & Medjerab, 2021).

On peut classer les différents indices de stress en deux types : des indices de stress basés sur les données météorologiques et d'autres qui sont basées sur les données satellitaires. Les indices basés sur les données météorologiques ciblent essentiellement la sécheresse éponyme, qui forme le premier stade d'évolution d'une période de sécheresse (Vicente-Serrano *et al.*,

2020). Ainsi, ces indices ne tiennent pas en compte des caractéristiques spatialisées du sol et de la croissance des cultures, et donc des impacts ou adaptations différenciées des surfaces à la pénurie d'eau. D'autre part, au niveau des pays les moins développés, la disponibilité des données météorologiques reste bien souvent limitée aux rares stations de mesure du réseau synoptique. D'un autre côté, les indices basés sur les données satellitaires présentent un apport conséquent grâce à leur large couverture spatiale et leur revisite régulière, qui permet de suivre l'évolution de l'état de surface durant les cycles culturaux. Par ailleurs, les données satellitaires offrent des informations diverses qui concernent le développement et l'état de la végétation, les précipitations, la température de surface, l'humidité du sol, et ce sur une large gamme de résolutions temporelles et spatiales (Jiao *et al.*, 2021; Anderson *et al.*, 2016). En effet, l'apparition des capteurs à haute et moyenne résolution permet d'envisager le suivi des propriétés des surfaces à partir de différentes longueurs d'onde : visible, proche infrarouge, infrarouge thermique et micro-ondes. Les réflectances obtenues dans les domaines visible et proche infrarouge sont sensibles à la quantité de végétation verte (Asrar *et al.*, 2019) et ont permis les premières applications pour l'identification du biomasse, la densité de végétation et le type d'occupation des sols (Bannari *et al.*, 1995). Un indice basé sur la réflectance optique des végétations mesure par ailleurs le degré de développement de la végétation ou une sénescence précoce résultante d'un déficit hydrique marqué auparavant dans le sol. Il existe alors un décalage entre le début de stress hydrique et la réponse en termes de réflectance optique des végétations en condition de stress hydrique. Ce déphasage de réponse forme l'une des limites majeures pour la détection de stress hydrique par le biais des capteurs satellitaires dans le visible (Smith *et al.*, 2019). En effet, en condition de déficit hydrique, la plante va fermer ses stomates qui ont un rôle primordial dans la régularisation des flux (Monteny & Lhomme, 1980). Ce fait va provoquer une diminution d'évapotranspiration et ainsi de l'absorption de CO<sub>2</sub>, ce qui va affecter par conséquent l'activité photosynthétique et la production de biomasse de la plante. La diminution de l'évapotranspiration va être compensée par un accroissement de flux de chaleur sensible pour dissiper la fraction d'énergie absorbée par le rayonnement solaire (Monteny, 1987). D'autre part, la fermeture des stomates va engendrer une élévation des gradients thermiques à la surface (entre la culture et l'air) (Monteny & Lhomme, 1980). Ce gradient thermique peut être détecté par les capteurs en infrarouge thermique. Par ailleurs, les indices basés sur la télédétection en infrarouge thermique permettent de caractériser les interactions énergétiques à la surface des terres (Zhang *et al.*, 2010). Ainsi, les variables de bilan d'énergie et notamment l'évapotranspiration forment une composante clé des échanges surface-atmosphère. Cette variable relie en effet, le cycle de l'eau (par évaporation), le cycle d'énergie (flux de chaleur latente) et le cycle de carbone (compromis entre la transpiration et la photosynthèse) (Fisher *et al.*, 2017). L'évapotranspiration est par ailleurs une variable principale pour la gestion intégrée de l'eau par la détermination des besoins de la végétation en fonction de la demande atmosphérique (Allen *et al.*, 1998; Anderson *et al.*, 2011b). Ainsi, elle présente un moyen efficace pour la détection des périodes de sécheresse (Anderson *et al.*, 2016) et spécifiquement les sécheresses rapides (Otkin *et al.*, 2013).

Une estimation précise des flux d'énergie entre le sol et la végétation est alors très utile pour la détermination de l'intensité des échanges d'énergie, et notamment l'évapotranspiration. Cette variable peut être simulée à partir des modèles de bilan d'énergie à double source.

Ces modèles fournissent des estimations plus robustes de l'évapotranspiration ainsi que du stress hydrique (Chirouze *et al.*, 2014). En effet, ces modèles sont alimentés par des données météorologiques et des données sur la couverture végétale, issues de la télédétection, ce qui leur permet de prendre en compte les interactions entre le sol et la végétation. Ils simulent les composantes d'évaporation du sol et de transpiration du flux total. Comme ils résolvent le bilan d'énergie instantané au passage du satellite, les modèles de bilan d'énergie à double source ont besoin d'une information météorologique instantanée précise, donc infra-journalière. D'autre part, l'utilisation directe des indices de stress ne présente que l'état hydrique ou la quantité de biomasse verte. Toutefois, pour une étude de sécheresse, c'est l'écart de ces indices par rapport à la situation "normale" observée sur de longues séries temporelles, ce qu'on désigne par anomalie, qui est nécessaire pour caractériser le degré de sécheresse. Cependant, les réseaux d'observation sont généralement peu représentatifs (peu de stations, souvent mal situées, une courte période d'observation et la possibilité d'avoir des données manquantes). Le besoin d'avoir une information météorologique en infra-journalier, bien étendue dans le temps nous amènent à recourir aux données à large échelle (réanalyses) qui combinent des observations et des modèles globaux de circulation atmosphériques afin de fournir une information spatialement et temporellement complète et cohérente (Hersbach *et al.*, 2018). En effet, avoir une base des données historique locale assez longue et assez représentative nous permet d'analyser l'état climatique actuel et son impact notamment sur les ressources hydriques. Les ré-analyses sont disponibles pour une longue période dans le passé (de 1950 à nos jours). Toutefois, leur résolution spatiale est trop basse (31 km) (Hersbach *et al.*, 2018). Cette dimension est bien évidemment trop élevée et inappropriée pour les modèles d'impact sur le bilan d'énergie qui sont très sensibles aux variations de petite échelle du climat (Maraun *et al.*, 2010). Cette différence entre les besoins en information de fine échelle spatiale et ce que peuvent fournir les modèles, explique le recours à une démarche de "descente d'échelle" ou le "downscaling".

## Objectifs de la thèse

L'objectif majeur de la thèse est d'étudier la pertinence de différents indices issus de la télédétection spatiale moyenne résolution pour suivre et anticiper la sécheresse agricole en Afrique du Nord. La finalité opérationnelle de ces travaux est de fournir aux gestionnaires et planificateurs des indicateurs spatialisés sur la mobilisation des ressources en eau en Tunisie et sur la bonne manière de gérer les périodes de sécheresses, c'est-à-dire les quantifier et les détecter d'une façon précoce. Cette démarche d'alerte précoce et de caractérisation de ce phénomène, permet de limiter l'impact négatif sur la production agricole par l'établissement des bonnes décisions de gestion des ressources et garantir une disponibilité durable des ressources en eau pour les différents usages à l'échelle de la parcelle, du bassin versant et aussi de la région. Ce genre d'étude, qui s'intéresse à une région typiquement semi-aride au centre de la Tunisie et qui forme un pilier de la production agricole nationale, aide à évaluer une disponibilité limitée des ressources en eau couplée avec des contraintes climatiques et des pratiques des agriculteurs qui ne sont pas nécessairement adéquates aux conditions existantes. Par ailleurs, une projection historique des informations concernant la disponibilité en

eau en réponses à l'évolution des conditions climatiques, et par conséquent, l'occurrence des futures périodes de sécheresse, s'avère indispensable pour mieux comprendre le mécanisme d'apparition de cet aléa. Cependant, la définition d'une période de sécheresse reste relative par rapport au secteur d'intérêt (météorologique, agronomique, hydrologique ou économique). D'autre part, la détermination du début, de la fin, de l'intensité de la sécheresse, etc. reste toujours subjective par rapport aux seuils et paramètres d'analyse qu'on va définir préalablement. Ce fait amplifie davantage la difficulté de ce type d'analyse, mais qui reste à envisager afin d'atténuer les dégâts causés par la sécheresse. Pour répondre à cet objectif global, on doit répondre à une succession d'objectifs spécifiques :

- Une faible densité du réseau d'observations météorologiques nous fait recourir à l'utilisation des données des modèles, ou ré-analyses lorsque l'on s'intéresse au passé. Le premier objectif de la thèse est alors, d'appliquer une méthode de descente d'échelle afin de rendre exploitable les données à large échelle à des fins de précision hydrologique nécessitant une résolution spatiale et temporelle fine. Pour ce faire, nous avons adapté un générateur stochastique de conditions météorologiques se basant sur des données de ré-analyses et qui est inspiré par une approche développée par Richard Chandler (Chandler, 2015). Ce générateur nous permet de simuler des scénarios spatio-temporels des différentes variables climatiques afin d'effectuer le comblement des données manquantes et d'établir une projection de la série climatique dans le passé.
- Le deuxième objectif est de comparer ces scénarios multi-variables avec une autre méthode statistique largement utilisée dans la littérature (méthodes classiques de correction de biais), avec les données de réanalyses brutes et avec les données ponctuelles issues des stations de mesure. Cette évaluation est effectuée en termes de critères statistiques et en termes d'impact pour l'étude des ressources en eau. Pour ce faire, ces différents scénarios vont être introduits comme un forçage climatique dans le modèle de bilan d'énergie à double source SPARSE "Soil Plant Atmosphere and Remote Sensing Evapotranspiration model" (Boulet *et al.*, 2015), et vérifier si ces scénarios peuvent restituer la simulation des variables de bilan d'énergie quand le modèle est forcé par les observations, et particulièrement le stress hydrique exprimé en fonction des flux de chaleur latente simulés par le modèle.
- Le troisième objectif est d'utiliser le forçage climatique ainsi généré par la méthode statistique et simuler à partir du modèle SPARSE des données d'évapotranspiration, nécessaires pour calculer cet indice de stress sur une période d'étude de 2000 jusqu'au fin 2019. L'utilisation des produits thermiques comme forçage satellitaire pour SPARSE engendre des discontinuités d'acquisitions des données, à cause de l'enuagement qui peut nuire à la fiabilité de l'information. De même, ces données ne sont fournies qu'à l'échelle instantanée, qui coïncide à l'heure du passage de satellite. Aussi, un travail supplémentaire d'extrapolation journalière et d'interpolation saisonnière de l'évapotranspiration est effectué avant d'utiliser cette variable pour le calcul de l'indice de stress thermique.
- Le quatrième objectif est de comparer l'indice de sécheresse ainsi généré à partir de l'information thermique, avec d'autres indices issus d'autres longueur d'ondes (visible/proche

infrarouge et micro-onde) en termes de pertinence, de réalisme, de cohérence et de précocité pour la détection d'une sécheresse agronomique. Pour ce faire, une standardisation des indices est effectuée, afin d'avoir des valeurs enlevées entre 0 et 1, mieux interprétables en termes d'occurrence et d'intensité. De même, nous avons transformé les différents indices en des indices "Rapid Change Index", en nous basant sur la méthode de Otkin *et al.* (2013), pour mieux illustrer les changements rapides de la différence standardisée des indices ( $\Delta$ Indice), sur un intervalle de temps court (décadaire). Cette transformation nous permettra d'évaluer la rapidité des réponses des différents indices pour un déficit de précipitations.

## Organisation du manuscrit

Le manuscrit s'organise de la façon suivante :

1. **Partie A** : La partie A présente, une synthèse bibliographique des indices de sécheresse/stress hydrique habituellement utilisés. On distingue deux catégories d'indices. On présente par ailleurs, une sélection des indices représentatifs de chaque type : une première catégorie qui utilise les données météorologiques et une deuxième catégorie qui utilise les données satellitaires. Par la suite, on fait une synthèse sur quelques études comparatives des indices de sécheresses effectuées au niveau de la Méditerranée afin de voir leurs réponses et cohérences pour des contextes climatiques similaires aux conditions climatiques de notre zone d'étude. On présente aussi les méthodes de simulation de l'évapotranspiration à partir des modèles de bilan d'énergie en détaillant les deux types d'approches utilisées dans la littérature : contextuelle et résiduelle (single source). Finalement, on présente les méthodes de descente d'échelle, les différentes méthodes, leurs avantages et limitations afin de justifier le choix de la méthode qu'on a utilisée.

Ce chapitre contient aussi une présentation de notre zone d'étude (contexte géographique, climatique, hydrique, pédologique, etc.). Ensuite, on présente les données utilisées pour forcer le modèle de bilan d'énergie "SPARSE", données météorologiques et satellitaires, afin de simuler l'évapotranspiration et le stress hydrique.

2. **Partie B** : La partie B s'intéresse à la présentation de l'approche statistique développée, appelée "générateur stochastique", afin de gérer la disparité d'échelle qui existe entre les informations issues des réanalyses et celles dont les modèles d'impact ont besoin. La première partie est présentée sous forme d'un article. Il détaille la méthode de construction du générateur, notamment les modèles de régression de chaque variable météorologique. Ceci implique un choix précis de co-variables et de distribution de probabilité, permettant de restituer la variabilité en infra-journalier des différentes variables climatiques. La validation est effectuée en termes de comparaison avec d'autres méthodes très utilisées dans la littérature (méthodes de correction de biais), ainsi, qu'en termes d'application du générateur pour constituer le forçage climatique pour notre modèle d'impact SPARSE. Au niveau de l'article, on se restreint à présenter les quatre variables qui vont être utilisées pour la simulation de l'évapotranspiration (température de l'air, humidité relative

de l'air, vitesse du vent et rayonnement global). Par ailleurs, une deuxième partie du chapitre présente des résultats complémentaires, qui concernent la simulation d'autres variables climatiques : les précipitations et la pression atmosphérique, très utiles pour les études du changement climatique.

3. **Partie C** : La partie C présente une évaluation de l'indice de stress simulé à partir de SPARSE en utilisant des données thermiques, comparé avec d'autres indices largement utilisés dans la littérature, le SWI et le NDVI issus des domaines micro-onde et visible/proche infra-rouge respectivement. Cette inter-comparaison est présentée sous forme d'un article. Une deuxième partie contient des résultats supplémentaires concernant les méthodes d'interpolation de l'évapotranspiration, afin d'avoir des données journalières et construire une information à une échelle saisonnière. Les méthodes d'interpolation englobent des méthodes classiquement utilisées, basées sur l'utilisation d'un support d'interpolation facilement accessible chaque jour (rayonnement solaire et précipitations). D'autre part, afin de profiter de l'utilisation d'un modèle de bilan d'énergie à double source permettant de simuler l'évaporation du sol et la transpiration de la végétation d'une façon séparée, nous avons simulé l'indice de stress calculé seulement à partir de la transpiration. En effet, une sécheresse agronomique doit être détectée selon l'état hydrique de la plante et non de la surface composite plante-sol intercalaire. Par ailleurs, nous avons testé la réponse de l'indice de stress thermique, pour une année sèche et sur un pixel de céréale en sec caractérisé par un LAI élevé et un apport d'eau lié aux conditions climatiques (apport des précipitations), pour voir au mieux la réponse de cet indice par rapport à une situation du stress.
4. **Conclusion et perspectives** : la conclusion générale contiendra la synthèse des principaux résultats de cette thèse et l'apport des méthodes utilisées, s'attachera à décrire ses limitations et esquissera en perspective quelques pistes de travail susceptibles d'améliorer la démarche qu'on avait suivie et d'assurer le développement des résultats extraits de ce travail.

## 2.1 Etat de l'art

Pour caractériser la sécheresse comme déviation de l'indice de stress par rapport à une situation "moyenne" (anomalie), il est important de considérer une base de donnée de profondeur et de résolution temporelle adéquate des variables climatiques et thermo-hydriques, à savoir une série temporelle suffisamment longue (robustesse de la détection d'anomalies) et données infra-journalières pour alimenter le modèle de bilan d'énergie. La nécessité d'avoir une longue base de données météorologique nous oblige souvent à exploiter les données de ré-analyses présentant une faible résolution spatiale. Le besoin d'affiner le plus possible les informations météorologiques, explique notre recours à l'utilisation des méthodes de descente d'échelle largement utilisées pour alimenter les modèles d'impact. Les étapes nécessaires pour la simulation de ET et du stress hydrique sont définies au niveau de l'organigramme présenté dans le graphique 2.1.

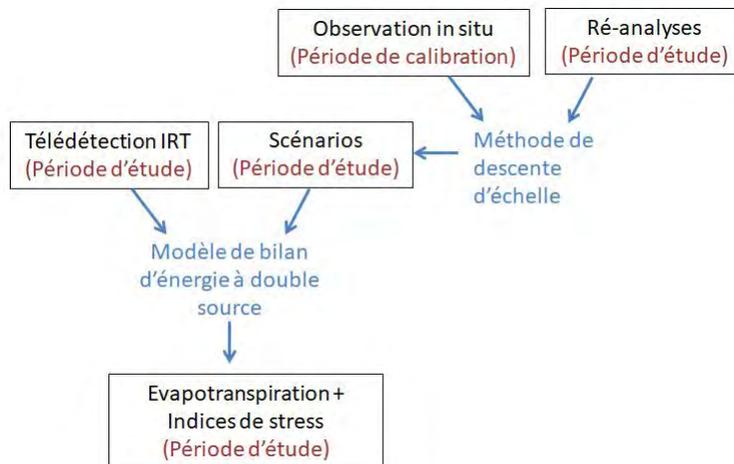


Figure 2.1 – Organigramme des étapes à suivre pour la simulation de l'évapotranspiration et l'indice du stress issu du thermique

La première section de ce chapitre présente un état de l'art avec un focus particulier sur les méthodes qui seront implémentées par la suite :

- Les indicateurs de sécheresse habituellement utilisés dans la littérature,
- Les différents modèles de bilan d'énergie pour la détermination de l'évapotranspiration, nécessaire pour calculer un indice de stress thermique,

- Les méthodes de descentes d'échelle pour la prolongation des données météorologiques.

### 2.1.1 Indices et indicateurs usuellement utilisés pour la détection du stress hydrique

L'estimation des périodes de sécheresse hydrique nécessite une quantification précise des évènements de stress. Pour ce faire, on a recours à des indices de sécheresse qui sont d'après l'Organisation Météorologique Mondiale (OMM), "des représentations numériques de l'intensité des sécheresses, que l'on calcule à partir des valeurs climatiques ou hydrométéorologiques. Ils mesurent l'état qualitatif d'une sécheresse à un emplacement donné pour une période donnée" (Svoboda & Fuchs, 2016). Ces indices utilisent par ailleurs, des variables hydrométéorologiques (les précipitations, la température, l'écoulement fluvial, l'humidité du sol...) qui servent à décrire son intensité, sa répartition géographique et sa durée. Du fait de leur relation avec le statut hydrique de la plante, ses paramètres biophysiques et physiologiques (photosynthèse, conductance stomatique, teneur en chlorophylle, potentiel hydrique, etc.), ces indices peuvent être utilisés comme des indicateurs de l'état hydrique de la végétation (Narasimhan & Srinivasan, 2005). Plusieurs indices basés sur ces variables biophysiques ont été développés pour quantifier la sécheresse. Ces indices peuvent être classés selon deux catégories : indices de sécheresse basés sur des variables climatiques ou bien basés sur des observations satellitaires des états de surface.

#### 2.1.1.1 Indices basés sur les données météorologiques

Les indices de sécheresse les plus anciens sont basés essentiellement sur des variables météorologiques fournies à partir des observations in-situ. La majorité de ces indices utilisent les données de précipitations, seules ou en combinaison avec d'autres variables météorologiques (Mishra & Singh, 2010). L'indice météorologique le plus utilisé est le Standardized Precipitation Index (SPI) (McKee *et al.*, 1993), défini comme la différence entre la valeur des précipitations et sa moyenne pour une période donnée, divisée par son écart type calculé sur la même période (McKee *et al.*, 1993), voir le tableau 2.1. Le SPI est conçu pour quantifier le déficit des précipitations sur différentes échelles temporelles (1 mois, 3 mois ou plus), définies selon le domaine d'application. Les courtes durées sont plus appropriées pour les applications agricoles, et les durées plus longues sont utilisées pour la gestion de l'eau à long termes (Guttman, 1998). Le calcul de l'indice SPI est basé sur un historique des précipitations sur une longue durée (minimum 20 à 30 ans). SPI repose sur le calcul de la probabilité de la pluie pour donner une idée sur l'intensité des précipitations sur une période donnée (McKee *et al.*, 1993). Ensuite, cette longue série de précipitations est transformée en une distribution normale. Ainsi, McKee *et al.* (1993), ont pu définir un système de classification des valeurs de l'indice SPI pour attribuer l'intensité des épisodes de sécheresse à la valeur de l'indice, tel que, les valeurs positives de l'indice SPI indiquent des précipitations supérieures à la médiane et les valeurs négatives, indiquent des précipitations inférieures à la médiane. Étant donné que l'indice est normalisé, SPI se révèle efficace pour analyser de la même manière, les périodes humides et les périodes sèches, et d'assurer par conséquent la surveillance de la sécheresse,

l'analyse de sa fréquence et l'identification de ses différentes classes. (OMM N°1090, 2012). Par ailleurs, SPI présente l'avantage d'être à la fois un indice puissant et simple à calculer, ayant les données de précipitations comme le seul paramètre à introduire. Il est exploité dans plus de 70 pays (Daki *et al.*, 2016).

L'indice Palmer drought severity PDSI (Palmer, 1965) est également largement utilisé pour la détection des périodes de sécheresse ou humides prolongées (Narasimhan & Srinivasan, 2005). Cet indice utilise les données mensuelles ou hebdomadaire de précipitations, de températures de l'air et la teneur en eau disponible au niveau du sol afin de résoudre des équations de bilan hydrique et déterminer l'évapotranspiration, le taux de recharge du sol, le ruissellement et la perte d'eau de la couche superficielle du sol (Alley, 1984). Ces variables permettent d'estimer l'état d'humidité du sol, et constituent la variable Z "moisture anomaly index", nécessaire pour le calcul de l'indice PDSI (voir l'équation au niveau du tableau 2.1). Palmer (1965) a ajusté l'équation de calcul de cet indice, présentée dans le tableau 2.1. Il est calibré par des facteurs de pondération, relatives aux conditions normales des variables climatologiques (précipitations normales, évapotranspiration potentielle, taux de recharge potentiel, etc.). La détermination de l'intensité de sécheresse est effectuée selon des classes pré-définies de l'indice par Palmer (1965), et qui ont été déterminées à partir de l'observation de sécheresses extrêmes historiques pour sa zone d'étude. Une sécheresse sévère par exemple correspond à un  $PDSI = -4$ . Cet indice présente néanmoins plusieurs limitations selon Alley (1984), notamment dues aux choix arbitraires de ses classes de sécheresses ajustées seulement selon des régions de la zone d'étude (Iowa et Kansas).

L'indice surface water supply index (SWSI) est un indicateur normalisé (voir le tableau 2.1), permettant de comparer différentes régions, pour des périodes mensuelles. Il a été développé par Shafer & Dezman (1982), en tant qu'indice complémentaire de l'indice PDSI Palmer (1965), afin de décrire les différents régimes hydrauliques. SWSI est développé comme un indicateur de disponibilité en eau pour les régions où la fonte des neiges est la principale source d'écoulement fluvial (régions montagneuses généralement, où la neige contribue par 65 % à 85 % au débit annuel du fleuve) (Shafer & Dezman, 1982). En effet, en printemps, la couche de neige se transforme en ruissellements et rejoint les ressources en eau disponibles (précipitations liquides et réservoirs) (Shafer & Dezman, 1982) avec un temps de latence par rapport à la date d'apport neigeux, délai à prendre en compte pour les régions à fort régime nival. Par ailleurs, cet indice est calculé à l'échelle du bassin hydrographique d'une façon différente en fonction des saisons. Il est calculé à partir du manteau neigeux, du débit des précipitations et du stockage dans le réservoir (voir le tableau 2.1) pendant les saisons d'hiver. En revanche, pendant les saisons d'été, le manteau neigeux est remplacé par l'eau de ruissellement. Narasimhan & Srinivasan (2005) trouvent que cet indice a une bonne performance pour identifier les impacts des sécheresses hydrologiques sur la demande urbaine et industrielle ainsi que le potentiel hydroélectrique, mais il n'est pas adéquat pour la détection de la sécheresse agronomique à cause de la période de latence nécessaire à prendre en compte pour identifier un déficit de précipitations par rapport au bilan des ressources d'eau disponibles superficielles ou souterraines qui sont utilisées pour l'agriculture. On cite aussi, le Rainfall Anomaly Index (RAI) (Van Rooy, 1965) qui se base seulement sur les données de précipitations. Olukayode Oladipo (1985) a montré que les indices basés seulement sur les

précipitations ont une performance très comparables aux autres indices météorologiques plus complexes. On présente les équations de chaque indice présenté au niveau de ce paragraphe, ses variables d'entrées ainsi que son échelle temporelle au niveau du tableau 2.1.

Malgré la facilité de leur utilisation et leur efficacité pour identifier les périodes de sécheresse, les indices météorologiques montrent plusieurs limites : D'une part, ils sont incapables de prendre en compte la variabilité spatiale des variables climatiques à cause de la disponibilité insuffisante des stations synoptiques notamment dans les pays les moins développés (Mishra & Singh, 2010). Ainsi, les données météorologiques issues des stations synoptiques sont disponibles à large échelle en général. D'autre part, les données météorologiques comportent des périodes d'observations manquantes qui peuvent dépasser le mois (AghaKouchak, 2017).

Table 2.1 – Indices de sécheresses météorologiques habituellement utilisés.

Indices	Variables d'entrée	Échelle de temps	Formules	Références.
Standardized Precipitation Index (SPI)	Précipitations	mensuelle	$SPI = \frac{X_i - X_{moyen}}{X_{ecarttype}}$ Avec, $X_i$ est le cumul mensuel de Précipitations, $X_{moyen}$ est la moyenne sur toute la période considérée et $X_{ecarttype}$ est son écart type	McKee <i>et al.</i> (1993)
Palmer drought severity (PDSI)	Précipitations, Température, teneur en eau	mensuelle	$PDSI_i = 0.897PDSI_{i-1} + (z_i/3)$ avec, $PDSI_{i-1}$ est la valeur du mois dernier, $z_i$ est l'indice de l'anomalie en humidité	Palmer (1965)
Rainfall Anomaly Index (RAI)	Précipitations	mensuelle	$RAI_i = \frac{3(p-\bar{p})}{\bar{m}-\bar{p}}$ pour les anomalies positives, $RAI_i = \frac{-3(p-\bar{p})}{\bar{x}-\bar{p}}$ pour les anomalies négatives, avec, $p$ est le taux des précipitations actuelles, $\bar{p}$ est le taux des précipitations moyennes, $\bar{m}$ moyenne des 10 précipitations les plus élevées, $\bar{x}$ moyenne des 10 précipitations les plus faibles	Van Rooy (1965)
surface water supply index (SWSI)	Précipitations, réservoir, écoulement fluvial et manteau neigeux	mensuelle	$SWSI = \frac{(aPN_{SP})+(bPN_{PCP})+(cPN_{RS})-50}{12}$ avec, $a$ , $b$ et $c$ les poids pour chaque composant, $PN$ : probabilité de non dépassement et $SP$ , $PCP$ et $RS$ sont respectivement : neige, précipitations et réservoir	Shafer & Dezman (1982)

### 2.1.1.2 Indices basés sur les données satellitaires

Comme présenté au début du chapitre, les indices basés sur les données météorologiques sont souvent issus soit de données météorologiques acquises par des réseaux de station in-situ peu denses, soit par des sorties de modèles mésoéchelle à maille très grossière. La télédétection spatiale permet d'obtenir un suivi des états de surfaces en tout point du globe pour une large gamme de résolutions spatiales et temporelles (Zhang *et al.*, 2013). Les caractéristiques des surfaces observées par satellite sont nombreuses : couverture végétale, humidité du sol ou bien température de surface (Kogan, 1995; Le Page & Zribi, 2019). Ces variables biophysiques sont obtenues à partir de différents domaines spectraux : domaine visible/proche infrarouge, micro-ondes (radiomètre passif ou actif) et infrarouge thermique (voir figure 2.2).

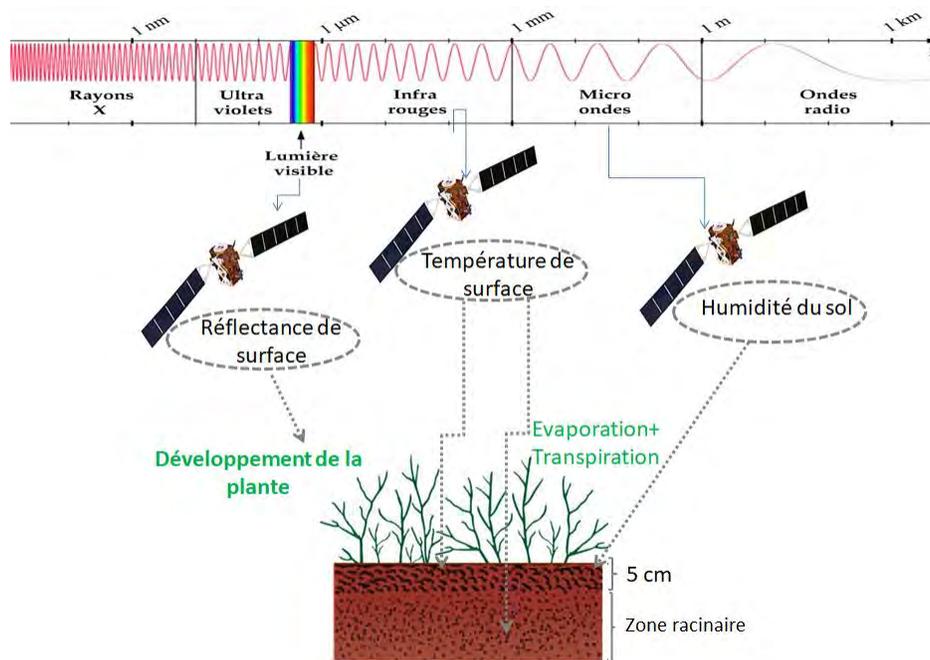


Figure 2.2 – Indicateurs usuellement utilisés selon les différentes longueurs d'onde.

#### 1. Indicateurs issus du visible/Proche infra rouge

Les bandes rouge (R) et proche infrarouge (IR) sont particulièrement utiles à l'étude de la couverture végétale. En effet, ces bandes contiennent la majorité des informations utiles relatives à la végétation (Bannari *et al.*, 1995). Les mesures quantitatives issues des différentes combinaisons de ces canaux forment les indices de végétation (Bannari *et al.*, 1995). Ces indices permettent de détecter les changements des occupations des sols, de l'évaluation de la densité du couvert végétal et l'estimation des types des cultures (Baret *et al.*, 1986). En effet, ces indices utilisent la réflectance mesurée à la fois dans le visible et dans l'infrarouge, afin de déterminer la quantité de biomasse verte de la végétation, son activité photosynthétique et par conséquent son fonctionnement. Comme illustré

au niveau de la figure 2.3, le sol n'a pas le même taux de réflectivité en bande R et IR, d'une végétation saine (bien développée et verte), qui montre un taux de réflectivité plus important en bande IR et moins importante en bande R qu'une végétation malade (moins développée, moins de verdure,...). Par ailleurs, les indicateurs qui utilisent les données de l'indice de végétation forment un moyen pertinent pour le suivi de la variation inter-annuelle de l'état du couvert végétal et de son état hydrique (AghaKouchak, 2017; Kennedy, 1989).

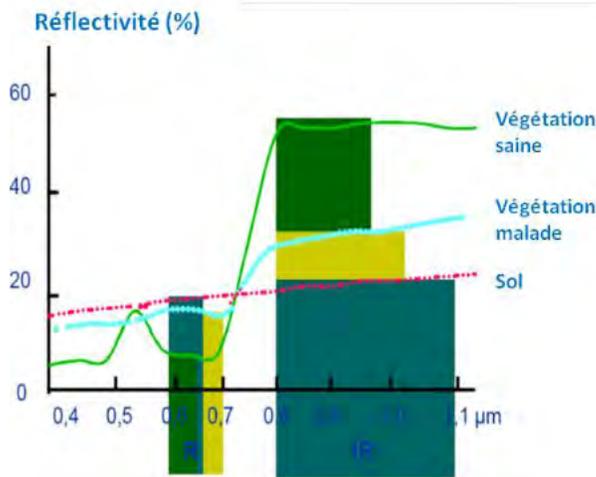


Figure 2.3 – Réflectance de la végétation au niveau du bande rouge (R) et proche infrarouge (IR) en fonction de l'état de la végétation (saine, malade).

Dans la littérature, La majorité des travaux utilisent les données des capteurs optiques, tels que Advanced Very High Resolution Radiometer (AVHRR), SPOT-Vegetation et Moderate-Resolution Imaging Spectroradiometer (MODIS), afin d'identifier les anomalies dans la croissance de la végétation (retard d'émergence, sénescence précoce, etc.). Ces capteurs ont une résolution spatiale qui varie entre 250 et 1000m. L'indice Normalized Difference Vegetation Index (NDVI) (Tucker, 1979) est l'indice de végétation principalement utilisé, en raison de sa forte corrélation avec l'activité photosynthétique du couvert végétal. Le NDVI met en valeur la différence entre la bande visible du rouge (R) et celle du proche infrarouge (PIR) ( $NDVI = (PIR - R) / (PIR + R)$ ). Cependant cet indice est démontré déficient pour la surveillance des sécheresses dans les régions semi arides de la Tunisie (Kennedy, 1989). En effet, un pourcentage de couverture végétale faible engendre un signal relativement faible et peu représentatif dans les pixels de l'image (Kennedy, 1989; Kogan, 1990). Par exemple, l'utilisation des observations satellitaires de SPOT pour la surveillance des sécheresses requiert une couverture végétale supérieure à 30%-40%. Ce seuil est nécessaire pour garantir une corrélation significative avec l'humidité du sol (Chahbi Bellakanji *et al.*, 2018). Ainsi, les mesures du NDVI présentent certaines limites, essentiellement dues à la brillance de fond du sol qui affecte les mesures du NDVI (Bausch, 1993; Huete *et al.*, 1985).

Comme expliqué avant, la détection de stress hydrique en se basant sur les réflectances

optiques de la végétation dans le visible, reste limitée à l'apparition et à la bonne détection des impacts du stress sur le fonctionnement photosynthétique de la végétation. Ce fait forme un inconvénient majeur pour une détection précoce d'un stress hydrique et pour la prévention des dégâts de la sécheresse sur la production végétale, particulièrement au niveau des zones arides à semi-arides (Smith *et al.*, 2019).

## 2. Indicateurs issus du micro-ondes

Durant les deux dernières décennies, la télédétection par micro-ondes a montré un apport considérable pour l'estimation des propriétés diélectriques du sol et par conséquent, l'estimation de l'humidité du sol (Mohanty *et al.*, 2017). Cette variable est l'une des variables du sol les plus importantes. En effet, elle établit une liaison directe entre la surface du sol et les processus atmosphériques (Babaeian *et al.*, 2019). Cette variable contrôle la partition des énergies de flux de chaleur sensible et le flux de chaleur latente essentiellement au niveau des régions caractérisées par un régime évaporatif limité par la disponibilité en eau (Entekhabi *et al.*, 2010). L'humidité du sol a souvent été considérée comme un indicateur de sécheresse agronomique. En effet, ce type de sécheresse est défini comme un déficit d'humidité dans la zone racinaire, cette dernière étant reliée directement à l'humidité de surface. D'autre part, sous des conditions de stress hydrique, les anomalies de la teneur en eau précèdent l'apparition des anomalies des indices de végétation, ce qui est primordial pour une détection précoce de la sécheresse agronomique (Babaeian *et al.*, 2019).

On distingue deux types de capteurs pour la mesure de l'humidité du sol : des capteurs passifs, lorsque la source d'énergie provient d'émetteur à la surface, et actifs lorsque la source d'énergie provient d'émetteur radar (voir figure 2.4). Les deux types de capteurs (actifs et passifs) sont sensibles uniquement aux premiers centimètres du sol (5 cm) (Pellarin *et al.*, 2006). Cependant, les applications agronomiques nécessitent d'avoir des mesures au niveau de la zone racinaire, afin de quantifier la teneur en eau disponible, nécessaire pour la transpiration et la production de biomasse (photosynthèse par exemple) (Babaeian *et al.*, 2019). Il est alors nécessaire de recourir à l'estimation du profil de l'humidité du sol à partir de celle mesurée à la surface du sol.

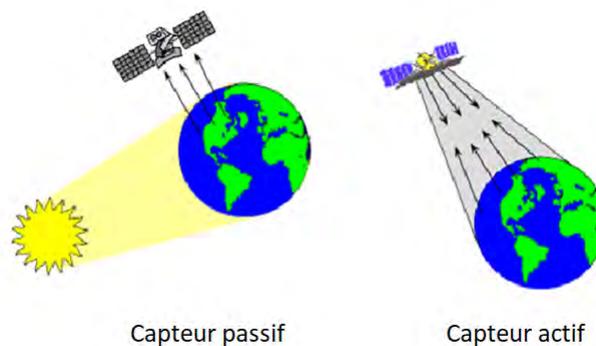


Figure 2.4 – schématisation d'un capteur actif et passif.

Les capteurs passifs et actifs offrent des mesures à différentes échelles spatiales : à l'échelle parcellaire, à l'échelle du bassin versant, à l'échelle régionale et ainsi à l'échelle continentale et mondiale (Babaeian *et al.*, 2019). Parmi les capteurs passifs, on peut citer le satellite SMOS (Soil Moisture and Ocean Salinity) (Kerr *et al.*, 2001), dédiée à la cartographie de l'humidité de surface à une très faible résolution spatiale (40 Km) (Wigneron *et al.*, 2017; Babaeian *et al.*, 2019). SMOS est lancé en Novembre 2009 et présente une revisite comprise entre 1 à 3 jours. Il permet un suivi global de la salinité des océans et l'humidité de surface en se basant sur la température de brillance ( $T_b$ ) sur la bande L (de 15 à 30 cm) (Kerr *et al.*, 2001). On cite aussi la mission Soil Moisture Active Passive (SMAP), lancée en 2015 et qui forme le capteur micro-onde le plus récent pour la mesure et la cartographie de l'humidité de surface (Babaeian *et al.*, 2019). SMAP utilise à la fois un capteur passif et actif sur la bande L, pour fournir des mesures globales de rétrodiffusion radar et de  $T_b$  à 36 km de résolution spatiale et une revisite de 1 à 3 jours (Burgin *et al.*, 2017; Babaeian *et al.*, 2019). D'autre part, on trouve la technique active (Radar) qui a apporté une forte contribution dans l'estimation de l'état hydrique du sol avec sa large gamme de résolution spatiale (Amri *et al.*, 2012). On cite par exemple European Meteorological Operational (MetOp-A, MetOp-B, et MetOp-C) avec le capteur Advanced Scatterometer (ASCAT) en bande C (de 3.75 à 7.5 cm de longueur de pénétration). MetOp-A et B ont été lancés respectivement en 2006 et 2012. MetOp-C est lancé récemment, en 2018. L'instrument a été largement utilisé pour les mesures de l'humidité de surface grâce à sa haute précision radiométrique et de calibration, à une résolution spatiale de 25 km et une revisite temporelle de 1-2 jours (Babaeian *et al.*, 2019). Ainsi, ASCAT fournit des estimations précises d'humidité de surface pour les sols nus et pour un couvert végétal épars, ce qui met en exergue l'utilité de l'utilisation des courtes longueurs d'onde (bande C) (Baghdadi *et al.*, 2012). Des résolutions spatiales encore beaucoup plus fines sont disponibles par la micro-onde active à des échelles régionales : ALOS-2 PALSAR et Sentinel-1 ayant une résolution décimétrique (6 et 10 mètres respectivement), ce qui leur confère à accéder à des échelles intéressantes pour les applications agrohydrologiques, celles de la parcelle ou du petit bassin versant. En revanche, ces satellites, comme Sentinel-1, a une période de revisite de l'ordre de 3-12 jours, ce qui n'est pas adéquat pour une analyse chronologique d'humidité du sol (Mohanty *et al.*, 2017). Le tableau 2.2, présente un résumé des avantages et inconvénients des capteurs actifs et passifs.

ASCAT fournit un indice global et journalier basé sur l'humidité de surface du sol, Soil Water Index (SWI), à une résolution spatiale de 12.5 km, utilisant les données de MetOp disponibles initialement à une résolution spatiale de 25 km. Paulik *et al.* (2014) ont évalué les estimations de SWI aux États-Unis pendant une période de 5 ans, et ils ont trouvé des bonnes corrélations avec les mesures in situ d'humidité du sol issu de International Soil Moisture Network (ISMN). De plus, ce produit a été validé à l'échelle globale par Paulik *et al.* (2012). Cet indice, développé par EUMETSAT, a été utilisé aussi pour fournir une estimation d'humidité du sol au niveau de la zone racinaire. Ce produit représente la teneur en eau au niveau de la zone racinaire le long du premier mètre du sol sous forme d'unité relative comprise entre le point de flétrissement (SWI=0) et la

Table 2.2 – Comparaison entre les capteurs passifs et actifs du micro-onde.

	capteurs actifs	capteurs passifs
Avantages	<ul style="list-style-type: none"> <li>• Disponibilité d'une large gamme de résolution spatiale</li> <li>• Disponibilité à l'échelle globale</li> <li>• Forte corrélation entre l'humidité de surface et la rétrodiffusion</li> </ul>	<ul style="list-style-type: none"> <li>• Résolution temporelle importante</li> <li>• Large disponibilité spatiale</li> </ul>
Inconvénients	<ul style="list-style-type: none"> <li>• Résolution temporelle faible</li> <li>• Signal perturbé par la rugosité de surface et la végétation</li> </ul>	<ul style="list-style-type: none"> <li>• Résolution spatiale faible, n'est pas convenable à l'échelle parcellaire</li> <li>• Signal perturbé par la rugosité de surface et la végétation</li> </ul>

capacité au champ (SWI=1). Cet indice est obtenu à partir de l'humidité de surface du sol, à l'aide d'un modèle de filtrage temporel exponentiel. Ce modèle décrit la relation entre la teneur volumétrique de l'humidité de surface ( $\theta_{surface}$ ) à une profondeur ( $Z=0$ ) et le profil d'humidité du sol à une profondeur  $L$ , en fonction du temps ( $t$ ) en se basant sur une simple équation de bilan de masse (Wagner *et al.*, 1999) (voir équation 2.1), et en tenant compte de coefficient de diffusivité ( $C$ ), qui dépend des propriétés du sol (Babaeian *et al.*, 2019).

$$L \frac{d\theta}{dt} = C(\theta_{surface} - \theta) \quad (2.1)$$

Un autre indice développé par (Narasimhan & Srinivasan, 2005), le Soil Moisture Deficit Index (SMDI), utilisant les données hebdomadaires de l'humidité du sol au niveau de la zone racinaire, est estimé à partir du ruissellement de surface simulé à partir d'un modèle hydrologique (SWAT). Cet indice a été comparé avec d'autres indices comme PDSI (Palmer, 1965) et SPI (McKee *et al.*, 1993). Il a montré une corrélation élevée avec ces deux derniers et une capacité à bien détecter les périodes de stress (Narasimhan & Srinivasan, 2005). Cependant, l'humidité du sol est difficile à évaluer avec précision à cause de sa forte variabilité spatiale et temporelle, particulièrement sous un couvert végétal (Dorigo *et al.*, 2017; Zhang *et al.*, 2013). En effet, l'humidité du sol dépend de la répartition irrégulière des précipitations, de la végétation, de la texture du sol, de la topographie de la zone étudiée et la variation de l'angle solaire (Entekhabi *et al.*, 2010). D'autre part, l'humidité du sol intègre les précipitations et l'évaporation sur des périodes de quelques jours à plusieurs semaines, de sorte qu'elle peut être moins capable de détecter un début de stress hydrique (Van Zyl *et al.*, 2003).

### 3. Indicateurs issus de l'Infra Rouge Thermique

L'utilisation de la télédétection infrarouge thermique (TIR) offre des voies prometteuses pour la détection du stress hydrique. Elle présente l'avantage de la facilité d'acquisition des données, et de fournir des produits à une résolution spatiale plus fine (Tomlinson *et al.*, 2011), ce qui est favorable pour un diagnostic hydrique. La température de surface

(LST), mesurée dans la bande infrarouge (8 et 15  $\mu\text{m}$ ), est la température radiative à la surface terrestre. Elle est estimée à partir des températures de brillance et de l'émissivité de surface sous forme de rayonnement électromagnétique (Tomlinson *et al.*, 2011). Au niveau d'une surface avec une couverture végétale dense, le LST présente la température du couvert végétal. En revanche, au niveau d'une surface avec une couverture végétale éparse, le LST présente la température moyenne de la couverture végétale et la surface du sol (Dash *et al.*, 2001).

Le LST montre généralement une variabilité spatiale et temporelle très importante, du fait de sa forte dépendance au rayonnement solaire mais aussi de la turbulence. En effet, cette dernière varie considérablement selon l'insolation et donc la couverture nuageuse, la charge d'aérosol et la vapeur d'eau présentes au niveau de l'atmosphère, (Westermann *et al.*, 2011). Elle peut également comporter de brusques variations liées aux phénomènes turbulents (tourbillons etc). Il est donc important d'acquérir les données de LST à une résolution spatiale fine pour échantillonner la variabilité spatiale mais pas trop fine pour ne pas souffrir de l'impact des phénomènes turbulents localisés (Lagouarde & Boulet, 2016). Le produit de LST peut être obtenu à partir des différents capteurs satellitaires au niveau de l'Infrarouge thermique, avec différentes résolutions spatiales et visites temporelles. On présente au niveau du tableau 2.3 une synthèse des différents satellites et capteurs associés permettant l'obtention de la variable LST.

Table 2.3 – Résolution temporelle et spatiale de chaque satellite et capteur associé effectuant des mesures de LST au niveau de l'infra rouge thermique.

Capteurs	Satellites	Résolution spatiale (m)	Revisite	Disponibilité temporelle
AVHRR	NOAA	1100	2/jour	depuis 1979
AVHRR	MetOp	1100	29 jours	depuis 2006
Thematic Mapper (TM)	Landsat 4 et 5	120	16 jours	1982-2013
Enhanced Thematic Mapper (ETM+)	Landsat 7	60	16 jours	depuis 1999
GOES Imager	GOES network	4000	géostationnaire	depuis 1974
MODIS	TERRA	1000	2/jour	depuis 2000
MODIS	AQUA	1000	2/jour	depuis 2002
ASTER	TERRA	90	16 jours	depuis 1999
AASTR	Envisat Environment	1000	35 jours	depuis 2004
SÉVÉRI	Meteosat-8	3000	géostationnaire	depuis 2005

D'après ce tableau, on observe des capteurs avec une résolution spatiale très importante comme le Landsat 7 (60 mètres), mais qui présente une répétitivité temporelle très faible de l'ordre de 16 jours. De même pour le capteur ASTER, il présente une haute résolution spatiale (90 mètres) mais une résolution temporelle faible (16 jours). D'autre part, on trouve des capteurs géostationnaires (SEVERI ou GOES imager par exemple) qui offrent une très haute résolution temporelle au dépend d'une résolution spatiale très faible et une couverture restreinte. D'autres satellites, comme MetOp et

Envisat Environment, présentent une résolution spatiale moyenne, kilométrique, mais avec une répétitivité temporelle très faible (29 et 35 jours pour MetOp et Envisat Environment respectivement). Cependant, on observe les capteurs MODIS et AVHRR NOAA présentant une résolution adéquate, à la fois sur l'échelle spatiale et temporelle, pour un suivi de stress hydrique. L'acquisition des données LST à partir de la télédétection reste limitée par le temps du passage du satellite et la restriction de l'utilisation des images acquises que pendant les jours clairs (sans couverture nuageuse), afin d'avoir des informations fiables. Ce fait a pour conséquence de diminuer le nombre des jours d'acquisition des données (Tomlinson *et al.*, 2011).

La chaleur latente forme la façon la plus efficace de dissiper l'énergie reçue par la surface (Chirouze *et al.*, 2014). Toutefois, si la disponibilité en eau est limitée, le couvert réduit son évapotranspiration et la dissipation d'énergie va être effectuée d'une manière moins efficace par le biais du flux de chaleur sensible seulement. Cette variable est quasi proportionnelle à l'écart entre la température de surface et de l'air. L'accroissement des valeurs du flux de chaleur sensible au cours d'un épisode de stress hydrique, va générer par ailleurs une hausse conséquente de la température de surface. L'écart résultant entre ces deux températures et en particulier l'élévation de la température de surface est donc un bon marqueur de stress hydrique accessible par télédétection (Lagouarde & Boulet, 2016). La température de surface (LST) est par conséquent un paramètre clé pour la surveillance intégrée de la sécheresse. Cette variable a mené par ailleurs au développement de plusieurs indices de stress hydrique en se basant sur la relation existante entre la température du couvert et son état hydrique.

Les premiers indices thermiques sont basés seulement sur la LST vue sa liaison directe au statut hydrique de la plante. On cite à titre d'exemple, l'indice Canopy Temperature (CT) (Berliner *et al.*, 1984) qui forme l'un des premiers indices thermiques utilisé pour l'évaluation de l'état hydrique de la végétation. Sa détermination nécessite uniquement la mesure de la température par radiométrie. La LST étant très corrélée au statut hydrique, cet indice l'est également, cependant, les variations de LST sont aussi dépendantes des variables météorologiques, ce qui limite la robustesse et la fiabilité de cet indice. Dans ce contexte, Idso *et al.* (1977) et Jackson *et al.* (1977) ont utilisé la différence entre la température du feuillage ( $T_c$ ) et la température de l'air ( $T_a$ ), pour développer un indice nommé Stress Degree Day (SDD). Ce dernier permet de suivre l'état hydrique de la culture et constitue le premier indice tenant compte de la correspondance entre l'évapotranspiration et la température de surface (Luquet, 2002). En effet, un stress hydrique provoque une fermeture des stomates, ce qui fait réduire la transpiration et induire un réchauffement du couvert végétal (Lili *et al.*, 1991). Ce fait va engendrer des valeurs élevées de température de la végétation par rapport à la température de l'air ambiant. Par ailleurs, des valeurs négatives ou proches de 0 de cet indice, témoignent d'une disponibilité suffisante en eau. Cependant, des valeurs supérieures à 0 de cet indice indiquent l'existence d'un déficit hydrique. Cet indice forme un bon indicateur sur le statut hydrique, mais il n'est pas suffisant comme un indicateur quantitatif (Jackson *et al.*, 1977).

Jackson *et al.* (1981) ont développé un indice Crop Water Stress Index (CWSI), basé aussi sur la différence de la température du feuillage ( $T_c$ ) et la température de l'air  $T_a$ . Cet indice relie l'équation de Penman Monteith et la différence de ces températures afin d'estimer une température maximale et minimale du feuillage. Ces températures sont associées respectivement à un minimum et un maximum du taux de transpiration. Ensuite, la comparaison de ces températures extrêmes de feuillage avec leurs mesures de températures actuelles, permet l'estimation du rapport de  $ET_r$  et  $ET_p$  (voir équation 2.2) et déduire par conséquent, l'état de stress hydrique (Moran *et al.*, 1996). En effet, les besoins en eau des cultures sont estimés comme la quantité qui permet de compenser les pertes d'eau par évapotranspiration réelle ( $ET_r$ ) (Diarra *et al.*, 2017), qui représente la quantité d'eau perdue par la plante par transpiration (Brochet, 1977). En condition de stress hydrique, la plante réduit sa transpiration  $ET_r$  par rapport aux conditions climatiques (Moran, 2004). L'intensité de stress peut être alors bien estimée à partir du rapport entre  $ET_r$  et ET potentielle ( $ET_p$ ) qui représente la quantité d'eau susceptible d'être perdue dans les mêmes conditions quand l'eau n'est pas un facteur limitant (Brochet, 1977).

$$CWSI = 1 - \frac{ET_r}{ET_p} \cong \frac{(T_c - T_a)_{min} - (T_c - T_a)_{actuelle}}{(T_c - T_a)_{min} - (T_c - T_a)_{max}} \quad (2.2)$$

L'utilisation du CWSI a obtenu un succès considérable aux États-Unis (Luquet, 2002). Cependant, cet indice est sensible à la présence du sol nu. Son application est par ailleurs, entravée par la difficulté de mesurer la température des couverts hétérogènes (Moran *et al.*, 1996). En effet, les mesures des températures de surface par les capteurs thermiques englobent aussi la température du sol  $T_s$  et non seulement  $T_c$ . Ainsi, l'amplitude des variations de LST dépend de la proportion du sol nu dans le pixel, le sol ayant une hausse de température plus importante lorsqu'il s'assèche que la végétation. Ce fait induit des fausses indications sur l'état du stress hydrique en l'utilisant pour une surface hétérogène (Luquet *et al.*, 2004). Il est par ailleurs important d'ajouter aux informations issues du thermique et de la météorologie, une information supplémentaire qui concerne la végétation.

L'indice Water Deficit Index (WDI) développé par Moran *et al.* (1994), forme une amélioration de l'indice CWSI (Jackson *et al.*, 1981) pour qu'il soit mieux applicable au niveau des régions ayant une couverture végétale éparse. Opérationnellement, WDI est l'équivalent de CWSI pour les surfaces avec une couverture végétale dense, où LST est équivalente à  $T_c$  (Moran *et al.*, 1994). Il s'agit d'un indice qui fait partie de l'approche triangle. Cette approche établit un lien relatif simple entre les températures observées au sein d'un paysage mises à l'échelle avec les températures extrêmes observées sur la même image, tels que, les pixels les plus chauds qui correspondent aux zones les plus sèches et les pixels les plus froids correspondent aux zones les plus humides. Par ailleurs, cette méthode suppose l'existence sur une même image des conditions hydriques et de couverture végétale contrastées, ce qui nous permet d'identifier la distribution relative de l'évapotranspiration et par conséquent du stress hydrique des plantes sur chaque

image (Lagouarde & Boulet, 2016). WDI est déterminé en se basant sur la construction d'un trapèze, avec les quatre sommets correspondent aux conditions extrêmes de la surface en termes du taux de couverture (exprimée par un indice de végétation comme NDVI par exemple), et de températures ((1) couvert complètement couvrant bien alimentée en eau, (2) couvert complètement couvrant sec, (3) sol nu bien alimenté en eau, (4) sol nu sec), comme présenté au niveau de la figure 2.5. Les valeurs de  $LST - T_a$  et le taux de couverture végétale varient linéairement tout au long des conditions extrêmes de la surface. Cependant, les conditions intermédiaires reliant  $LST - T_a$  et le taux de couverture végétale sont supposées incluses au niveau du trapèze (Luquet *et al.*, 2004). Pour un point C présenté au graphe 2.5, WDI est calculé comme présenté au niveau de l'équation 2.3.

$$WDI = \frac{AC}{AB} = 1 - \frac{ET_r}{ET_{pot}} \quad (2.3)$$

WDI présente des sources d'erreurs qui peuvent prévenir à partir des mesures météorologiques (Moran *et al.*, 1994). Ainsi, cette méthode ne permet pas directement de séparer les estimations des composantes de transpiration et d'évaporation à l'exception pour certaines méthodes où on introduit les mesures de l'humidité de surface du sol (Luquet *et al.*, 2004; Moran *et al.*, 1996).

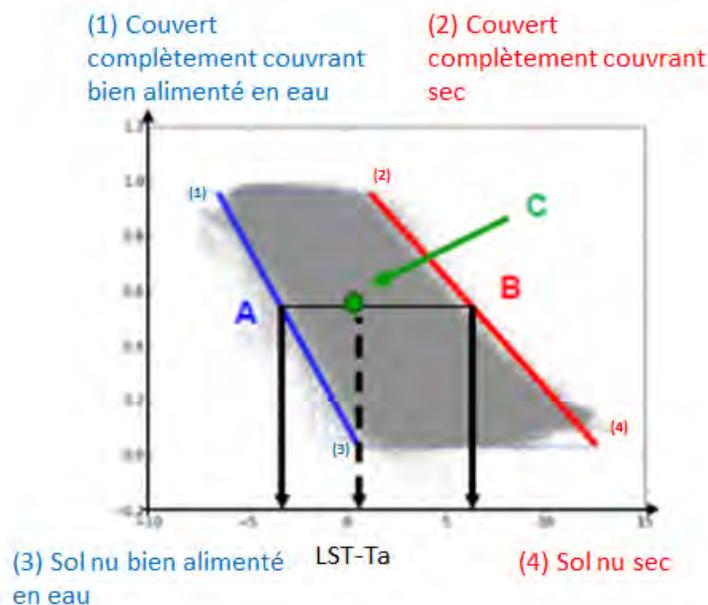


Figure 2.5 – illustration graphique pour le modèle WDI (Water Deficit Index) qui estime un facteur de stress pour le point C en fonction des températures limites A et B, identifiées par gamme de taux de couverture de la végétation. (Adaptée de Lagouarde & Boulet (2016))

Plusieurs autres indices de sécheresse qui intègrent une estimation de ET basée sur la télédétection ont été appliqués avec succès pour la surveillance de la sécheresse. On cite comme exemple, l'indice de stress évaporatif (ESI), défini par Anderson *et al.*

(2016). Cet indice est calculé à partir du rapport  $ET_r$  et ET de référence ( $ET_0$ ) en se basant sur la même formulation utilisée pour le calcul des indices CWSI et WDI (Anderson *et al.*, 2007).  $ET_r$  est simulée à partir d'un modèle de bilan d'énergie à double source Atmospheric Land Exchange Inverse (ALEXI) (Anderson *et al.*, 2007), et  $ET_0$  est calculée à partir de la méthode de Penman-Monteith, décrite par Monteith (1965) selon Allen *et al.* (1998).  $ET_0$  ne dépend que des conditions climatiques (température de l'air, humidité relative de l'air, vitesse du vent et rayonnement global). ESI a été validé en tant qu'un bon indicateur de sécheresse agricole et a été testé au Brésil, au sud des Etats-Unis d'Amérique, dans les régions méditerranéennes de l'Europe ainsi qu'en Afrique (Anderson *et al.*, 2016, 2011b). Cet indice a montré une bonne performance pour la détection précoce des périodes de sécheresse, en particulier pour les périodes de stress rapides "flash drought events", qui peuvent avoir des répercussions très sévères malgré leur courte durée (Otkin *et al.*, 2013). On trouve également l'indice de stress présenté dans Boulet *et al.* (2015), défini comme le rapport entre  $ET_r$  et  $ET_p$ . Les deux variables  $ET_r$  et  $ET_p$  sont simulées à partir d'un modèle de bilan d'énergie à double source "Soil Plant Atmosphere and Remote Sensing Evapotranspiration" (SPARSE), en utilisant respectivement un mode inverse (l'état hydrique du sol et de la végétation est inconnu) et un mode prescrit (l'état hydrique du sol et de la végétation est connu) (Boulet *et al.*, 2015). Contrairement à  $ET_0$ ,  $ET_p$  dépend à la fois des conditions météorologiques et aussi du type de la culture et son stade végétatif. Cet indice a montré une efficacité pour l'identification du stress hydrique au niveau des zones semi-arides, avec une précision sur le stress hydrique de l'ordre de 0.2 (Boulet *et al.*, 2015; Saadi *et al.*, 2018).

## 2.1.2 Indices basés sur les anomalies

L'utilisation directe des indices présentés au niveau de la section précédente 2.1.1, peut constituer un problème dans l'analyse et l'interprétation de leurs valeurs, certaines zones étant en permanence assez sèches et/ou de végétation peu dense. C'est plutôt un écart pour ces valeurs d'indice à une situation définie comme "normale" que nous essayons de mettre en évidence. De ce fait, une standardisation/normalisation de ces indices sera plus utile pour l'étude des périodes de sécheresse. Ces indices standardisés ou bien anomalies sont obtenues par une transformation statistique des valeurs initiales en se basant sur plusieurs méthodes :

- Une normalisation standard qui consiste à soustraire l'indice de sa moyenne et à le diviser par l'écart type. Ce type de standardisation suppose que l'indice provient d'une distribution normale standard. La distribution résultante a par conséquent une moyenne de 0 et un écart type de 1.
- Une normalisation basée sur les valeurs minimales et maximales consiste à amener les indices sur l'échelle de 0 à 1 en soustrayant le minimum et en divisant par le maximum de toutes les observations.
- Une normalisation par la fonction de répartition empirique consiste à amener chaque valeur d'indice à son percentile empirique et avoir par conséquent des valeurs entre 0

et 1. L'avantage de cette méthode est que chaque valeur d'indice a une interprétation relativement claire, c'est la probabilité de non dépassement de cette valeur.

Les deux premières méthodes de normalisation (normalisation standard et normalisation par le min-max) sont très utilisées dans la littérature. On présente alors quelques exemples des indices normalisés issus de ces deux méthodes.

### 2.1.2.1 Indices issus d'une normalisation basée sur le minimum et le maximum

On cite le "Vegetation Condition Index" (VCI) (Kogan, 1995), qui normalise le minimum et le maximum des valeurs inter-annuelle de NDVI dans un site donné. Il forme l'un des indices normalisés les plus utilisés pour le suivi de la sécheresse (Kogan, 2001; Seiler *et al.*, 2000). Cet indice a été développé par Kogan (1995) pour suivre les sécheresses du Mid-Ouest américain, et il a été testé sur plusieurs zones (Ex-Union soviétique, Argentine, États Unis, Chine, Afrique du Sud et Inde) afin de détecter les sécheresses à grande échelle (Kogan, 2001; Unganai & Kogan, 1998; Singh *et al.*, 2003). Ces études ont montré que cet indice normalisé est très efficace pour détecter le début d'une sécheresse, son intensité, sa durée et son impact sur la végétation. Ainsi, le VCI identifie bien les sécheresses prolongées, ayant une étendue spatiale importante, mais également, dans une moindre mesure, les sécheresses localisées apparues sur un court terme (Kogan, 1995). D'autre part, Kogan *et al.* (2004) ont utilisé le VCI pour évaluer la biomasse des pâturages et le rendement des cultures de maïs en Chine. Le VCI apporte aussi une amélioration par rapport à l'utilisation du NDVI, dans l'analyse de la condition de la végétation au niveau des zones non homogènes, caractérisées par une couverture végétale éparse (Kogan, 1990; Quiring & Ganesh, 2010). Cependant, cet indice normalisé est incapable de différencier si un stress végétal est dû à une disponibilité limitée ou bien excessive de l'eau. Il a besoin en effet, d'autres informations auxiliaires (Kogan, 1995).

Kogan (1995) a développé l'indice Temperature Condition Index (TCI) comme une amélioration de l'indice VCI, afin de permettre de différencier entre le stress de la plante causé par une humidité excessive ou bien causé par une dégradation de la disponibilité en eau. En effet, cet indice normalisé permet d'évaluer la réponse de la végétation à la température et par ailleurs à l'état hydrique. TCI est calculé en se basant sur la même formule utilisé pour VCI, en normalisant à l'aide du minimum et du maximum de la variable LST (équation 2.4).

$$TCI = 100 \frac{LST_{max} - LST}{LST_{max} - LST_{min}} \quad (2.4)$$

Avec,  $LST$ ,  $LST_{max}$  et  $LST_{min}$  sont respectivement, la température moyenne hebdomadaire, sa valeur moyenne maximale et minimale inter-annuelle

L'application de ce type des indices basés sur la normalisation des valeurs minimales et maximales de la variable de température pour TCI ou de la variable NDVI comme pour VCI est fortement liée aux valeurs d'images acquises pour une certaine période étudiée. Ce fait représente un inconvénient de l'application de cette méthode de normalisation. Notamment pour TCI, le calcul de cet indice en se basant sur les valeurs maximales et minimales de la température suppose que cet écart représente la variation maximale possible pour une période d'étude donnée. Cette hypothèse est irréaliste. En effet, la réduction de l'évaporation du

sol nu génère des températures plus élevées que pour un stress hydrique de la plante, donc la différence de LST max-min dépend du taux de couverture végétal. Ainsi, l'application des indices issus de LST doit être combinée à la prise en compte de l'impact du taux de couverture (représenté par le NDVI par exemple) sur la différence de LST.

### 2.1.2.2 Indices issus d'une normalisation standard

Amri *et al.* (2012) propose une standardisation de l'indice SWI par l'indice normalisé MAI «Moisture Anomaly Index» obtenu sur la base des statistiques effectués sur les séries temporelles de SWI. Cet indice peut fournir une représentation quantitative de l'intensité de la sécheresse et de sa durée. MAI est efficace particulièrement pour les régions non équipées de réseaux pluviométriques, essentiellement dans les régions semi-arides où un maximum inter-annuel de précipitations avec un nombre limité d'événements pourrait être en contradiction avec le stock d'eau dans le sol (Amri *et al.*, 2012).

On trouve aussi, l'indice normalisé Vegetation Anomaly Index (VAI) développé ainsi par Amri *et al.* (2011), qui est basé sur les statistiques des séries temporelles du NDVI. Cet indice a montré une bonne capacité à détecter les anomalies du couvert végétal au niveau du centre de la Tunisie, qui forme une région typiquement semi-aride (Amri *et al.*, 2011). Amri *et al.* (2011) ont montré que le VAI est plus performant que le VCI pour la détection de l'intensité de la sécheresse. En effet, VAI permet de calculer un indice normalisé par rapport à l'écart type de NDVI de chaque mois, calculé sur une longue série temporelle de NDVI (13 ans dans ce cas). De ce fait, les valeurs de cet indice sont indépendantes de la variation inter-saisonnière ou du type d'occupation du sol. Ainsi, le VAI pourrait fournir une description plus réaliste de la dynamique du couvert végétal, sans avoir besoin à d'autres informations auxiliaires, contrairement au VCI, qui est calculé seulement à partir des valeurs minimales et maximales de NDVI, qui peuvent varier selon les régions et selon le type du couvert végétal (Amri *et al.*, 2011). Toutefois, le VAI présente certaines limites liées essentiellement à sa sensibilité potentielle à d'autres effets autres que la sécheresse comme par exemple, le type d'occupation du sol. En effet, VAI est moins sensible à la sécheresse pour des zones présentant une occupation du sol d'olivier (Amri *et al.*, 2011). En effet, la verdure des feuilles de l'olivier n'est jamais influencée par l'état de stress hydrique grâce à leur résistance élevée pour un déficit hydrique. En cas de stress, cet arbre réduit sa surface foliaire (enroulement des feuilles, etc.) (Trigui, 1987), ce qui est peu visible sur des images à une résolution kilométriques.

### 2.1.3 Indices et anomalies combinés

Cette catégorie englobe les indices qui combinent différents indices ou anomalies issus de différentes longueurs d'onde. L'indice combiné le plus connu est celui élaboré par Kogan (1995). Cet indice combine les indices normalisés VCI et TCI (décrits plus haut), afin de former un indice plus robuste et plus informatif. En effet, VCI est incapable d'identifier la nature du stress hydrique, s'il est causé par un déficit hydrique ou apport excessif de l'eau. L'indice combiné, utilise alors de l'information supplémentaire, issue de TCI pour évaluer la température émise par la végétation et déterminer par ailleurs le type de stress. La somme

pondérée de ces deux indices normalisés a donné le Vegetation Health Index (VHI) défini par l'équation 2.5.

$$VHI = \alpha VCI + (1 - \alpha)TCI \quad (2.5)$$

Le poids accordé à chacun des deux indices peut différer selon les auteurs. Dans un premier temps, Kogan (1995) a choisi un poids plus important pour le VCI (0,7 contre 0,3 assigné pour le TCI). Selon lui, le NDVI utilisé dans le VCI reflète à la fois de l'information pour les températures et les précipitations. Ses travaux ont montré que cet indice combiné a été efficient pour déterminer deux types de sécheresse observées en même temps aux Etats Unis, l'une au nord midwest provoquée par des précipitations excessives qui a duré 4 mois, et l'autre au sud, provoquée par un déficit de précipitations. Cependant, d'autres études ont utilisé des poids égaux ( $\alpha=0,5$ ) pour les deux indices (Bhuiyan *et al.*, 2006; Kogan, 1997; Park *et al.*, 2018). Une étude récente avait pour objectif d'estimer les poids pour la région méditerranéenne Bento *et al.* (2018). Ils ont montré que les contributions relatives de VCI et TCI dépendaient du type de couvert végétal, avec un poids plus important pour VCI dans les régions semi-arides dépendant de la disponibilité en eau, et un poids plus important pour TCI au niveau des régions humides où l'énergie solaire présente un facteur limitant pour le développement de la végétation. Comme expliqué avant, ce type des indices n'utilise pas NDVI pour discriminer une forte différence de LST reflétée par un sol nu ou une végétation à un niveau de stress égal.

Zribi *et al.* (2021) ont testé aussi la combinaison de l'information issue de MAI et VAI dans le but d'avoir un indice plus robuste et plus complet pour la détection de la sécheresse agronomique. Cet indice "Global Drought Index" (GDI), est défini par la normalisation de la somme des indices MAI et VAI (voir équation 2.7), en fonction du niveau moyen de la végétation ou d'humidité pour un mois  $i$ . Ceci signifie par exemple, d'attribuer un poids plus élevé à l'anomalie de la végétation si la végétation est bien développée. Cet indice combine alors l'information sur la couverture végétale et sur l'humidité du sol pour définir un statut de stress hydrique.

$$Ind_i = \alpha_i VAI_i + \beta_i MAI_i \quad (2.6)$$

$$GDI_i = \frac{Ind_i - (Ind_i)_{moyenne}}{\sigma_{ind,i}} \quad (2.7)$$

Avec,  $\alpha$  et  $\beta$  sont les poids associés respectivement pour VAI et MAI pour un mois  $i$ , et  $\sigma_{ind,i}$  est l'écart type de l'Ind estimé pour un mois  $i$  durant 13 années.

Ce type d'association a été évoqué aussi par d'autres études (Mu *et al.*, 2013; Jiménez-Donaire *et al.*, 2020). Jiménez-Donaire *et al.* (2020) ont proposé de combiner le SPI, l'anomalie de l'indice de l'humidité du sol et le NDVI, pour une analyse au sud de l'Espagne. L'objectif de l'élaboration de cet indice est d'effectuer une analyse de relation cause-effet entre le déficit hydrique et le rendement agricole. En effet, cette relation identifie différents stades de sécheresse : Un manque de précipitations (phase 1), identifié par le SPI, provoque un déficit de la teneur en eau dans le sol (phase 2), identifié par l'indice basée sur la teneur en eau. Sur un long terme, ce déficit va provoquer à son tour un stress de la végétation reflété par NDVI

(phase 3), ce qui va engendrer finalement une réduction en production agricole (phase 4). Cet indice a pour rôle d'identifier quelle relation cause-effet a provoqué un évènement de sécheresse.

#### **2.1.4 Mise en regard de certains indices et anomalies appliqués aux régions semi-arides de la Méditerranée**

Les régions méditerranéennes font parties des zones les plus touchées par la pénurie d'eau dans le monde (Tramblay & Somot, 2018). Ces régions, caractérisées par un climat semi-aride, font alors l'objet de nombreux travaux pour évaluer certains indices de stress en termes de leur capacité à détecter des périodes de sécheresse. Les formules des indices sont présentées à la fin (tableau 2.4). On cite à titre d'exemple, l'application de l'indice de végétation VAI (Amri *et al.*, 2011) (voir équation dans le tableau 2.4), pour évaluer le stress hydrique au niveau de la région centrale de la Tunisie (Kairouan). Cet indice a été ainsi comparé avec d'autres indices comme VCI (Kogan, 1995) (voir 2.4). Les deux indices sont calculés sur une base mensuelle, durant la saison de pluie (entre Décembre et Mai) et pour 3 types de cultures différentes : Olivier pluvial, Pâturage et culture annuelle. Les deux indices montrent une bonne corrélation. Cependant, VAI se révèle avoir une meilleure performance en termes de mesure de l'intensité de sécheresse essentiellement pour les cultures annuelles. En effet, comme expliqué avant, VCI ne tient en compte que des valeurs minimales et maximales de NDVI et présente une différence bien significative entre les NDVI des saisons sèches et humides pour les cultures annuelles, ce qui engendre des écarts entre la valeur statistique moyenne de l'indice de végétation NDVI et le niveau moyen calculé pour l'indice VCI. D'autre part, Amri *et al.* (2012) ont comparé, au niveau de la même zone, l'utilisation de l'indice MAI (voir équation au niveau du tableau 2.4), avec l'indice SPI (McKee *et al.*, 1993) (voir équation dans le tableau 2.1) calculé à partir des stations de mesure. Une très bonne corrélation entre les deux indices a été observée particulièrement en utilisant SPI sur une période de 3 mois. Toutefois, certaines différences ont été enregistrées, qui peuvent être dues à la faible fréquence des estimations de l'humidité du sol ou aux caractéristiques du climat semi-aride qui font que certains évènements à faible précipitation n'ont pas un effet considérable sur l'humidité du sol (Amri *et al.*, 2012). Dans un autre contexte, Saadi *et al.* (2018) ont testé au niveau de la même zone du centre de la Tunisie, un indice thermique "Stress Factor" (SF), calculé à partir des simulations d'évapotranspirations en utilisant le modèle SPARSE (Boulet *et al.*, 2015), voir équation dans le tableau 2.4. Cet indice a montré une très bonne performance en comparaison avec l'indice dérivé des mesures de Scintillomètre xlas menées au niveau de la zone d'étude.

Le Page & Zribi (2019) ont réalisé une analyse spatiale plus globale au Nord-Ouest de l'Afrique, en utilisant le NDVI, le SWI et la LST issue d'un produit de composite mensuel. Une très forte corrélation inter-annuelle est observée entre le NDVI et SWI pour la plupart des régions de la zone étudiée. Cette corrélation devient plus faible au niveau des régions forestières, où les mesures satellitaires d'humidité de surface du sol ne sont pas précises, à cause de la densité de couverture végétale au niveau de ces zones, ainsi qu'au niveau des régions du sud caractérisées par une couverture végétale dispersée. Les analyses mensuelles illustrent les mêmes variations, avec plus de corrélations observées pendant la saison de culture

correspondant aux mois de Mars et d'Avril, caractérisés par une humidité du sol maximale et un développement de végétation maximal. Le Page & Zribi (2019) ont trouvé une corrélation négative entre le NDVI et LST. En effet, une évapotranspiration élevée grâce à une végétation bien verte et bien développée, induit une baisse de température de surface. A l'échelle mensuelle, une faible corrélation entre ces deux variables est observée pendant les mois d'hiver. Cette corrélation augmente davantage au niveau des régions forestières expliquée par la baisse des températures et le bon développement de la végétation (un NDVI élevé) pendant cette saison froide. De même, LST et SWI sont aussi anti-corrélées, une efficacité d'humidité du sol est accompagnée par une élévation rapide des températures de surface. Au niveau de ce travail, l'utilisation d'une donnée LST (MOD11C3) à faible résolution temporelle (pas de temps mensuel) et spatiale (0.05°) d'une part, et issue d'une synthèse d'un autre produit journalier de LST (MOD11C1) d'autre part, a été une limitation pour l'analyse de la sécheresse. En effet, comme expliqué au niveau de la sous-section 2.1.1.2, LST est caractérisée par une très forte variabilité temporelle et spatiale, d'où le besoin d'utiliser des données à résolution suffisamment fine (Westermann *et al.*, 2011). Ainsi, le produit de synthèse de LST est sensible à un effet de rétro-action de la température de l'air, ce qui amplifie l'anticorrélation entre les deux indices SWI et LST. En effet, ce produit de synthèse mensuelle de LST répond de manière lente sans rendre compte d'un dessèchement qui peut surgir pour quelques jours. Ainsi, il est plutôt corrélé avec une dynamique lente, telle que l'on peut observer pour la température de l'air par exemple : une sécheresse prolongée va induire un dessèchement et un réchauffement de l'air, et c'est plutôt cet effet qu'une synthèse mensuelle de LST va mettre en évidence.

L'indice VAI a été testé dans un autre contexte, au niveau de la Méditerranée occidentale (Zribi *et al.*, 2021). L'évaluation a été effectuée en termes d'occupation des sols pour voir l'effet de l'activité anthropique sur cet indice, essentiellement l'effet de l'irrigation, le choix des cultures et la répercussion des décisions des agriculteurs face à des périodes de sécheresse. On montre une faible corrélation entre les mesures de VAI et les zones agricoles ayant subi évidemment une intervention humaine c'est-à-dire, plus des effets anthropiques. Par exemple, pendant les périodes très sèches, les agriculteurs peuvent abandonner leurs champs au bétail, ce qui accentue la baisse du NDVI et par conséquent les impacts de la sécheresse. D'autre part, au niveau des zones agricoles, on observe une très forte corrélation de VAI au niveau des zones pluviales ou forestières et une faible corrélation au niveau des zones irriguées. Dans le cadre du même travail, Zribi *et al.* (2021) a comparé aussi la réponse de VAI et celle de MAI et GDI par rapport à différents contextes (différents saisons, occupation du sol et présence ou absence de l'irrigation). On montre une corrélation importante entre ces trois indices durant la saison pluvieuse, une faible corrélation durant la saison d'été caractérisée par une faible densité de végétation et faible stock d'eau, ainsi, une faible corrélation au niveau des zones irriguées. En effet, au niveau des zones irriguées, un apport d'eau par l'irrigation va toujours subvenir aux besoins de la végétation et maintenir une teneur en eau adéquate au niveau de la zone racinaire et par ailleurs un indice MAI bien élevé. Cependant, cet état ne peut pas être reproduit par un indice de végétation (VAI) qui va refléter une réponse en fonction de la densité de couverture végétale et leur développement qui va être important que pendant la saison de culture (généralement au printemps).

Table 2.4 – Indices de sécheresses utilisés pour la détection du stress au niveau des régions semi arides de la Méditerranée.

Longueurs d'onde	Indices	Variables d'entrée	Capteurs et résolutions	Echelle de temps	Formule.
Optique	Vegetation Anomaly Index (VAI) (Amri <i>et al.</i> , 2011)	NDVI	SPOT-VEGETATION (1000m)	mensuel	$VAI_i = \frac{NDVI_i - (NDVI_i)_{mean}}{\sigma_i}$ , avec $NDVI_i$ est le du mois $i$ ; $(NDVI_i)_{mean}$ est la valeur moyenne du NDVI au cours du mois $i$ , $\sigma$ est l'écart type des NDVI au mois $i$ , calculé sur une période de 13 ans
	Vegetation Condition Index (VCI) (Kogan, 1995; Vicente-Serrano, 2007)	NDVI	NOAA-AVHRR (1000m)	journalier	$VCI = \frac{100(NDVI_i - NDVI_{i,min})}{NDVI_{i,max} - NDVI_{i,min}}$ , avec NDVI est le NDVI lissé par semaine; $NDVI_{i,min}$ et $NDVI_{i,max}$ sont NDVI minimal et maximal observé sur plusieurs années
micro-onde	Normalized Difference Vegetation Index (NDVI) (Tucker, 1979; Le Page & Zribi, 2019)	Reflectances	MODIS MOD13Q1 (250m)	16 jours	$NDVI = \frac{PIR-R}{PIR+R}$ , avec PIR et R sont respectivement les réflectances des canaux en rouge et proche infrarouge
	Moisture Anomaly Index (MAI) (Amri <i>et al.</i> , 2012)	SWI	ASCAT ME-TOP (12.5km)	décadaire	$MAI = \frac{SWI_i - (SWI_i)_{mean}}{\sigma_i}$ , avec $SWI_i$ est l'humidité SWI pour le mois $i$ , $(SWI_i)_{mean}$ est la valeur moyenne de SWI pour le mois $i$ , $\sigma$ est l'écart type des SWI au mois $i$ , calculé sur une période de 20 ans
Thermique	Soil Water Index (SWI) (Le Page & Zribi, 2019)	humidité de surface du sol	ASCAT ME-TOP (12.5km)	journalier	$SWI(t) = \frac{\sum_{ssm(t_i)} e^{-(t-t_i)/T}}{\sum e^{-(t-t_i)/T}}$ , Avec $m_s$ est l'humidité de surface du sol estimée à l'instant $t_i$ , $T$ (=20jours) représente l'échelle de temps des variations d'humidité du sol en unités de temps.
	Stress Factor (SF) (Saadi <i>et al.</i> , 2018)	LST, albedo et NDVI	MODIS MOD11A1, MOD13A2 et MCD43B (1000 m)	journalier	$SF = 1 - \frac{LE}{LE_{pot}}$ avec LE est le flux de chaleur latente simulé par le modèle SPARSE (Boulet <i>et al.</i> , 2015), $LE_{pot}$ est le flux de chaleur latente simulé aux conditions non stressées
	Land Surface Temperature (LST) (Le Page & Zribi, 2019)	Reflectance	MODIS MOD11C3 (0.05°)	mensuel	

### 2.1.5 Modèle de bilan d'énergie pour la détermination de l'évapotranspiration

comme l'indice ESI, le rapport  $ET_r/ET_p$  représente les conditions réelles du statut hydrique de la surface et par conséquent les conditions d'un stress hydrique. On choisit par conséquent de simuler un indice de stress basé sur le rapport ( $ET_r/ET_p$ ) comme présenté au niveau des travaux de Saadi *et al.* (2018). La majorité des méthodes d'estimation de ET sont basées sur la résolution du bilan d'énergie de surface (Anderson & Kustas, 2008). Ce bilan est exprimé en fonction de la température de surface (LST) mesurée par la télédétection et le flux de chaleur latente. En effet, en partant des échanges d'énergie sur la surface d'une végétation qui se réduisent aux échanges radiatifs, convectifs (entre la surface et l'atmosphère) et aux échanges par conduction par transferts thermiques dans le sol, on peut estimer l'évapotranspiration par échange d'énergie sur le couvert végétal. Le taux d'évapotranspiration est obtenu en appliquant le principe de la conservation d'énergie. Le bilan s'exprime comme décrit par l'équation 2.8

$$LE = Rn - G - H \quad (2.8)$$

Avec,  $Rn$  est le rayonnement net,  $H$  et  $LE$  les flux de chaleur sensible et de chaleur latente respectivement, et  $G$  le flux de conduction dans le sol.

Les échanges radiatifs s'effectuent à la fois dans les domaines des courtes longueurs d'onde (ou domaine solaire entre 0,3 et 5,0  $\mu\text{m}$  environ) et des grandes longueurs d'onde (ou domaine infrarouge thermique IRT entre 4 et 50  $\mu\text{m}$  environ). Le rayonnement solaire atteignant la surface résulte d'une composante directe provenant du soleil, et d'une composante diffuse (par les gaz, les aérosols, etc). Aussi, le rayonnement solaire est en partie réfléchi avec un coefficient  $a$  : albédo, qui dépend de plusieurs paramètres : nature de la végétation, structure de la canopée, type de sol, présence de neige, etc. (Lagouarde & Boulet, 2016). D'autre part, dans le domaine de l'IRT, le sol émet selon la loi de Stefan-Boltzmann un rayonnement  $L \uparrow = \epsilon\sigma T_s^4$ , avec  $T_s$  est la température de surface,  $\epsilon$  est l'émissivité et  $\sigma$  est la constante de Stefan-Boltzmann. De ce fait, tenant compte de domaine solaire et thermique, le bilan radiatif disponible au niveau de la surface s'écrit comme présenté dans l'équation 2.9 et aussi d'une façon schématique au niveau de la figure 2.6.

$$Rn = Rg(1 - a) + \epsilon(L \downarrow - \sigma T_s^4) \quad (2.9)$$

Les échanges convectifs caractérisent les transferts de chaleur et de masse par transfert turbulent. Le flux de chaleur sensible ( $H$ ) (voir équation 2.10) dépend de la température de l'air au sein de la canopée (température aérodynamique) et de la résistance aérodynamique entre le niveau aérodynamique et la hauteur de référence  $Z_a$  au niveau de laquelle les variables météorologiques sont observées ou simulées (généralement à 2 m pour les couverts herbacés, plus haut pour les ligneux évidemment). Le flux de chaleur latente ( $LE$ ) est l'énergie correspondant au flux d'évapotranspiration, i.e. la densité de flux de vapeur multipliée par la chaleur latente de vaporisation de l'eau  $L$ . Cette énergie représente la quantité nécessaire au changement d'état et au transport à la fois de l'évaporation de l'eau présente sur les espaces physiques (sol, feuilles, eau libre) et de la transpiration foliaire (Luquet, 2002). Elle est

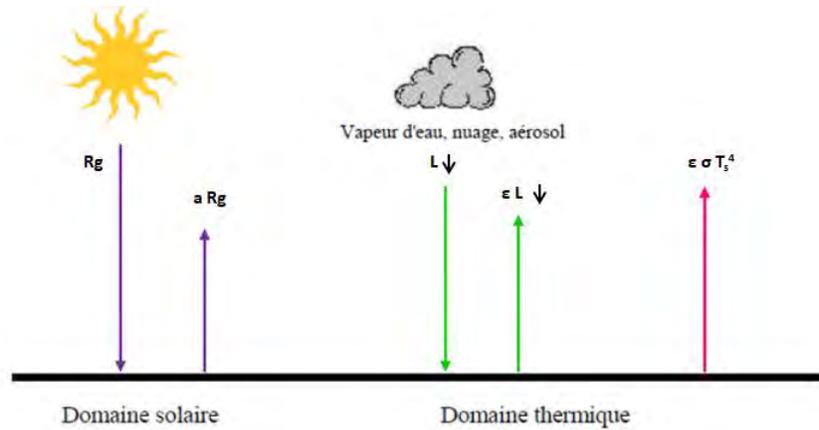


Figure 2.6 – Echanges radiatifs simplifiés terre-atmosphère.

exprimée par l'équation 2.11.

$$H = \rho C_p \frac{T_0 - T_a}{r_a} \quad (2.10)$$

Avec,  $\rho$  et  $C_p$  désignent respectivement la masse volumique ( $\text{kg m}^{-3}$ ) et la chaleur spécifique ( $\text{J kg}^{-1} \text{ } ^\circ\text{C}^{-1}$ ) de l'air,  $T_a$  et  $T_0$  sont la température de l'air au niveau de référence et au niveau aérodynamique respectivement, exprimé en  $^\circ\text{K}$  et  $r_a$  est la résistance aérodynamique ( $\text{s m}^{-1}$ ).

$$LE = \frac{\rho C_p}{\gamma} \frac{e_s(T_0) - e_a}{r_a + r_c} \quad (2.11)$$

Avec,  $e_s(T)$  représente la courbe de pression saturante de la vapeur d'eau (Pa) en fonction de la température  $T$ ,  $e_a$  est la pression partielle de vapeur de l'air (Pa) au niveau de référence choisi au-dessus du couvert,  $\gamma$  est la constante psychrométrique voisine de  $66 \text{ Pa } ^\circ\text{C}^{-1}$ .  $r_c$  et  $r_a$  désignent la résistance au transfert de vapeur d'eau et la résistance aérodynamique respectivement.

Finalement, sachant que le flux de chaleur dans le sol dépend essentiellement de la fraction du rayonnement solaire interceptée par la végétation, ce dernier est exprimé proportionnellement au rayonnement net échangé au niveau du sol par la fraction  $\xi$  (voir l'équation 2.12) qui représente le rapport  $G$  sur  $R_n$ .

$$\xi = f_c A_c + (1 - f_c) A_s \quad (2.12)$$

Où,  $f_c$  est le taux de couverture végétale (dédié généralement du NDVI),  $A_c$  le rapport typique pour un couvert dense et  $A_s$  celui pour un sol nu.

Ainsi, l'équation 2.8, peut être exprimée en fonction de la température de surface en remplaçant chacun des termes dans l'équation 2.8 par les expressions des variables du transfert radiatif, convectif et conductif. L'équation du bilan d'énergie peut être présentée comme

montré dans l'équation suivante :

$$LE = (1 - \xi)(Rg(1 - a) + \epsilon(L \downarrow - \sigma T_s^4)) - \rho C_p \frac{T_0 - T_a}{r_a} \quad (2.13)$$

Les méthodes de bilan d'énergie pour l'estimation de l'évapotranspiration reposent sur deux types d'approches : les approches "single pixel" et les approches "contextuelles" (Lagouarde & Boulet, 2016). La méthode de triangle présentée par le graphique 2.5 au niveau de la section 2.1.1.2, fait partie des approches contextuelles très bien exploitées pour des couverts épars et du sol nu.

### 2.1.5.1 Approches contextuelles

Les méthodes contextuelles forment les approches qui utilisent les contrastes thermiques d'une image satellitaire, en parcourant les différents pixels, pour fournir une estimation instantanée du condition de stress hydrique au passage du satellite (Chirouze *et al.*, 2014). Par ailleurs, ce type d'approche va nous permettre de tenir en compte de l'impact de la température de surface en fonction d'une large gamme de végétations, des propriétés de surface observées au sein de la même image satellitaire (densité de végétation, rugosité, nature de la surface, etc.), et évidemment de différents contrastes d'humidité du sol (Chirouze *et al.*, 2014). D'autre part, la plupart des méthodes contextuelles utilisent assez peu de variables d'entrée. En particulier, ces méthodes n'utilisent pas la température de l'air dont l'échelle spatiale représentative (mesure locale ou au contraire estimation régionale des réanalyses) est rarement compatible avec l'échelle de la température de surface, ce qui réduit par conséquent l'erreur absolue associée aux températures (Chirouze *et al.*, 2014). Les méthodes contextuelles sont par ailleurs moins sensibles à la précision absolue du champ de variables biophysiques (LST, réflectances etc) car basées sur des différences au sein de l'image (Lagouarde & Boulet, 2016). Les modèles contextuels les plus répandus sont : "Surface Energy Balance Algorithm for Land model" (SEBAL) (Bastiaanssen *et al.*, 1998), "Mapping EvapoTranspiration with High-resolution and Internalized Calibration" (METRIC) (Allen *et al.*, 2007), "Triangle method" (Carlson, 2007) et "Simplified Surface Energy Balance Index" (S-SEBI) (Roerink *et al.*, 2000). Toutefois, les modèles contextuels sont moins adaptés pour l'utilisation pour des paysages homogènes et uniformes (Chirouze *et al.*, 2014). En effet, cette méthode suppose l'existence de toutes les conditions d'humidité du sol au sein d'une même image, ce qui n'est pas toujours vrai. Cette hypothèse n'est pas valide aussi lorsque l'image est acquise en tout début ou en toute fin de ressuyage après une pluie (Lagouarde & Boulet, 2016). Ainsi, les sols nus humides ou une végétation entièrement stressée ne sont pas toujours présents sur une seule image, en particulier dans les zones agricoles irriguées avec un approvisionnement en eau suffisant (Chirouze *et al.*, 2014). D'autre part, en utilisant des images avec des résolutions spatiales moyennes (kilométriques par exemple), l'image est moins contrastée. Un seul pixel peut contenir plusieurs parcelles, ce qui fait dissoudre l'information réellement observée. Tous ces faits, ne permettent pas d'avoir suffisamment de contrastes pour calculer un stress relatif fiable et font que ce modèle soit difficilement applicable pour des basses résolutions. En revanche, on trouve que Rajasekaran *et al.* (2017) ont montré dans leurs travaux pour l'estimation de l'évaporation en utilisant la méthode de triangle que l'utilisation des données

de température de surface désagrégées a engendré plus d'erreurs par rapport à l'utilisation des données LST disponibles originellement à une résolution kilométrique. En effet, l'utilisation de LST désagrégée pour la méthode de triangle a perturbé la position relative des pixels dans le contexte de liaison LST/NDVI.

### 2.1.5.2 Approches de single source

Contrairement aux modèles contextuels, les modèles de single pixel utilisent l'information de chaque pixel indépendamment des autres pixels de l'image. Ils résolvent l'équation du bilan d'énergie pour chaque pixel d'une façon séparée (Chirouze *et al.*, 2014). Ces méthodes simulent l'évapotranspiration comme résidu du bilan d'énergie (voir l'équation 2.8). L'avantage de ces modèles tient au fait qu'ils sont adaptés aux milieux homogènes et à l'utilisation des informations à basses résolutions. En effet, les données à basse résolution contiennent des pixels mixtes qui couvrent différentes couvertures végétales et par ailleurs différents niveaux de contraste de NDVI et d'humidité du sol. Ces caractéristiques permettent de générer des moyennes effectives d'humidité et de végétation au niveau du pixel, ce qui permet de les utiliser pour produire des cartes globales d'évapotranspiration (Chirouze *et al.*, 2014).

On distingue deux types de modèles : modèle "mono source" et modèle "double source".

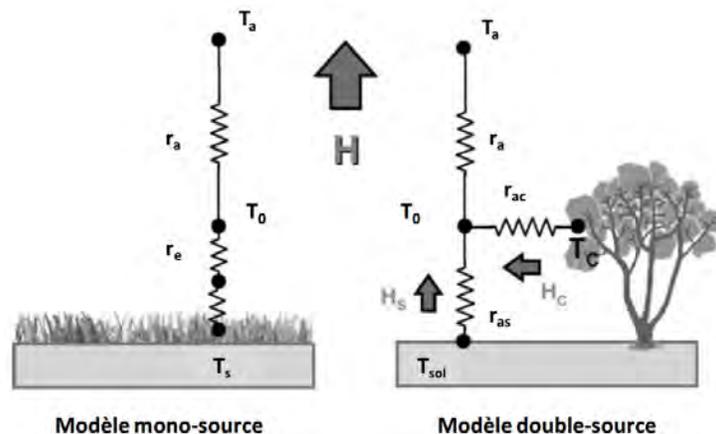


Figure 2.7 – Schémas résistifs mono- et double-source :  $r_{as}$  et  $r_{ac}$  forment les résistances aérodynamiques des transferts respectivement du composite sol/air et végétation/air.  $T_{sol}$  et  $T_C$  représentent les températures du sol et de la canopée, successivement.  $H_s$  et  $H_c$  forment les flux de chaleur sensible issus du sol et de la canopée, successivement.

#### 1. Modèles Mono source

Le type "mono source" considère la surface dans son ensemble pour le calcul de l'évapotranspiration (Lagouarde & Boulet, 2016). Pour ces modèles, ET est calculée en tant qu'une variable résiduelle de l'équation du bilan d'énergie 2.8. Il s'agit plutôt de calculer la variable de flux de chaleur sensible ( $H$ ) en se basant sur l'équation 2.14. Ce modèle utilise des relations simples et robustes pour relier la température aérodynamique et la

température de surface LST (Boulet *et al.*, 2012). Cependant, cette dernière est en général plus élevée que la température aérodynamique ( $T_0$ ), parce qu'elle intègre tous les éléments de surface, en particulier le sol nu qui est plus chaud. Substituer LST à  $T_0$  revient soit à pondérer H par un facteur correctif  $\alpha$ , soit à rajouter une résistance additionnelle  $r_e$  (voir équation 2.14), appelée aussi le kB-1 radiométrique (Boulet *et al.*, 2012).

$$H = \rho C_p \alpha \frac{LST - T_a}{r_a} = \rho C_p \frac{LST - T_a}{r_a + r_e} \quad (2.14)$$

Plusieurs modèles proposent une paramétrisation de  $\alpha$  et  $r_e$  en fonction notamment de l'indice foliaire. C'est le cas du modèle Surface Energy Budget System SEBS (Su, 2002), par exemple. Ces méthodes sont appliquées pour les couverts homogènes sans la distinction entre l'évaporation du sol et la transpiration des plantes (Lagouarde & Boulet, 2016). Cependant, pour de nombreuses applications éco-hydrologiques ou agronomiques, une évaluation des composantes «transpiration» et «évaporation» est utile.

## 2. Modèles à double source

Les modèles "double source" permettent d'estimer d'une façon séparée les composantes transpiration des plantes (T) et évaporation du sol (E). En effet, ces modèles traitent ET comme la somme des flux de E et T (Norman *et al.*, 1995). Ils utilisent les températures d'équilibre du sol ( $T_s$ ) et de la végétation ( $T_v$ ), considérées comme les principales sources de chaleur. Cette estimation séparée des deux composantes E et T forme un apport considérable pour l'estimation des flux d'énergie au niveau des zones semi-arides qui sont caractérisées par un couvert végétal épars. Ainsi, les modèles à double source fournissent une description plus réaliste des flux d'eau et de chaleur pour ces couverts, contrairement aux modèles mono source qui estiment le kB-1 radiométrique d'une façon empirique la plupart du temps. D'autre part, les modèles à double sources estiment  $T_s$ , puis  $T_v$  pour ensuite estimer  $T_0$ , ce qui est plus robuste selon Chirouze *et al.* (2014).

Cependant, résoudre deux bilans, l'un pour le sol et l'autre pour la végétation en utilisant une seule information de température (température moyenne de la surface) produit un problème d'indétermination. Le modèle Two-Source Energy Balance (TSEB) (Kustas & Norman, 1999; Norman *et al.*, 1995) propose de rajouter une hypothèse afin de réduire le nombre d'inconnus. Ce modèle suppose que la végétation a suffisamment d'eau dans la zone racinaire. A ce stade, on utilise la température de surface pour résoudre l'équation de bilan d'énergie pour la composante du sol et estimer par conséquent le flux de chaleur latente du sol  $LE_s$ . Si la végétation présente des conditions de stress hydrique, l'évaporation résultante  $LE_s$  diminuera à des niveaux irréalistes (flux négatifs) afin de maintenir la même température de surface. Dans ce cas, on assume que le sol est sec et on lui accorde une valeur minimale ( $LE_s = 0$ ). Ensuite, le bilan d'énergie est résolu pour la composante de la végétation  $LE_v$ . Si le flux de chaleur latente correspondant à la végétation est aussi négatif, dans ce cas, on assume une condition de stress à la fois pour le sol et pour la végétation (Boulet *et al.*, 2015).

Une généralisation du modèle TSEB appelé SPARSE : "Soil Plant Atmosphere and Remote Sensing Evapotranspiration" est développée par Boulet *et al.* (2015). Ce modèle apporte une linéarisation de l'ensemble des équations du bilan énergétique et des expressions de résistances aérodynamiques. SPARSE forme un modèle mécaniste pour l'estimation à chaque pas de temps des principaux transferts verticaux au niveau de la plante et du sol d'une façon séparée.

### 2.1.6 Méthodes de descente d'échelle

Pour une analyse des caractéristiques des périodes de sécheresse, on aura besoin d'analyser une série des données assez profondes et suffisantes afin de comprendre le mécanisme d'occurrence de ce phénomène climatique. La simulation de notre indice de stress est effectuée par les modèles de bilan d'énergie à double source. Ce dernier aura besoin des données météorologiques à une petite échelle temporelle et spatiale. En effet, les processus hydrologiques se produisent généralement à des échelles spatiales fines (Maraun *et al.*, 2010). Les données météorologiques disponibles au niveau des réseaux de mesure, sont la plupart du temps insuffisantes pour les besoins des modèles d'impact (Ayar *et al.*, 2016). On a recours alors aux modèles climatiques globaux qui offrent une disponibilité temporelle importante, bien appropriée pour ce genre d'analyses. Les modèles climatiques forment des outils importants dans les travaux de recherche afin de déterminer l'évolution du climat actuel (Herrera *et al.*, 2006; Maraun *et al.*, 2010). Ces modèles simulent des processus physiques à haute résolution en se basant sur la dynamique à grande échelle qui inclut les océans, les processus de surface terrestre et d'interface océan-glace, ainsi que des paramétrisations physiques (essentiellement des équations hydrostatiques ou non hydrostatiques de la mécanique des fluides) (Ayar *et al.*, 2016; Herrera *et al.*, 2006). Toutefois, ces modèles sont très coûteux en temps de calcul, et la plupart de ces modèles climatiques fournissent des informations à une résolution spatiale de plusieurs dizaines de Km (Maraun *et al.*, 2010). Ces résolutions spatiales restent encore limitées et inadéquates pour les modèles d'impact qui sont souvent très sensibles aux variations locales du climat. Aussi, les données issues des modèles climatiques sont généralement biaisées par rapport aux observations locales (Ayar *et al.*, 2016). Par ailleurs, une étape intermédiaire nommée la "descente d'échelle" doit être mise en oeuvre, afin de simuler des variables à une résolution adéquate pour les modèles d'impacts, à partir des données des modèles climatiques (voir figure 2.8).

L'approche de descente d'échelle consiste à établir une relation statistique entre les variables à large échelle (les prédicteurs), issues généralement des sorties des modèles climatiques globaux ou régionaux, et l'ensemble des variables locales (les prédicteurs) qui sont dans notre cas les données météorologiques issues des observations à une échelle locale (Herrera *et al.*, 2006; Wilby *et al.*, 2004). Cette relation est supposée valide pour n'importe quelle période temporelle et non pas seulement sur la période où la relation est calibrée (Ayar *et al.*, 2016). Cette relation est ensuite utilisée pour générer des données climatiques locales bien étendues dans le temps et ayant une petite échelle spatiale.

Au cours des dernières décennies, plusieurs études de recherche ont été menées sur les méthodes et les applications de la descente d'échelle (Maraun *et al.*, 2010). Généralement,

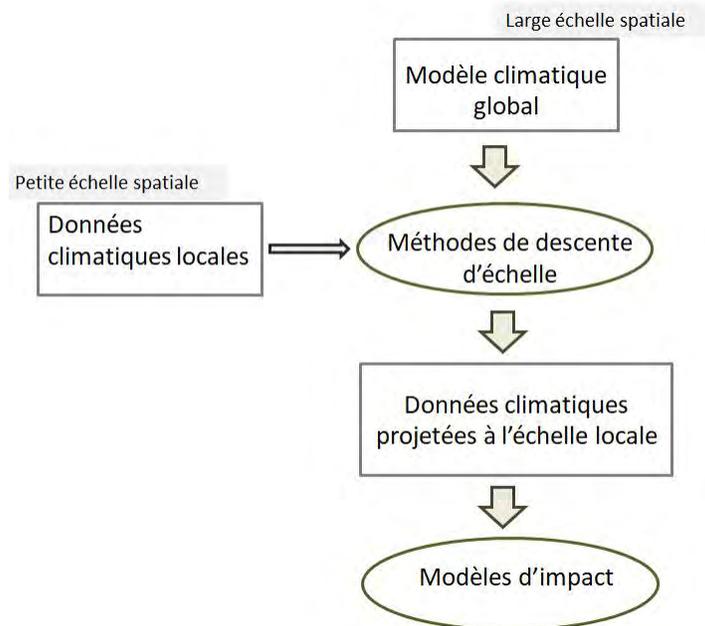


Figure 2.8 – Schématisation de l'application des méthodes de descente d'échelle pour fournir des données à une échelle adéquate pour les modèles d'impact.

ces méthodes se scindent en deux familles : Méthodes de *perfect-prog* et Méthode de *model output statistics*.

### 2.1.6.1 Méthodes Perfect-prog

Les méthodes Perfect-prog établissent des relations entre les variables à large échelle disponibles sur une longue période historique et les variables locales, en conservant la synchronicité temporelle observée originalement (Klein, 1982; Carter *et al.*, 1989). Parmi ce type d'approches, on trouve essentiellement : les fonctions de transfert, les méthodes des analogues et les générateurs des conditions météorologiques (Wilby *et al.*, 2004).

1. **Fonctions de transfert** Parmi les approches statistiques classiquement utilisées par les hydrologues se trouve les méthodes de fonction de transfert. Ces méthodes permettent une modélisation "simple" des relations entre la variable à échelle locale ( $Y$ ) et la variable à large échelle ( $X$ ) en utilisant des fonctions déterministes exprimées par l'équation 2.15, avec,  $Y$  est la variable à simuler à partir des  $N$  variables à large échelle  $X_j$ , et  $f_j$  est la fonction déterministe à appliquer pour modéliser la relation entre  $X_j$  et  $Y$  c'est-à-dire, entre la variable à large échelle et la variable locale.

$$E(Y|X_1...X_N) = \sum_{j=1}^N f_j(X_j) \quad (2.15)$$

Dans la littérature, plusieurs fonctions ont été appliquées pour exprimer la relation entre les variables  $X$  et  $Y$ . On trouve par exemple les méthodes de régression linéaire ou non

linéaire (Wilby *et al.*, 1998), qui sont parmi les méthodes les plus préliminaires (Wilby & Wigley, 1997a). Les méthodes de régression non linéaires sont particulièrement utilisées quand la relation entre les variables climatiques à grande échelle et les variables locaux est non linéaire (Herrera *et al.*, 2006). On trouve aussi, l'analyse en composantes principales (Friston *et al.*, 1993; Wackernagel, 1995), pour l'analyse statistique multivariée des données atmosphériques. On cite aussi, l'analyse de corrélation canonique (Von Storch *et al.*, 1993) qui forme une méthode complexe et peu utilisée en météorologie mais qui a été utilisée dans les travaux de Von Storch *et al.* (1993), pour relier les pluies d'hiver observées dans la Péninsule Ibérique, avec le niveau de pression de la mer mesuré au nord de l'Atlantique. On trouve aussi la méthode de krigeage (Wackernagel, 1995), basée sur la structure spatiale de la covariance estimée des données observées, et les méthodes de réseaux de neurones artificiels (RNA) (Liu *et al.*, 2008), qui sont en plein essor grâce à leur performance pour modéliser plusieurs variables reliées par des relations non linéaires (Herrera *et al.*, 2006), les méthodes de RNA sont généralement plus puissantes que les autres techniques, bien que l'interprétation de ses relations dynamiques soit moins facile. D'une façon globale, les méthodes du fonction du transfert sont simples à manipuler et à appliquer. Cependant, elles ont tendance à sous estimer la variance (Wilby *et al.*, 2002; Ayar *et al.*, 2016).

2. **Méthode des analogues** Cette méthode fait partie des méthodes les plus simples pour la descente d'échelle. Elle a été intégrée dans le domaine de descente d'échelle par Zorita & von Storch (1997). Cette méthode repose sur l'hypothèse que la circulation générale de l'atmosphère est un mécanisme physique unique, ayant un cycle de développement continu et dépendant des conditions initiales des données météorologiques observées. Par ailleurs, on suppose que si on trouve deux situations historiques semblables, leurs évolutions météorologiques devrait être aussi semblables (Herrera *et al.*, 2006). D'une façon plus concrète, chaque variable journalière issue des données à large échelle est associée à un jour analogue observé historiquement parmi les observations locales, ayant des conditions similaires selon des critères de champs atmosphériques prédéfinis (Zorita & von Storch, 1997). La difficulté de l'approche repose sur le choix de l'analogue et sa validation, ainsi, la nécessité d'avoir une série temporelle des observations assez longue et suffisante (Zorita & Von Storch, 1999). D'autre part, en contexte du changement climatique, cette méthode peut ne pas se rendre compte d'un signal de changement climatique probable qui n'a pas été observé dans la période passée (Ayar *et al.*, 2016). En effet, les méthodes des analogues sont le résultat d'un re-échantillonnage, elles sont incapables alors, de simuler des situations qui diffèrent de celles des observations. Cela peut être un inconvénient majeur lorsque les périodes d'observation ne sont pas assez longues pour contenir tous les évènements climatiques.
3. **Générateur stochastique** La méthode du générateur stochastique est basée sur la modélisation de la distribution des variables météorologiques à petite échelle. Son objectif est de produire des séries temporelles en tenant compte des propriétés statistiques et de la variabilité des observations (Wilks & Wilby, 1999). En effet, les générateurs de conditions météorologiques stochastiques fournissent des scénarios à haute résolu-

tion temporelle et spatiale basés sur des informations à basse résolution provenant des modèles climatiques globaux (sur des périodes futures ou passées) ou des réanalyses (sur des périodes passées). Ces modèles sont habituellement utilisés quand les données météorologiques sont inadéquates quant à leur profondeur temporelle, complétude ou extension spatiale. Ils sont également utilisés pour combler les valeurs manquantes au niveau d'une série climatique observée (Wilby *et al.*, 2004). On trouve par exemple le modèle "WGEN" de Richardson (1981) qui est le plus utilisé pour les études des impacts climatiques (Wilby & Wigley, 1997a). Ce modèle simule les taux des précipitations journalières, les températures minimales et maximales et la radiation solaire en se basant sur l'occurrence de la pluie. Ce modèle "WGEN" a été adapté pour des nombreuses études de changements climatiques et étude d'impacts, comme par exemple dans Wilks (1989). Plus récemment, Chandler (2020) a réussi à reproduire des séries climatiques journalières multi-variables et multi-sites par son modèle "Rglimclim". L'approche est basée sur des modèles linéaires généralisés (GLM). Chaque variable est simulée selon sa propre distribution en conservant la dépendance inter-variable et inter-site afin d'assurer une représentation réaliste et cohérente des séries simulées.

La méthode du générateur stochastique peut générer non seulement une extension temporelle des variables locales, mais aussi une extension spatiale (Wilby & Wigley, 1997a). Également, des informations à faible résolution provenant des modèles globaux ou de réanalyses, peuvent être introduites sous forme de covariables qui agissent sur les fonctions de distribution. Cependant, les générateurs stochastiques sont caractérisés par une mise en oeuvre qui peut s'avérer complexe (Wilby & Wigley, 1997a).

### 2.1.6.2 Model output statistics MOS

Les simulations des modèles climatiques présentent des biais qui empêchent leur application directement pour les études d'impacts (Turco *et al.*, 2017). La correction de ces biais qui sont les différences systématiques entre une variable d'intérêt à grande échelle et une variable locale (Cannon, 2018), est une étape primordiale avant l'application des modèles d'impact, par le biais des méthodes de "Model Output Statistics" (MOS). Ces méthodes établissent une relation statistique entre les variables à large échelle et les variables locales observées afin de corriger les propriétés statistiques des variables à large échelle, pour qu'elles se rapprochent des propriétés statistiques des variables locales (moyenne, variance, probabilité de distribution) (Maraun *et al.*, 2010; Turco *et al.*, 2017; Ayar *et al.*, 2016). Les méthodes MOS utilisent l'information sur la distribution de la variable observée (moyenne et variance) pour bien reproduire la distribution de la variable simulée (Maraun *et al.*, 2010). Ainsi, ces méthodes, contrairement aux méthodes de Perfect-prog peuvent être appliquées directement sur les sorties des modèles climatiques (Ayar *et al.*, 2016), comme un post-traitement de ces simulations, sans chercher à reproduire la synchronicité temporelle (Carter *et al.*, 1989).

De nombreux travaux de recherche ont utilisé les méthodes MOS pour corriger les variables à large échelle et les rendre exploitables pour une évaluation des impacts (Guo & Senior, 2006; Themeßl & Gobiet, 2010). Les méthodes les plus répandues sont les méthodes de correction de biais et les méthodes de type "delta" ou "factor change" (Arnell & Reynard,

1996). La méthode delta consiste à superposer le changement climatologique moyen simulé par les données à large échelle aux observations actuelles. Cette méthode assume qu'un biais  $\delta$ , qui existe entre les simulations et les observations, est identique entre les conditions climatiques futures et actuelles (Willems & Vrac, 2011). Cette méthode permet de prendre en compte uniquement le changement de moyenne des variables à large échelle, et pas les changements de variabilité et de distribution (Boe, 2007). D'autre part, les méthodes de correction de biais sont appliquées pour transformer les données à large échelle en une nouvelle série climatique ne comportant pas, ou peu de biais statistique par rapport aux variables locales observées (Vrac & Friederichs, 2015). Les biais sont calculés entre une variable d'intérêt à grande échelle et une variable locale dans une période de référence et vont être supprimés de toute la période d'étude (Cannon, 2018), de sorte que la simulation corrigée soit plus précise et plus robuste (Xi *et al.*, 2017). Cette approche cherche alors à opérer une correction en distribution en réduisant le biais statistique des données simulées (Vrac & Friederichs, 2015). Par exemple, si la variable de faible résolution spatiale a une moyenne systématiquement plus faible que la variable observée, on corrigera la variable de faible résolution spatiale pour que sa moyenne corresponde à celle des observations. Pour un exemple pratique, figure 2.9, supposons que la moyenne de température de la variable observée est de  $22^{\circ}\text{C}$  et que celle de la variable à large échelle est de  $18^{\circ}\text{C}$ , la correction à appliquer à la variable à basse résolution est de  $22^{\circ} - 18^{\circ} = 4^{\circ}\text{C}$ . Ce faisant, la temporalité de cette variable n'est pas affectée, la série est simplement augmentée du biais de  $4^{\circ}\text{C}$ .

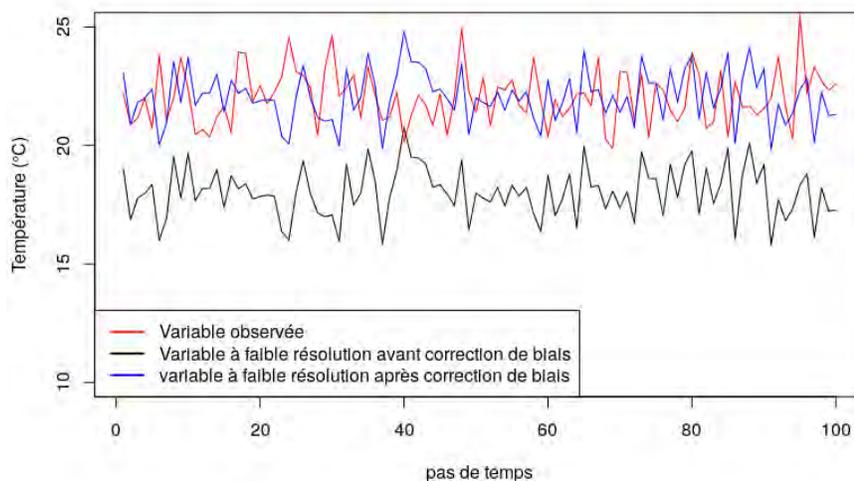


Figure 2.9 – Exemple illustratif de l'effet de correction de biais sur une température à faible résolution spatiale par rapport à une température observée.

Des efforts considérables effectués ont été récemment pour le développement des méthodes de correction de biais (Vrac & Friederichs, 2015; Cannon, 2018; Michelangeli *et al.*, 2009). On trouve essentiellement les méthodes de correction univariée (une seule variable climatique en un seul site) ou multivariée (plusieurs variables climatiques et/ou plusieurs sites).

Dans le cadre univarié, les méthodes de correction de biais ont été appliquées et discutées depuis de nombreuses années. Toutefois, si on opère une correction univariée variable par variable, on risque de perdre la cohérence et la dépendance inter-variable (Cannon, 2018). Cependant, les méthodes de correction multivariées assurent le transfert de toutes les caractéristiques statistiques d'une distribution multivariée observée (comme la corrélation entre les variables observées) vers les variables simulées correspondantes (Cannon, 2018).

Les méthodes MOS de descente d'échelle sont efficaces pour la correction de l'intensité de biais entre la variable locale et celle à large échelle et pour l'ajustement des fréquences des évènements (nombre des jours humides par exemple). Toutefois, ces méthodes sont incapables d'améliorer la structure temporelle, comme par exemple pour la détermination de la longueur exacte des périodes humides (Maraun *et al.*, 2010). Ceci forme un inconvénient majeur pour l'utilisation de ces méthodes pour des études agronomiques et hydrologiques, où le facteur temporelle est primordial pour la croissance végétative et l'estimation de l'humidité de surface (Zhang, 2013).

### 2.1.6.3 Revue sur les travaux de comparaison des différentes méthodes

Durant les dernières décennies, de nombreux travaux de recherche sur les méthodes de descente d'échelle ont été effectués (Maraun *et al.*, 2010). Toutefois, on trouve rarement des travaux qui comparent toutes les méthodes à la fois. Dans ce contexte, on trouve les travaux de Ayar *et al.* (2016), qui ont comparé plusieurs méthodes de descente d'échelle issues des différentes familles, de la famille de Perfect-prog (méthodes de transfert, générateur stochastique de conditions météorologiques et méthode des analogues) et des méthodes de la famille MOS. Les précipitations forment une variable d'intérêt pour différents travaux dans la littérature, vu qu'elle forme la variable la plus importante pour les modèles d'impacts (pour l'étude des inondations, l'estimation des rendements, etc.) et elle est caractérisée par une variabilité temporelle et spatiale très importante. Pour cette comparaison, les auteurs ont utilisé la même information à large échelle (les données ERA interim), appliquée sur la même région et sur la même période d'étude qui s'étale entre 1989 et 2008. L'évaluation de ces méthodes est assurée en termes de capacité à reproduire différents aspects : l'occurrence, l'intensité et la distribution spatiale et temporelle des précipitations par rapport aux données d'observation. Pour l'occurrence, les auteurs ont évalué seulement la méthode des analogues et autres méthodes issues des fonctions de transfert. Les résultats montrent que les méthodes des fonctions du transfert surestiment le pourcentage des jours pluvieux. Cependant, les méthodes des analogues semblent être la meilleure méthode pour une bonne estimation d'occurrence de la pluie et la détermination par ailleurs, de la moyenne des jours secs et humides. Cependant, cette méthode n'était pas efficace pour l'estimation de l'occurrence des différents types des jours (sec ou humide). En termes d'intensité, la méthode du générateur stochastique et des analogues ont été efficaces pour la simulation de la moyenne de précipitations, ainsi que les valeurs extrêmes. La méthode de CDF-t de la famille MOS est moins performante et plus instable pour les estimations futures. En effet, cette méthode sur-estime souvent la variance. D'autre part, les méthodes des analogues ont mieux reproduit la distribution spatiale. Ceci revient au fait que cette méthode est basée sur une procédure de ré-échantillonnage, elle

conserve donc la structure spatiale observée initialement. Cependant, cette méthode n'a pas réussi à reproduire la synchronicité de ces structures spatiales. Finalement, la distribution temporelle a été évaluée en se basant sur deux critères : la variabilité inter-annuelle et la saisonnalité. La méthode du générateur stochastique a échoué à reproduire la variabilité inter-annuelle. Ce fait peut être expliqué par la nature stochastique de cette méthode qui peut simuler des faibles ou des fortes valeurs par rapport aux observations, ou bien le manque d'information sur la variabilité inter-annuelle au niveau des prédicteurs.

Dans un autre contexte, Zorita & Von Storch (1999) ont comparé aussi la méthode des analogues, différentes méthodes de régression linéaire et une méthode de régression non linéaire pour la simulation des précipitations, au sud-ouest de l'Europe. Ce travail montre aussi que la méthode des analogues a généré des résultats un peu mieux pour reproduire le cumul journalier des précipitations et son occurrence, par rapport aux autres méthodes. D'autre part, Wilby *et al.* (1998) et Wilby & Wigley (1997b) ont comparé des méthodes issues des fonctions de transfert, du générateur de conditions météorologiques, et des analogues, pour la simulation des précipitations journalières. L'étude est menée au niveau de différents sites en Amérique du Nord. Les résultats montrent que les générateurs de conditions météorologiques produisent une bonne estimation de l'occurrence des journées humides. Cependant, ils sont moins performants pour reproduire la variabilité inter-annuelle contrairement aux fonctions de transferts.

Un nombre très limité de travaux ont testé la performance des techniques de descente d'échelle sur les sorties des modèles hydrologiques. On trouve à titre d'exemple les travaux de Wilby *et al.* (2000), qui a alimenté le modèle hydrologique "Precipitation-Runoff Modeling System" (PRMS) par des données des précipitations journalières et de température de surface minimale et maximale afin de simuler le ruissellement. Ces données sont issues des réanalyses brutes ou corrigées à partir de différentes méthodes statistiques, faisant partie des méthodes de Perfect-prog. Cette étude a permis d'identifier l'amélioration des sorties du modèle hydrologique, en utilisant les méthodes de descente d'échelle pour les entrées du modèle. Ainsi, ce travail a relevé la sensibilité d'un modèle hydrologique par rapport aux différentes méthodes de descente d'échelle utilisées. Dans le même contexte, une étude plus récente menée par Grouillet *et al.* (2015), a cherché à tester la capacité des différentes méthodes de descente d'échelle à produire des simulations hydrologiques précises pour les études d'impacts de changements climatiques. Trois méthodes ont été comparées : la méthode des analogues, la méthode de générateur stochastique et la méthode de correction de biais univariée CDF-t pour simuler la température de l'air et les précipitations à une échelle réduite. Cette étude montre que les modèles hydrologiques sont sensibles par rapport à la méthode de descente d'échelle utilisée. Ainsi, les simulations issues des forçages climatiques des données après la descente d'échelle sont meilleures que les simulations issues d'un forçage avec les données brutes issues du modèle climatique. Finalement, d'après cette analyse, la méthode des analogues a montré les meilleures performances pour reproduire les variables hydrologiques, par rapport à des simulations issues des observations. Le choix de la méthode de descente d'échelle reste facteur de la nature des variables qu'on va utiliser, de la résolution temporelle et spatiale à reproduire et de la profondeur temporelle à estimer. Le tableau 2.5 présente un résumé des avantages et des limitations des méthodes de descente d'échelle usuellement utilisées.

Table 2.5 – Avantages et limitations des méthodes de descente d'échelle.

Méthodes	Avantages	Limitations
Méthodes de régression	<ul style="list-style-type: none"> <li>• Simplicité de calcul</li> <li>• Nécessite moins de sources informatiques</li> </ul>	<ul style="list-style-type: none"> <li>• Capacité à ne reproduire que les basses fréquences des séries temporelles</li> <li>• Incapacité à reproduire des événements extrêmes</li> <li>• Sous estimations de la variabilité des variables climatiques</li> </ul>
Générateur stochastique	<ul style="list-style-type: none"> <li>• Capacité à reproduire les caractéristiques statistiques des séries observées</li> <li>• Capacité à produire des simulations en infra-journalier</li> <li>• Capacité à reproduire la variabilité spatiale</li> <li>• Capable d'assurer à la fois une extension temporelle et un gap-filling des données manquantes</li> <li>• Introduction des co-variables qui peuvent améliorer et mieux ajuster la simulation</li> </ul>	<ul style="list-style-type: none"> <li>• Sa mise en oeuvre peut s'avérer complexe et la compréhension des approches demandent un bon niveau en statistiques</li> </ul>
Méthode des analogues	<ul style="list-style-type: none"> <li>• Une bonne production des événements extrêmes tant qu'ils sont parmi les observations</li> <li>• Conservation de la dépendance spatiale et la cohérence des champs atmosphériques</li> <li>• Les modèles vont être pilotés par la distribution des variables locales</li> </ul>	<ul style="list-style-type: none"> <li>• Incapacité de simuler des situations qui diffèrent de celles des observations</li> <li>• Besoin d'un grand nombre d'observation</li> <li>• La difficulté du choix de l'analogue</li> </ul>
Méthodes MOS	<ul style="list-style-type: none"> <li>• Appliquée directement sur les variables des modèles climatiques</li> <li>• Mise en oeuvre relativement rapide</li> </ul>	<ul style="list-style-type: none"> <li>• Difficulté de bien reproduire les valeurs extrêmes</li> <li>• Incapables d'améliorer la structure temporelle</li> </ul>

## 2.2 Matériel et Méthodes

### 2.2.1 Zone d'étude

#### 2.2.1.1 Contexte géographique

Le bassin versant Merguellil est situé au centre de la Tunisie. La partie amont de Merguellil ( $1\,200\text{ km}^2$ ) est relativement montagneuse, avec une altitude comprise entre 200 et 1200 m, et une occupation du sol très variée. Cependant, la partie aval (latitude  $9^{\circ}30'E$ - $10^{\circ}15'E$ , et de longitude  $35^{\circ}N$ - $35^{\circ}45'N$ ), qui s'étend sur  $3000\text{ km}^2$ , est principalement plate avec une altitude comprise entre 50 et 200 m. Cette partie constitue la plaine de Kairouan qui est fortement exploitée par l'agriculture. La plaine de Kairouan forme une large cuvette de 100 km de longueur (Nord-Sud) et de 40 Km de largeur (Est-Ouest). Elle est entourée par les collines du Sahel à l'Est et par des montagnes à l'Ouest (Djebel Ouesslat, Djebel Cherichira, Djebel Touila). Entre les deux parties amont et aval, on trouve le barrage d'El Haouareb (voir figure 2.10), qui est construit en 1989 pour la protection de la ville des inondations, après les inondations millénaires de 1969 (Massuel & Riaux, 2017), et alimenter certains périmètres irrigués de la région. Le barrage a une capacité maximale de 90 millions de  $m^3$ , mais n'a jamais été totalement rempli (25 % s'évapore, 63 % part en recharge des nappes, et le reste pour l'irrigation) (Leduc *et al.*, 2007). La plaine de Kairouan se trouve au débouché de trois oueds : Zeroud, Merguellil et Nebhana qui avait un rôle important pour la recharge de l'aquifère de la plaine de Kairouan étant la seule ressource en eau disponible (Leduc *et al.*, 2007).

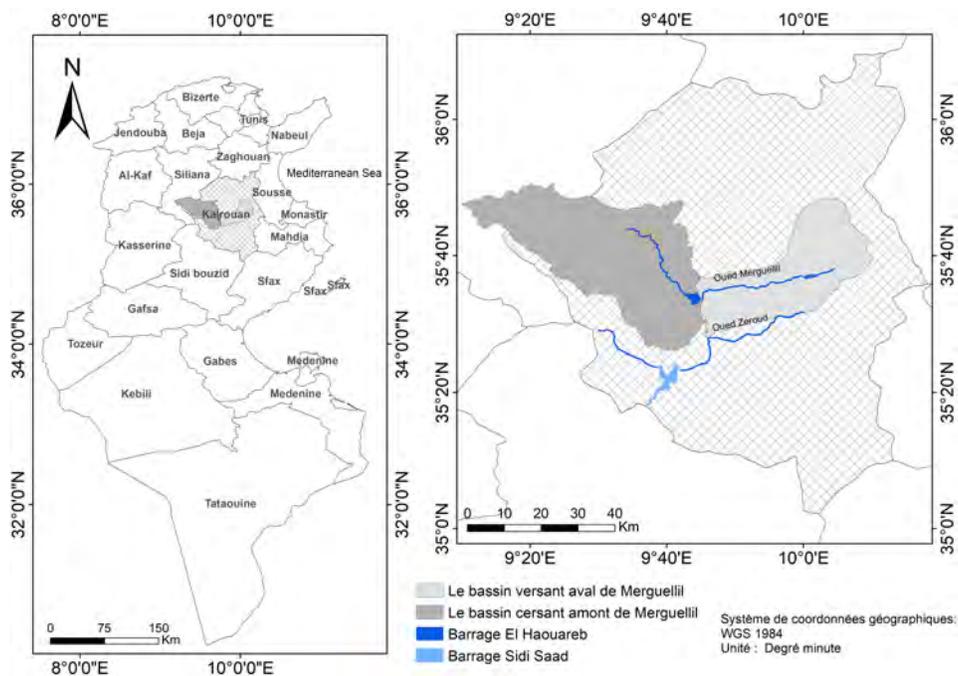


Figure 2.10 – Localisation géographique du bassin versant de Merguellil.

### 2.2.1.2 Contexte climatique

La plaine de Kairouan fait partie de l'étage aride à hiver tempéré. En effet, la Tunisie centrale se situe entre le bassin occidental de la méditerranée au Nord, et le plein désert, au Sud. De ce fait, le bassin du Merguellil se trouve soumis aux influences simultanées, d'une part, de la zone à tendance humide caractérisée par le climat de la méditerranée à vents humides, et d'autre part, de la zone pré-saharienne caractérisée par le climat saharien sec et humide (Slimani *et al.*, 2007).

1. **Pluviométrie** Le bassin versant de Merguellil présente des caractéristiques emblématiques des milieux semi-arides. Il est caractérisé par des courtes saisons de pluie qui s'étend de Septembre à mai (Alazard *et al.*, 2015), une forte variabilité spatiale et temporelle de précipitations avec la présence récurrente d'évènements extrêmes tels que les inondations et les sécheresses intenses. La région est caractérisée par des étés secs et des hivers humides. Les précipitations annuelles mesurées entre 1970 et 2021 à la station météorologique de Kairouan, sont présentées au niveau de la figure 2.11. Dans cette figure, les moyennes des cumuls de précipitation annuels calculées pour les années hydrologiques (de septembre à août) est de l'ordre de 300 mm/an présentées en ligne rouge au niveau de la figure 2.11. On peut distinguer deux types des années : (1) des années humides présentant un cumul annuel au dessus de la moyenne et (2) des années sèches avec un cumul de précipitations annuels inférieur à 300 mm/an. 2000-2001 forme l'année hydrologique la plus sèche sur toute la période, avec une pluviométrie annuelle totale d'environ 150 mm. L'année hydrologique la plus humide est l'année 1989-1990, avec un cumul d'environ 570 mm de précipitations.

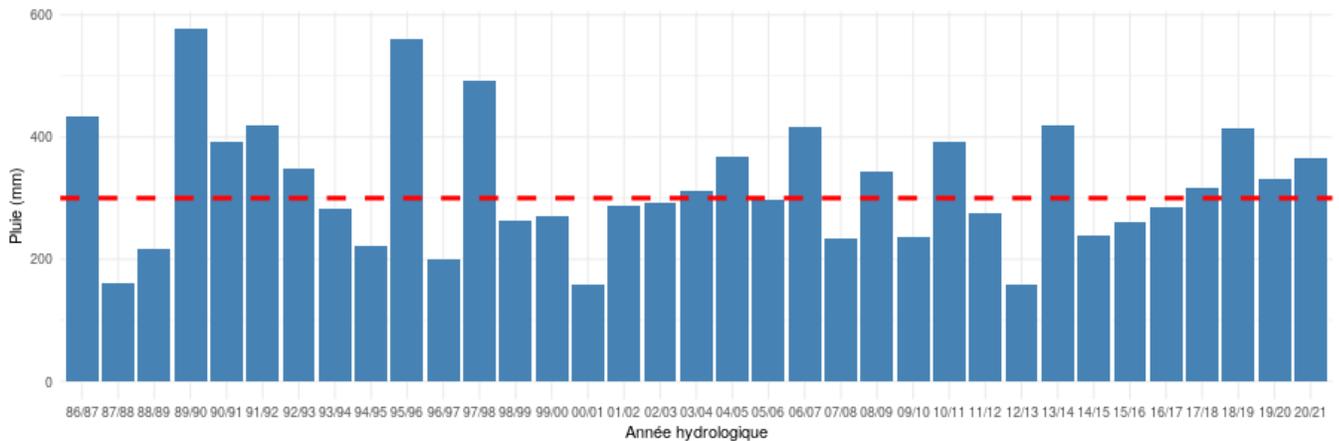


Figure 2.11 – Le cumul des précipitations annuelles sur la période 1986-2021, mesurées à la station synoptique de Kairouan.

2. **Température** La région de Kairouan est caractérisée par des grandes amplitudes thermiques (Alazard *et al.*, 2015), bien identifiées au niveau de la figure 2.12. Ce graphique montre le cycle annuel de la température aux Kairouanais, calculé à partir d'une série de

35 ans, allant de 1986 jusqu'à 2021. L'analyse de ces données montre que la température mensuelle moyenne oscille entre 11°C et 30°C. Les valeurs de température élevées, observées essentiellement en été, témoignent l'aridité climatique de la région.

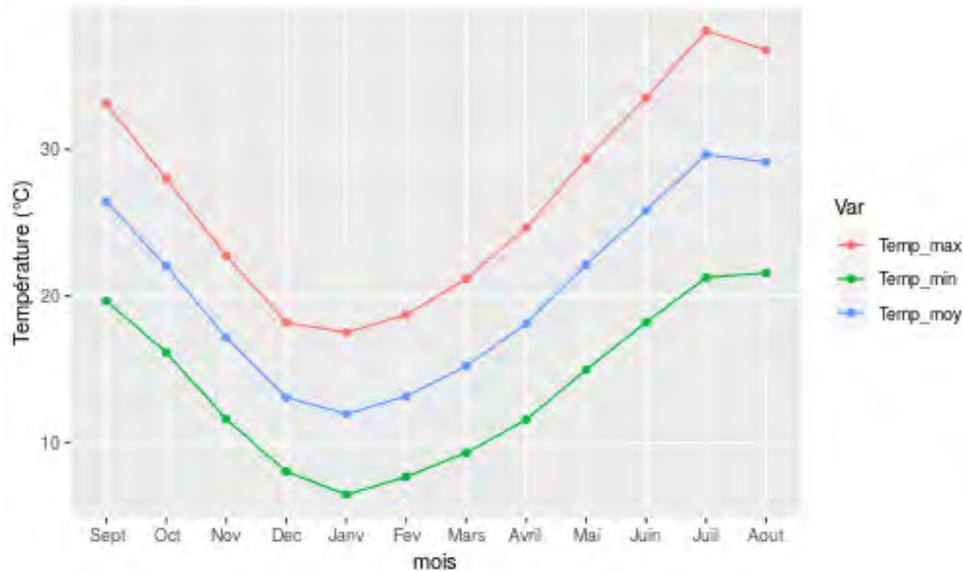


Figure 2.12 – Les moyennes, les minimums et les maximums mensuelles des températures journalières mesurées à la station de Kairouan entre 1986 et 2021.

On présente dans la figure 2.13, le diagramme ombrothermique, connu aussi sous le nom de diagramme climatique élaboré pour la ville de Kairouan à partir des données de température et des précipitations moyennes mesurées entre 1986 et 2021. Ce diagramme montre que généralement, Janvier forme le mois le plus froid, avec une température moyenne mensuelle de 12°C et Juillet forme le mois le plus chaud, avec une température moyenne mensuelle de 30°C. Ainsi, les précipitations mensuelles moyennes présentées font apparaître la variation inter-mensuelle de cette variable durant toute l'année. On observe ainsi que les pluies les plus intenses sont enregistrées en automne, essentiellement entre Septembre et Octobre, Les pluies d'hiver sont généralement de faible intensité et les pluies d'été sont très rares.

### 2.2.1.3 Contexte hydrique

la plaine de Kairouan dispose l'un des plus importants réservoirs aquifères en Tunisie. Ce réservoir est contenu dans les alluvions de la plaine. Deux grands barrages (El Haouareb et Sidi Saad, figure 2.10) ont été construits sur les principaux oueds (Zéroud, Merguellil et Nebhana) afin de lutter contre les crues et pour alimenter certains périmètres irrigués de la région. La nappe alluviale est en communication avec d'autres nappes latérales, comme les autres nappes de l'amont. L'infiltration des eaux de l'oued lors des crues est probablement la principale source d'alimentation de la nappe alluviale (Leduc *et al.*, 2007).

Différents ouvrages de conservation des eaux et des sols ont été réalisés au niveau du bassin versant de Merguellil, comme les lacs, les barrages collinaires et des banquettes afin

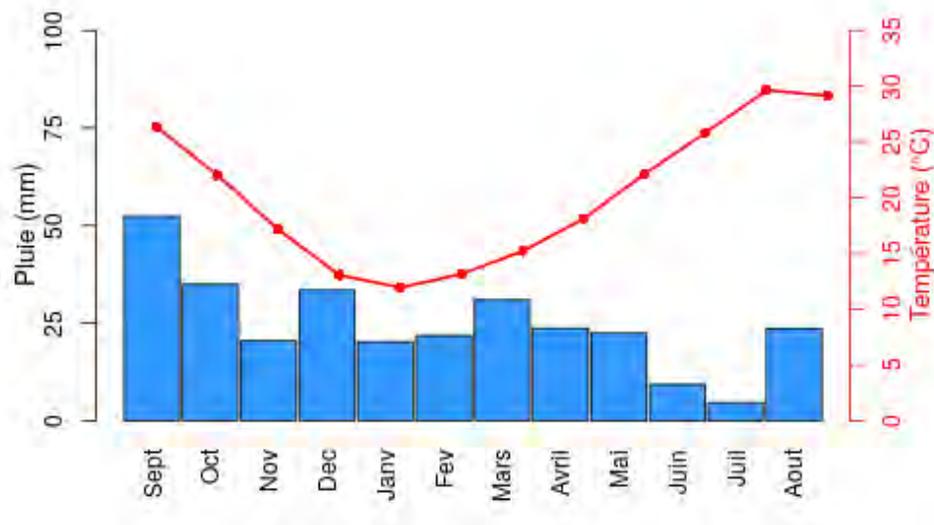


Figure 2.13 – Diagramme Ombrothermique : Précipitations et températures mensuelles moyennes mesurées à la station de Kairouan de 1986 à 2021.

de diminuer le processus de l'érosion. Cependant, les processus hydrologiques ont été fortement modifiés par ces activités humaines (Alazard *et al.*, 2015), notamment les travaux de conservation des sols et de l'eau et les grands et les petits barrages qui ont été destinés pour approvisionner de l'eau pour les périmètres irrigués. La surexploitation des ressources, l'augmentation de la population, le développement industriel, et notamment l'intensification de l'agriculture depuis la fin des années 60s, ont engendré une réduction en alimentation de la nappe aquifère et par conséquent, une baisse considérable de son niveau (environ 1.5 m par an) (Massuel & Riaux, 2017). Par ailleurs, l'avenir du développement agricole devient dépendant de la maîtrise et la bonne gestion des ressources souterraines (Leduc *et al.*, 2007).

#### 2.2.1.4 Contexte pédologique et occupation du sol

la couverture pédologique est très hétérogène à la fois à l'échelle régionale et parcel-laire, pour des raisons géologiques et pédogénétiques. Les sols sont généralement stratifiés et argileux en surface dû à la présence des alluvions avec la présence des encroûtements calcaires (Guyot, 1964). La présence des calcaires dans les sols, les conditions climatiques et les ressources hydriques agissent comme des facteurs limitant pour le choix de culture au niveau de la zone. Malgré ces conditions restrictives, la plaine de Kairouan représente une zone de plus fort développement agricole potentiel en Tunisie (Le Goulven *et al.*, 2009).

Les inondations de 1969 ont influencé énormément le type d'occupation des sols au niveau de la plaine de Kairouan. En effet, après cette crue dévastatrice et la construction du barrage El Haouareb, il y'a eu un développement de l'agriculture irriguée d'une façon intensive afin d'évacuer la quantité d'eau disponible. Ce fait a induit un changement du système d'irrigation au niveau de la plaine, d'un système pluvial vers un système irrigué (Massuel & Riaux, 2017).

On présente au niveau de la figure 2.14 une carte saisonnière d'occupation du sol, effectué

sur la saison hiver-printemps (de Septembre à Mai) de l'année hydrologique 2019-2020, dans le cadre des travaux de thèse de Hayfa Zayani. Cette carte a été établie à partir d'une série temporelle de NDVI dérivée des images Sentinel-2 à une résolution spatiale de 10 m, à partir des indices spectraux (NDVI, IB, etc.). Elle est élaborée par la méthode de classification supervisée, en se basant sur l'algorithme Random Forest avec un indice de Kappa de l'ordre de 0.81. La classe majoritaire est la classe des terres jachères. Ensuite, on trouve l'arboriculture avec un pourcentage de 24.3%, avec une dominance de l'oléiculture qui est une culture typique des zones arides. On trouve aussi un pourcentage élevé de céréaliculture, de l'ordre de 12.2%, qui fait parti des classes dominantes pendant le printemps.

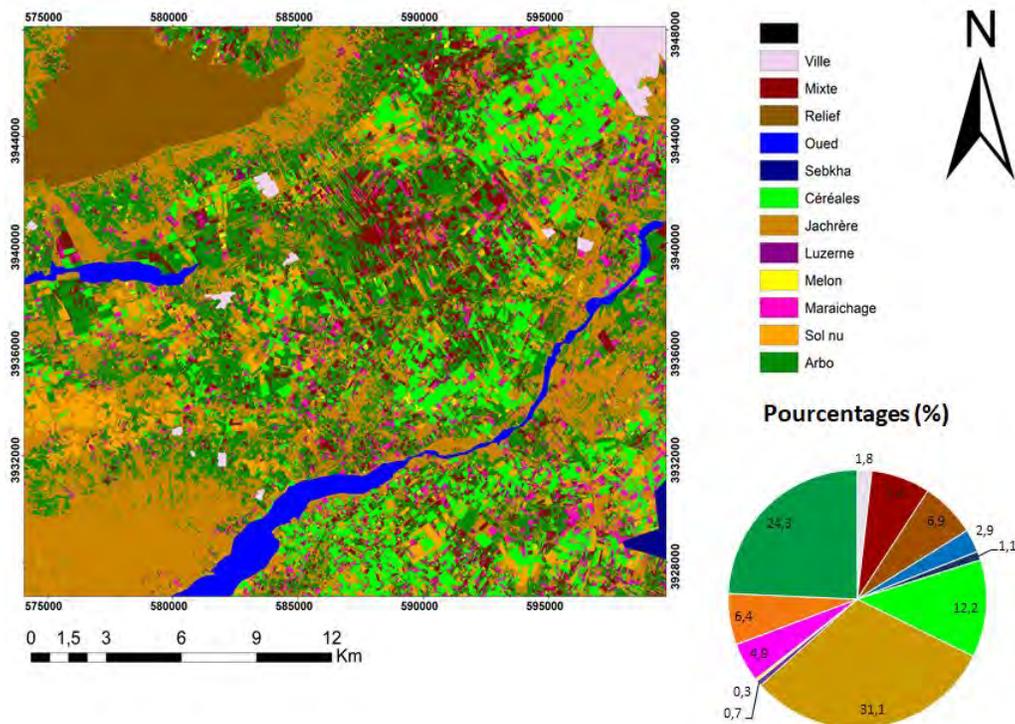


Figure 2.14 – Carte d'occupation du sol de la plaine de Merguellil (hiver-printemps 2019-2020) effectuée dans le cadre des travaux de thèse de Hayfa Zayani en utilisant une série temporelle de NDVI dérivée des images Sentinel-2.

## 2.2.2 Données climatiques

Les données climatiques forment les variables clés pour la simulation des variables de bilan d'énergie, particulièrement l'évapotranspiration. En effet, l'estimation de l'évapotranspiration nécessite un calage aux conditions climatiques locales (Xu & Singh, 2001; Valipour, 2015). Ainsi, leur profondeur temporelle, résolution spatiale et temporelle sont de grande importance pour les modèles d'impacts afin de simuler des valeurs d'évapotranspiration assez réalistes et conformes aux conditions climatiques. Deux types de données climatiques sont utilisées pour ce travail : des variables météorologiques issues des stations de mesure, et des données

météorologiques issues des sorties des modèles climatiques régionaux : les ré-analyses.

### 2.2.2.1 Données climatiques ponctuelles

Trois stations de mesure sujettes de notre étude sont implantées au niveau de la plaine : Ben Salem, Barrouta, et Chebika (voir la figure 2.15). La station de Ben Salem, installée depuis 2011 (latitude=35,550086° et longitude= 9,922119°), est notre station de référence, c'est-à-dire plus conforme aux normes des directives de l'OMM (OMM, 2010). La station de Chebika est installée depuis la fin de l'année 2010 et Barrouta est la station la plus récente, installée depuis 2014. Toutes ces stations sont automatisées, alimentées par un panneau solaire et une batterie. Elles mesurent la température de l'air à 2 m, l'humidité relative de l'air, le rayonnement global, la vitesse et la direction du vent à 2 m, la pression atmosphérique et les précipitations à l'aide d'un pluviomètre à auget basculant. Chaque station est équipée par une centrale d'acquisition pour le calcul et le stockage des valeurs moyennes des différentes variables météorologiques sur un intervalle de temps de 30 minutes. Ainsi, le transfert des données est assuré d'une façon manuelle ou automatique par transmission GPRS.

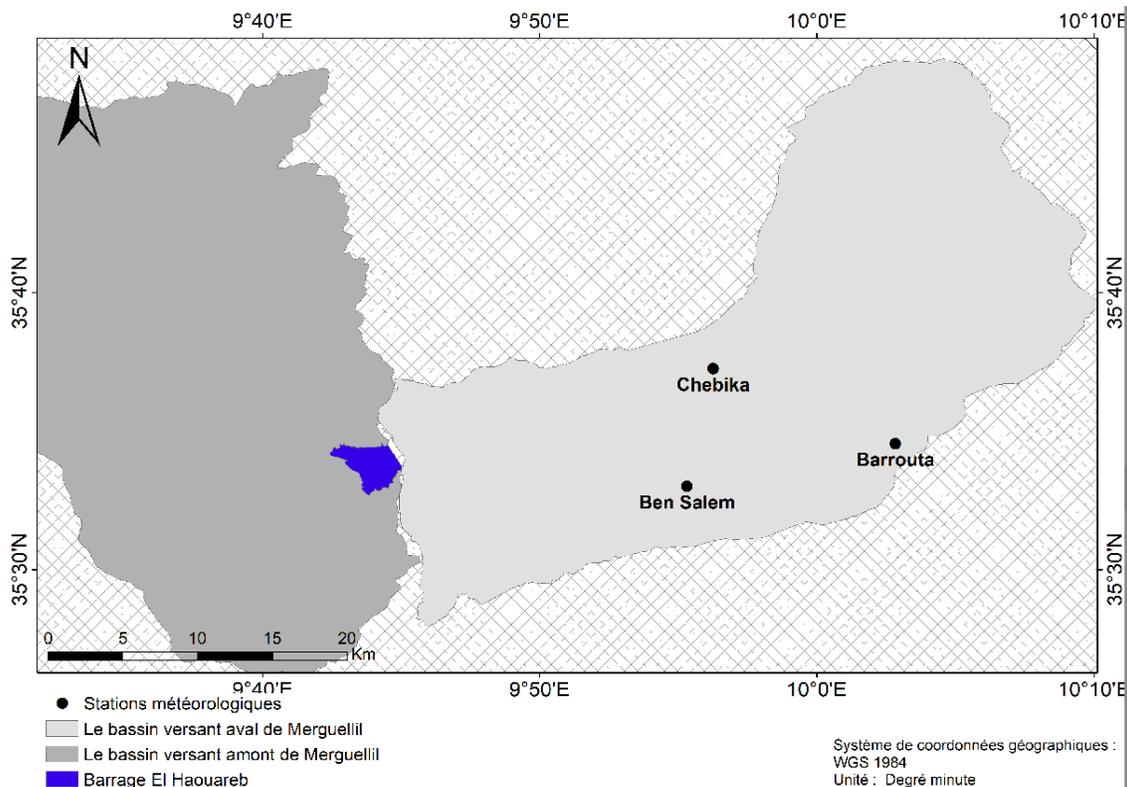


Figure 2.15 – Localisation des stations météorologiques au niveau de la plaine de Merguellil.

Les séries temporelles observées au niveau de chaque station présentent un certain nombre des données manquantes dues à des problèmes techniques au niveau des capteurs ou de la centrale d'acquisition. De ce fait, un traitement et une analyse de la base des données disponible sont indispensables avant de s'en servir. Le traitement consiste à vérifier la continuité temporelle, la synchronicité et la qualité de ces données afin d'enlever les valeurs aberrantes.

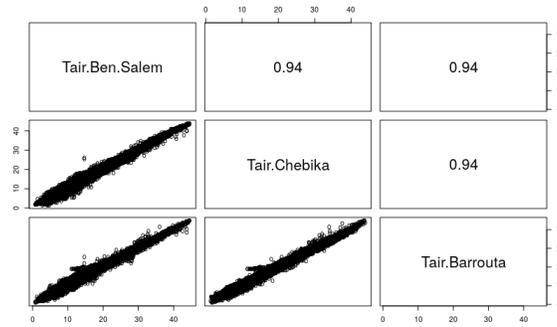
Ces stations de mesure présentent des informations climatiques similaires en raison de leur proximité géographique (7,9 km entre Ben Salem et Chebika, 11,7 km entre Ben Salem et Barrouta et 11 km entre Chebika et Barrouta). Comme montré au niveau des graphiques de dépendance inter-stations (figure 2.16), les variables mesurées au niveau des différentes stations sont bien corrélées. Cette figure montre les coefficients de corrélation non linéaire, par rang "Kendall" (Granger & Lin, 1994). Ce coefficient permet de déterminer le degré de liaison des variables mesurées au niveau de plusieurs sites. Ce coefficient est compris entre -1 et +1 et s'interprète comme suit : plus il s'approche de 1, plus on est certain qu'il existe une corrélation positive (variations dans le même sens des variables étudiées) et plus il est proche de -1, plus on peut supposer l'existence d'une corrélation négative. Cependant, si le coefficient de Kendall est proche de zéro, la probabilité qu'il n'existe aucune liaison entre les deux variables comparées est forte. Le coefficient de "Kendall" permet de mesurer les relations non linéaires entre deux variables, contrairement au coefficient de "Pearson" qui est restreint aux mesures des corrélations linéaires.

Au niveau des figures 2.16a et 2.16c, on observe une corrélation très importante entre la température de l'air et le rayonnement global mesurées au niveau des différentes stations, avec un tau de Kendall qui dépasse 0.94. L'humidité relative présente aussi des coefficients de corrélation inter-sites très importantes (figure 2.16b). Au niveau de la figure 2.16d, on voit que la vitesse du vent mesurée au niveau de Chebika présente une corrélation moins importante avec les deux autres stations. Ceci est expliqué par la présence des brises vent aux alentours de cette station, ce qui lui confère des caractéristiques un peu différentes par rapport aux autres stations.

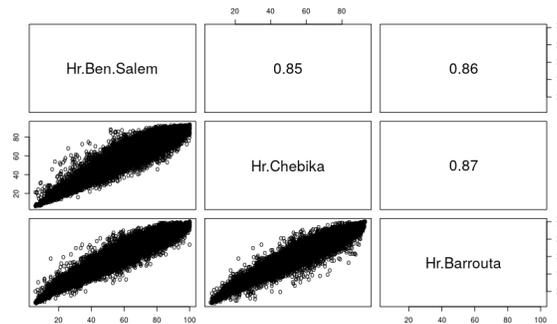
### 2.2.2.2 Données climatiques à large échelle : la ré-analyse

La ré-analyse météorologique est une méthode qui permet de reproduire le climat passé en se basant sur des données globales observées, de surface et d'altitude obtenues sur plusieurs années. Les jeux de données générés par les ré-analyses sont d'une grande valeur pour la recherche atmosphérique. En effet, elles fournissent une analyse spatialisée et des enregistrements cohérents de la circulation atmosphérique mondiale. Elles visent à uniformiser et homogénéiser les données passées pour améliorer la prévision météorologique et l'étude des événements passés. Ces données couvrent des longues périodes avec une résolution spatiale et temporelle plus ou moins fine sur des grilles à trois dimensions. L'objectif de ré-analyse est de produire un enregistrement homogène de l'évolution atmosphérique passé. Les ré-analyses constituent un apport clé aux démarches climatologiques classiques grâce à la connaissance des erreurs des prévisions et des observations (Balsamo, 2012). De ce fait, les simulations du modèle du climat basées sur la ré-analyse diminuent les biais systématiques et cherchent à reproduire la séquence réelle des événements météorologiques jour par jour. Ceci permettra d'injecter la synchronisation entre les simulations fournies par les modèles de climat et les observations (Maraun *et al.*, 2010).

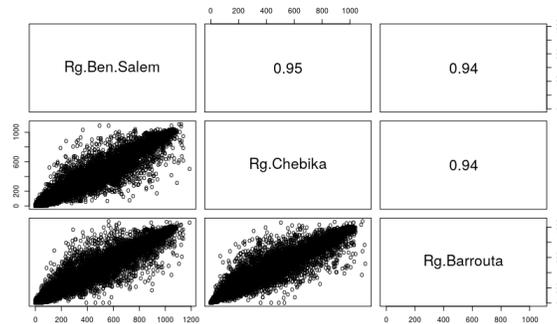
ERA5 est la cinquième génération du Centre Européen pour les Prévisions Météorologiques à Moyen Terme (cepmmt ou ecmwf en anglais pour European Centre for Medium-range Weather Forecasts). Le premier produit de réanalyse est le produit FGGE dans les années 1980s,



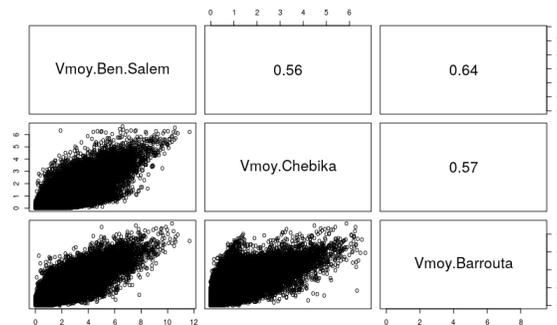
(a) Température de l'air



(b) Humidité relative de l'air



(c) Rayonnement global



(d) Vitesse du vent

Figure 2.16 – Dépendance inter-sites des différentes variables climatiques pendant la période entre 08/12/2014 et 31/12/2016.

suivi par ERA-15 (125 km de résolution horizontale), ERA5-40 ayant la même résolution, ERA-Interim (75 km de résolution), et récemment ERA5 (Hersbach, 2016). ERA5 présente une amélioration sur les différents aspects techniques de la ré-analyse, essentiellement la résolution spatiale qui passe de 75 km à 31 km ainsi que la résolution temporelle qui passe d'un produit pour tous les trois heures en un produit horaire. Ainsi, elle couvre une très longue période qui s'étend de 1950 à nos jours. La réanalyse combine des données modélisées avec des observations du monde entier en un ensemble de données globalement cohérent et complet utilisant les lois de la physique. Le modèle de ré-analyse ERA5 présente un système de prévision intégré configuré selon une grille de 0.3 par 0.3 degré (31 km) qui couvre toute la planète horizontalement. En altitude, la couverture va de 1000 hPa à 1 hPa avec 137 niveaux sur 80 km (Hoffmann *et al.*, 2019).

Une seule maille ERA5 couvre les trois stations météorologiques implantées au niveau de la plaine. La maille utilisée pour l'extraction des données est présentée au niveau de la figure 2.17. Les variables téléchargées sur la période qui s'étale entre Décembre 1999 et Décembre 2019, sont énumérées au niveau du tableau 2.6.

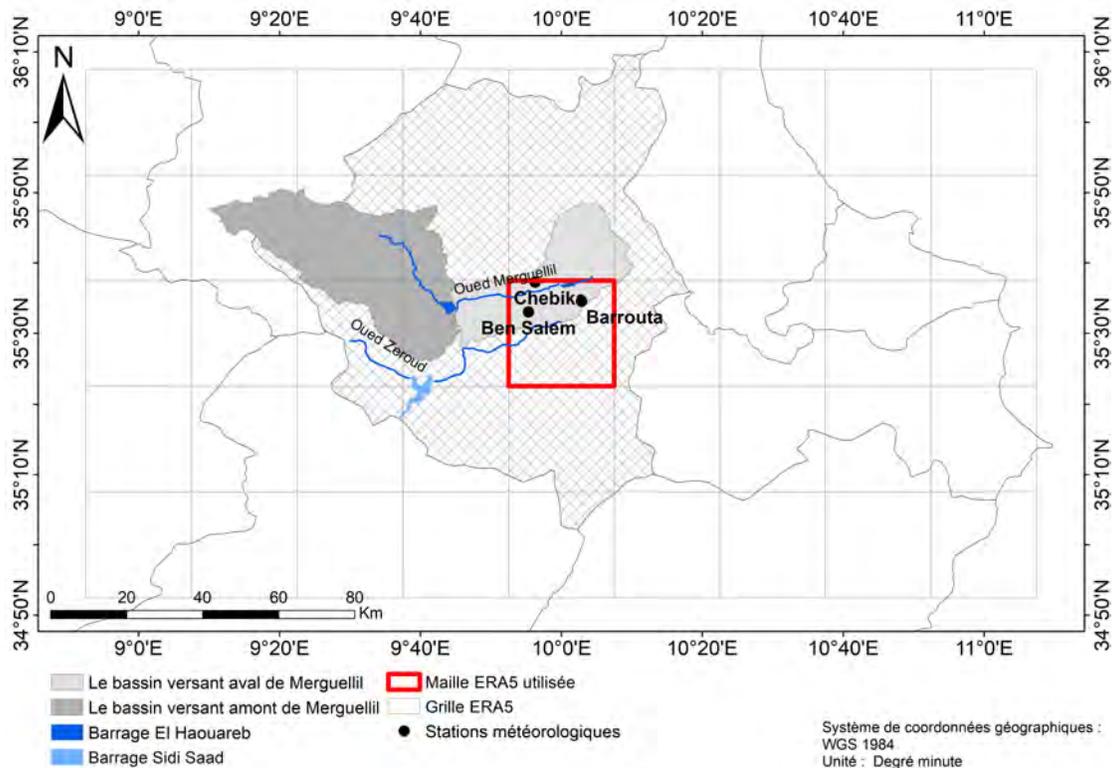


Figure 2.17 – Localisation de la maille ERA5 utilisée pour extraire les variables à large échelle nécessaires pour notre étude.

Certaines variables comme mslp, t2m, ssrd, tcc et tp sont utilisées directement. D'autres variables comme u10m, v10m et d2m sont servis pour extraire d'autres variables climatiques (humidité relative de l'air et vitesse du vent). L'humidité relative de l'air est calculée selon l'équation 2.16 définie au niveau du bulletin FAO 56 (Allen *et al.*, 1998), en utilisant la

Table 2.6 – Variables issues des ré-analyses ERA5 utilisées dans le cadre du travail.

Abréviation	Nom du variable
mslp	Pression au niveau de la mer (Pa)
u10m	Composante horizontale de la vitesse du vent ( $m s^{-1}$ )
v10m	Composante verticale de la vitesse du vent ( $m s^{-1}$ )
T2m	Température de l'air à 2m (K)
Tdew	Température de rosée à 2m (K)
ssrd	Rayonnement solaire sur les courtes longueurs d'onde ( $J m^2$ )
tcc	Couverture nuageuse
tp	Précipitations totales (m)

température de l'air et la température de rosée.

$$Hr = \frac{esat_{T_{dew}}}{esat_{T_{moy}}} * 100 \quad (2.16)$$

Avec,  $esat_{T_{dew}}$  est la pression de vapeur saturante à la température de rosée (KPa), exprimée selon l'équation 2.17 et  $esat_{T_{moy}}$  est la pression de vapeur saturante à la température moyenne (KPa) exprimée par l'équation 2.18.

$$esat_{T_{dew}} = 0.6108 * \exp\left(\frac{17.27 * T_{dew}}{T_{dew} + 237.3}\right) \quad (2.17)$$

$$esat_{T_{moy}} = 0.6108 * \exp\left(\frac{17.27 * T_{moy}}{T_{moy} + 237.3}\right) \quad (2.18)$$

Ainsi, la vitesse du vent est calculée à partir des deux composantes horizontale et verticale U10m et V10m. Un rabattement de niveau a été appliqué pour ces composantes en appliquant la formule 2.19, définie aussi au niveau du bulletin FAO 56 (Allen *et al.*, 1998). Ce rabattement a pour but, de rendre la variable de la vitesse du vent issue d'ERA5 comparable à celle issue des observations. Ensuite, la vitesse du vent est exprimée en fonction de leur distance euclidienne.

$$U_2 = U_z \left( \frac{4.87}{\log(67.8 * Z - 5.42)} \right) \quad (2.19)$$

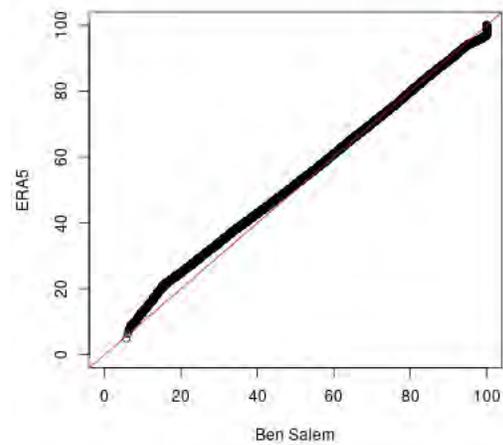
Avec,  $U_2$  est la vitesse du vent mesurée à 2 mètres au dessus de surface du sol en m/s et  $U_z$  est celle mesurée à un niveau Z par rapport à la surface du sol.

Nous avons élaboré une analyse quantile-quantile (figure 2.18) nous permettant d'établir une appréciation graphique de l'ajustement de la distribution observée. Au niveau de ces graphiques, l'axe des abscisses porte les quantiles  $x_i$  de la distribution observée à la station de Ben Salem (la référence) et l'axe des ordonnées porte les quantiles  $x_{i*}$  de la distribution estimée (ERA5).

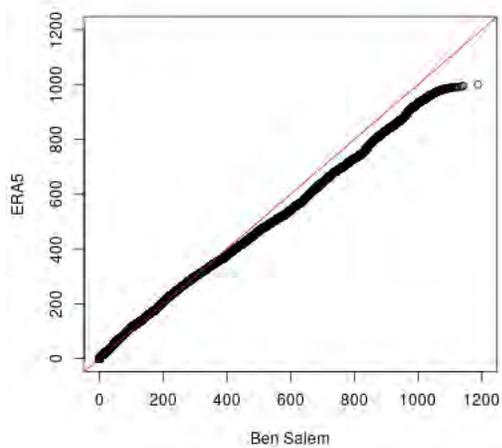
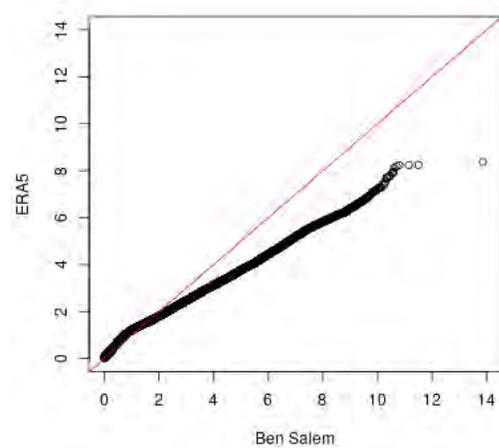
Ces graphiques montrent des points bien alignés suivant la bissectrice, présentés par la ligne rouge. Ceci indique la présence d'une même loi de distribution entre les deux variables. Cependant, pour le rayonnement global, figure 2.18c, on trouve que les valeurs fortes observées à Ben Salem sont sous-estimées par les réanalyses. Ainsi, La vitesse du vent issue des ré-analyses sous-estime les valeurs enregistrées au niveau de la station de mesure (figure 2.18d).



(a) Température de l'air (°C)



(b) Humidité relative de l'air (%)

(c) Rayonnement global ( $W/m^2$ )

(d) Vitesse du vent(m/s)

Figure 2.18 – Graphiques quantile-quantile des données climatiques observées à Ben Salem et les données climatiques issues des réanalyses.

### 2.2.3 Données satellitaires

MODIS (The Moderate Resolution Imaging Spectroradiometer) est conçu pour le suivi de la terre, l'océan et l'atmosphère. Cette composante d'imagerie combine les caractéristiques du radiomètre à ultra haute résolution avancé (AVHRR) et Landsat Thematic Mapper en ajoutant des bandes spectrales dans le long et moyen infrarouge. Il fournit une résolution spatiale de 250 m, 500 m et 1 km. Des canaux spectraux avec des meilleures caractérisations de l'atmosphère et des nuages sont conçus afin de permettre l'élimination de l'effet de l'atmosphère sur les observations de surface et la fourniture des mesures d'atmosphère (Justice *et al.*, 1998). Le capteur MODIS est une source usuelle d'imagerie de télédétection utilisée pour la surveillance globale, en raison de sa disponibilité gratuite, de son large éventail et de ses capacités régulières de révision (fréquence quotidienne). Les images MODIS sont composées de 36 bandes spectrales, dont deux bandes de résolution spatiale de 250 m (bandes 1 et 2), cinq bandes de 500 m (bandes 3-7) et 29 bandes de 1 km. Les sept premières bandes ont été largement utilisées pour la surveillance du LCLU (couverture terrestre / utilisation du sol) (Wang *et al.*, 2015).

Le même capteur MODIS est embarqué à bord des satellites Terra et Aqua de la National Aeronautical and Space Administration (NASA), lancés respectivement en 1999 et 2002. Le satellite Terra occupe une orbite polaire héliosynchrone avec un temps de passage 10h30 (en descendant) et 22h30 (en ascendant). Le satellite Aqua forme une paire supplémentaire d'observation à 01h30 (en descendant) et à 13h30 (en ascendant) heure solaire (Giglio *et al.*, 2006). Les données satellitaires issues de Terra et Aqua sont disponibles gratuitement sur le site de NASA (<http://modis.gsfc.nasa.gov/>). Pour le présent travail, on utilise les données de : la température de surface terrestre (LST), émissivité de surface ( $\epsilon$ ) et angle de vue ( $\phi$ ) (MOD11A1 / MYD11A1 pour Terra et Aqua, respectivement), NDVI (MOD13A2 / MYD13A2 pour Terra et Aqua, respectivement) et albédo (MCD43A3). La tuile MODIS téléchargée initialement en projection sinusoidale, est reprojétée en UTM. Ensuite, les sous-images de 396km × 22 km correspondant à la zone d'étude (Figure 2.19) ont été extraites à partir de 30/03/2000 pour TERRA et 08/07/2002 pour AQUA, jusqu'à la fin de l'année 2019 à 1 km de résolution spatiale.

La température de surface MODIS et l'angle de vue sont acquis d'une façon journalière. Cependant, l'albédo et le NDVI sont obtenus en synthèses de 8 et 16 jours respectivement. Les données MODIS qui vont servir comme variables d'entrée pour le modèle SPARSE contiennent des données manquantes ou non fiables. De ce fait, un masque de contrôle qualité utilisant les données "QCDay", issu de MOD11A1 et MYD11A1, a été appliqué. L'indice foliaire (LAI) est obtenu à partir des données du NDVI selon l'équation 2.20, développée par Clevers (1989).

$$LAI = -\frac{1}{k} \ln\left(\frac{NDVI_{max} - NDVI_i}{NDVI_{max} - NDVI_{min}}\right) \quad (2.20)$$

Avec,  $k=1.13$  forme le facteur d'extinction calibré,  $NDVI_{max} = 0.97$  forme le NDVI d'une canopée bien développée,  $NDVI_{min} = 0.05$  correspond à un NDVI d'un sol nu (Chirouze *et al.*, 2014) et  $NDVI_i$  est le NDVI qui correspond à chaque pixel MODIS.

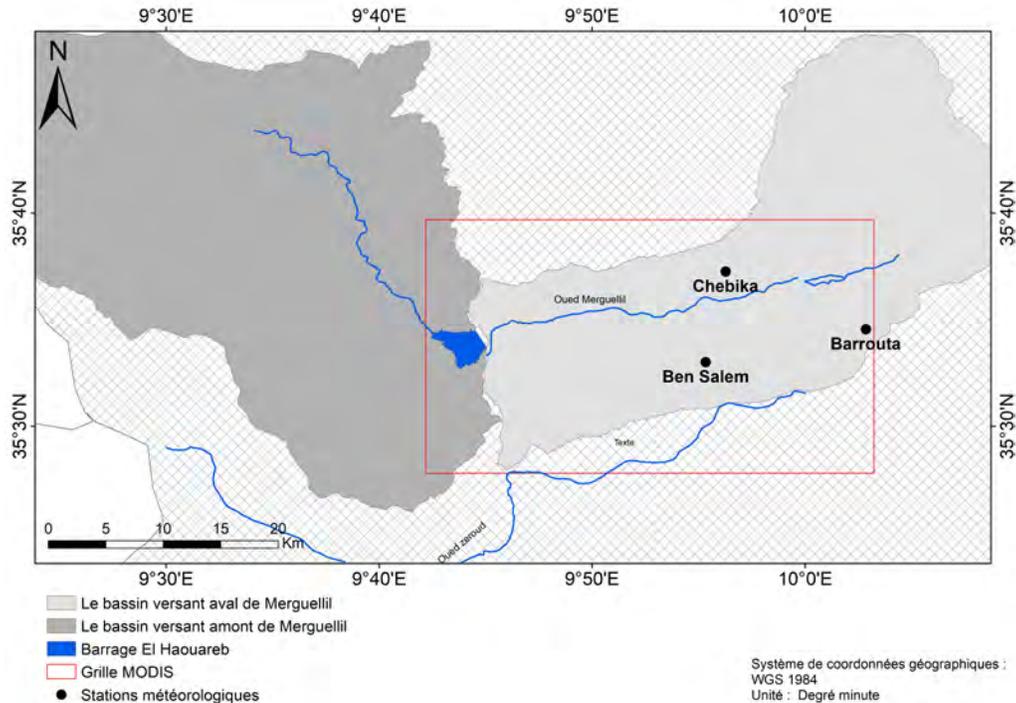


Figure 2.19 – Grille MODIS téléchargée.

## 2.2.4 Données d'occupation du sol

On dispose des cartes d'occupation des sols OS de la plaine de Kairouan, élaborées dans le cadre des différents travaux effectués au sein de notre laboratoire. Une carte pour l'année 2008-2009 effectuée par Shabou *et al.* (2015), 2011-2012 par Chahbi *et al.* (2014), 2012-2013 par Chahbi *et al.* (2015), 2013-2014 et 2014-2015 par Chahbi *et al.* (2016), 2016-2017, 2017-2018, 2018-2019 sont effectuées dans le cadre des travaux de Kassouk *et al.* (2018). Toutes ces cartes sont élaborées par la méthode de classification supervisée, en se basant sur l'algorithme Random Forest.

Une analyse préliminaire de la variation de la classe "céréale", souvent liée aux études des sécheresses, est présentée au niveau du graphique 2.20. Nous avons effectué une extraction des informations issues de ces cartes, qui concernent notre zone d'étude, en se basant sur la délimitation de la grille MODIS téléchargée (figure 2.19). Le graphique 2.20 contient aussi la variation annuelle du pourcentage des sols nus qui doit remplacer les cultures de céréales dans les années sèches. Ce graphique montre une variation anti-corrélée de ces deux types d'occupation du sol jusqu'à l'année 2013. Au delà de cette année, ces deux occupations du sol ne montrent plus une variation corrélée. Ceci est dû majoritairement, à l'extension des surfaces irriguées au niveau de notre zone d'étude pendant les dernières années.

D'autre part, nous avons transformé ces cartes de OS en des cartes de pourcentages selon la grille MODIS. Pour ce faire, un travail de découpage et ré-échantillonnage a été appliqué pour aboutir à des cartes de pourcentage majoritaire (OS majoritaire au niveau de chaque pixel) ou bien spécifique (extraire le pourcentage de céréale en sec au niveau de chaque pixel) (voir schéma 2.21). Ces analyses sont de grandes utilités pour l'étude de la réponse des indices

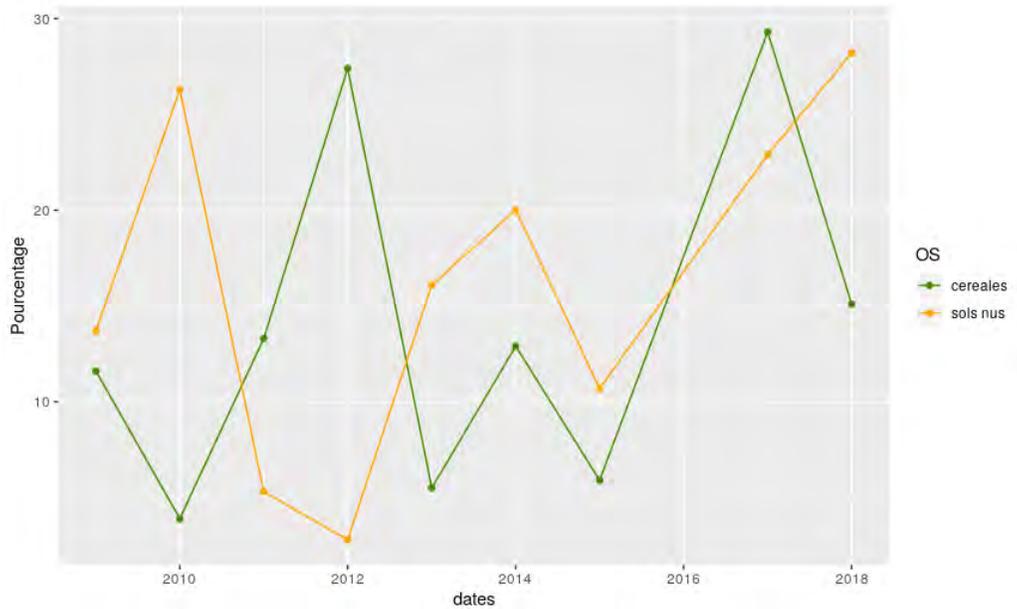


Figure 2.20 – Évolution annuelle de la variation des pourcentages des céréales et la portion du sol nu sur la zone d'étude.

de stress en fonction du pourcentage du céréale en sec observé au niveau de chaque pixel.

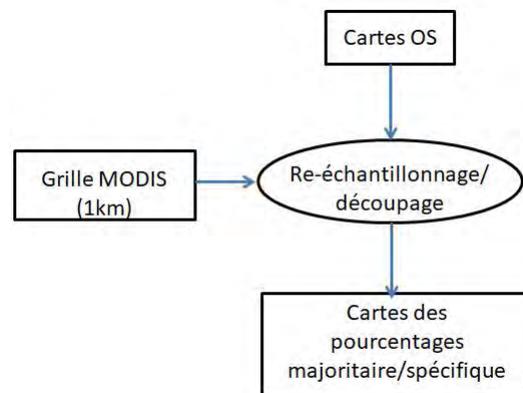


Figure 2.21 – schéma pour illustrer les étapes de transformation des cartes OS en cartes des pourcentages selon les pixel de MODIS.

## 2.3 Conclusion partielle et synthèse

Ce chapitre présente l'état de l'art des différents indices de sécheresses habituellement utilisés d'une façon générale et particulièrement sur les zones de la méditerranée, des méthodes de simulation de l'évapotranspiration à partir des modèles du bilan d'énergie et finalement, les différentes méthodes de descente d'échelle qui sont très utiles pour la génération d'une série

climatique assez fine à la fois à l'échelle spatiale et temporelle. Par ailleurs, nous avons présenté au niveau de ce chapitre le contexte géographique, climatique, hydrique et pédologique de notre zone d'étude. Finalement, ce chapitre contient aussi une présentation des données météorologiques et satellitaires que nous avons utilisées pour contraindre le modèle de bilan d'énergie à double source utilisé dans la suite du travail.

### 3.1 Introduction

L'étude d'un système agricole repose sur l'utilisation d'un modèle d'impact simulant son fonctionnement en réponse à un forçage météorologique. Il est par ailleurs nécessaire d'avoir des scénarios climatiques avec une profondeur temporelle et une résolution spatiale adéquate pour ce type d'analyses. Ce chapitre présente la méthode de descente d'échelle appliquée sur les réanalyses, que nous avons élaborée afin de prolonger les données météorologiques disponibles au niveau de la zone d'étude, et avoir par conséquent des données assez longues et assez fines à l'échelle spatiale et temporelle. Ce chapitre contient deux sections, la première présente l'approche statistique développée, qui nous a permis de simuler les variables nécessaires pour contraindre le modèle de bilan d'énergie à double source SPARSE (température de l'air, humidité relative de l'air, vitesse du vent et rayonnement global). Cette section est présentée sous format d'un article. La deuxième contient des résultats complémentaires à cet article, qui s'articulent sur la simulation d'autres variables météorologiques, pression atmosphérique et précipitations, qui n'ont pas été utilisées pour le forçage du modèle du bilan d'énergie, mais qui sont d'une grande importance pour des études de changement climatiques. La présentation de la méthode et sa validation au niveau de ce chapitre est effectuée sur la période d'observation, qui s'étale entre 2011 et 2016.

### 3.2 Développement de la méthode statistique

L'objectif majeur de cette section est de développer une méthode statistique qui permet de combler les lacunes et d'étendre sur une période passée les séries d'observations des variables météorologiques qui vont être utilisées comme entrée du modèle de bilan d'énergie à double source SPARSE. En effet, ce modèle a besoin des données météorologiques à l'heure du passage du satellite. D'où le besoin d'une information précise en infra-journalier (horaire ou demi-horaire). Pour ce faire, nous avons adapté un générateur stochastique de conditions météorologiques "MetGen", basé sur des variables à large échelle (ERA5) et qui permet de simuler des scénarios multi-variables et régional à une échelle infra-journalière. Le développement de cette approche fait l'objet d'une publication à "Environmental Modelling and Software" qui est présentée ci-dessous.

### 3.2.1 Résumé de l'article

Dans l'optique de surveiller les périodes de sécheresse dans notre zone d'étude et aider les décideurs à établir de bonnes pratiques de gestion de la ressource en eau, nous nous intéressons à l'analyse des besoins en eau de la plante au travers des indices de stress hydrique. À cette fin, un modèle de bilan d'énergie à double source (Boulet *et al.*, 2015) permet, en combinant l'information satellitaire kilométrique, et l'information météorologique provenant d'un réseau d'observations in situ, de simuler l'évapotranspiration ainsi que le stress hydrique. Or, les réseaux d'observations météorologiques sont peu représentatifs de la variabilité spatio-temporelle au niveau de la zone étudiée (faible densité des sites instrumentés, disparité entre les sites au niveau des variables mesurées et périodes d'observation courtes et souvent non-concomitantes). L'objectif de cet article est alors de développer une approche statistique nous permettant d'effectuer l'imputation des données manquantes et une projection de la série climatique existante sur une période passée. Pour ce faire, une source d'information importante à exploiter provient des réanalyses qui fournissent ces variables mais à une faible résolution spatiale (environ 31 km (Hersbach *et al.*, 2018)).

Nous avons développé un générateur stochastique de conditions météorologiques "MetGen", adapté au contexte de l'estimation du stress hydrique en région semi-aride, en se basant sur les données de ré-analyses. "MetGen" est disponible gratuitement sous forme d'une librairie R, (<https://CRAN.R-project.org/package=MetGen>). Notre générateur stochastique permet de combler les lacunes présentes sur les séries des observations et de projeter des scénarios des différentes variables sur une longue période passée. "MetGen" est basé sur des modèles linéaires généralisés (GLM) en choisissant une loi de probabilité appropriée pour chaque variable (Normal, Gamma ou Binomiale). Les données de ré-analyses sont exploitées en tant que covariables dans "MetGen" afin de fournir des scénarios à l'échelle locale et à une résolution infra-journalière (horaire ou demi-horaire). D'autres covariables sont utilisées afin d'introduire les dépendances inter-variables, les effets déterministes (informations géographique, cycles saisonniers et diurnes) et la persistance temporelle. Notre générateur exploite l'information issue de trois stations implantées au niveau de la plaine, présentant une homogénéité due à leur proximité géographique. L'utilisation des données à partir de différentes stations confère à la fois, une base de données plus large à exploiter et une information représentative de la zone étudiée.

"MetGen" est inspiré du générateur développé par Chandler (2015), "Rglimclim", présenté au niveau de la section 2.1.6.1. Les adaptations mises en oeuvre par notre travail en comparaison aux travaux de Chandler (2015) concernent principalement, le passage du pas de temps journalier au pas de temps infra-journalier, l'identification de jeux de covariables appropriés pour notre contexte climatique ainsi qu'une procédure de validation rigoureuse. Cette procédure comprend deux volets : l'évaluation de la capacité (1) à bien reproduire les variables météorologiques et (2) à bien restituer les variables de bilan d'énergie.

La construction des modèles est basée sur le choix de la loi de probabilité et la sélection des covariables en se basant au début sur un nombre réduit des covariables préliminaires. Ensuite, on ajoute les covariables optionnelles qui permettent d'améliorer les critères de validation de nos modèles. La validation du modèle est basée sur une procédure de double validation croisée. La performance est évaluée quantitativement et essentiellement sur la base

de l'erreur de distribution, calculée à partir de l'erreur quadratique moyenne des graphiques quantiles-quantiles. D'une façon complémentaire, on génère des cycles annuels et diurnes pour une meilleure inspection visuelle, des graphiques d'autocorrélation temporelle ainsi que des graphiques d'inter-dépendance des variables pour une évaluation qualitative des simulations. Nous comparons aussi la méthode de générateur stochastique avec une méthode de la famille MOS (voir la section 2.1.6.2), qui est la méthode de correction de biais qui exploite également les données de ré-analyses. Nous considérons une méthode de correction de biais univariée (correction variable par variable) et multivariée (correction de toutes les variables à la fois). Nous avons utilisé la fonction CDFt (Michelangeli *et al.*, 2009) en univarié, et en multivarié l'approche MBCn (Cannon, 2018). Pour ces méthodes, nous avons proposé de travailler plutôt sur les anomalies en enlevant les fluctuations systématiques des cycles diurnes et saisonniers (supposés constants pour chaque saison).

Par ailleurs, afin d'étudier la nature stochastique de notre générateur stochastique, nous avons comparé 50 réalisations de "MetGen", avec les deux séries issues des deux types de corrections de biais (univariée et multivariée), une série issue des données ERA5 brutes, avec une série issue des observations. Ces approches statistiques sont utilisées pour générer des scénarios climatiques dans la région de Kairouan, qui est notre zone d'étude.

Sur le premier volet de validation, les résultats montrent que les analyses de ces différentes séries, menées avec les données des stations météorologiques du système d'observations du Merguellil, ont permis de valider le générateur stochastique. En effet, le générateur a pu corriger le biais observé initialement en utilisant les données ERA5 brutes. Toutefois, "MetGen" et aussi bien pour les deux méthodes de correction de biais n'ont pas réussi à corriger la distribution des valeurs fortes de la vitesse du vent. En effet, cette dernière forme la variable la plus turbulente et engendre des différences au niveau de ses valeurs simulées selon la méthode utilisée. Ces différences ont influencé par conséquent la simulation des variables de bilan d'énergie qui sont particulièrement sensibles aux valeurs de vitesse du vent. Par ailleurs, sur les 50 simulations de "MetGen", nous avons présenté les valeurs élevées et faibles des indices de stress (les quantiles 2.5% et 97.5% respectivement) qui reflètent à leur tour les faibles et les fortes valeurs de la vitesse du vent successivement. La présentation des valeurs d'indice de stress simulées à partir de ces deux quantiles a formé une sorte d'intervalle de confiance qui semble assez fiable pour les faibles valeurs de stress. Ces faibles valeurs représentent le début de stress hydrique. Cette information est particulièrement utile pour la détection des périodes de sécheresse et l'atténuation de ses conséquences dramatiques. De ce fait, le générateur est validé aussi sur le deuxième volet.

### 3.2.2 Publication

# Regional sub-daily stochastic weather generator based on reanalyses for surface water stress estimation in central Tunisia

Nesrine Farhani<sup>a,b,\*</sup>, Julie Carreau<sup>c,d</sup>, Zeineb Kassouk<sup>a</sup>, Bernard Mougenot<sup>b</sup>, Michel Le Page<sup>b</sup>, Zohra Lili-Chabaane<sup>a</sup>, Rim Zitouna-Chebbi<sup>e</sup>, Gilles Boulet<sup>b</sup>

<sup>a</sup>*Université de Carthage, Institut National Agronomique de Tunisie, Lr GREEN-TEAM, 43 avenue Charles Nicolle, Tunis, Tunisie*

<sup>b</sup>*Centre d'Études Spatiales de la Biosphère, Univ. de Toulouse, CNRS, CNES, IRD, UPS, INRAE, Toulouse, France*

<sup>c</sup>*HydroSciences Montpellier, Univ. de Montpellier, CNRS, IRD, Montpellier, France*

<sup>d</sup>*Polytechnique Montréal, Montréal, Canada*

<sup>e</sup>*LR VENC, INRGREF, University of Carthage, Rue Hedi Karray 2080, Ariana, Tunisia*

---

## Abstract

We present **MetGen** : a sub-daily multi-variable stochastic weather generator implemented as an R library that can be used to perform gap-filling and to extend in time meteorological observation series. **MetGen** is tailored to provide surrogate series of air temperature, relative air humidity, global radiation and wind speed needed for surface water stress estimation that requires sub-daily resolution. Multiple gauged stations within a homogeneous area can be used for the calibration. The approach relies on Generalized Linear Models that use, among their covariates, large-scale variables derived from ERA5 reanalyses. **MetGen** aims at preserving key features of the meteorological variables along with inter-variable dependencies. We illustrate the abilities of **MetGen** using a case study with three stations in central Tunisia. We consider as alternatives a univariate and a multivariate bias correction techniques along with the un-processed large-scale variables.

*Keywords:* stochastic weather generator, bias correction, surface water stress estimation, sub-daily resolution, ERA5 reanalyses

---

## 1. Introduction

In semi-arid areas, water is a major limitation factor for agricultural production. Indeed, these areas are characterized by short rainy seasons and strong variability of precipitation events in time and space [1]. Natural variations in the water cycle affect the availability of water, leading to irregularities in agricultural production [2] and constitutes the main driver of agricultural

---

\*Corresponding author

Email addresses: [farhani.nesrine@gmail.com](mailto:farhani.nesrine@gmail.com) (Nesrine Farhani), [julie.carreau@ird.fr](mailto:julie.carreau@ird.fr) (Julie Carreau)

droughts. The vegetation health status being generally representative of water availability [3], an important issue concerns the detection of surface water stress and the estimation of evapotranspiration (ET). Water stress may be deduced from ET with energy balance models. At satellite overpass time, energy balance models compute instantaneous ET as the residual term of the land surface energy balance equation, once net radiation, soil heat flux and sensible heat flux are derived from remotely sensed surface temperature [4, 5, 6]. Such water stress estimates are particularly informative for the detection of incipient plant stress during early stages of drought development compared to estimates derived from other wave lengths (microwave or visible) and allow to launch early drought alerts [7].

Energy balance models use as inputs satellite data (normalized difference vegetation index, albedo and surface temperature) and in-situ meteorological observations (air temperature **AirT**, relative air humidity **Rh**, global radiation **GR** and wind speed **WS**) as provided by gauged networks. ET and water stress estimates computed from the instantaneous surface energy budget constrained by the surface temperature require meteorological observations acquired at the satellite overpass time. To ensure precise timing with satellite information, in-situ meteorological observations must be available at sub-daily resolution. In this work, we use satellite data provided by the latest MODIS collection (<http://earthexplorer.usgs.gov>) that has a 1 km spatial resolution. ET and water stress are estimated over a region covered by several MODIS grid cells. This region is defined so that it can be considered to be homogeneous in terms of climate and weather. As a consequence, a single multi-variable meteorological series representative of this homogeneous region is needed. Nevertheless, there may be several gauged stations in the region. Moreover, it is often the case that the observed meteorological series are available over too short periods of time and present numerous gaps. Therefore, an important task is to develop a rigorous way to obtain a representative sub-daily multi-variable meteorological surrogate series in which gaps are filled and that extends in time the original series by exploiting the information provided by all the gauged stations in the region.

Stochastic Weather Generators (SWG) are stochastic models based on statistical approaches for simulating, at high spatial resolution, surrogate meteorological series that are similar to observation series in terms of distributional properties, preserving both systematic and random variations [8]. SWGs are thus very useful models to perform coherent gap filling and to generate realistic surrogate series over periods for which no observations are available. In the aforementioned surface water stress application, the sub-daily series of the four meteorological variables (AirT, Rh, GR and WS) display both annual and diurnal cycles. Once these primary systematic variations are accounted for, there will likely remain some random variations. There are two main strategies in SWGs to model systematic and random variability which are applicable in the case of multiple meteorological variables at several gauged stations : the weather type approach that breaks down weather into classes (or types or states) of typical recurring meteorological situations (e.g., clear blue sky, cloudy, rainy, *etc...*) and the stochastic regression approach that captures the variability of weather smoothly by using the proper covariates.

In the weather type approach, the underlying assumption is that each time step belongs to one weather type and all the time steps belonging to the same weather type can be modeled with a relatively simple statistical approach. In other words, the temporal sequence is grouped into blocks, each block being associated to one weather type [8, 9]. One of the earliest weather type SWG proposed by [10] models precipitation as a two-state Markov chain that corresponds to two simple weather types : wet and dry types. [11] builds on the latter model and represents the intensity of precipitation, for the wet weather type as an exponential probability distribution. The other three variables (maximum/minimum temperature and solar radiation) are modeled with a multivariate normal distribution whose means and standard deviations change according to the wet or dry types. In more recent approaches, more general weather types may be obtained automatically as an unsupervised classification problem. They thus are defined as the classes resulting from the clustering of time steps with each time step characterized by climatic or meteorological features [12]. Weather types may also be defined indirectly as the states of a latent variable using, for instance, hidden markov models [13]. The analog approach in SWGs can be seen as pushing the weather type strategy to the limit where each time step constitutes a weather type [14, 15]. Analog-based SWGs may be entirely non-parametric and they may succeed in reproducing complex patterns between variables and between sites. However, non-parametric analog-based SWGs are essentially resampling schemes that are unable to simulate values and patterns that differ from those present in the observations. This may be a serious drawback when observation periods are not long enough to contain all potential patterns. To account for annual cycles, weather types may be modeled separately for each season, with the difficulty that the definition of seasons might be somewhat arbitrary [12, 16]. Except in the case of rainfall [17], sub-daily weather typing for variables such as temperature, humidity and solar radiation requires a suitable model of the daily cycle. It is not yet clear how to adapt the weather type strategy in order to account for the presence of diurnal cycles [8]. Stochastic regression or, equivalently conditional distribution modeling, is another widely used strategy that can account for both systematic and random variability in SWGs. For instance, [18] links the two parameters of the gamma distribution that models the intensity of rainfall and the probability of rainfall to information on the rainfall pattern on the preceding day, the time of the year, *etc* ... with a one-hidden-layer feed-forward neural network. More generally, instead of focusing on estimating the conditional mean as is the case in conventional regression, stochastic regression seeks to estimate the full conditional distribution from which simulations can be drawn thereby allowing to reproduce the observed variability. Covariates that carry temporal and spatial information can be introduced letting the parameters of the conditional distribution vary in time and space. As an alternative to neural networks, Generalized Linear Models (GLMs) have been used within SWGs for the past 20 years or so, see for instance [19, 20] and the references therein. In GLMs, the conditional distribution belongs to the exponential family that encompasses the gaussian distribution. The link between the parameters and the covariates is established with a potentially transformed (e.g., with a logarithm) linear regression [21]. Routines to implement GLMs are readily available in standard statistical software (e.g., R [22]). Spatial dependence may be accounted for by modeling the dependence structure of the residuals, e.g., with gaussian

processes [20]. If informative enough covariates are used, GLM-based SWGs can reproduce very accurately both systematic and random variability [23]. In particular, despite that GLM-based SWGs generally operate at the daily resolution, sub-daily resolution modeling may be achieved by introducing covariates carrying sub-daily information. Most GLM-based SWGs simulate only one or two meteorological variables (very often, precipitation and temperature as in [20]). One notable exception is [23] who proposes a simple scheme to model jointly several meteorological variables based on the decomposition of the multivariate density into a product of conditional univariate densities.

In this work, we introduce **MetGen**, an SWG based on GLM, hence relying on stochastic regression, that extends the approach described in [23] to the sub-daily resolution. Its implementation is publicly and freely available as an R library (<https://CRAN.R-project.org/package=MetGen>). In **MetGen**, the scheme proposed in [23] to model jointly several meteorological variables is adapted to the four meteorological variables (AirT, Rh, GR and WS) required for surface water stress estimation for which inter-variable dependencies are rather strong. In contrast to [23] who has proposed a way to model spatial dependence, inter-site dependence is not explicitly modeled in **MetGen**. The proposed SWG works in a manner similar to the so-called *regional approach* developed in hydrology [24]. Indeed, several stations within the region of interest, which is assumed to be homogeneous, may be used to calibrate the SWG to augment the size of the data set. In addition to the covariates proposed in [25], special covariates are considered to enable the reproduction of diurnal cycles, based on pairs of sines and cosines, similarly as for annual cycles. An important category of covariates used to carry sub-daily information albeit at a large-scale (horizontal resolution of 31 km) are the meteorological reanalyses provided by ERA5, available at hourly resolution [26].

Since **MetGen** makes use of reanalyses in its covariates and because there are no other, to our knowledge, publicly and freely available multi-variable sub-daily SWG, we resorted to two *statistical downscaling* methods as comparative approaches. Statistical downscaling aims to bridge the gap between low resolution and potentially biased simulations from global climate models and the high resolution series required for impact studies such as observation series from gauged networks [27, 28]. Although often used to obtain climate change scenarios over future periods, statistical downscaling methods may be applied to reanalysis products in order to generate surrogate series over past periods [27]. An active area of research in statistical downscaling concerns the so-called bias correction methods [29]. Bias correction aims at transforming the low resolution series of meteorological variables such as provided by reanalyses so as to match, in terms of distributional properties (e.g., in terms of means), the high resolution series such as measured at gauged stations. In order to assess whether explicitly accounting for inter-variable dependencies is essential, we include as comparative approaches a univariate (**CDF-t** developed in [30]) and a multivariate bias correction method (**MBCn** proposed by [31]). Both methods provide fast, non-parametric (i.e., without strong distributional assumptions) alternatives to **MetGen** and are implemented as publicly available R libraries.

The paper is organized as follows. Section 2 presents our study area, the Merguellil plain in central Tunisia together with the meteorological data provided by gauged stations and derived from ERA5 reanalyses. The multi-variable sub-daily GLM-based SWG **MetGen** is described in section 3 along with the two aforementioned bias correction methods and their adaptation to enable their application at the sub-daily resolution. Section 4 is devoted to the comparison of the statistical methods at generating surrogate meteorological series both in terms of the ability to fill gaps in the observation series and in terms of the ability to extend in time the observation series. Section 5 reports an application of the surrogate meteorological series for surface water stress estimation. In section 6, a discussion is presented followed by conclusions and research perspectives in section 7.

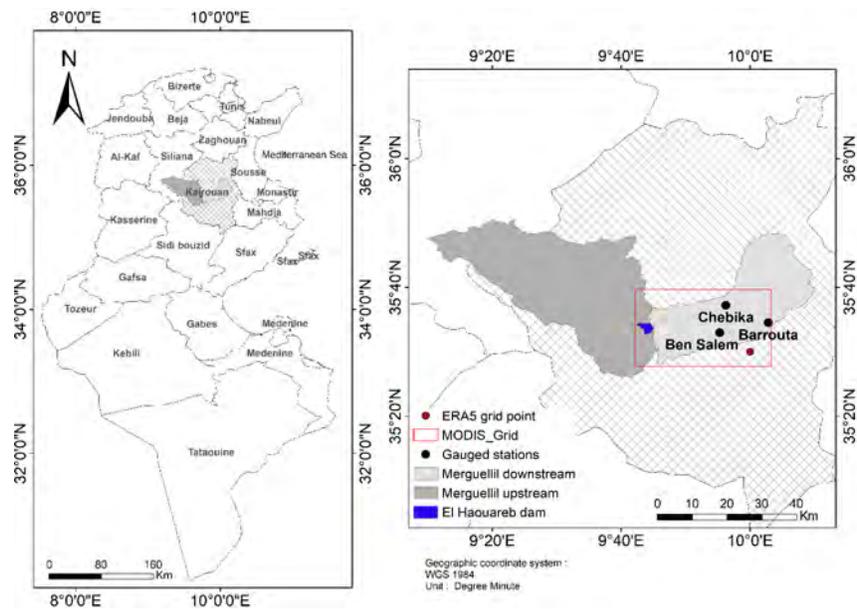
## 2. Study area and meteorological data

### 2.1. Study area : the Merguellil plain

The study area is part of the downstream plain of the Merguellil catchment called the Merguellil plain, see Fig 1. Lying in a semi-arid region located in central Tunisia, the catchment is characterized by a relatively mountainous upstream area (1200 km<sup>2</sup>) and by a downstream alluvial plain (676 km<sup>2</sup>). The upstream area presents a hilly topography (altitude between 200 and 1200 m with a median elevation of 500 m) [32]. In the plain, the landscape is mainly flat, and the vegetation is typical of semi-arid regions : rainfed agriculture (olive tree and cereals) and summer vegetables (melons, peppers and tomatos). Downstream farms are composed mainly of small cultivated areas [33]. The upstream and downstream areas are separated by the El Haouareb dam (Fig 1), which was built in 1989 to protect villages from inundations and to store irrigation water for the plain [34]. The study area is influenced both by the Mediterranean climate (dry subhumid) and the pre-Saharan climate (arid) [1]. It is characterized by the inter-annual irregularity of precipitation, with an average of annual rainfall of about 300 mm per year, and by a high evaporative demand of about 1600 mm per year. Water supply is by far insufficient to meet water demand which is rising steadily. The rise is due to the increase in population and industrial development and, most importantly, to the intensification of agriculture, which is the main water consumer (around 80 %) [35].

### 2.2. Meteorological observations

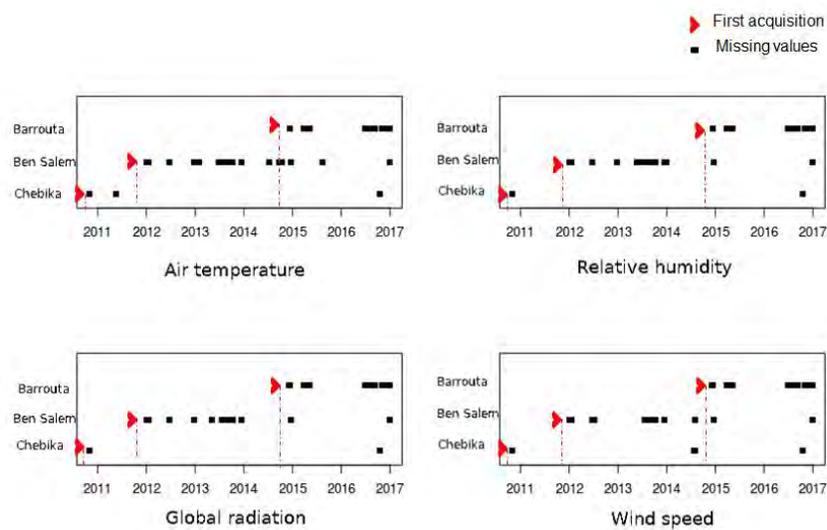
Hourly observation series of the four meteorological variables (air temperature **AirT**, relative air humidity **Rh**, global radiation **GR** and wind speed **WS**) needed for surface water stress estimation are collected from the three gauged stations, Ben Salem, Chebika and Barrouta, located in the Merguellil plain (see Fig. 1). The observation period, approximate number of observations and approximate percentage of missing values are given in Table 1. In addition, Fig. 2 illustrates the observation period and the positions of the gaps in the series.



**Figure 1:** Localisation of gauged stations : the plain located downstream of the Merguellil catchment in central Tunisia.

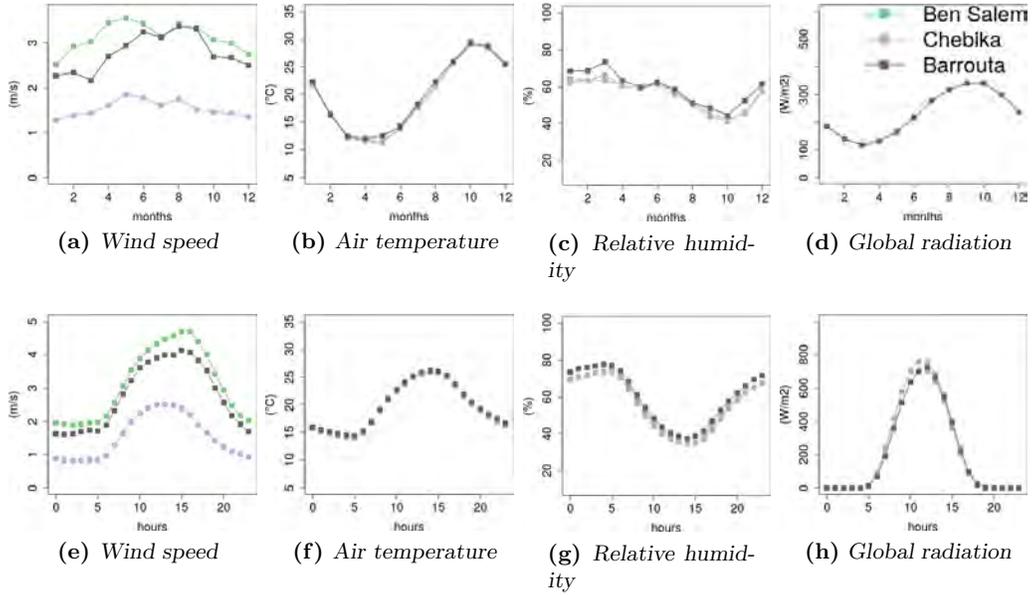
**Table 1:** Observation series at the three stations in our study area (see Fig. 1). The number of observations and of missing values may vary slightly depending on the meteorological variable.

Station	Obs. period	# Obs.	% Miss. values
Chebika	2011 - 2016	54 181	0.01
Ben Salem	2012 - 2016	43 818	0.2 - 1.2
Barrouta	2014 - 2016	18 106	4



**Figure 2:** Observation period and lengths of the gaps for each meteorological variables at the three gauged stations in the Merguellil plain.

The three gauged stations provide similar meteorological information owing to their geographical proximity (7.9 km between Ben Salem and Chebika, 11.7 km between Ben Salem and Barrouta and 11 km between Chebika and Barrouta). As shown by the annual and diurnal cycles in Fig. 3, these stations share the same climatic variability with the exception of the wind speed observed at Chebika. Lower wind speed values are caused by the presence of a windbreak in the vicinity of the station. In addition, inter-station pair plots (not shown) confirmed the strong relationship in the meteorological information provided by the three stations. We therefore assume, in what follows, that the region where the three stations are located is homogeneous and that the distribution of the meteorological variables is the same at all the stations apart for the wind for which a special covariate will be introduced (see details in § 3.1). Ben Salem is selected as the *reference station* as it complies best with the meteorological standards according to the WMO guidelines [36].



**Figure 3:** Annual (top row) and diurnal (bottom row) cycles for each observed meteorological variable at the three gauged stations in the Merguellil plain.

### 2.3. Meteorological reanalyses (ERA5)

Reanalyses combine forecast models and observations through data assimilation schemes thereby providing a multivariate, spatially complete and coherent record, without gaps, of atmospheric, land and oceanic climate variables [37, 26]. In particular, ERA5 reanalyses are available for a long period in the past, from 1950 till now [26]. Despite being available at hourly resolution, the ERA5 spatial resolution is low (horizontal resolution of 31 km [26]) and thus local-scale variability might not sufficiently be accounted for [38]. Besides the mismatch in spatial resolution, several limitations affected the quality of previous reanalyses such as ERA-Interim which were improved with respect to most aspects for ERA5 [26].

The three gauged stations from the Merguellil plain lie in the same ERA5 grid cell whose center is shown in Fig. 1. ERA5 reanalyses were extracted at this grid cell and were combined to obtain the six large-scale meteorological variables listed in Table 2. In most cases, the large-scale variables correspond to raw reanalysis products. There are two exceptions. The first one concerns the wind speed that was derived by taking the Euclidean norm of the 10 m vertical and horizontal wind components. The second exception concerns the relative humidity that was derived based on 2 m temperature and 2 m dewpoint temperature ERA5 products, according to the procedures defined in [39].

The six large-scale variables from Table 2 serve as covariates in the statistical methods described in section 3 to obtain surrogate meteorological series for the Merguellil plain. Among these, the first four are the large-scale counterpart of the meteorological variables needed for the surface water stress application. To evaluate the quality of the reanalysis products, these four large scale variables are used without further processing as one of the candidate surrogate meteorological series. It is expected that the statistical methods should be able to correct departures in terms of distributional properties of the large-scale variables. Therefore, other large-scale data, whether reanalyses or remote sensed, could be used instead of ERA5.

**Table 2:** Large-scale meteorological variables deduced from ERA5 reanalyses at the grid cell encompassing the three gauged station from the Merguellil plain. The second column indicates when the large-scale variable is considered as the large-scale counterpart of one of the meteorological variable needed in the surface water stress application.

Large-scale meteo. var.	counterpart for
wind speed - 10 m (derived)	WS
air temperature - 2 m (raw)	AirT
relative humidity - 2 m (derived)	Rh
surface solar radiation downwards (raw)	GR
total cloud cover (raw)	
mean sea level pressure (raw)	

### 3. Statistical methods

#### 3.1. *MetGen* : a regional multi-variable sub-daily GLM-based SWG

We focus on *MetGen* implementation for the surface water stress application in central Tunisia. The required steps, summarized in Fig. 4, can be adapted in principle to any study area and to any other meteorological variables. *MetGen* is regional in the sense that the observations from several gauged stations can be used in the calibration to increase the sample size. As discussed in § 2.2, we may assume that the three gauged stations from the Merguellil plain lie in a homogeneous region. Once calibrated, *MetGen* simulates series at all the gauged stations. However, for our surface water stress application, a single series, representative of the region, is needed. To this end, we make use of the series corresponding to Ben Salem as it is our reference station, see § 2.2.

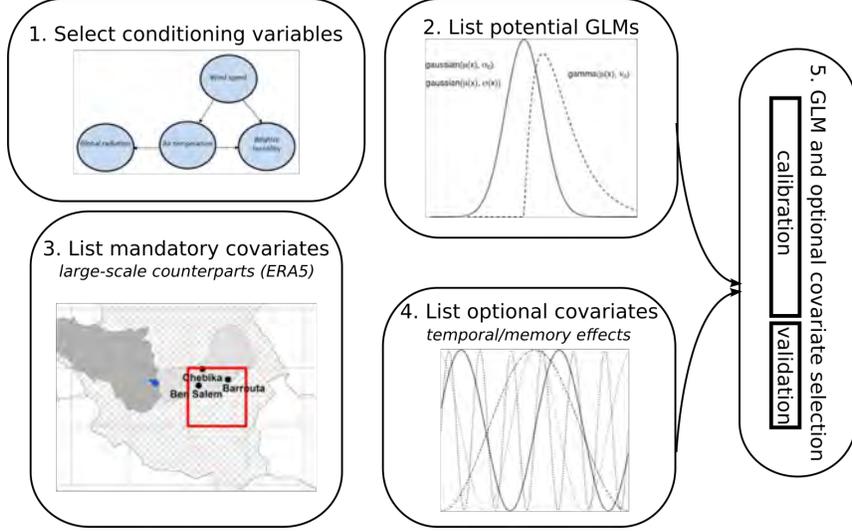


Figure 4: *MetGen* implementation steps.

### 3.1.1. Multi-variable modeling : conditioning variables

The first step to implement *MetGen* consists in modeling the inter-variable dependencies by means of conditioning variables (see Fig. 4). This follows the proposal of *RGlimclim* [25] by which a multivariate distribution can be decomposed with the product rule into conditional univariate distributions. To determine the order of the decomposition in the product rule and to reduce the number of conditioning variables, we rely on the dependence graph shown in Fig. 5. It has been adapted from the one made in the *HydEF* project (<https://www.imperial.ac.uk/media/imperial-college/research-centres-and-groups/environmental-and-water-resource-engineering/UCL15Feb2012.pdf>) to apply *RGlimclim* in the UK. More precisely, the multivariate distribution of the four meteorological variables needed for the surface water stress application boils down to modeling four conditional univariate distributions (one for each meteorological variable) and including in the covariates the appropriate conditioning variables :

$$\mathbb{P}(WS|\mathbf{x}) \tag{1}$$

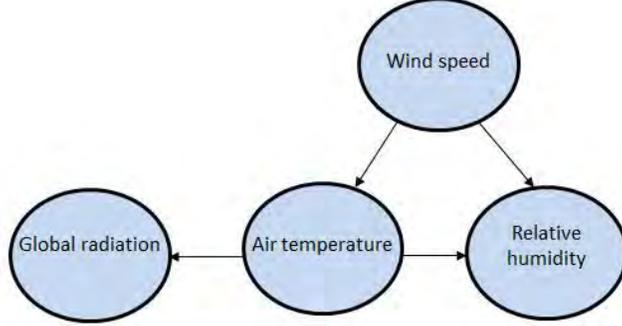
$$\mathbb{P}(AirT|WS, \mathbf{x}) \tag{2}$$

$$\mathbb{P}(Rh|AirT, WS, \mathbf{x}) \tag{3}$$

$$\mathbb{P}(Gr|AirT, \mathbf{x}). \tag{4}$$

where  $\mathbf{x}$  are additional covariates to be described in § 3.1.3. The choice of the conditional distribution model, the selection of covariates and the calibration can be performed separately for each meteorological variable. The simulation of the multi-variable surrogate series proceeds following the order dictated by the dependence graph in Fig. 5 : wind speed is simulated first, then air temperature is simulated including among the covariates the series simulated for wind speed, relative humidity is simulated afterwards with the previously simulated series for air temperature and wind speed included in the covariates and finally, global radiation is simulated

conditionally on the series simulated for air temperature.



**Figure 5:** Inter-variable dependency graph yielding the conditional univariate distributions in (1)-(4) which allows to model the dependencies among the four meteorological variables in *MetGen*.

### 3.1.2. Conditional univariate distribution models : Generalized Linear Models (GLMs)

In the second step of *MetGen*, potential conditional univariate distribution models, which are from the Generalized Linear Model (GLM) family, for each meteorological variable must be defined (see Fig. 4). At present, three possible choices of probability distributions for the GLMs are available in *MetGen* : the gaussian distribution with constant (homoscedastic) or non-constant variance (heteroscedastic) and the gamma distribution. In the GLMs, the parameters of the probability distributions may vary according to covariates. In *MetGen*, we made the following choices to link the covariates to the probability distribution parameters. Let  $\mathbf{x}_\mu$  and  $\mathbf{x}_\sigma$  be two covariate vectors, let  $\beta_\mu$ , and  $\beta_\sigma$  be regression coefficient vectors of the same length as  $\mathbf{x}_\mu$  and  $\mathbf{x}_\sigma$  respectively and let  $\mu_0$ ,  $\sigma_0 > 0$  and  $\nu_0 > 0$  be three constants. Then, the parameters of the conditional distributions are provided as follows for each of the three possible choices :

$$\text{homoscedastic gaussian} : \begin{cases} \mu(\mathbf{x}_\mu) = \mathbf{x}'_\mu \beta_\mu + \mu_0 & \text{Location param.} \\ \sigma_0 & \text{Scale param.} \end{cases} \quad (5)$$

$$\text{heteroscedastic gaussian} : \begin{cases} \mu(\mathbf{x}_\mu) = \mathbf{x}'_\mu \beta_\mu + \mu_0 & \text{Location param.} \\ \sigma(\mathbf{x}_\sigma) = \exp(\mathbf{x}'_\sigma \beta_\sigma + \sigma_0) & \text{Scale param.} \end{cases} \quad (6)$$

$$\text{gamma} : \begin{cases} \mu(\mathbf{x}_\mu) = \exp(\mathbf{x}'_\mu \beta_\mu + \mu_0) & \text{Location param.} \\ \nu_0 & \text{Shape param.} \end{cases} \quad (7)$$

Each of these models may be fitted by maximizing the log-likelihood (with the `glm` function in the base package of `R` for (5) and (7) and with the package `lmvar` for (6)).

Some preprocessing is performed on the raw observed series before model fitting. First, time steps for which global radiation is assumed to be zero (i.e., during the night) are determined based on the time of the sunrise and of the sunset at the coordinates of the station and for the given day of the year (see `R` package `insol`). Model fitting and simulation for global radiation is performed only on identified diurnal time steps. Second, preliminary transformations are

defined for three meteorological variables (WS, Rh and GR) so as to remove range constraints and make them more likely to be suitably modeled by the gaussian distribution, see Table 3. For each meteorological variable, either two (the homo- or heteroscedastic gaussian, see (5)-(6)) or three choices of probability distributions (including also the gamma, see (7), for WS and GR that take only positive values) with a preliminary transformation when necessary are considered as potential models for each meteorological variable, see the complete list in Table 3.

**Table 3:** *Potential conditional distribution models, with a preliminary transformation if necessary, considered for each meteorological variable.  $\Phi^{\leftarrow}(\cdot)$  indicates the quantile function of the standard Normal distribution and  $1360.4 \text{ W/m}^2$  is the solar constant.*

Meteo. var.	Range constraint	Transformation	Prob. distr.	Model
WS	$WS > 0$	$\ln(\exp(WS) - 1)$	Homo. gaussian (5)	$\mathcal{M}_1^{WS}$
		$\ln(\exp(WS) - 1)$	Hetero. gaussian (6)	$\mathcal{M}_2^{WS}$
		$\boldsymbol{\times}$	Gamma (7)	$\mathcal{M}_3^{WS}$
AirT	$\boldsymbol{\times}$	$\boldsymbol{\times}$	Homo. gaussian (5)	$\mathcal{M}_1^{AirT}$
			Hetero. gaussian (6)	$\mathcal{M}_2^{AirT}$
Rh	$0 < Rh < 1$	$\Phi^{\leftarrow}(Rh)$	Homo. gaussian (5)	$\mathcal{M}_1^{Rh}$
GR	$0 < GR < 1360.4$	$\Phi^{\leftarrow}(GR/1360.4)$	Hetero. gaussian (6)	$\mathcal{M}_2^{Rh}$
		$\Phi^{\leftarrow}(GR/1360.4)$	Homo. gaussian (5)	$\mathcal{M}_1^{GR}$
		$\boldsymbol{\times}$	Hetero. gaussian (6)	$\mathcal{M}_2^{GR}$
			Gamma (7)	$\mathcal{M}_3^{GR}$

### 3.1.3. Mandatory and optional covariates

This corresponds to steps 3 and 4 in Fig. 4. Mandatory covariates, which are always included in the models defined in (5)-(7), are set as follows. For the location parameter  $\mu(\boldsymbol{x}_\mu)$ , either of the gaussian distributions in (5)-(6) or of the gamma distribution in (7),  $\boldsymbol{x}_\mu$  includes the conditioning variables and the large-scale variables listed in the column 2 and 3 respectively of Table 4. To limit model complexity, the mandatory covariates included in  $\boldsymbol{x}_\sigma$  for the scale parameter of the heteroscedastic gaussian distribution,  $\sigma(\boldsymbol{x}_\sigma)$  in (6), only include the large-scale variables, listed in the 3<sup>rd</sup> column of Table 4.

**Table 4:** *Mandatory covariates used in MetGen for each meteorological variable needed in the surface water stress application : conditioning variables to introduce inter-variable dependencies based on the dependence graph from Fig. 5 and large-scale variables obtained from ERA5 reanalyses (see Table 2).*

Meteo. var.	conditioning var.	large-scale variables
WS	$\boldsymbol{\times}$	10 m wind speed & mean sea level pressure
AirT	WS	2 m temperature
Rh	AirT, WS	relative humidity
GR	AirT	surface solar radiation downwards & total cloud cover

In addition to these mandatory covariates, other covariates may be optionally included in  $\boldsymbol{x}_\mu$  to model systematic temporal variability (annual and diurnal cycles) and to account for temporal persistence (memory effects). Let  $Y_{t,s}$  be the meteorological variable of interest (either wind speed, air temperature, relative humidity or global radiation), at time step  $t$  and at site  $s \in \{1, \dots, S\}$ . The following optional covariates are considered :

- pairs of cosines and sines with annual oscillations :

$$\text{Annual cycle covariates : } \cos\left(\frac{2\pi d}{k_d}\right) \quad \sin\left(\frac{2\pi d}{k_d}\right) \quad (8)$$

with  $1 \leq d \leq 366$ , the day of the year associated to time step  $t$  and  $k_d \in (365, 183, 91, 30)$  ;

- pairs of cosines and sines with diurnal oscillations :

$$\text{Diurnal cycle covariates : } \cos\left(\frac{2\pi h}{k_h}\right) \quad \sin\left(\frac{2\pi h}{k_h}\right) \quad (9)$$

with  $1 \leq h \leq 24$ , the hour of the day associated to time step  $t$  and  $k_h \in (24, 12, 6)$ .

- lagged values of the meteorological variable :

$$\text{Var.lagk : } Y_{t-k,s} \quad k \geq 1 \quad (10)$$

- lagged values of the spatial average of the meteorological variable :

$$\text{SA.lagk : } \frac{1}{S} \sum_{s=1}^S Y_{t-k,s} \quad k \geq 1 \quad (11)$$

- lagged values of  $b$ -moving averages, with  $b > 1$  :

$$\text{MAb.lagk : } \frac{1}{b} \sum_{j=1}^b Y_{t-j+1-k,s} \quad k \geq 1 \quad (12)$$

- lagged values of spatial  $b$ -moving averages, with  $b > 1$  :

$$\text{SMAb.lagk : } \frac{1}{bS} \sum_{j=1}^b \sum_{s=1}^S Y_{t-j+1-k,s} \quad k \geq 1. \quad (13)$$

### 3.1.4. Selection of the conditional distribution model and of optional covariates

This is the last step to implement **MetGen**, step 5 in Fig. 4. For each meteorological variable, a conditional distribution model, see Table 3, must be selected and additional optional covariates may be included in the covariate set. Statistical tests that are conventionally used to perform model and covariate selection, e.g., based on p-values or likelihood ratios, rely on the assumption that observations are conditionally independent which is likely not the case in our application. To circumvent this issue, model and covariate selection is performed with a calibration-validation scheme in which the data is split into a calibration period, used for model fitting, and a validation period, used to assess model performance. The performance criteria are detailed in Table 5.

The first stage consists of selecting the conditional distribution model for a given meteorological variable. All the potential models, see Table 3, with the covariate set restricted to the mandatory

covariates, see Table 4, are fitted on the calibration period. The choice of conditional distribution model yielding the best performance computed on the validation period is retained for the subsequent stages. In the following stages, the conditional distribution model is thus fixed and optional covariates, among those in (8)-(13), are added to the covariate set gradually. The model with the larger covariate set is fitted on the calibration period. The optional covariates are retained when the performance criteria computed on the validation period are improved.

**Table 5:** *Performance criteria for the selection of a conditional distribution model and of optional covariates.  $y_{(i),s}$  and  $\hat{y}_{(i),s}$  with  $i = 1, \dots, n$  are resp. the sorted observations and sorted surrogates of a given meteorological variable at site  $s$ .*

$$\text{Main criterion} \quad \sum_s \sum_{i=1}^n (y_{(i),s} - \hat{y}_{(i),s})^2 \quad (14)$$

Visual criteria	<ul style="list-style-type: none"> <li>• annual and diurnal cycles plots</li> <li>• temporal auto-correlation plots</li> <li>• inter-variable dependence plots</li> </ul>
-----------------	---

The strategy adopted for performance assessment is based on the minimization of (14), the mean squared error of quantile-quantile plots (qq-plots) that assesses how well the distribution is reproduced, as long as no lacks of fit are detected from the plots (see the visual criteria in Table 5). An example of addition of optional covariates that arose was when a lack of auto-correlation in the surrogate series was noticed that led to the inclusion of memory effects, see (10)-(13). Another example is the presence of the windbreak at Chebika station that generates differences in the diurnal and annual cycles, see Fig. 3a and 3e. This led to the inclusion of a special binary covariate (0 indicates no windbreak while 1 indicates the presence of a windbreak) that allows the intercept term in  $\mathbf{x}'_{\mu}\beta_{\mu} + \mu_0$  (see (5)-(7)) to take on a different value according to whether there is a windbreak or not. Some interactions among the covariates were tested but none of them brought significant performance improvements in our application so that none were retained.

### 3.2. Sub-daily bias correction techniques

#### 3.2.1. A univariate and a multivariate bias correction techniques

Initial bias correction techniques such as the quantile-matching method [40] are univariate, i.e., they seek to transform a single series (representing a single meteorological variable at a single location). The quantile-matching method relies on a transformation that combines the cumulative distribution functions (CDFs) of the high resolution and the low resolution series estimated over a calibration period. This transformation ensures that the CDF of the corrected series matches the high resolution series' CDF accurately over the calibration period. [30] built on the quantile-matching method to propose a transformation, called CDF-t, that incorporates additionally the CDF of the low resolution series over the study (or validation) period. More recent bias correction methods are multivariate, i.e., they correct jointly multiple series (either from

several locations or for several meteorological variables or both) seeking to reproduce, in addition to univariate distributional properties, the dependence structures present in the series [31, 41]. One such recent multivariate bias correction method is the N-dimensional probability density function transform (**MBCn**) proposed by [31]. The **MBCn** method looks iteratively for linear combinations of the variables and performs bias correction with a univariate bias correction method such as quantile-matching or **CDF-t** on the linear combinations rather than on each variable separately. These two bias correction techniques, **CDF-t** and **MBCn** are described next.

As previously, let  $Y_{t,s}$  be the meteorological variable of interest (either wind speed, air temperature, relative humidity or global radiation), at time step  $t$  and at given site  $s$ . Let  $X_{t,m}$  be its large-scale counterpart provided by the ERA5 reanalyses, as listed in the first four rows of Table 2, at the same time step  $t$  and at the grid cell  $m$  that contains the site  $s$ . In other words,  $Y_{t,s}$  is the high resolution meteorological variable from the gauged station and  $X_{t,m}$  is its large-scale version obtained from the reanalyses. Let us assume that there is a period used for calibration for which both  $Y_{t,s}$  and  $X_{t,m}$  are available and a period used for validation for which only  $X_{t,m}$  is available. **CDF-t** estimates  $Y_{t,s}$ , a single meteorological variable at a single site, over the validation period as :

$$\hat{y}_{t,s} = \tilde{F}_{X_m}^{\leftarrow} (F_{X_m} (F_{Y_s}^{\leftarrow} (\tilde{F}_{X_m}(x_{t,m})))) \quad (15)$$

where  $x_{t,m}$  is the value of the large-scale variable  $X_{t,m}$  that actually occurred on time  $t$  of the validation period,  $F_Z$  and  $F_Z^{\leftarrow}$  denote respectively the empirical cumulative distribution function of the random variable  $Z$  and its inverse, the quantile function, and  $F_Z(\tilde{F}_Z)$  indicates the empirical distribution function estimated over the calibration (validation) period (see [30] for more details).

In contrast to **CDF-t** which needs to be applied separately to each variable meteorological variable, **MBCn** works directly with all the meteorological variables for which bias correction needs to be performed. Let  $\mathbf{Y}_{t,s}$  be the vector of four meteorological variables at site  $s$  and time step  $t$ . Similarly, let  $\mathbf{X}_{t,m}$  ( $\tilde{\mathbf{X}}_{t,m}$ ) be the 4-dimensional vector of large-scale meteorological variables at the grid cell containing the gauged-station for the calibration (validation) period. **MBCn** relies on random orthogonal matrices  $\mathbf{R}$ . A univariate bias correction technique (in the implementation of **MBCn**, the quantile delta mapping is used, see [42] for detailed explanations) is applied separately by working on each element of the rotated vectors  $\mathbf{X}_{t,m}\mathbf{R}$ ,  $\tilde{\mathbf{X}}_{t,m}\mathbf{R}$  and  $\mathbf{Y}_{t,s}\mathbf{R}$ . Then the bias corrected large-scale variable vectors are rotated back. These steps are summarized as follows, with  $T(\cdot)$  denoting the univariate bias correction operator applied elementwise :

$$\mathbf{X}_{t,m} \leftarrow T(\mathbf{X}_{t,m}\mathbf{R})\mathbf{R}^{-1} \quad (16)$$

$$\tilde{\mathbf{X}}_{t,m} \leftarrow T(\tilde{\mathbf{X}}_{t,m}\mathbf{R})\mathbf{R}^{-1}. \quad (17)$$

This procedure is iterated with new random matrices  $\mathbf{R}$  until the multivariate distribution of

$\mathbf{X}_{t,m}$  matches the one of  $\mathbf{Y}_{t,s}$ . Then  $\tilde{\mathbf{X}}_{t,m}$  contains the bias corrected series in the validation period.

### 3.2.2. Working with anomalies at a single station

As our goal for the surface water stress application is to obtain a single surrogate series for each of the four meteorological variables that is representative of a homogeneous area, the two bias correction techniques are applied to the observation series at a single station, namely Ben Salem which is the reference station (see § 2.2). In other words, the large-scale variables obtained from ERA5 reanalyses are corrected, either with **CDFt** or with **MBCn**, to fill the gaps and extend in time the observation series at Ben Salem. **CDFt**, being a univariate approach, is applied separately for each meteorological variable while **MBCn** is applied to all four meteorological variables at once.

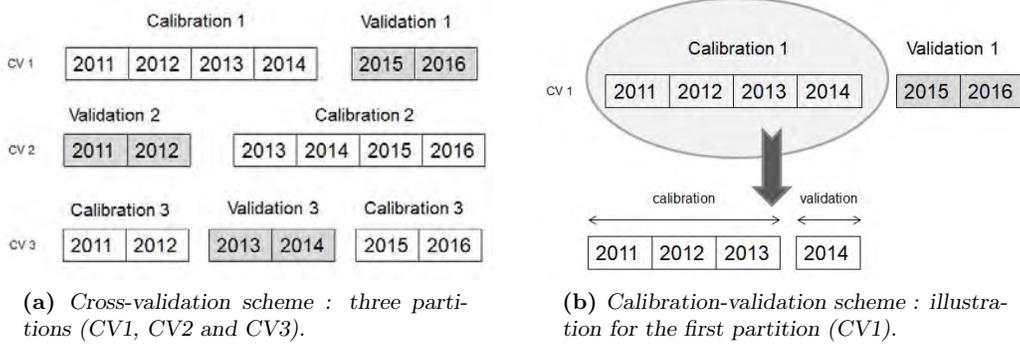
The following procedure is adopted to apply the two bias correction techniques at the sub-daily resolution. A conventional way to deal with the presence of annual cycles is to split the year into seasons and to apply bias correction separately on each season, see for instance [27]. However, the four meteorological variables used in the energy balance model (WS, AirT, Rh, GR) also display clear diurnal cycles, see Fig. 3. As this strategy (splitting the year into seasons) is not straightforward to extend to deal with diurnal cycles, we propose instead to work on anomalies of diurnal cycles with the diurnal cycle that is allowed to change with the season. More precisely, diurnal cycles are computed for three seasons : summer (June to August), winter (November to March) and inter-season (the remaining months). Observed (large-scale) anomalies are computed by subtracting the observed (large-scale) diurnal cycles from the observation (large-scale) series. Working with anomalies allows to remove systematic fluctuations from the meteorological variables and to focus on random fluctuations around the diurnal cycles. Bias corrected meteorological series are obtained by adding the observed diurnal cycles for each season to the bias corrected anomalies.

## 4. Evaluation and comparison

### 4.1. Cross-validation scheme

A cross-validation scheme is used to evaluate the statistical methods described in section 3 in terms of their ability to extend in time the original series, i.e., to simulate on periods for which no observations are available. Indeed, cross-validation is convenient for small data sets and is frequently used to evaluate out-of-sample performance [43, 44]. The cross-validation scheme is made of three temporal partitions of the observations at all the stations, CV1, CV2 and CV3, as presented in Fig. 6a. In each partition, the observation period (from 2011 to 2016) is split into a calibration period made of four years used for model fitting and a validation period consisting of two years for model evaluation and comparison. The statistical models thus simulate out-of-sample surrogate series over the 2 year validation set of each partition. To account for the stochastic aspect of **MetGen**, the series are replicated 50 times, i.e., for each time

step, 50 values are drawn from the conditional models. The replications are limited to 50 to keep the computation times reasonable while still permitting to explore the uncertainty captured by the models. Out-of-sample surrogate series are generated over the complete observation period by putting together the validation periods of the three partitions.



**Figure 6:** Performance evaluation and comparison. A second calibration/validation split within the calibration period of each partition of the cross-validation scheme is introduced to perform model and covariate selection for *MetGen*.

For each partition of the cross-validation scheme, the conditional distribution model and optional covariates must be selected for *MetGen*, as described in § 3.1.4. To this end, the calibration set of the partition, which has four years, is split into a smaller calibration period of three years and a validation period of one year, see Fig. 6b. The selected conditional distribution model with the selected optional covariate set and the mandatory covariates is then calibrated anew over the whole calibration set (four years) of the partition. Note that different selections of model and of optional covariates may occur for each of the three partitions of the cross-validation scheme.

The selections of conditional distribution models and optional covariates for each of the three partitions (CV1, CV2 and CV3) are as indicated in Table 6. For all the partitions, the conditional distribution model selected for all four meteorological variables is the heteroscedastic gaussian distribution, plus a preliminary transformation when needed (models  $\mathcal{M}_2^{WS}$ ,  $\mathcal{M}_2^{AirT}$ ,  $\mathcal{M}_2^{Rh}$  and  $\mathcal{M}_2^{GR}$  from Table 3). In addition to the mandatory covariates from Table 4, the optional covariates included in the final set of covariates are listed in Table 6. For a given meteorological variable, different covariate sets may be selected for each partition of the cross-validation scheme as the selection was performed separately for each partition. Nevertheless, the covariate sets are very similar in most cases. For the wind speed, for example, the special binary covariate indicating the presence of a windbreak at one of the station was deemed necessary for all partitions. Besides, no optional covariates related to seasonal or diurnal cycle for the relative humidity and no optional covariates related to memory effects for the global radiation were included for any partition as the improvement in performance was not significant.

**Table 6:** Model and covariate selection results for *MetGen* : for each meteorological variable conditional distribution models selected (see Table 3) and optional covariates (see (8)-(13)). For WS, the special binary covariate indicating the presence of a windbreak is included for each partition.

Partition	Meteo. var. & model	Annual oscill. (days)	Diurnal oscill. (hours)	Memory effects lagk = longest lag
CV1	WS : $\mathcal{M}_2^{WS}$	30, 365	12, 24	$\times$
	AirT : $\mathcal{M}_2^{AirT}$	183	24, 12	$\times$
	Rh : $\mathcal{M}_2^{Rh}$	$\times$	$\times$	$\times$
	GR : $\mathcal{M}_2^{GR}$	365, 183	24, 12	$\times$
CV2	WS : $\mathcal{M}_2^{WS}$	30, 365	12, 24	$\times$
	AirT : $\mathcal{M}_2^{AirT}$	183, 365	6, 12, 24	SA.lag3, MA.lag8, SMA.lag8, Var.lag3
	Rh : $\mathcal{M}_2^{Rh}$	$\times$	$\times$	SA.lag3, MA.lag3, SMA.lag7, Var.lag1
	GR : $\mathcal{M}_2^{GR}$	$\times$	24, 12	$\times$
CV3	WS : $\mathcal{M}_2^{WS}$	30	12, 24	$\times$
	AirT : $\mathcal{M}_2^{AirT}$	183	24, 12	SA.lag3, MA.lag3, SMA.lag1, Var.lag3
	Rh : $\mathcal{M}_2^{Rh}$	$\times$	$\times$	$\times$
	GR : $\mathcal{M}_2^{GR}$	$\times$	24, 12	$\times$

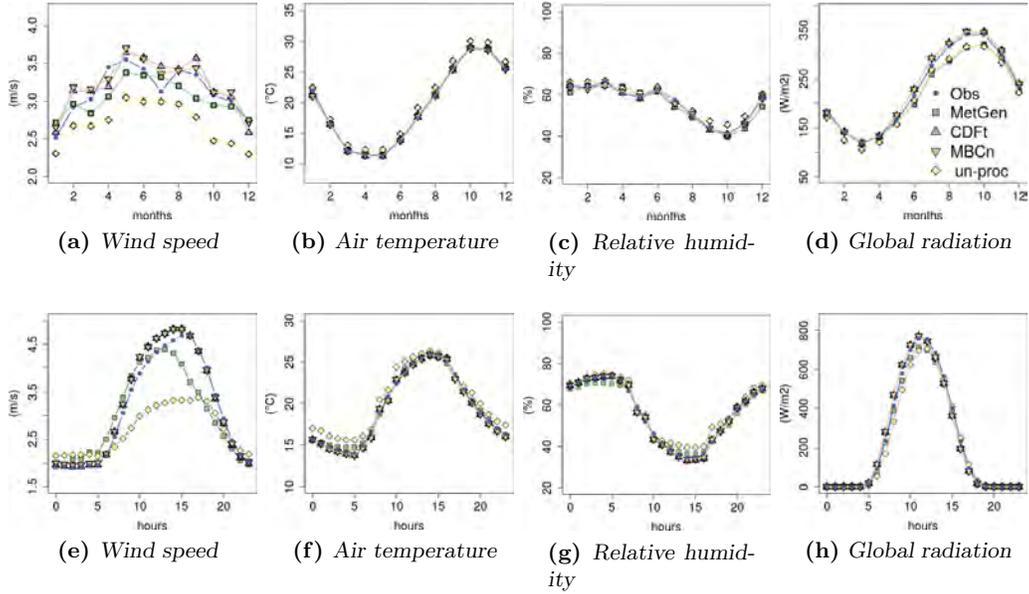
#### 4.2. Cross-validation evaluation

In what follows, we report the evaluation and comparison of the three statistical models, the SWG *MetGen* and the two bias correction techniques CDFt and MBCn (see § 3) based on the surrogate meteorological series simulated with the cross-validation scheme described in § 4.1. Recall that, for *MetGen*, the surrogate series is replicated 50 times and that only the series related to the Ben Salem station is kept in this evaluation. In addition to the surrogate series from the statistical models, the un-processed large-scale variables obtained from the ERA5 reanalyses, see Table 2, are included in the comparison.

##### 4.2.1. Annual and diurnal cycles

Fig. 7 presents the annual (top row) and diurnal (bottom row) cycles for each of the four meteorological variables. Annual cycles are computed by averaging values in each month over the observation period. Similarly, diurnal cycles are obtained as hourly averages.

The cycles of the un-processed large-scale variables (yellow diamonds in Fig. 7) accurately reproduce the observations' cycles (blue dots in Fig. 7) for most meteorological variables. This is the case for air temperature (Fig. 7f and Fig. 7b), relative humidity (Fig. 7g and Fig. 7c) and global radiation (Fig. 7h and Fig. 7d). However, the wind speed cycles are under-estimated in the un-processed large-scale variables series : the afternoon peak present in the diurnal cycle (Fig. 7e) is too low (about 3 m/s instead of about 4.5 m/s in the observation cycle) and the annual cycle values (Fig. 7a) are consistently below the observed ones. Despite relying on this information through its covariates, *MetGen* is able to correct fairly well the under-estimation shown in the un-processed large-scale series and to reproduce much more accurately annual and diurnal cycles of the wind speed variable. The series produced by the bias correction techniques are bias corrected anomalies to which observed diurnal cycles are added (see § 3.2). Since the



**Figure 7:** Annual (top row) and diurnal (bottom row) cycles for each meteorological variable at the Ben Salem station. Comparison between observed (in blue) and surrogate series (see color legend). Note that the surrogate series from the two bias correction techniques (CDFt and MBCn) are the sum of the observed diurnal cycles plus the corrected anomalies hence the good adequation with the observed cycles (see § 3.2).

root-mean-squared errors (not reported) of the bias corrected anomalies are low (below one in all cases), it follows that the cycles of the corresponding series are well reproduced.

#### 4.2.2. Goodness-of-fit of the whole distribution and of extreme values

Quantile-quantile plots (qq-plots) are used instead of scatter plots to assess whether the distribution of the observation series is well reproduced by the surrogate series. Indeed, all the statistical methods considered to generate surrogate series aim at providing a correction of the distributional properties of the large-scale variables rather than reproducing the chronology of the observed series. Similarly as in (14), let  $y_{(i)}$  and  $\hat{y}_{(i)}$ , with  $i = 1, \dots, n$ , be respectively the sorted observations and the sorted surrogates of a given meteorological variable for one of the approach (either one of the three statistical methods or the un-processed large-scale variables) corresponding to the Ben Salem station. Root Mean-Squared Errors (RMSEs) of the qq-plots relative to the standard deviation of the observations are defined as follows :

$$\frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_{(i)} - \hat{y}_{(i)})^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_{(i)} - \frac{1}{n} \sum_{i=1}^n y_{(i)})^2}}. \quad (18)$$

The relative RMSE is near zero when the qq-plot is well aligned on the first bisector which means that the distribution of the observations is well reproduced by the surrogates. It is below (above) one when the RMSE is smaller (greater) than the standard deviation, i.e. when the surrogate series is better (worse) than the empirical average at reproducing the observations.

The relative RMSEs for each meteorological variable and for each type of surrogate series are reported in Table 7. For **MetGen**, the median of the relative RMSEs of the 50 replications is reported. The relative RMSE is computed either on all the quantiles (indicated as 100 %) or the 1 % highest quantiles to focus on how extreme values are reproduced. On the complete distribution, all three statistical methods performed quite well (relative RMSEs are all rather close to zero). The multivariate bias correction technique, MBCn, performed best in general although often not by much. The un-processed large-scale variables always have the poorest performance (relative RMSEs are higher) especially for the wind speed. This indicates that all three statistical methods improved upon the distributional properties of the un-processed large-scale variables. Nevertheless, their performance is rather decent as the relative RMSEs are always lower than one. On the extreme values (i.e., the 1 % highest quantiles), the relative RMSEs give a very different picture. **MetGen** outperforms the two bias correction techniques for the wind speed and the air temperature. The un-processed large-scale variables relative RMSEs may be quite high, higher than one in three instances and sometime by a rather large factor. The extreme values of the relative humidity variable were the most difficult to reproduce in all surrogate series and is the only case in which all statistical methods do worst than the un-processed large-scale variables. This might be caused by the upper bound on the values taken by the relative humidity.

**Table 7:** Goodness-of-fit of the whole distribution and of extreme values. The relative RMSEs, see (18), are computed for all the quantiles (100%) and for the 1 % highest quantiles. For **MetGen**, the median of the relative RMSE of the 50 replications is provided. The best performance (lowest value) for each meteorological variable is indicated in *italic font* and values above one are indicated in **bold font**.

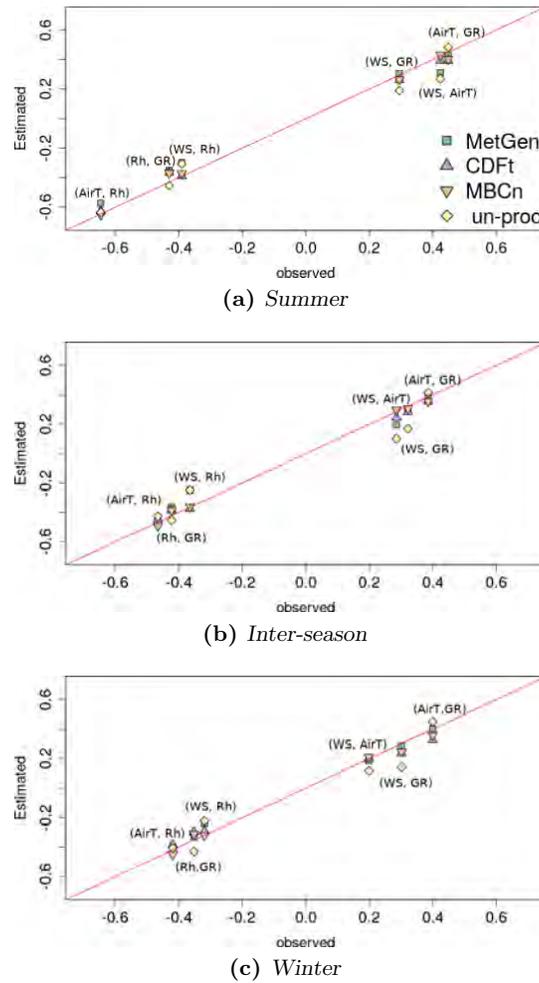
Quantiles	Meteo. var.	MetGen	un-proc.	CDFt	MBCn
100%	WS	0.12	0.4	0.08	<i>0.07</i>
	AirT	0.03	0.12	<i>0.02</i>	<i>0.02</i>
	GR	0.08	0.11	<i>0.02</i>	<i>0.02</i>
	Rh	0.05	0.13	0.04	<i>0.03</i>
1% highest	WS	<i>0.71</i>	<b>2.3</b>	<b>1.34</b>	0.97
	AirT	<i>0.16</i>	0.3	0.41	0.28
	GR	0.94	<b>3.88</b>	<i>0.44</i>	0.8
	Rh	<b>2.23</b>	<b>1.14</b>	<b>1.23</b>	<b>1.28</b>

#### 4.2.3. Inter-variable dependencies

Accurately reproducing inter-variable dependencies is particularly important since surface water stress is generally triggered by a combination of meteorological factors. In this evaluation, inter-variable dependence is summarized by Kendall’s  $\tau$ , a non-parametric correlation coefficient based on ranks, that is suitable for non-gaussian distributions (as opposed to the Pearson correlation coefficient) [45]. Positive values of Kendall’s  $\tau$  indicate that both variables tend to increase or decrease simultaneously while negative values indicate that they tend to vary in an opposite manner. A value near zero signals a lack of dependence. The comparison of correlation coefficients is carried out for each of the three seasons considered for the sub-daily bias correction techniques (see § 3.2) : summer (June to August), winter (November to March) and inter-season

(the remaining months).

In Fig. 8, the correlation coefficients computed from the observations are on the x-axis while those derived from the surrogate series are on the y-axis. There is an overall relatively good alignment along the first bisector (the red line) showing that the inter-variable dependence strength is rather well preserved using the different surrogate series and for all three seasons. Nevertheless, Kendall's  $\tau$  coefficients computed from the large-scale variables series tend to be less tightly aligned, especially when the wind speed (WS) variable is involved. For example, the correlation coefficient in winter (Fig. 8c) between the wind speed (WS) and the air temperature (AirT) in the observation series is about 0.2 whereas it is about 0.1 for the un-processed large-scale variable series. The correlation is always improved with the bias corrected series and, in most cases, with the **MetGen** series. A relatively strong negative dependence between the air temperature (AirT) and the relative humidity (Rh) is preserved in the different surrogate series and in the three seasons, especially in summer as it reaches -0.65 (Fig. 8a).



**Figure 8:** Inter-variable dependencies : comparison of Kendall's  $\tau$  according to seasons (summer, winter, inter-season) for each pair of meteorological variables from the observation series on the x-axis and from the surrogate series (see color legend) on the y-axis.

### 4.3. Gap-filling exercise

**MetGen** can run in gap-filling mode. As the simulation proceeds step by step, when observations are found missing, surrogate values are simulated by **MetGen** and the covariates introducing memory effects (see (10)-(13)) are updated based on the simulated values. Simulations of **MetGen** in gap-filling mode can be repeated to account for the uncertainty. The two bias correction techniques may also be used to perform gap-filling. However, in contrast to **MetGen**, the surrogate series are produced in the same way as for a validation period. A gap-filling exercise is carried out visually by inspecting the chronological plots of the surrogate series over one day, December 26th 2014, for which the observed values of all four meteorological variables from the Ben Salem station were removed artificially.

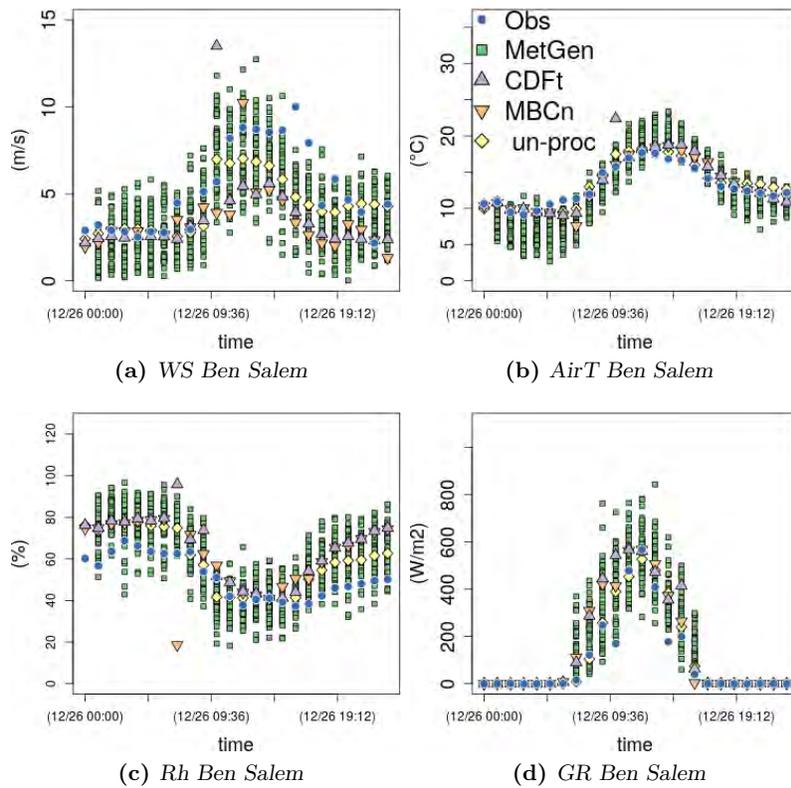
Fig. 9 presents the observed series (in blue) at the Ben Salem station over the day selected for the gap-filling exercise. The surrogate series of **MetGen** produced in gap-filling mode (50 replications), of the two bias correction techniques and of the un-processed large-scale variable are superimposed (see color legend). In Fig. 9, we observe that, the simulated values from the three statistical methods reproduce rather well the original values observed at the Ben Salem station. The 50 simulations from **MetGen** are generally centered around the large-scale series and, most importantly, their spread covers the observed series. We also note that **MetGen** is able to rectify values that are too low (e.g., the wind speed in Fig. 9a) or too high (e.g., the relative humidity in Fig. 9c) that are present in the un-processed large-scale variable series.

## 5. Surface water stress application

### 5.1. SPARSE : a dual-source energy balance model

Surface water stress may be deduced from evapotranspiration ( $ET$ ) using energy balance models. At satellite overpass time, energy balance models compute instantaneous latent heat flux ( $LE$ ), expressed in  $W/m^2$ , as the residual term of the land surface energy balance equation [4, 5, 6]. In this application of surface water stress estimation, we use the dual-source model Soil Plant Atmosphere and Remote Evapotranspiration (SPARSE) [46] which is based on the same rationale as TSEB (Two-Source Energy Balance model) [5]. SPARSE derives from the remotely sensed surface temperature ( $T_{surf}$ ) separate estimates of the instantaneous fluxes of the soil (subscript s) and vegetation (subscript v) components of the energy budget at the satellite overpass time. SPARSE can be run under the two following modes :

- A prescribed mode which simulates evaporation and transpiration rates for known stress levels (for instance, the two extremes of the water status spectrum : fully stressed or maximum moisture, i.e., potential conditions). The prescribed mode provides an estimate of the potential latent flux for the soil and the vegetation ( $LE_{spot}$  and  $LE_{vpot}$  respectively).



**Figure 9:** Gap-filling exercise : chronological plots over one day (December 26th 2014) of the four meteorological variables observed at the Ben Salem station. Superimposed are the surrogate series (50 replications) of *MetGen* ran in gap-filling mode along with the surrogate series of the two bias correction techniques (*CDFt* and *MBCn*) ran in validation mode and the un-processed large-scale variable series (see color legend).

- A retrieval mode which simulates actual evaporation and transpiration : the respective stress levels (between non evaporating/transpiring and potential rates) correspond to two unknown which are solved from the single piece of information ( $T_{surf}$ ) [46].

The surface water stress index ( $SI$ ) is derived from the actual and potential evapotranspiration rates simulated from the retrieval and the prescribed mode respectively at the time of the satellite overpass.  $SI$  can be defined so as to describe the water status of a single component : either of the soil or of the vegetation (using  $LEs$  or  $LEv$ ). In this application, we used rather the definition of  $SI$  to describe the water status of the soil-vegetation composite :

$$SI = 1 - \frac{LEv + LEs}{LEvpot + LEspot}. \quad (19)$$

The stress index values obtained directly with (19) may contain negative values due to enhanced turbulence in unstable conditions. As negative values cannot theoretically occur, stress index values below -0.5 are replaced by zeros. Besides  $SI$ , daily evapotranspiration ( $ETd$ ) is computed from an extrapolation algorithm in order to reconstruct its sub-daily variations by assuming the self preservation of the evaporative fraction [47]. SPARSE is only ran when remote sensing data are available (i.e., on clear-sky days). The implementation of SPARSE in the Matlab environment is freely available online (<http://tully.ups-tlse.fr/gilles.boulet/sparse>).

### 5.2. Remote sensing data (MODIS)

In addition to meteorological information (air temperature, relative air humidity, global radiation and wind speed), SPARSE uses as inputs satellite data (Normalized Difference Vegetation Index (NDVI), albedo and surface temperature) that provide a description of the initial conditions and of the characteristics of the surface cover. To this end, we relied on remotely sensed data from the latest collection 6 of MODIS (<http://earthexplorer.usgs.gov>) that are available from the year 2000. More precisely, we used the temporal 16-day composite series of MODIS NDVI (MOD13A2), daily Land Surface Temperature (LST), surface emissivity and viewing angle from (MOD11A1) and 8-day of albedo series (MCD43A3) having a spatial resolution of 500 m. These data are acquired for the observation period available at the Ben Salem station (2012-2016) at the resolution of the MODIS sensor (1 km). We extracted a sub-image covering the whole Merguellil plain, see Fig. 1. In addition, we performed a temporal interpolation of albedo and NDVI data to have daily information at the time of the satellite overpass. Last, NDVI information is used to compute remotely sensed leaf area index.

### 5.3. Estimation of $SI$ and $ETd$

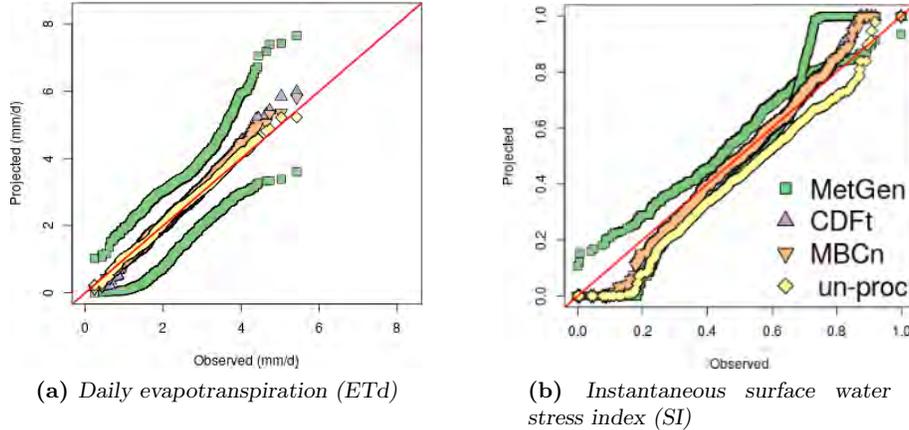
We compare various estimates of instantaneous surface water stress index  $SI$  and of daily evapotranspiration  $ETd$  by constraining SPARSE, on one hand, with the aforementioned MODIS data and, on the other hand, with different choices of meteorological information. Our ground truth

is the estimates of  $SI$  and  $ETd$  obtained when the observed meteorological series at Ben Salem station are used. Other estimates of  $SI$  and  $ETd$  are obtained when these observed meteorological series are replaced by the surrogate meteorological series produced with the cross-validation scheme (see § 4.1) for the three statistical methods (**MetGen**, **CDFt** and **MBCn**) and by the un-processed large-scale variable series. Out of the 50 replications of **MetGen**, we designed two surrogate series corresponding to meteorological conditions leading to lower or higher surface water stress. The low stress conditions consist of high humidity levels set to the 97.5 % quantile of  $Rh$  and low levels of  $WS$ ,  $AirT$  and  $GR$  set to 2.5 % quantiles. The quantile levels are reversed to obtain the high stress conditions.

In Fig. 10, the comparison between the various estimates of  $SI$  and  $ETd$  is first carried out in terms of distribution with qq-plots. On the x-axis,  $ETd$  (Fig. 10a) and  $SI$  (Fig. 10b) are the ground truth estimates (i.e., when the observed meteorological series at Ben Salem station are used to constrain SPARSE). On the y-axis of both panels of Fig. 10, the estimates are obtained by replacing the observed meteorological series with one of the surrogate meteorological series. The distribution of  $ETd$ , see Fig. 10a, is overall well reproduced by all the estimates computed with the surrogate meteorological series. The low and high stress condition surrogate series from **MetGen** form a sort of confidence band around the first bisector.

More pronounced differences are observed in the comparison of the qq-plots of  $SI$  in Fig. 10b. As explained in § 5.1, a truncation of  $SI$  estimates is performed to reduce the importance of negative values which are theoretically not realistic. This creates an atom, i.e., a concentration of values, at zero in the distribution of  $SI$  estimates which explains the shape of the qq-plot close to zero. The atom is especially important when the un-processed large-scale variable series are used as meteorological information to obtain the  $SI$  estimates (the atom makes a horizontal line of zero values starting at 0 in the yellow curve of the qq-plot in Fig. 10b). The atom is also present for the high stress condition series from **MetGen** and the two bias corrected series (the plotting symbols are covered partially). The  $SI$  estimates from the un-processed large-scale variables are globally too low (the yellow curve of the qq-plot is well under the first bisector). The  $SI$  estimates deduced with the surrogate series from the two bias correction techniques (**CDFt** and **MBCn**) display a much milder under-estimation that concerns mostly the lower values, indicative of high stress levels. The  $SI$  estimates computed with the low and high stress condition series from **MetGen** can be thought of as forming a sort of confidence band for the lower index values. However, their behavior for the higher  $SI$  values is much harder to interpret and would require further investigation.

In order to translate differences in distribution as visualized by discrepancies from the first bisector in the qq-plot from Fig. 10b into a more hydrologically interpretable analysis, we propose a comparison based on the probability that the  $SI$  estimate exceeds a given threshold, so-called *exceedance probability*. In Fig. 11, threshold values are represented on the x-axis while the exceedance probabilities are on the y-axis, both ranging from 0 to 1. Black dots represent the exceedance probabilities computed from the ground truth  $SI$  estimates along with an empirical



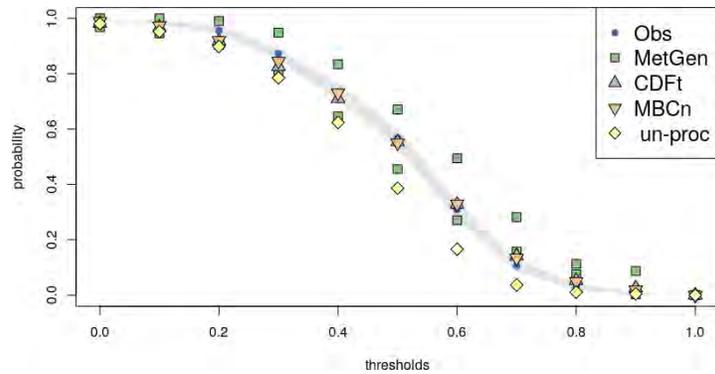
**Figure 10:** Qq-plots of the outputs of the SPARSE energy balance model, see § 5.1, when constrained by the meteorological information from the observed series on the x-axis and from the surrogate series obtained with the cross-validation scheme (see § 4.1) on the y-axis (see color legend). For **MetGen**, a low and high stress condition series are computed out of the 50 replicated surrogate series.

95% confidence band in gray based on binomial proportion confidence intervals :

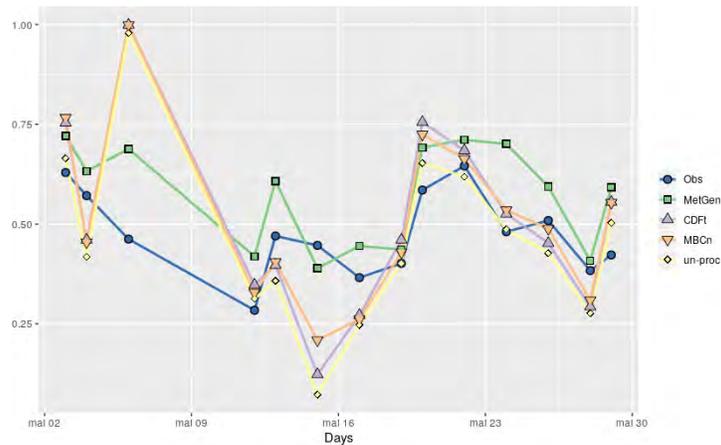
$$\sqrt{\frac{p(1-p)}{n}} \times 1.96, \tag{20}$$

where  $p$  is the probability to have a  $SI$  estimate value that exceeds the threshold and  $n$  is the number of the available time steps. The exceedance probabilities computed from the other  $SI$  estimates (when relying on the surrogate series for the meteorological information constraining the SPARSE model) are as indicated in the color legend in Fig. 11. With the  $SI$  estimates obtained when using the un-processed large-scale variable series, the exceedance probability (yellow diamonds in Fig. 11) falls below the grey confidence band for almost all threshold values. This is coherent with the under-estimation detected from the qq-plot in Fig. 10b. With the  $SI$  estimates obtained when using the surrogate series from the two bias correction techniques, the exceedance probability also tend to fall slightly below the confidence band but only for the lower thresholds. The  $SI$  estimates based on the low and high stress condition series from **MetGen** form a band that overlaps the empirical confidence band for most threshold values except the larger ones.

A final analysis is carried out to illustrate the fluctuations of the  $SI$  estimates chronologically over a short period. Unlike the previous analyses that were based on the whole observation period, Fig. 12 presents an extract of the  $SI$  estimates during one month, May 2016. The estimates are derived at the satellite overpass times during this month with the different choices of meteorological information (either the observed series or one of the surrogate series) used to constrain the SPARSE energy balance model. For **MetGen**, the high stress condition series was used as it follows more closely the ground truth estimates. In Fig. 12, we observe that the  $SI$  estimates derived with the surrogate series generated by the bias correction techniques or the un-processed large-scale variables tend to over- or under-estimate the ground truth estimates.



**Figure 11:** Exceedance probabilities for increasing threshold values computed from the SI estimates as computed by the SPARSE energy balance model when constrained with the observed meteorological series (in black) along with a 95% empirical confidence band (in gray) and when constrained with the surrogate series (see the color legend). For MetGen, a low and high stress condition series are computed out of the 50 replicated surrogate series.



**Figure 12:** Chronological fluctuations of SI estimates at the satellite overpass times during May 2016 with the different choices of meteorological information (either the observed series or one of the surrogate series) used to constrain the SPARSE energy balance model (see color legend). For MetGen, a high stress condition series is computed out of the 50 replicated surrogate series.

## 6. Discussion

We proposed an adaptation of the GLM-based SWG developed in [25], called **MetGen**, to the sub-daily resolution. Indeed, sub-daily resolution is necessary when meteorological observations are used as inputs in energy balance models to estimate surface water stress in order to ensure a precise timing with satellite overpass time. By decomposing the joint distribution into a product of conditional univariate distributions, **MetGen** can model any number of meteorological variables simultaneously. GLMs are a flexible family to model the conditional distributions of meteorological variables from which surrogate series can be simulated stochastically thereby allowing to take into account uncertainty. Although the inter-station dependence is not modeled, **MetGen** can exploit the observations of several gauged stations as long as they belong to a sufficiently homogeneous region. This allows to increase the sample size similarly as in the regional approach frequently used in hydrology. In addition, for the surface water stress application, surrogate series at a single station are needed hence inter-site dependence is not an issue. The framework of **MetGen**, available freely as an R library (<https://CRAN.R-project.org/package=MetGen>), can be useful in many cases since consistent gap-filling and the extension in time of a multi-variable sub-daily observation series is a frequent issue [48, 49].

Since observations at successive time-steps and at neighboring gauged stations during the same time-steps are likely to be dependent, standard model selection procedure that relies, for instance, on the bayesian information criterion [50] and covariate selection that relies on p-values of the coefficients cannot be used for **MetGen**. For this reason, we put forward a model selection procedure that relies entirely on out-of-sample data through a double cross-validation scheme suitable for the size of our observation period. Model selection proceeds iteratively, by first selecting the type of conditional distribution model using the mandatory covariates only. At the successive stages, optional covariates may be introduced to improve goodness-of-fit criteria computed on the validation set. In addition to the mean-squared error from the qq-plot, we included several visual criteria that help to assess additional important features that may not be reflected in the mean-squared error. One notable example of such primary features are the annual and diurnal cycles that lead to the design and the inclusion of the special covariate used to account for the presence of a windbreak at one of the stations. This covariate proved essential in order to reproduce correctly the different shapes of the annual and diurnal cycles of the wind speed at each station. This shows that the **MetGen** framework can handle a certain level of heterogeneity among the stations used for calibration.

Our analyses showed that relying directly on the un-processed large-variables obtained from ERA5 reanalyses [26] as surrogate series may lead to some considerable biases. This is especially true for the wind speed variable as can be seen in the comparison of the annual and diurnal cycles and in the relative RMSE of the qq-plots. Bias correction techniques are devised to correct this kind of systematic departures in terms of distribution of the large-scale variables. Although based on a completely different kind of statistical approaches than SWGs, bias correction techniques can easily be adapted to the task of gap-filling and temporal extension for which **MetGen** is

designed. These techniques are more straightforward to apply since no model selection is needed. Nevertheless, in contrast to **MetGen**, they are not stochastic (a single replication of the surrogate series is produced). In the gap-filling exercise, from the replicated surrogate series simulated by **MetGen**, confidence bands could be obtained by computing a low and a high quantiles (such as 2.5% and 97.5% for a 95% confidence level) for each time step. In addition, despite the development of multivariate approaches, the bias correction techniques are not devised explicitly to exploit meteorological data from neighboring gauged stations lying in the same ERA5 grid cell. In **MetGen**, we made the choice to include, among the mandatory covariates in the GLMs, the large-scale variables in order to allow the simulated series to be guided by the non-stationary behavior present in these covariates, e.g., to follow trends and cycles. Owing to the inclusion of these large-scale covariates, **MetGen** may also be thought of as performing, in some sense, a form of bias correction [51, 52].

The mechanisms that allow the SWG **MetGen** and the bias correction techniques to reproduce seasonal and diurnal behavior are very different. The surrogate series produced by the two bias correction techniques are constructed by adding the observations' diurnal cycles, computed separately for three seasons, to the bias corrected anomalies. The fact that the cycles, annual and diurnal, computed from these surrogate series are very similar to the observations' cycles for all meteorological variables essentially means that the bias corrected anomalies are zero-mean. The construction of these surrogate series also explains why inter-variable correlations can suitably vary from one season to another. This explains also likely why there are no large differences between the univariate and the multivariate bias correction techniques as there is probably not much residual inter-variable dependencies left in the anomalies (in most cases, the plot symbols of **CDFt** are hidden by those of **MBCn**). In contrast, the surrogate series simulated by **MetGen** are able to mimic seasonal and diurnal behavior by exploiting the information in the covariates included in the GLMs. Even when no covariates dedicated to the cycles are included, such as for the relative humidity variable, information on seasonal and diurnal behavior can be drawn from other covariates such as the large-scale variables. In particular, the reproduction of the annual and diurnal cycles and of the seasonal variation of the inter-variable correlations of the surrogate series simulated by **MetGen** is only due to the information present in the covariates. This mechanism is, in our opinion, more flexible as, for instance, it allows the starting and ending dates of each season to vary seamlessly from year to year instead of resorting to a specialized model [53].

In our analyses, one of the main differences among the surrogate meteorological series concerns the wind speed variable. This can be explained by the fact that the wind is the most turbulent of the meteorological variables being modeled. The turbulence yields more frequent and important random fluctuations that are challenging for the statistical models. As **MetGen** relies on stochastic regression, it can better cope with the wind speed fluctuations than the bias correction techniques. These differences in the wind speed surrogate series might be related to the differences in the resulting surface water stress index estimates. Indeed, the SPARSE energy balance model is particularly sensitive to wind speed, especially high wind speed values

that can lead to low  $SI$  estimates. These lower  $SI$  values are particularly important as they are indicative of incipient water stress. When the low and high stress condition series derived from the replicated series of **MetGen** are used as inputs to the SPARSE model, a form of confidence band is obtained that seems particularly reliable for the lower  $SI$  values. More work on the influence of each meteorological variable on the  $SI$  estimates is required to improve these low and high stress condition series in order to yield proper confidence bands. Besides, the daily evapotranspiration estimates,  $ET_d$ , are not very sensitive to which meteorological variables are used as inputs in the SPARSE model. This lack of sensitivity may be explained by the fact that an extrapolation scheme, that might act as a form of filter of the differences, is applied to obtain daily values from instantaneous ones. However, the sensitivity of the SPARSE model should be explored further.

## 7. Conclusion

The framework of **MetGen** provides a solution to the task of obtaining a representative sub-daily multi-variable meteorological series which mimic the observation series but in which gaps are filled and whose simulation period can extend the observation period. It relies on GLMs to model the conditional distributions with large-scale variables used as covariates. These large-scale variables, derived from ERA5 reanalyses in our application, are useful to provide a temporal dynamic of the weather. They may present, in some cases, important biases that can be reduced thanks to the GLMs and the introduction of other covariates. **MetGen** was compared to two bias correction techniques applied on the anomalies of seasonal diurnal cycles. In contrast to the bias correction techniques, **MetGen** can exploit the observations series from nearby gauged-stations having a relatively homogeneous behavior and can generate several replications of the same series to allow uncertainty assessment. Nevertheless, **MetGen** requires a careful selection of the conditional distribution models and of the covariate sets and the simulation can be slow while the bias correction techniques are more straightforward to implement and fast to run.

The analyses performed in this work provide a two-pronged evaluation of the surrogate series generated by **MetGen** and the two bias correction techniques. Firstly, the evaluation is carried out in terms of the ability to reproduce several statistical properties of the meteorological variables and secondly, in terms of surface water stress estimation when the series serve as inputs in the SPARSE energy balance model. Although the performance of the two bias correction techniques was similar for some criteria to **MetGen**, the evaluation in terms of the surface water stress application tends to indicate that the surrogate series generated by **MetGen** may be used to deduce reliable confidence intervals, especially for the lower  $SI$  values that are important for early drought detection.

Perspectives for this work include the study of the sensitivity of **MetGen** to the choice of the dependence graph that serves to simplify the decomposition of the multivariate distribution into a product of conditional univariate distributions. In addition, the introduction of a mechanism

to explicitly model spatial dependence would certainly be a useful development. Also, it would be interesting to evaluate **MetGen** in diverse climatic conditions (e.g., a coastal area) and for other meteorological variables (e.g., precipitation). The two main challenges for **MetGen** implementation are to find representative enough large-scale variables and to perform a thorough selection of the GLMs and of the optional covariates. Last, more thorough analyses pertaining to agricultural drought monitoring should be performed.

### **Acknowledgements:**

The authors gratefully acknowledge financial support by the LMI NAILA, the PHC Utique Amande (a French-Tunisian program), ERANETMED CHAAMS and the European PRIMA project ALTOS. We would also wish to thank all the technical teams of the IRD and INAT for their strong collaboration and support.

### **References**

- [1] H. Baccour, H. Feki, M. Slimani, C. Cudenneq, Interpolation de l'évapotranspiration de référence en tunisie par la méthode de krigeage ordinaire, *Science et changements planétaires/Sécheresse* 23 (2) (2012) 121–132.
- [2] S. Saadi, G. Boulet, M. Bahir, A. Brut, E. Delogu, P. Fanise, B. Mougenot, V. Simonneaux, Z. Chabaane, Assessment of actual evapotranspiration over a semi arid heterogeneous land surface by means of coupled low-resolution remote sensing data with an energy balance model: comparison to extra-large aperture scintillometer measurements, *Hydrology and Earth System Sciences* 22 (4) (2018) 2187–2209.
- [3] J. Sheffield, E. F. Wood, *Drought: past problems and future scenarios*, Routledge, 2012.
- [4] J. Hoedjes, A. Chehbouni, F. Jacob, J. Ezzahar, G. Boulet, Deriving daily evapotranspiration from remotely sensed instantaneous evaporative fraction over olive orchard in semi-arid morocco, *Journal of Hydrology* 354 (1-4) (2008) 53–64.
- [5] J. M. Norman, W. P. Kustas, K. S. Humes, Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature, *Agricultural and Forest Meteorology* 77 (3-4) (1995) 263–293.
- [6] W. J. Timmermans, W. P. Kustas, M. C. Anderson, A. N. French, An intercomparison of the surface energy balance algorithm for land (sebal) and the two-source energy balance (tseb) modeling schemes, *Remote Sensing of Environment* 108 (4) (2007) 369–384.
- [7] J. A. Otkin, M. C. Anderson, C. Hain, I. E. Mladenova, J. B. Basara, M. Svoboda, Examining rapid onset drought development using the thermal infrared-based evaporative stress index, *Journal of Hydrometeorology* 14 (4) (2013) 1057–1074.

- [8] P. Ailliot, D. Allard, V. Monbet, P. Naveau, Stochastic weather generators: an overview of weather type models, *J. de la Société Française de Statistique* 156 (1) (2015) 101–113.
- [9] D. S. Wilks, R. L. Wilby, The weather generation game: a review of stochastic weather models, *Progress in Physical Geography: Earth and Environment* 23 (3) (1999) 329–357. doi:10.1177/030913339902300302.  
URL <https://doi.org/10.1177/030913339902300302>
- [10] R. W. Katz, Precipitation as a chain-dependent process, *Journal of Applied Meteorology and Climatology* 16 (7) (1977) 671 – 676. doi:10.1175/1520-0450(1977)016;0671:PAACDP;2.0.CO;2.  
URL [https://journals.ametsoc.org/view/journals/apme/16/7/1520-0450\\_1977\\_016\\_0671\\_paacdp\\_2\\_0\\_co\\_2.xml](https://journals.ametsoc.org/view/journals/apme/16/7/1520-0450_1977_016_0671_paacdp_2_0_co_2.xml)
- [11] C. W. Richardson, Stochastic simulation of daily precipitation, temperature, and solar radiation, *Water Resources Research* 17 (1) (1981) 182–190. doi:<https://doi.org/10.1029/WR017i001p00182>.
- [12] C. Flecher, P. Naveau, D. Allard, N. Brisson, A stochastic daily weather generator for skewed data, *Water Resources Research* 46 (7) (2010).
- [13] P. Ailliot, C. Thompson, P. Thomson, Space–time modelling of precipitation by using a hidden markov model and censored gaussian distributions, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 58 (3) (2009) 405–426.
- [14] F. Oriani, J. Straubhaar, P. Renard, G. Mariethoz, Simulation of rainfall time series from different climatic regions using the direct sampling technique, *Hydrology and Earth System Sciences* 18 (8) (2014) 3015–3031.
- [15] P. Yiou, Anawege: a weather generator based on analogues of atmospheric circulation, *Geoscientific Model Development* 7 (2) (2014) 531–543.
- [16] F. Garavaglia, J. Gailhard, E. Paquet, M. Lang, R. Garçon, P. Bernardara, Introducing a rainfall compound distribution model based on weather patterns sub-sampling, *Hydrology and Earth System Sciences* 14 (6) (2010) 951–964.
- [17] L. Benoit, M. Vrac, G. Mariethoz, Nonstationary stochastic rain type generation: accounting for climate drivers, *Hydrology and Earth System Sciences* 24 (5) (2020) 2841–2854.
- [18] P. M. Williams, Modelling seasonality and trends in daily rainfall data, in: *Advances in neural information processing systems*, 1998, pp. 985–991.
- [19] R. E. Chandler, On the use of generalized linear models for interpreting climate variability, *Environmetrics* 16 (7) (2005) 699–715.

- [20] A. Verdin, B. Rajagopalan, W. Kleiber, G. Podesta, F. Bert, A conditional stochastic weather generator for seasonal to multi-decadal simulations, *Journal of Hydrology* 556 (2018) 835 – 846.
- [21] P. McCullagh, J. A. Nelder, *Generalized linear models*, Monographs on statistics and applied probability, Chapman and Hall, 1989.
- [22] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2020).  
URL <https://www.R-project.org/>
- [23] R. E. Chandler, Multisite, multivariate weather generation based on generalised linear models, *Environmental Modelling and Software* 134 (2020) 104867.
- [24] J. R. M. Hosking, J. R. Wallis, *Regional frequency analysis: an approach based on L-moments*, Cambridge University Press, 2005.
- [25] R. E. Chandler, *A multisite, multivariate daily weather generator based on generalized linear models, User guide : R package* (2015).  
URL <https://www.ucl.ac.uk/~ucakarc/work/glimclim.html>
- [26] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, J.-N. Thépaut, The era5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society* 146 (730) (2020) 1999–2049.
- [27] P. V. Ayar, M. Vrac, S. Bastin, J. Carreau, M. Déqué, C. Gallardo, Intercomparison of statistical and dynamical downscaling models under the euro-and med-cordex initiative framework: present climate evaluations, *Climate Dynamics* 46 (3-4) (2016) 1301–1329.
- [28] D. Maraun, F. Wetterhall, A. M. Ireson, R. E. Chandler, E. J. Kendon, M. Widmann, S. Brienen, H. W. Rust, T. Sauter, M. Themeßl, V. K. Venema, K. P. Chun, C. M. Goodess, R. G. Jones, C. Onof, M. Vrac, I. Thiele-Eich, Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, *Reviews of Geophysics* 48 (3) (2010).
- [29] B. François, M. Vrac, A. J. Cannon, Y. Robin, D. Allard, Multivariate bias corrections of climate simulations: which benefits for which losses?, *Earth System Dynamics* 11 (2) (2020) 537–562.
- [30] P. A. Michelangeli, M. Vrac, H. Loukos, Probabilistic downscaling approaches: Application to wind cumulative distribution functions, *Geophysical Research Letters* 36 (11) (2009).

- [31] A. J. Cannon, Multivariate quantile mapping bias correction: an n-dimensional probability density function transform for climate model simulations of multiple variables, *Climate dynamics* 50 (1-2) (2018) 31–49.
- [32] C. Leduc, S. Ben Ammar, G. Favreau, R. Beji, R. Virrion, G. Lacombe, J. Tarhouni, C. Aouadi, B. Zenati Chelli, N. Jebnoun, M. Oi, J. L. Michelot, K. Zouari, Impacts of hydrological changes in the mediterranean zone: environmental modifications and rural development in the merguellil catchment, central tunisia / un exemple d'évolution hydrologique en méditerranée: impacts des modifications environnementales et du développement agricole dans le bassin-versant du merguellil (tunisie centrale), *Hydrological Sciences Journal* 52 (6) (2007) 1162–1178.
- [33] F. Molle, P. Wester, *River basin trajectories: societies, environments and development*, Vol. 8, IWMI, 2009.
- [34] S. Ben Ammar, K. Zouari, C. Leduc, J. M'barek, Caractérisation isotopique de la relation barrage-nappe dans le bassin du merguellil (plaine de kairouan, tunisie centrale), *Hydrological sciences journal* 51 (2) (2006) 272–284.
- [35] C. Leduc, R. Calvez, R. Beji, Y. Nazoumou, G. Lacombe, C. Aouadi, Evolution de la ressource en eau dans la vallée du merguellil (tunisie centrale), in: *Séminaire sur la modernisation de l'agriculture irriguée*, IAV Hassan II, 2004, pp. 10–p.
- [36] OMM, *Guide des instruments et des méthodes d'observation météorologiques*, OMM 8, Geneve, 2010.
- [37] D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balsameda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J. J. Morcrette, B. K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J. N. Thépaut, F. Vitart, The era-interim reanalysis: Configuration and performance of the data assimilation system, *Quarterly Journal of the royal meteorological society* 137 (656) (2011) 553–597.
- [38] J. Hooker, G. Duveiller, A. Cescatti, A global data set of air temperature derived from satellite remote sensing and weather stations, *Scientific data* 5 (2018) 180246.
- [39] R. G. Allen, L. S. Pereira, M. Smith, D. Raes, J. Wright, Fao-56 dual crop coefficient method for estimating evaporation from soil and application extensions, *Journal of irrigation and drainage engineering* 131 (1) (2005) 2–13.
- [40] D. Déqué, Frequency of precipitation and temperature extremes over france in an anthropogenic scenario: Model results and statistical correction according to observed values, *Global and Planetary Change* 57 (1) (2007) 16 – 26, extreme Climatic Events.

- [41] M. Vrac, P. Friederichs, Multivariate—intervariable, spatial, and temporal bias correction, *Journal of Climate* 28 (1) (2015) 218–237.
- [42] A. J. Cannon, Neural networks for probabilistic environmental prediction: Conditional Density Estimation Network Creation and Evaluation (CaDENCE) in R, *Computers & Geosciences* 41 (2012) 126–135.
- [43] P. Bruce, A. Bruce, P. Gedeck, *Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python*, O’Reilly Media, 2020.
- [44] P. Filzmoser, B. Liebmann, K. Varmuza, Repeated double cross validation, *Journal of Chemometrics: A Journal of the Chemometrics Society* 23 (4) (2009) 160–171.
- [45] H. Joe, *Multivariate models and multivariate dependence concepts*, CRC Press, 1997.
- [46] G. Boulet, B. Mougenot, J. Lhomme, P. Fanise, Z. Lili-Chabaane, A. Olioso, M. Bahir, V. Rivalland, L. Jarlan, O. Merlin, et al., The sparse model for the prediction of water stress and evapotranspiration components from thermal infra-red data and its evaluation over irrigated and rainfed wheat, *Hydrology and Earth System Sciences* 19 (11) (2015) 4653–4672.
- [47] E. Delogu, G. Boulet, A. Olioso, B. Coudert, J. Chirouze, E. Ceschia, V. Le Dantec, O. Marloie, G. Chehbouni, J. P. Lagouarde, Reconstruction of temporal variations of evapotranspiration using instantaneous estimates at the time of satellite overpass, *Hydrology and earth system sciences* 16 (2012) 2995–3010.
- [48] S. Er-Raki, A. Chehbouni, S. Khabba, V. Simonneaux, L. Jarlan, A. Ouldbba, J. Rodriguez, R. Allen, Assessment of reference evapotranspiration methods in semi-arid regions: Can weather forecast data be used as alternate of ground meteorological parameters?, *Journal of Arid Environments* 74 (12) (2010) 1587–1596.
- [49] C. Leauthaud, B. Cappelaere, J. Demarty, F. Guichard, C. Velluet, L. Kergoat, T. Vischel, M. Grippa, M. Mouhaimouni, I. Bouzou Moussa, I. Mainassara, B. Sultan, A 60-year reconstructed high-resolution local meteorological data set in central sahel (1950–2009): evaluation, analysis and application to land surface modelling, *International Journal of Climatology* 37 (5) (2017) 2699–2718.
- [50] G. Schwarz, Estimating the dimension of a model, *Ann. Statist.* 6 (2) (1978) 461–464. doi:10.1214/aos/1176344136.
- [51] N. Fodor, I. Dobi, J. Mika, L. Szeidl, Applications of the mvwg multivariable stochastic weather generator, *The Scientific World Journal* 2013 (2013).
- [52] J. Chen, F. P. Brissette, R. Leconte, A daily stochastic weather generator for preserving low-frequency of climate variability, *Journal of hydrology* 388 (3-4) (2010) 480–490.
- [53] T. Carey-Smith, J. Sansom, P. Thomson, A hidden seasonal switching model for multisite daily rainfall, *Water Resources Research* 50 (1) (2014) 257–272.

## 3.3 Résultats complémentaires

Le générateur stochastique "MetGen" assure la simulation des scénarios multi-sites et multi-variables à une échelle infra-journalière. Comme expliqué avant, le générateur est basé sur la construction des modèles linéaires généralisés (GLM) afin de reproduire la variabilité temporelle et spatiale des variables. Les GLM permettent de généraliser la régression classique en s'appuyant sur une régression linéaire, qui forme une transformation simple d'un modèle linéaire (e.g., logarithme). Ce fait nous permet ainsi, d'avoir recours à des lois autres que la loi normale (e.g., les lois gamma, binomiale ou normale avec dispersion non-constante), ce qui rend cette approche applicable pour différentes variables climatiques. Au niveau des analyses présentées dans la section précédente (3.2.2), nous avons réduit notre étude sur les quatre variables météorologiques qu'on a utilisés pour contraindre le modèle de bilan d'énergie à double source SPARSE (température de l'air, humidité relative de l'air, vitesse du vent et rayonnement global). Outre ces variables, "MetGen" est aussi capable de simuler d'autres types des variables climatiques. Au niveau de cette section, nous allons présenter comme exemple, l'ajustement et la simulation des précipitations et la pression atmosphérique. Ces variables font partie des variables mesurées par les stations de mesure implantées au niveau de notre zone d'étude, et elles sont très utiles pour les analyses du changement climatique.

### 3.3.1 Ajustement des variables

"MetGen" se base sur une succession d'étapes, bien détaillées au niveau de l'article présenté précédemment, afin de construire les modèles de régression de chaque variable : (1) la définition de l'ordre de simulation, (2) le choix de la loi de probabilité de chaque variable et finalement (3) un choix d'une sélection adéquate des covariables en se basant sur la procédure d'une validation croisée. Nous reprenons ces étapes pour simuler la pression atmosphérique et les précipitations.

#### 3.3.1.1 Dépendance inter-variables

Suite à l'ajout de ces deux variables : pression atmosphérique et précipitations, la première étape à effectuer est de redéfinir les relations d'inter-dépendance. Ce fait est très important afin d'assurer la cohérence entre les variables et bien reproduire les relations d'inter-dépendance initialement observées. On se base par ailleurs, sur le graphique 3.1, présenté ci-dessous, qui explique l'ordre de la simulation des variables et réduit le nombre des variables dépendantes. Dans ce cas, on suppose que la pression atmosphérique est la variable qui ne dépend d'aucune autre variable. La pluie est cependant la dernière variable à simuler. On conserve par conséquent les relations définies avant, au niveau de l'article, pour le reste des variables. Ces variables dépendantes seront introduites parmi les covariables choisies pour la simulation de chaque variable météorologique.

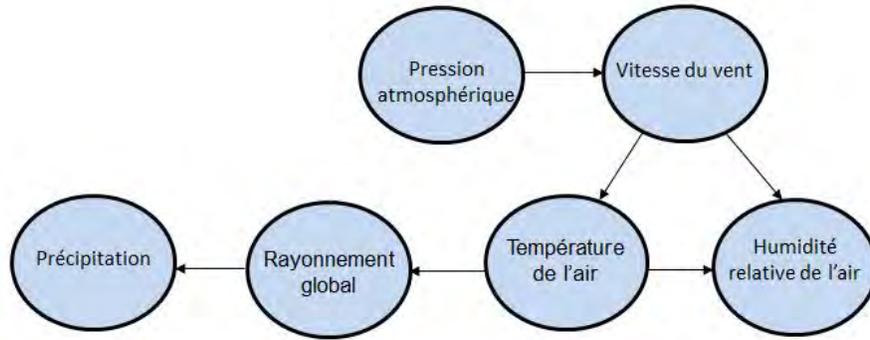


Figure 3.1 – Graphique complet des relations d'inter-dépendance des variables.

### 3.3.1.2 Lois de probabilité choisies

L'utilisation des GLM nous offre l'avantage, par rapport à la régression linéaire classique, d'utiliser d'autres lois de probabilité, autre que la distribution normale et qui sont plus adéquates par rapport à la nature des variables à simuler. Comme par exemple, la distribution gamma qui est destinée pour une variable qui ne prend que des valeurs positives et la loi binomiale pour estimer la probabilité d'occurrence d'un certain évènement. Nous avons choisi alors une distribution normale pour la pression atmosphérique, une distribution gamma pour l'intensité de précipitations et une distribution binomiale pour leurs occurrences. Nous présentons dans le tableau 3.1, un récapitulatif pour les paramètres et les propriétés des distributions utilisées ;

Table 3.1 – Paramètres et propriétés des différentes lois de probabilité utilisées.

Loi	Equations	Paramètres	Propriétés.
Normale	$Y \mathbf{x} \sim \mathcal{N}(\beta\mathbf{x}, \sigma^2)$	$\mu$ : moyenne, $\sigma$ : écart type	symétrie, forme en cloche, à support $\mathbb{R}$ ↗
Gamma	$Y \mathbf{x} \sim \text{Gamma}(\mu(\mathbf{x}), \nu)$ $\mu(\mathbf{x}) = \exp(\beta\mathbf{x})$	$\mu$ : moyenne, $\sigma$ : forme	Asymétrie, à support $\mathbb{R}^*$
Binomiale	$\mathbb{P}(Y = 1 \mathbf{x}) = (1 + e^{-\beta\mathbf{x}})^{-1}$	N : nbre d'essais, P : probabilité de succès	Discret, à support $[0, 1]$ spécifique de la loi de probabilité

### 3.3.1.3 Jeu de co-variables choisi

L'introduction des covariables va permettre aux distributions des variables à simuler et à varier spatialement et temporellement. Le jeu de covariables, comme pour les premières variables simulées dans la section 3.2.2, peut contenir (1) les variables dépendantes, (2) les variables à large échelle, (3) cycles annuels et (4) diurnes, (5) information géographique et (6) effet mémoire afin de tenir compte de la persistance temporelle. Ces effets mémoires sont

calculés sur la variable elle-même (Var.lag), sur la moyenne spatiale (SA.lag), sur la moyenne mobile (MA.lag), ou bien sur la moyenne spatiale et mobile à la fois (SMA.lag). Les variables à large échelle forment les principaux covariables utilisées dans le générateur, afin de mieux guider la simulation. Nous avons choisi pour ces deux variables, i.e., pression atmosphérique et précipitations, leur variable correspondante disponible au niveau des réanalyses, pression au niveau de la mer (mslp) et pluie totale (tp) respectivement. Nous présentons le jeu des covariables sélectionné pour les deux variables à simuler au niveau du tableau 3.2. Pour améliorer la simulation de la variable intensité de précipitations, nous avons ajouté au niveau des covariables supplémentaires, la saisonnalité (saison) pour identifier les mois pluvieux par rapport aux mois non pluvieux, et une autre covariable pour différencier les stations, autre que l'information géographique qui n'était pas significative pour cette variable.

### 3.3.2 Simulation des variables

Comme expliqué avant, MetGen permet d'assurer à la fois (1) une extension temporelle, lorsqu'il s'agit d'une simulation libre sur une longue période guidée par les valeurs des variables à large échelle et les effets déterministes prédéfinis pour chaque variable, ou (2) un gap-filling (imputation) lorsque la simulation est effectuée que pour les pas de temps manquants. En mode imputation, la simulation est fortement guidée par les valeurs observées au niveau des trois stations. On limite la présentation des résultats sur la station de référence, Ben Salem.

#### 3.3.2.1 Extension temporelle

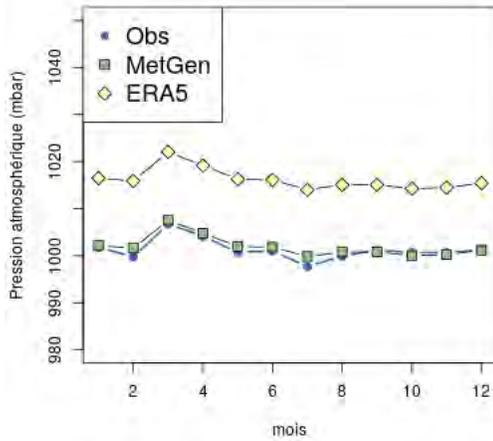
En mode d'extension temporelle, MetGen se base sur une période de buffer, généralement égale à un mois, en tant qu'une sorte de période d'entraînement avant le début de simulation basée sur la loi de distribution et les covariables choisies. En effet, la période du buffer nous permet de calculer les effets mémoires d'une façon arbitraire due à la non disponibilité d'un historique d'observations au début de la période de simulation. Dans le présent travail, on simule la pression atmosphérique qui ne dépend d'aucune autre variable en premier lieu. Cette variable est guidée essentiellement par la variable à grande échelle "mslp" issue d'ERA5. Ensuite, on continue à suivre l'ordre de simulation du reste des variables selon le graphe d'inter-dépendance (figure 3.1). La pluie sera alors la dernière variable à simuler en utilisant les valeurs simulées du rayonnement global parmi ses covariables.

Nous présentons au niveau de la figure 3.2, des cycles annuels et journaliers des deux variables pression atmosphérique et précipitations. On observe une bonne reproduction des cycles annuels et journaliers des valeurs observées, en utilisant les simulations issues du générateur stochastique pour la pression atmosphérique. En effet, MetGen qui utilise les données d'ERA5 a pu corriger le biais observé initialement dans les valeurs brutes de mslp (figures 3.2a et 3.2c). Pour les précipitations, MetGen a pu reproduire au mieux la variabilité des faibles moyennes de cette variable. Cependant, il n'a pas pu simuler les moyennes annuelles ou diurnes fortes, observées in-situ (figures 3.2b et 3.2d).

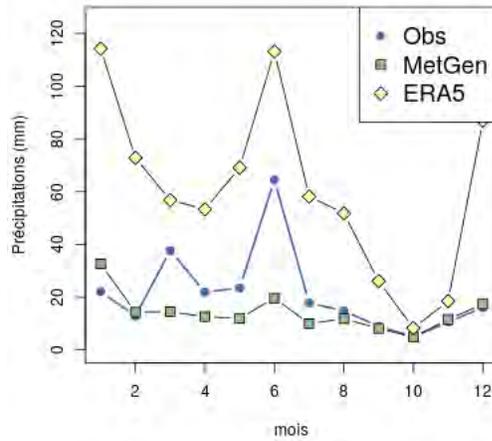
D'autre part, comme pour le reste des variables présentées dans la section 3.2.2, nous avons comparé la méthode de générateur stochastique avec une autre méthode classique,

Table 3.2 – Covariables sélectionnées pour la pression atmosphérique et les précipitations. Les périodes d'oscillations pour définir un cycle annuel sont spécifiées en jours, et pour un cycle diurne en des heures. Pour les effets mémoires la longueur de la plus longue période de lag est identifiée. La colonne autre contient tous les covariables supplémentaires qu'on a rajoutées pour améliorer la simulation des variables

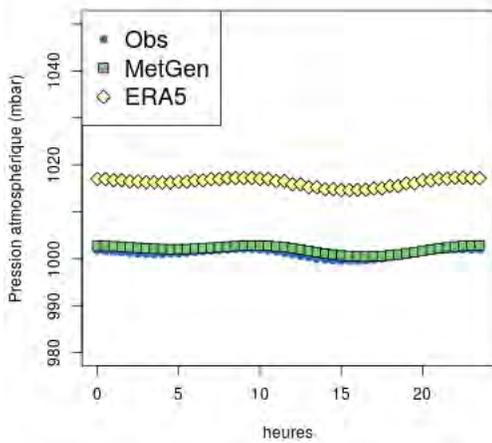
Variable	Variables dépendantes ERA5	Variables ERA5	Info géographique	Cycle annuel	Cycle diurne	effets mémoire	Autres
Pression atm	X	mslp	✓	X	X	SA.lag1, Var.lag3	X
Precip occ	Rg	tp	X	365, 183, 91, 30	24, 12, 6	SA.lag4, Var.lag1	X
Precip int	Rg	tp	X	X	X	SA.lag1, Var.lag1	saisons, stations



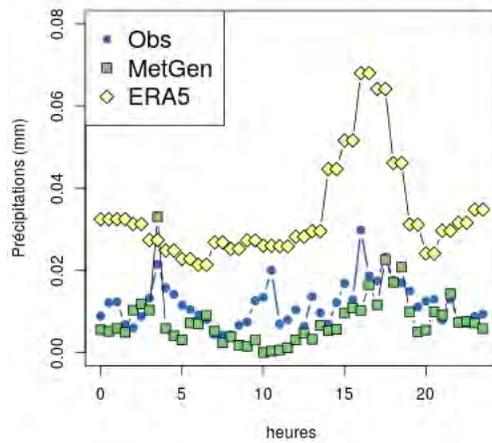
(a) Pression atmosphérique



(b) Précipitations



(c) Pression atmosphérique



(d) Précipitations

Figure 3.2 – Comparaison des cycles annuels (en haut) et journaliers (en bas) des variables Pression atmosphérique et Précipitations, observées au niveau de la station de Ben Salem, simulées par le générateur et issues des ré-analyses.

méthode de correction de biais dans sa forme univariée en utilisant la fonction CDFt (Miche-langeli *et al.*, 2009) et multivariée par la fonction MBCn (Cannon, 2018). Ces deux méthodes sont implémentées dans des bibliothèques de l'environnement R. Contrairement à l'approche suivie pour les quatre variables simulées avant (température de l'air, humidité relative de l'air, rayonnement global et vitesse du vent), la correction de biais pour la pression atmosphérique et les précipitations a été appliquée directement sur les variables elles mêmes et non pas sur leurs anomalies. En effet, on assume que ces deux variables ne présentent pas des cycles saisonniers ou diurnes constants comme pour le reste des variables.

La comparaison de ces différentes méthodes (générateur et corrections de biais) consiste à évaluer leur capacité, à reproduire la distribution des valeurs observées au niveau de la station de Ben Salem pour toute la période d'observation disponible (2011-2016), à partir des simulations en validation croisée. Les graphiques quantile-quantile présentés dans la figure 3.3, montrent une bonne reproduction de la distribution de la pression atmosphérique en utilisant les méthodes du générateur et de correction de biais, les droites sont bien alignées avec la pente égale à 1 (voir figure 3.3a). Cependant, ces méthodes n'ont pas été efficaces pour reproduire la distribution des précipitations, essentiellement pour les fortes valeurs de pluie. Le générateur et ERA5 fournissent des valeurs qui sous-estiment la distribution réellement observée par cette variable. Les méthodes de corrections de biais reproduisent au mieux la distribution des précipitations mais, ils sont aussi incapables à reproduire les fortes valeurs (figure 3.3b).

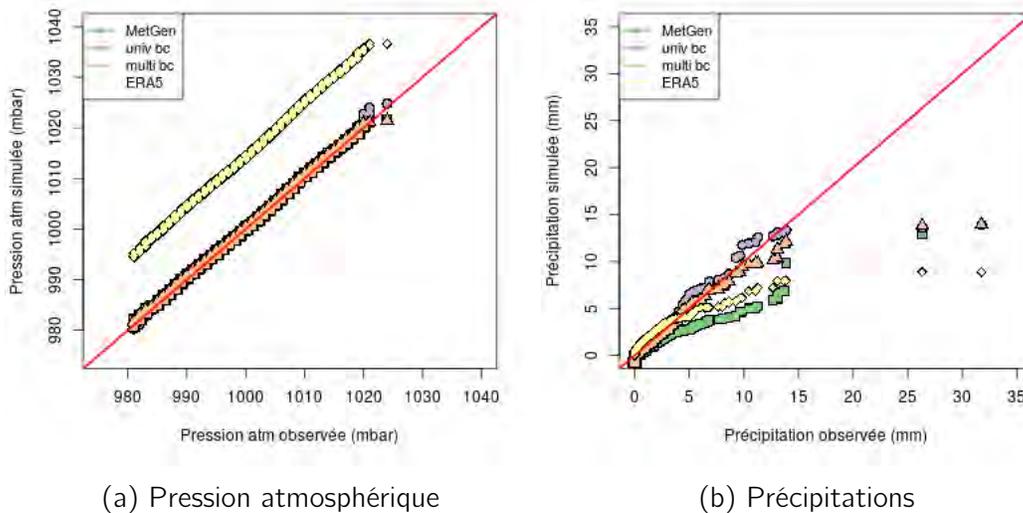


Figure 3.3 – Graphiques quantile-quantile des valeurs observées de la pression atmosphérique et des précipitations à Ben Salem (axe des x) et les valeurs simulées à partir des différents scénarios climatiques issus de MetGen, des deux méthodes de correction de biais (univariée et multivariée) et des ré-analyses brutes.

### 3.3.2.2 Imputation

Afin d'avoir une série des valeurs observées comparable à celle imputée, nous avons introduit des valeurs manquantes d'une manière aléatoire et artificielle, qui forment 15% de la totalité des pas de temps pour une période de temps entre 2014 et 2016 (période d'observation commune entre les trois stations S1, S2 et S3), comme présenté au niveau du schéma 3.4. Par ailleurs, les valeurs simulées peuvent être comparées aux valeurs réellement observées au niveau des différentes stations.

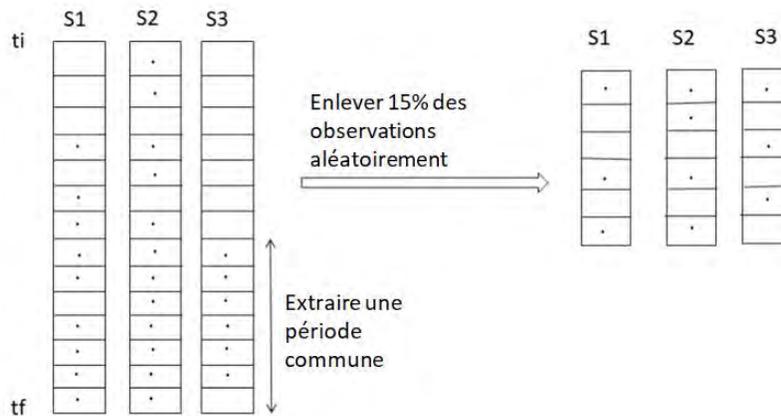


Figure 3.4 – Schéma représentatif de la méthode de validation de l'imputation.

Nous avons testé ensuite, la capacité des valeurs imputées à reproduire les valeurs observées en termes de distribution par le biais des graphiques quantile-quantiles, présentés au niveau de la figure 3.5b. En effet, nous avons tracé les valeurs observées de nos deux variables au niveau de l'axe des abscisses et seulement les valeurs imputées au niveau de l'axe des ordonnées. Nous observons que les valeurs gapfillées des deux variables ont bien reproduit la distribution initiale des valeurs observées, même pour les fortes valeurs de précipitations, voir la figure 3.5b.

## 3.4 Conclusion partielle et synthèse

Le développement d'un générateur stochastique "MetGen" nous permet d'assurer le comblement des données manquantes dans les séries d'observations, et d'effectuer une extension temporelle sur une longue période d'étude où les données météorologiques ne sont pas disponibles. MetGen repose sur la construction des modèles linéaires généralisés (GLM) en se basant sur un choix adéquat des lois de probabilité des variables, les transformations nécessaires à appliquer, la définition des relations d'inter-dépendance reflétant d'une façon réaliste la nature des interactions entre les variables, et finalement, un choix adéquat aussi pour les covariables à introduire. Le choix des modèles se fait d'une façon indépendante d'une variable météorologique à une autre.

Pour la validation de l'ajustement des modèles, nous avons effectué une validation croisée, sur la période d'observation 2011-2016. Après la construction des modèles pour chaque

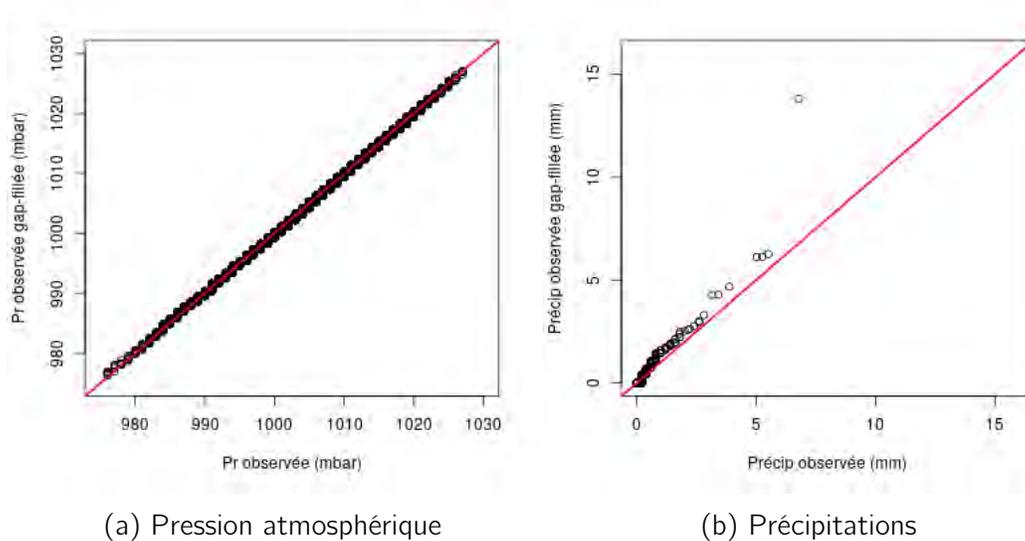


Figure 3.5 – Graphiques quantile-quantiles des valeurs observées de la pression atmosphérique et des précipitations à Ben Salem et leurs valeurs gap-remplées.

variable, la validation finale est effectuée en termes de projection et de gap-filling, par la simulation de plusieurs répliques (50 fois), afin d'évaluer ce modèle stochastique en fonction de ses différentes réalisations. Les simulations de "MetGen" sont aussi comparées avec les méthodes de corrections de biais (univariée et multivariée) et les données de ré-analyses brutes. Ces scénarios climatiques issus de ces différentes sources sont évalués en termes de leur capacité à bien reproduire les caractéristiques des variables climatiques observées in-situ, ainsi qu'en termes de leur capacité à reproduire la simulation de l'évapotranspiration et du stress hydrique avec le modèle de bilan d'énergie à double source SPARSE.

Les analyses menées dans ce chapitre ont démontré que, l'utilisation directe des données de ré-analyses brutes peut engendrer des biais considérables. Ce fait est observé essentiellement au niveau de la vitesse du vent, la pression atmosphérique et les précipitations, en regardant les cycles annuels et diurnes et les graphiques quantiles-quantiles. Cependant, les variables météorologiques sont bien reproduites en utilisant les méthodes de descentes d'échelle testées (générateur stochastique et les deux méthodes de correction de biais). La pression atmosphérique utilise l'information issue de mslp mesurée selon l'altitude de niveau de mer. MetGen et les deux méthodes de correction de biais ont été capables de corriger le biais initialement observé entre ces deux variables mesurées à des altitudes différentes. Toutefois, les précipitations et la vitesse du vent nécessitent une meilleure reproduction des valeurs extrêmes, notamment les fortes valeurs. Pour les précipitations par exemple, le choix de la distribution gamma peut être à l'origine de la sous-estimation des valeurs de pluie. Une distribution plus flexible comme présentée dans les travaux de Carreau & Vrac (2011), peut être suggérée, mais elle présente plus de complexité pour son application. Les dépendances inter-variables, mesurées par le coefficient de corrélation de Kendall, sont bien reproduites par tous les scénarios climatiques, même celui issu de la correction de biais univarié qui n'a pas un mécanisme qui tient en compte des relations d'inter-dépendance. Ceci, peut être expliqué par le fait qu'on a utilisé les cycles

des observations pour la reconstruction de la variable après l'application de la correction de biais que sur les anomalies (pour la température de l'air, l'humidité relative de l'air, vitesse du vent et rayonnement global). D'autre part, en utilisant le générateur stochastique, l'information sur les cycles est introduite au niveau des covariables, en définissant les variations des cycles annuels et saisonniers ou en utilisant l'information issue de la variable dépendante ou la variable à large échelle. On cite à titre d'exemple, la variable humidité relative de l'air qui n'était pas significative pour les effets des cycles annuels et diurnes. La reproduction de la variabilité annuelle et diurne au niveau de la variable simulée est par ailleurs effectuée grâce à l'information issue de la variable à large échelle et les variables dépendantes (vitesse du vent et température de l'air), introduites parmi ses covariables.

Cependant, nous avons observé des différences au niveau de la simulation des variables de bilan d'énergie en utilisant les différents scénarios climatiques. Ce fait peut être expliqué par les différences observées initialement dans la simulation de la variable de la vitesse du vent en utilisant les différentes méthodes statistiques. En effet le modèle de bilan d'énergie montre une sensibilité importante par rapport à cette variable caractérisée par une variation turbulente. Cependant, ces scénarios ont bien reproduit les valeurs faibles d'indice de stress, qui sont particulièrement informative par rapport à un début de stress. La validation de cette méthode sur la période d'observation (2011-2016), va nous permettre d'effectuer la projection des différentes variables utiles pour les études d'impact (température de l'air, humidité relative, vitesse du vent et rayonnement global) sur une période d'étude dans le passé, de 2000 jusqu'au 2020, afin d'étudier la variabilité temporelle des indices de stress.

## 4.1 Introduction

Après la présentation et la validation de la méthode du générateur stochastique (voir chapitre B) sur une période de validation (2011-2016), nous avons utilisé cette méthode pour simuler des données climatiques sur toute la période d'étude (2000-2019). Ainsi, nous avons utilisé cette série prolongée pour forcer le modèle d'énergie à double source SPARSE et simuler par ailleurs l'évapotranspiration sur toute la période d'étude. Cependant, comme rappelé précédemment, les méthodes basées sur les données optiques (visible, proche infrarouge et infrarouge thermique) ne peuvent pas fournir des informations journalières sur l'état de surface. En effet, seules les images acquises en conditions de ciel clair sont exploitables. Ainsi, une interpolation journalière s'avère nécessaire afin de générer une information saisonnière complète d'évapotranspiration et par la suite un indice de stress. Ce dernier est comparé avec d'autres indices de stress issus de différentes longueurs d'onde : Visible, proche infrarouge, infrarouge thermique et micro-onde afin de voir leur cohérence, pertinence et précocité. Rappelons également que seuls les indices issus du domaine micro-onde sont disponibles chaque jour car exploitables même en conditions de fort ennuagement. D'autre part, afin de mieux se focaliser sur la sécheresse agronomique qui représente un déficit hydrique au niveau de la zone racinaire de la plante, nous avons exploré la capacité du modèle de bilan d'énergie à simuler la transpiration et calculer aussi l'indice de stress qui en dérive.

La première section de ce chapitre contiendra une comparaison de quelques méthodes d'interpolation de l'évapotranspiration. Au niveau de la deuxième section, nous présentons une étude comparative des différents indices issus de différentes longueurs d'onde, présentée sous forme d'un article. La troisième section, contiendra la réponse de l'indice de stress thermique issu de la végétation. Finalement, nous présentons une comparaison entre l'indice thermique SI que nous avons utilisé et celui utilisé dans les travaux de Anderson *et al.* (2007), ESI.

## 4.2 Interpolation journalière de l'Evapotranspiration

L'utilisation des données satellitaires pour l'estimation de l'évapotranspiration forme un apport considérable dans le domaine hydrologique et agronomique. Toutefois, l'utilisation des données thermiques, qui bénéficient pourtant d'une revisite journalière, reste toujours contrainte par la présence du couvert nuageux qui limite le nombre d'acquisition des données et induit par conséquent des lacunes fréquentes dans la série simulée de ET. Il s'avère indispensable par conséquent, d'appliquer une interpolation temporelle de ET afin de pouvoir construire une information saisonnière continue, nécessaire pour une étude agronomique des

périodes de sécheresse.

La procédure classiquement utilisée pour effectuer une interpolation de l'évapotranspiration journalière consiste à interpoler linéairement le rapport entre l'évapotranspiration journalière ( $ET_j$ ) et un support d'interpolation facilement accessible chaque jour  $j$  ( $S_j = \sum S_t$ , tel que  $S_t$  est la valeur de ce support en infra-journalier). Notre but est alors de sélectionner un support ou une combinaison de supports, non directement issus des données de télédétection, à insérer entre deux dates d'acquisition des images successives (deux jours clairs), voir fig 4.1.

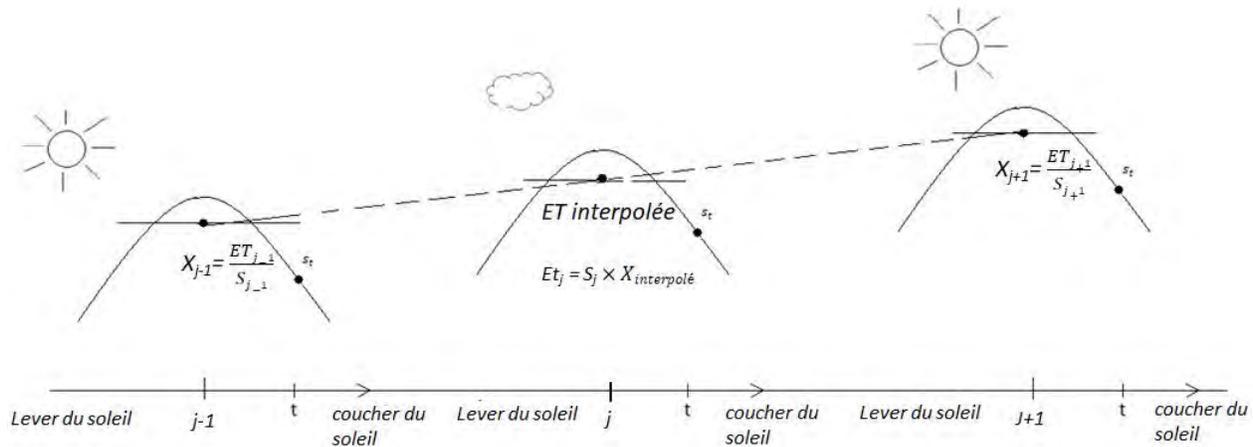


Figure 4.1 – interpolation de l'évapotranspiration (Delogu *et al.*, 2021)

Ce support de valeur connue entre deux dates d'acquisition constitue une information auxiliaire qui va servir à piloter l'interpolation. Dans ce contexte, Delogu *et al.* (2021) ont évalué plusieurs méthodes d'interpolation de ET sur des régions généralement de climat méditerranéen ou semi-aride, en utilisant différents supports disponibles à l'échelle locale d'une façon continue et qui ne dépendent pas de l'enneigement ou de la fréquence de revisite (le rayonnement global journalier mesuré (Rg), le rayonnement net calculé par la méthode FAO (Rn), le rayonnement ciel clair calculé (Rc), l'énergie disponible journalière reconstitué à partir du rayonnement global (AE), l'évapotranspiration de référence ( $ET_0$ ) et l'évapotranspiration potentielle ( $ET_{pot}$ ) simulée par SPARSE) afin de reconstituer une évapotranspiration journalière. D'autre part, des méthodes supplémentaires prenant en compte la pluie ont également été développées et testées au niveau de ce travail. Ces méthodes consistent à la simulation d'une fraction évaporative (EF) maximale (EF=1) lorsqu'une pluie d'une certaine intensité est mesurée localement, ou par la pondération de EF par l'indice des pluies antérieures (API) du jour  $j$ . Les auteurs de cette étude ont montré que le choix de support pour l'interpolation de ET dépend de l'objectif à atteindre. Ils ont trouvé que les méthodes d'interpolation simples (Rc ou Rg notamment) ont une très bonne performance pour interpoler ET sur une échelle saisonnière. Cependant, sur une échelle infra-saisonnière, pour une étude de gestion de l'irrigation ou la détection du stress hydrique, il est important de considérer les précipitations, afin de réduire la sous estimation de ET durant les longues périodes sans acquisition des images satellitaires. Les méthodes d'interpolation de ET basées sur l'information de la pluie, permettent de mieux reproduire les variations de  $ET_j$  induites par des phénomènes intervenants

entre deux passages de satellite (évènement pluvieux essentiellement). En effet, pendant les jours nuageux, les pertes par évapotranspiration deviennent très réduites en comparaison à ses valeurs durant les jours clairs, dues à la réduction de la demande atmosphérique par la présence des nuages ou à l'approvisionnement des besoins en eau par la pluie (Delogu *et al.*, 2021), ce qui entraîne une interpolation entre deux faibles valeurs de ET durant la période sans acquisition satellitaire. Ces faibles valeurs peuvent être alors corrigées par l'information issue des précipitations, qui tiennent en compte du statut hydrique de la surface réellement observé et qui permet d'éviter des fausses estimations de  $ET_j$ . Ce dernier point est particulièrement critique au niveau des zones semi-arides qui sont caractérisées par des périodes de ressuyage des sols courtes et une réponse brutale de la surface après chaque évènement pluvieux (Lohou *et al.*, 2014).

Dans ce contexte nous avons testé deux méthodes d'interpolation de ET.

- Une première méthode qui utilise un support d'interpolation simple, le rayonnement journalier ( $R_g$ ), très utile pour reproduire la variation annuelle.
- Une deuxième méthode plus complexe qui utilise l'indice des pluies antérieures API pour calculer la fraction évaporative (EF). Cette méthode est différente de la méthode utilisée dans Delogu *et al.* (2021). Elle nous permet de mieux suivre les variations de ET avant et après un évènement pluvieux.

#### 4.2.1 Interpolation basée sur le rayonnement global

C'est l'une des méthodes les plus classiquement utilisées. Elle repose sur une simple interpolation linéaire du rapport ( $X$ ), entre  $ET_j$  et le rayonnement global comme support d'interpolation. En effet, le flux de chaleur latente est très positivement corrélé au rayonnement solaire, particulièrement pour des teneurs élevées en humidité du sol. D'autre part, le rayonnement solaire forme la source majeure de l'énergie disponible, ce qui explique la forte corrélation avec la fraction évaporative et par la suite avec  $ET_j$  (Gentine *et al.*, 2007).

$$X = \frac{ET_j}{R_{g_j}} \quad (4.1)$$

Avec,

- $ET_j$  est l'évapotranspiration journalière en mm/j
- $R_{g_j}$  est la somme journalière du rayonnement global converti en mm par jour

#### 4.2.2 Interpolation basée sur API

La méthode se base sur la reconstitution de l'évapotranspiration à partir de l'énergie disponible  $AE_j$  et la fraction évaporative EF (Gallego-Elvira *et al.*, 2013). ET est alors reconstruite à partir de l'interpolation de ces deux composantes d'une façon séparée.  $AE_j$  est déterminée par une équation proposée par Jackson *et al.* (1983), qui suppose que AE présente la même variation diurne que  $R_g$  (voir équation 4.2). Les valeurs obtenues par cette équation vont être

obtenues que pour les jours clairs, les jours où nous avons acquis les données satellitaires. Les valeurs de  $AE_j$  sont alors interpolées afin d'avoir une série journalière complète.

$$AE_j = Rg_j * \frac{AE_t}{Rg_t} \quad (4.2)$$

où, t et j désignent respectivement une valeur instantanée au passage du satellite ou journalière.

Comme présenté au niveau des travaux de Delogu *et al.* (2021), EF peut être déterminée selon deux méthodes.

- (1) Aux jours clairs (jours avec acquisition satellitaire), EF est calculée à partir des simulations SPARSE. Dans ce cas, EF (notée  $EF_{sparse}$ ) est calculée en se basant sur l'équation 4.3.  $ET_j$  et  $AE_j$  sont obtenues à partir des valeurs instantanées de  $ET_t$  et  $AE_t$  respectivement.  $ET_j$  est reconstruite à partir de  $ET_t$  simulée par SPARSE à l'heure de passage du satellite, en appliquant une extrapolation journalière détaillée au niveau de la section 4.3.2. D'autre part,  $ET_j$  est obtenue à partir de  $AE_t$  à partir de l'équation 4.2.

$$EF_{sparse} = \frac{ET_j}{AE_j} \quad (4.3)$$

- (2) Aux jours nuageux, avec l'enregistrement d'une pluie journalière supérieure à 2 mm localement,  $EF_{api}$  qui est calculée à partir de API selon l'équation 4.4 ou bien à partir de  $EF_{max}$ , qui consiste à attribuer une valeur maximale de EF (EF=1) les jours de pluie (Delogu *et al.*, 2021). La valeur de 2 mm présente la valeur minimale qui fait basculer un pluviomètre à auget basculant lors d'un évènement pluviométrique. Une valeur de précipitations au dessous de 2 mm n'est pas considérée comme un évènement pluvieux. Il peut s'agir des gouttes de rosée par exemple ou d'une pluie non significative. Pour le calcul de  $EF_{api}$ , nous déterminons un maxAPI pour forcer EF à ne pas dépasser sa valeur maximale. Pour les deux méthodes, la taille des données disponibles pour interpoler ET dépend du nombre des évènements pluvieux. Ce fait augmente la base de données à utiliser et nous permet de tirer profit de l'information issue des précipitations pour effectuer l'interpolation.

$$EF_{api} = \min(API, \max API) / \max API \quad (4.4)$$

tel que, maxAPI est égale à 5 et API est calculé par l'équation 4.5. Un maxAPI égale à 5, correspond au maximum des 90% des valeurs de API. Utiliser cette valeur au lieu d'un maximum absolue de la série d'API, nous évite de diviser par des valeurs très fortes de API ayant une occurrence très faible, et avoir par conséquent des valeurs de  $EF_{api}$  très faibles.

$$API(j) = API(j - 1) * 0.85 + PP_j(j - 1) \quad (4.5)$$

Avec  $PP_j$  est la somme de la pluie journalière pour un jour  $j$ . La série de la pluie sur toute la période d'étude est issue à partir des simulations du générateur stochastique MetGen (voir la section 3.3).

La série de EF ainsi reconstruite à la fois (1) à partir de  $(EF_{sparse})$ , obtenues des simulations SPARSE durant les jours clairs, et (2) à partir de  $(EF_{api})$  ou  $(EF_{max})$ , calculée durant les jours nuageux ayant une pluviométrie qui dépasse 2 mm, est ensuite interpolée.  $ET_j$  est alors reconstruite à partir du produit de la série EF interpolée et la série interpolée de l'énergie disponible journalière  $(AE_j)$ , comme montré au niveau de l'équation 4.6.

$$ET_j = EF_{interp} * AE_{interp} \quad (4.6)$$

Au niveau du graphique 4.2, nous avons superposé le résultat de cette interpolation effectuée par la méthode de  $EF_{max}$  utilisée dans les travaux de Delogu *et al.* (2021), avec les données de ET avant interpolation, et des données de ET résiduelles issues des mesures du scintillomètre large ouverture XLAS effectuées dans le cadre des travaux de Saadi *et al.* (2018) et disponibles sur la période 2013-2015. Ces séries de ET sont présentées sur une période de l'année agricole 2013-2014 (du Novembre jusqu'au Mars) et sur un seul pixel choisi au hasard afin d'éviter de moyenner les valeurs de ET. Ainsi, au niveau de ce graphique (fig. 4.2), nous avons présenté les valeurs des précipitations simulées par le générateur stochastique et qui ont été utilisées pour l'interpolation, afin de voir la variabilité de ET interpolée en fonction de celle des précipitations. Cette figure montre que ET interpolée par la méthode de  $EF_{max}$  tient en compte de l'effet de la pluie, qui provoque une augmentation importante du taux d'évapotranspiration. Cette augmentation est aussi observée au niveau des données de ET issues des mesures in situ XLAS. Cependant, cette méthode est incapable de bien reproduire la variabilité de ET<sub>xl</sub> durant la période qui précède un évènement pluvieux, où ET<sub>xl</sub> présente des chutes importantes liées aux évènements pluvieux. Ces chutes correspondent à une réduction de ET enregistrés lors de l'approvisionnement des besoins en eau par les précipitations.

Par ailleurs, afin de mieux reproduire le réalisme de la variation de ET avant et après un évènement pluvial, nous avons guidé l'interpolation de ET lors des jours sans acquisitions satellitaires par l'information de EF issue à la fois des pluies antérieures pour mieux reproduire la variabilité de ET avant un évènement pluvial et  $EF_{max}$ , pour mieux reproduire la variation de ET après la pluie. La combinaison de ces deux informations  $EF_{api}$  et  $EF_{max}$  fait la différence entre la méthode utilisée au niveau des travaux de Delogu *et al.* (2021) et notre travail. La pondération de EF par l'indice de pluie API calculé à partir de l'équation 4.5, permet de retracer l'état réel du statut hydrique de la surface lors des jours pluvieux. En effet, pendant les jours nuageux sans enregistrement d'un apport de précipitations, API présente une variation décroissante jusqu'à atteindre une valeur très faible le jour où on enregistre des précipitations locales. Par conséquent, pendant les jours nuageux sans précipitations, EF est interpolée entre une valeur  $EF_{sparse}$  et  $EF_{api}$  très faible calculée lors de ce jour pluvieux à partir d'API. Ensuite, juste après cet évènement pluvieux, EF doit présenter sa valeur maximale traduisant une demande évaporative maximale  $EF_{max}$ , due à une humidité du sol très importante provoquée par cet apport d'eau. Par ailleurs, EF est forcée par sa valeur maximale ( $EF_{api} = 1$ ) lors du jour clair qui suit directement un évènement pluvieux. Puis, EF reprend ses valeurs calculées à partir des données simulées par SPARSE ( $EF_{sparse}$ ).

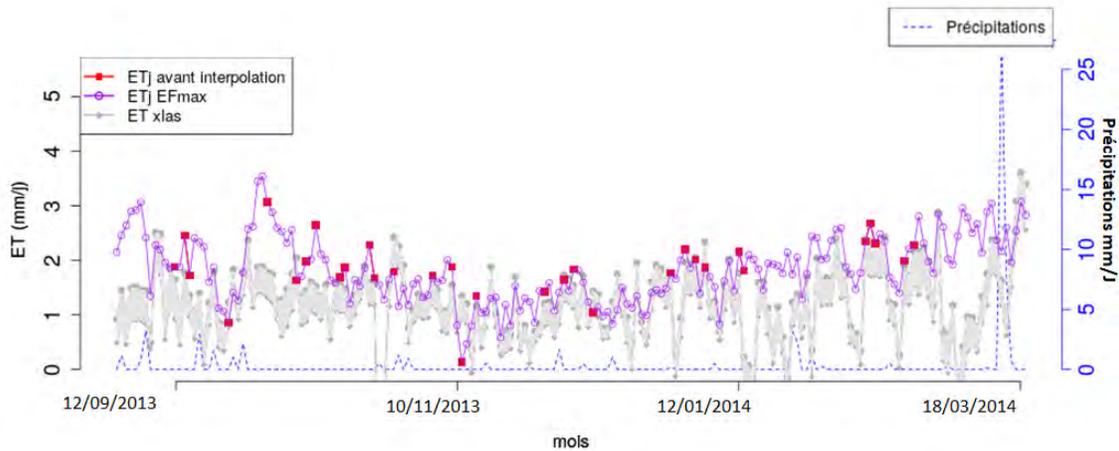


Figure 4.2 – Variation journalière des valeurs de  $ET_j$  sans interpolation,  $ET_j$  issue de l'interpolation par  $EF_{max}$ , suivant la méthode utilisée dans les travaux de Delogu *et al.* (2021), et les données de mesures de scintillomètre XLAS durant une période de l'année agricole 2013-2014 (forçage climatique issu du générateur stochastique).

Pour récapituler, nous disposons trois types de méthodes pour le calcul de EF, comme présenté au niveau du graphique 4.3. Ce graphique présente la variation de la variable EF en fonction des différentes sources de son calcul : (1)  $EF_{sparse}$  lors des jours où on dispose une acquisition satellitaire. Dans ce cas, EF reproduit la demande évaporative réellement observée à partir des conditions climatiques et de surface insérées en entrée du modèle SPARSE, (2)  $EF_{api}$  lors des jours pluvieux avec des précipitations supérieures à 2 mm et (3)  $EF_{max}$  lors du jour clair qui suit un évènement pluvieux. Ces différentes sources de calcul font varier EF d'une façon plus réaliste et plus adéquate aux caractéristiques du climat de notre zone d'étude.

Nous présentons au niveau de la figure 4.4, le résultat de l'interpolation de ET avec la méthode d'API. Nous observons, au niveau de ce graphique, une variabilité de ET interpolée plus cohérente avec ET-xlas. Avant chaque évènement pluvial,  $ET_j$ -API montre une décroissance progressive engendrée par la décroissance des valeurs de API. Après chaque apport en eau par les précipitations, ET augmente d'une façon intense suite au forçage de EF à une valeur maximale, reflétant une demande évaporative importante et une humidité du sol très importante. Cependant, nous observons parfois, au niveau de ce graphique, une discordance entre la variation de ET interpolée par API et les valeurs de ET-xlas. Cette discordance est expliquée par l'utilisation des données de précipitations simulées par le générateur stochastique, qui n'était pas performant pour la reproduction de cette variable, en termes d'occurrence et d'intensité. Cependant, les variations de ET-xlas sont dépendantes des mesures locales des précipitations, qui ne sont pas forcément reproduites par les simulations du générateur.

### 4.2.3 Comparaison des deux méthodes d'interpolation Rg/API

Au niveau de cette section, nous avons comparé la performance de différents supports d'interpolation utilisés Rg et API. Nous présentons au niveau de la figure 4.5, des séries

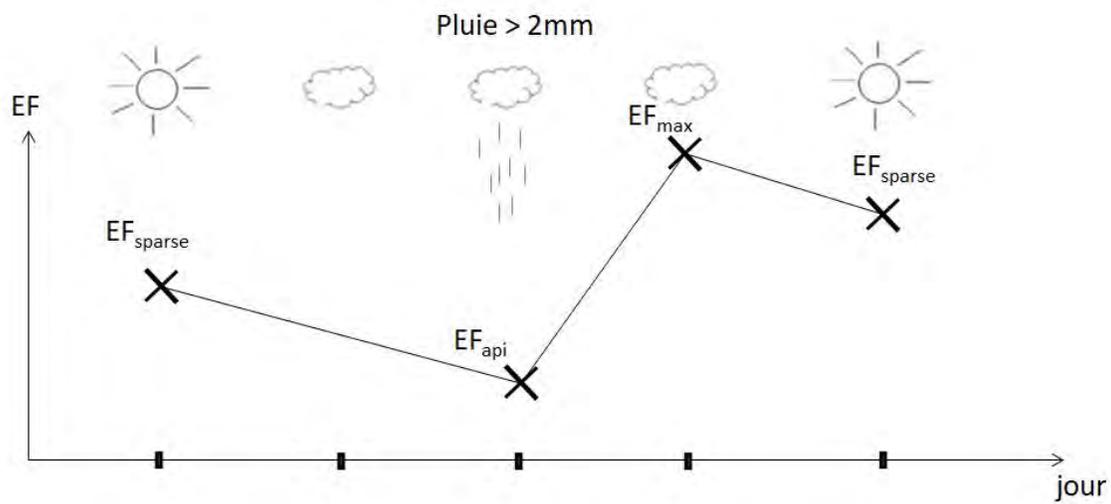


Figure 4.3 – Illustration de la variation de la variable fraction évaporative (EF) en fonction de la méthode de son calcul.

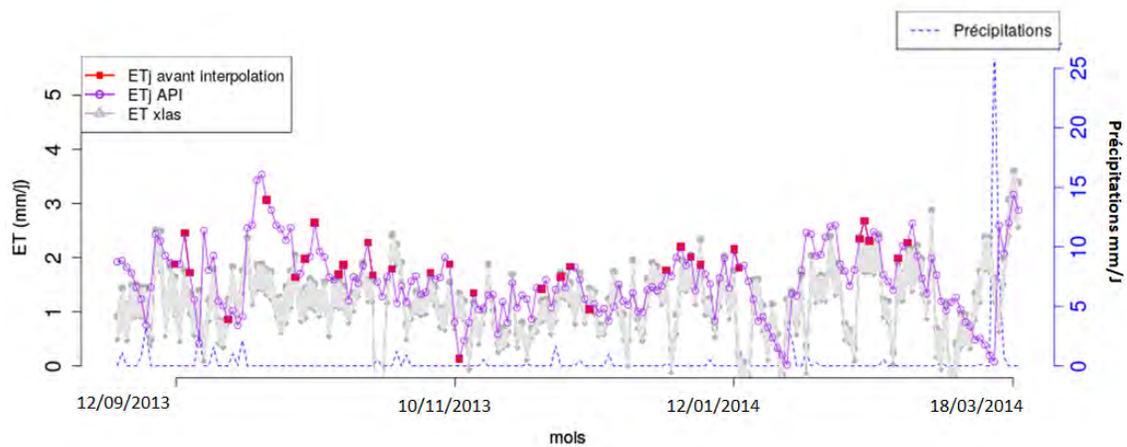


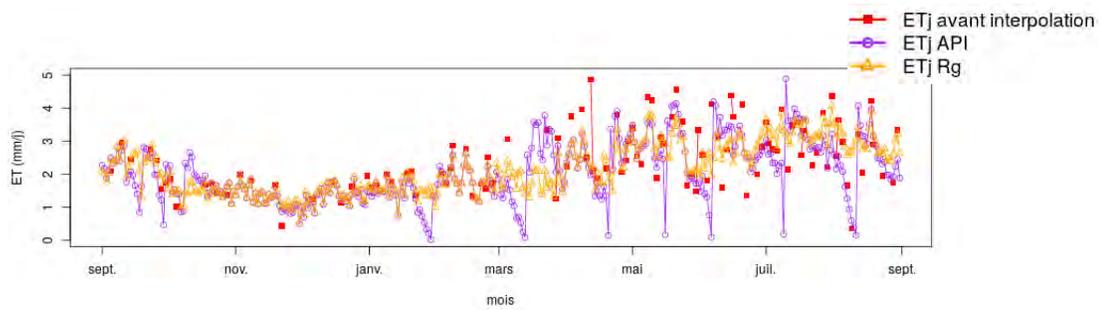
Figure 4.4 – Variation journalière des valeurs de  $ET_j$  sans interpolation,  $ET_j$  issue de la nouvelle méthode d'interpolation par API, et les données de mesures de scintillomètre XLAS durant une période de l'année agricole 2013-2014 (forçage climatique issu du générateur stochastique).

temporelles de l'évapotranspiration journalière ( $ET_j$ ), simulées à partir des différents forçages climatiques (SWG et ERA5) et différents forçages satellitaires (Terra et Aqua). Nous avons superposé au niveau de chaque graphe,  $ET_j$  avant l'interpolation,  $ET_j$  interpolée par le support Rg et  $ET_j$  interpolée par le support API. Pour cette comparaison, nous avons utilisé pour chaque graphe 4.5a, 4.5b, 4.5c et 4.5d, une combinaison de forçage climatique et satellitaire, pour l'année hydrologique 2013-2014, durant laquelle nous disposons les données de mesure de scintillomètre XLAS. L'interpolation temporelle est effectuée d'une façon indépendante au niveau de chaque pixel. Les valeurs d'interpolation présentées au niveau de cette figure forment les moyennes calculées pour chaque image, c'est-à-dire, pour chaque pas de temps.

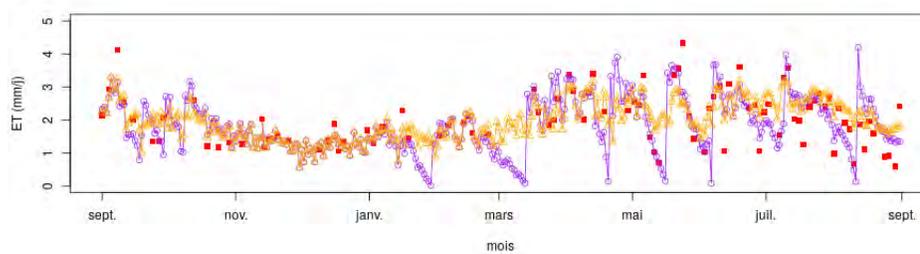
Les différences observées au niveau de ce graphique (4.5), entre les différentes séries de ET sont dues au fait de moyenniser les images de ET issue de différentes sources  $ET_j$  avant l'interpolation et ET interpolée utilisant les supports API et Rg. Au niveau de ce graphique, nous observons qu'en utilisant la méthode d'interpolation basée sur le support Rg, les valeurs de ET interpolées suivent correctement les valeurs de  $ET_j$  avant l'interpolation avec les différents forçages climatiques et satellitaires. Nous observons une fluctuation stable, qui tient en compte de la saisonnalité, avec des valeurs de ET plus faibles en hiver et plus élevées en été. L'effet des cycles saisonniers est introduit par le support d'interpolation utilisé Rg, qui forme une variable avec une variation diurne et saisonnière très marquée. Cependant, en utilisant le support d'interpolation API, on voit que l'effet saisonnier est moins marqué. Nous observons en revanche, des fluctuations plus accentuées typiques aux fluctuations de ET après un événement pluvial. Les chutes observées au niveau des valeurs de  $ET_j$  issues de l'interpolation par API sont observées davantage pendant la période d'été, où le dessèchement de la surface se fait d'une façon plus rapide en réponse à la demande évaporative élevée durant cette saison. Pour vérifier le réalisme de ces deux types d'interpolation, nous allons superposer ces deux séries de ET interpolées avec les données mesurées par le scintillomètre XLAS, au niveau du graphique 4.6, durant toute la période où nous disposons des données de mesures XLAS (de Mars 2013 jusqu'au Juin 2015).

Au niveau du graphique 4.6, nous présentons la variation journalière de ET issues de deux types d'interpolation, de ET avant l'interpolation et des données ET issues des mesures de XLAS sur toute la période où on dispose ces données. Ce graphique montre que, pendant les jours clairs,  $ET_j$  sans interpolation et  $ET_j$  interpolées avec les deux méthodes sont bien les mêmes. Pendant, les jours nuageux,  $ET_j$  issues de l'interpolation par API ou par Rg montrent une variation très semblable, même si celle basée sur l'API permet de simuler des diminutions de ET pendant les périodes sèches nuageuses. Pendant les jours nuageux avec l'enregistrement des précipitations supérieures à 2 mm/jour, nous observons une variation très différentes entre les deux méthodes d'interpolations.  $ET_j$ -API montrent des chutes brutales induites par la diminution accentuée de la demande évaporative. Cependant,  $ET_j$ -Rg montrent toujours une variation proportionnelle à la variation diurne de Rg. Les fluctuations observées de  $ET_j$ -API se rapprochent au mieux des fluctuations observées par les valeurs de ET issues de XLAS.

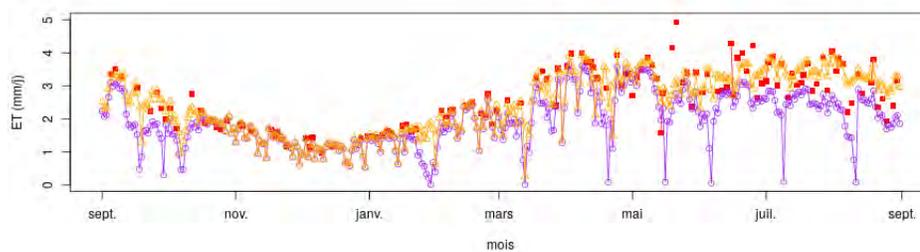
Une interpolation basée sur la pluie antérieure semble être une bonne piste pour la reconstruction journalière et saisonnière de l'évapotranspiration. Cependant, cette méthode nécessite encore une amélioration pour construire d'une façon plus réaliste la variation de



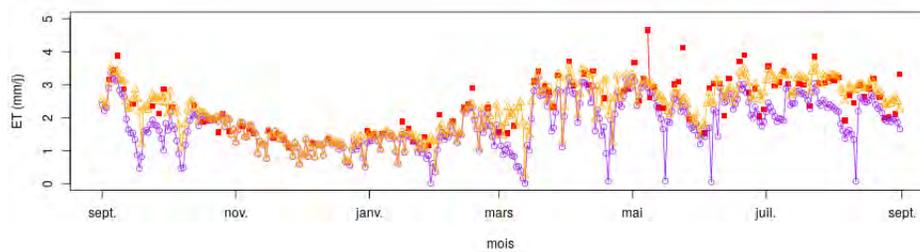
(a) Forçage climatique SWG, satellite Terra



(b) Forçage climatique SWG, satellite Aqua

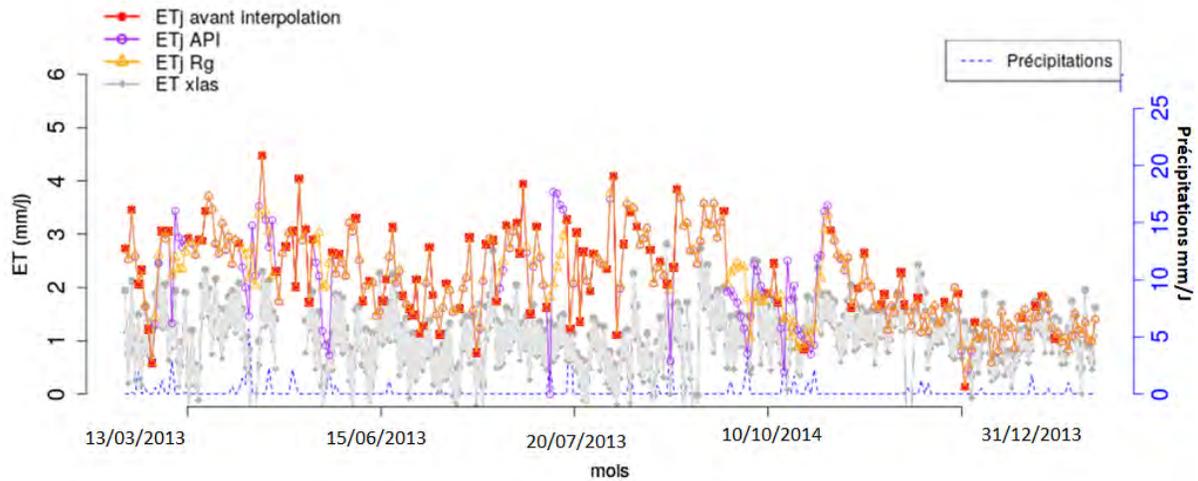


(c) Forçage climatique ERA5, satellite Terra

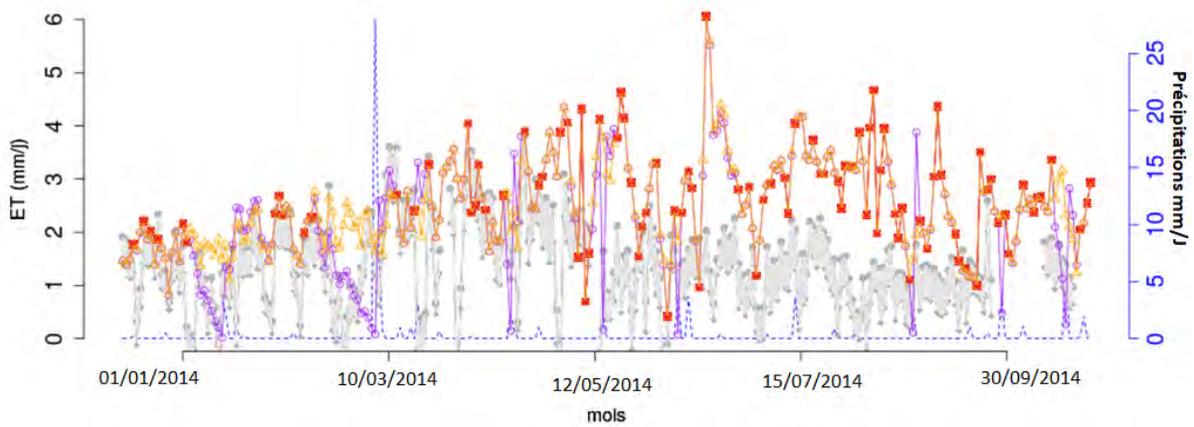


(d) Forçage climatique ERA5, satellite Aqua

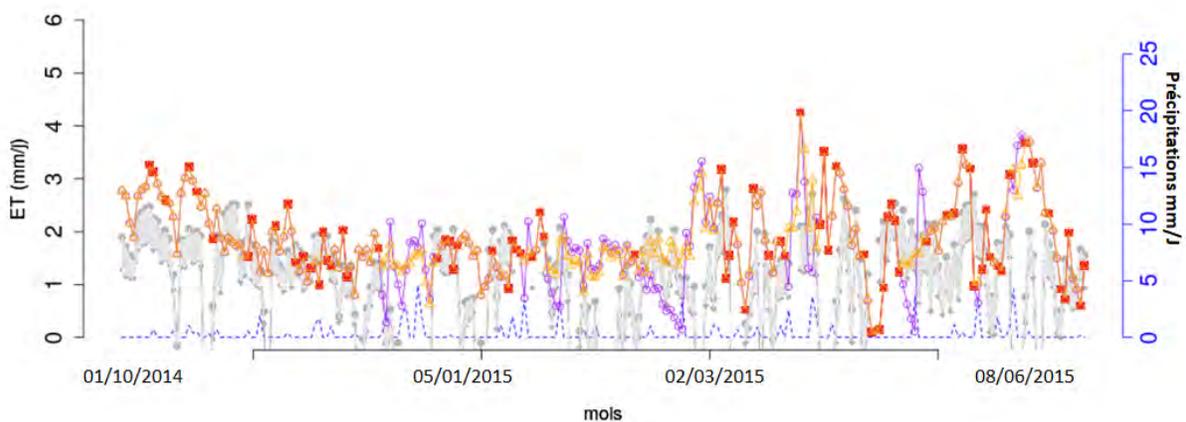
Figure 4.5 – Séries temporelles de l'évapotranspirations journalières et celles issues de deux méthodes d'interpolation testées, en utilisant différents forçages climatiques et satellitaires, sur l'année agricole 2013-2014.



(a) Partie1



(b) Partie2



(c) Partie3

Figure 4.6 – Variations journalières des valeurs de  $ET_j$  sans interpolation,  $ET_j$  issue de deux méthodes d'interpolation par API et par Rg (forçage climatique issu du générateur stochastique) et les données de mesures de scintillomètre XLAS durant toute la période où on dispose des mesures XLAS (de mars 2013 jusqu'au Juin 2015).

l'évapotranspiration lors des jours pluvieux tout en profitant de cette information. Un travail supplémentaire de paramétrage de l'équation, vérification avec des données de mesure et aussi une meilleure estimation des taux de précipitations s'avèrent indispensable pour rendre cette méthode plus robuste pour estimer l'évapotranspiration d'une façon journalière. Delogu *et al.* (2021) trouvent que la méthode d'interpolation basée sur la pluie antérieure sous-estime souvent les valeurs d'évapotranspiration durant les jours humides. Pour éviter ce problème de sous-estimation essentiellement durant des longues périodes de non acquisition satellitaire, nous avons finalement choisi d'utiliser le support  $R_g$  pour interpoler les valeurs de  $ET_j$  pour le reste des résultats montrés au niveau de ce travail.

## 4.3 Indices de sécheresse

La comparaison des différents indices de stress fait l'objet d'un article soumis à "Remote Sensing and environment" qui est présenté ci-dessous.

### 4.3.1 Résumé de l'article

La surveillance des sécheresses dans les régions arides et semi-arides est cruciale car les conséquences de sécheresses prolongées pour l'agriculture peuvent être dramatiques. Ainsi, une détection précoce des périodes de sécheresse s'avère très utile pour atténuer ses dégâts. Afin d'aider les décideurs à établir de bonnes pratiques de gestion des ressources en eau et d'atténuation du risque des sécheresses, nous nous intéressons à l'analyse des besoins en eau de la plante à travers des indices de stress hydrique. Le choix d'un indice efficient reste un enjeu majeur pour l'identification précoce et précise d'une période de sécheresse. D'autre part, le contexte climatique et anthropique, en termes d'utilisation de la terre agricole et les différentes pratiques usuellement observées au niveau de la région étudiée, doivent être pris en considération pour établir un choix adéquat de l'indice à utiliser. Pour répondre à ces questions, nous avons comparé un indice thermique (SI) qui, selon la littérature, forme un outil performant pour la détection du stress hydrique au niveau des régions semi arides, avec d'autres indices, NDVI et SWI issus du visible et micro-onde respectivement. Cette comparaison est effectuée pour étudier la pertinence de ces différents indices, mais également leur cohérence et précocité, c'est-à-dire détecter les plus faibles valeurs de stress hydrique d'une façon anticipée. Pour faciliter la comparaison et l'interprétation du comportement des indices, nous avons effectué une standardisation qui utilise la fonction de répartition empirique afin de renvoyer des valeurs entre 0 et 1 facilement interprétables en termes de probabilité de non dépassement pour chaque valeur de l'indice. D'autre part, l'utilisation des valeurs standardisées des indices de stress (anomalies) permet d'identifier le degré de sécheresse par rapport à la moyenne des valeurs historiquement observées.

L'indice thermique utilisé est simulé à partir du modèle de bilan d'énergie à double source SPARSE (Boulet *et al.*, 2015), à résolution spatiale kilométrique. Ce dernier combine à la fois des informations satellitaires et météorologiques. SPARSE a été contraint par la série climatique générée à partir du générateur stochastique simulée à une échelle locale et aussi par les données à large échelle ERA5 brute, afin de simuler deux indices thermiques  $SI_{SWG}$

et  $SI_{ERA5}$ , respectivement. L'utilisation de ces deux séries en tant que forçage climatique au niveau du modèle d'impact, nous permettra de voir la valeur ajoutée de l'application d'une méthode de descente d'échelle sur les modèles d'impact, ainsi que pour l'analyse des périodes de sécheresse. D'autre part, la comparaison de ces deux types de séries climatiques nous permettra d'évaluer l'utilisation directe d'un forçage issu des réanalyses en tant qu'alternative d'application d'une méthode statistique compliquée à mettre en oeuvre.

L'évaluation de ces indices est effectuée à l'échelle régionale (au niveau de la plaine) et à l'échelle locale (échelle kilométrique), en se focalisant sur la zone du transect XLAS ou bien sur les zones ayant un pourcentage élevé de céréales pluviales. En effet, notre zone d'étude forme une zone très irriguée à partir d'une nappe quasi fossile. Elle forme alors une source peu sensible à la sécheresse. Pour cela, nous avons focalisé une partie de ces analyses sur des cultures herbacées pluviales, les plus sensibles à la sécheresse.

Plutôt que le SPI, nous avons utilisé ici l'UPI comme indice standardisé des précipitations, considéré comme le signal déclencheur de la sécheresse agricole. Ces deux indices sont issus de deux méthodes de standardisation différentes (normalisation standard et normalisation par la fonction de répartition empirique, pour SPI et UPI respectivement). Toutefois, ces deux méthodes sont basées sur le même principe. En effet, SPI est standardisé en appliquant la fonction de répartition empirique et en calculant des quantiles de la loi normale associés à ces probabilités de non-dépassement. SPI applique alors une transformation supplémentaire à partir de UPI pour obtenir des valeurs distribuées selon la loi normale standard. Dans notre cas, nous avons utilisé la standardisation par la fonction de répartition empirique, pour ne pas nous restreindre à l'application de cette méthode que pour les variables ayant une loi de probabilité normale.

L'évaluation de ces indices à l'échelle régionale est effectuée sur différentes échelles temporelles (saisonnière et décadaire), essentiellement sur la période de croissance végétative de céréale (de Septembre jusqu'au Mai). A l'échelle saisonnière, nous observons une concordance au niveau de la variation des différents indices et une bonne performance à identifier simultanément les années sèches et humides. Cette corrélation est prouvée à l'aide d'un coefficient de corrélation non linéaire, le coefficient de Kendall, qui a montré non seulement une bonne corrélation entre la majorité des indices et UPI, mais également une très bonne corrélation entre les deux indices thermiques  $SI_{SWG}$  et  $SI_{ERA5}$ , issu de la même information satellitaire. A une échelle décadaire, nous observons une réponse plus rapide et cohérente entre UPI et les deux indices thermiques. D'autre part, nous observons une réponse très tardive de NDVI suite à un déficit hydrique. Cette interprétation est consolidée par l'établissement d'une analyse statistique qui nous permet de déterminer la probabilité d'identifier les conditions de stress (stress ou non stress) en concordance avec UPI ou avec les moyennes mobiles de UPI. Cette analyse est effectuée sur les 4 saisons de l'année, afin d'étudier la cohérence de ces indices avec UPI et son effet mémoire. La définition d'une période de stress est basée sur UPI, en fixant un seuil de stress, égal à 20% des valeurs de probabilité d'occurrence les plus faibles. Ce seuil est fixé en équivalence aux classes de stress définies par l'indice SPI, pour lesquelles SPI est inférieur à -1 d'après les classes identifiées dans McKee *et al.* (1993). En effet, la probabilité d'avoir des valeurs inférieures ou égales à -1 pour une loi Normale standard est environ de 16%, ce qui correspond approximativement à une probabilité de 20% pour une

fonction de répartition à partir de laquelle nous avons calculé nos indices. Ainsi, cette période de stress doit persister pour au moins deux décades successifs, afin d'éviter les périodes de stress incertaines. On trouve que les indices  $SI_{ERA5}$ ,  $SI_{SWG}$  et SWI présentent des résultats similaires avec une performance légèrement meilleure lorsque le générateur stochastique est utilisé. La concordance entre les indices et UPI diminuent en été. En effet, la saison d'été est caractérisée au niveau des régions semi aride, par la rareté des précipitations. Pour cela, nous observons plus des périodes de stress identifiées par UPI mais qui n'ont pas été identifiées par les autres indices vu l'introduction de l'irrigation qui devient l'unique source d'approvisionnement en eau pour cette saison. D'autre part, en introduisant des effets mémoires sur l'indice de précipitations, on voit une augmentation de la cohérence des réponses des indices par rapport à UPI et diminution des périodes de stress non identifiées par UPI. Ceci prouve qu'un temps de lag est nécessaire pour la végétation pour répondre à un stress hydrique. Cette période de lag diffère d'un indice à un autre.

A l'échelle locale, nous avons utilisé quelques pixels correspondant au transect XLAS et comparer par ailleurs les différents indices avec l'indice de stress issu des données de mesures. D'autre part, nous avons identifié quelques pixels avec un pourcentage élevé de céréale en sec et nous avons évalué la performance des indices à définir la classe de stress hydrique d'une année humide et une année sèche. Ces analyses ont montré l'efficacité des indices thermiques à identifier les périodes de stress, et particulièrement celui issu du générateur stochastique. La performance du générateur revient au fait qu'il était calibré par rapport aux données météorologiques observées localement, et qu'il présente une résolution spatiale plus fine par rapport à  $SI_{ERA5}$ , qui est simulée en utilisant des données à 31 km de résolution spatiale.

Ce travail nous a permis de voir la pertinence des indices thermiques en comparaison avec les autres indices et notamment celui issu du micro-onde, considéré comme un indicateur efficace de la sécheresse agronomique dans la littérature. La résolution spatiale fine des indices thermiques, essentiellement celui issu du générateur stochastique comme forçage climatique, a joué un rôle très important dans la précision et l'efficacité de cet indice pour la détection de l'état de stress hydrique. Ce constat est beaucoup plus nette sur les résultats de comparaison avec l'indice issu des mesures XLAS.

### 4.3.2 Publication

Article

# Analysis of Multispectral Drought Indices in Central Tunisia

Nesrine Farhani <sup>1,2</sup>, Julie Carreau <sup>3</sup>, Zeineb Kassouk <sup>1</sup>, Michel Le Page <sup>2</sup>, Zohra Lili Chabaane <sup>1</sup> and Gilles Boulet <sup>2</sup><sup>1</sup> Université de Carthage, Institut National Agronomique de Tunisie, LR17AGR01-GREEN-TEAM, Tunis, Tunisia<sup>2</sup> Centre d'Études Spatiales de la Biosphère, Université de Toulouse, CNRS, CNES, IRD, UPS, INRAE, Toulouse, France<sup>3</sup> Department of Mathematics and Industrial Engineering, Polytechnique, Montréal, Canada

\* Correspondence: nesrine.farhani@ird.fr; Tel.: +216-20-212-738

**Abstract:** Surface water stress remote sensing indices can be very helpful to monitor the impact of drought on agro-ecosystems and serve as early warning indicators to avoid further damages to the vegetation and the productivity. In this study, we compare indices from three different spectral domains: the plant water use derived from evapotranspiration retrieved using data from the thermal infrared domain, the root zone soil moisture at low resolution derived from the microwave domain using the Soil Water Index (SWI), and the active vegetation fraction cover deduced from the Normalized Difference Vegetation Index (NDVI) time series. The thermal stress index is computed from a dual-source model Soil Plant Atmosphere and Remote Evapotranspiration (SPARSE) that relies on meteorological variables and remote sensing data. In order to extend in time the available meteorological series, we compare the use of a statistical downscaling method applied to reanalysis data with the use of the unprocessed reanalysis data. Our study shows that thermal indices show comparable performance in overall compared to the SWI at better resolution. However, thermal indices are more sensitive for a drought period and tend to react quickly to water stress.

**Keywords:** Indicators, droughts, evapotranspiration, reanalyses

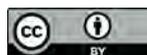
**Citation:** Farhani, N.; Carreau, J.; Kassouk, Z.; Le Page, M.; Chabaane, Z.L.; Boulet, G. *Remote Sens.* **2022**, *xx*, 0xx, 5.

Received:

Accepted:

Published:

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Droughts are a recurring natural climate event that result from a reduction in precipitation amount received over an extended period of time, such as a season or a year [1]. In arid and semi-arid areas characterized by a significant temporal and spatial climatic variability, the vulnerability advanced by recurrent droughts is considerable as it makes serious threats to agroecosystem health and productivity [2]. More specifically, in several Mediterranean countries, we observe during the last decade a significant warming trend, more pronounced in summer and a decrease in rainfall during the wet season [3]. This could have a strong impact on water resources and constitutes the main driver of agricultural droughts which appears as a result of a long-term period of precipitation deficiency and leads to lower soil moisture. Droughts at a sensitive stage of crop development (emergence, pollination, and grain filling) can lead to a significant damages [4,5] specifically on the rainfed agriculture. Thus, an important issue for rainfed agriculture is to receive adequate rainfall at the appropriate timing. In contrast, drought responses over irrigated area is related to surface or groundwater resource availability which can also suffer from a severe drop during prolonged periods of droughts. Therefore, it is crucial to improve the

41 monitoring of agricultural droughts and the prediction of their occurrence in the future [6].  
42 In this work, we are interested by agronomic drought identification.

43 Several drought indices have been developed to quantify agronomic drought periods.  
44 Initial drought indices rely essentially on meteorological variables. Most meteorological  
45 indices use precipitation either on its own or in combination with other meteorological  
46 variables [6]. The most commonly used meteorological index is the Standardized Precipitation  
47 Index (SPI) [7], which classifies precipitation intensities according to a probability  
48 distribution derived from historical records, over different durations (1 month SPI, 3  
49 months SPI, etc.). SPI is particularly useful for drought identification and frequency analysis  
50 [7–9]. However, meteorological indices bring information on meteorological droughts  
51 defined only in terms of rainfall shortage, but not specifically on their consequences on  
52 vegetation for instance, especially for agronomic droughts. On the other hand, remote  
53 sensing (RS) is an effective tool to provide information on whether vegetation functioning  
54 is optimal or sub-optimal. Indeed, RS has shown high potential to retrieve land surface  
55 properties related to plant functionality, such as vegetation cover from the visible/near  
56 infrared domain [10,11], surface soil moisture from the microwave remote sensing using  
57 passive or active radiometers [12–15], or surface temperature from the thermal infra-red  
58 (TIR) data that exploits surface energy balance components to identify land surface energy  
59 interactions [16].

60 RS observations of green biomass vegetation indices are relevant indicators of year-to-  
61 year drought conditions and vegetation greenness, especially over the past few decades  
62 thanks to the increasing number of multi and hyperspectral sensors [11,17], such as,  
63 Advanced Very High Resolution Radiometer (AVHRR), SPOT-Vegetation and Moderate-  
64 Resolution Imaging Spectroradiometer (MODIS). These sensors provide spatial resolution  
65 ranging from 250 to 1000 meters. Historically, the Normalized Difference Vegetation  
66 Index (NDVI) [18], built from the normalized difference between Near-InfraRed (NIR) and  
67 RED reflectances, has been used for decades to provide information on vegetation health  
68 due to its strong correlation with the photosynthetic activity of plant cover. A suite of  
69 drought indices based on NDVI were subsequently developed in order to provide more  
70 focused information, such as the vegetation condition index (VCI) [19]. VCI uses the  
71 NDVI information provided from each pixel normalized with the maximum and minimum  
72 statistical range over the available time series of images. It has been tested in several regions  
73 of the world, showing a relevant efficiency to characterize period of droughts [20,21]. The  
74 Vegetation Anomaly Index (VAI) developed by [22], forms also a standardization of the  
75 NDVI time series. Although RS based vegetation observations indices have proven their  
76 efficiency to assess drought impacts on vegetation growth, they are not able to detect  
77 incipient stress because of the time lag needed for the plant to reflect drought-induced  
78 physiological changes. Therefore, a delayed response to prolonged stress will be provided  
79 using these indices [23]. Response of vegetation to drought is a gradual process altering  
80 soil moisture, that lead to a time lag effect [24].

81 Surface soil moisture (SSM), which is inferred from microwave RS, is another efficient  
82 key parameter for agricultural planning and water resources management [12,25]. Never-  
83 theless, for drought monitoring and agricultural modeling a representation of root-zone  
84 soil moisture is needed [26]. A number of soil moisture-based indices have been developed  
85 and widely used for drought monitoring, such as the Soil Water Index (SWI) [27]. This  
86 index is retrieved from the Advanced SCATterometer (ASCAT) SSM estimates in C band,

87 which was found more suitable in bare soils or sparse vegetation cover [28], to determine  
88 Soil Moisture (SM) distributions within the soil profile using an exponential relationship  
89 that links surface and subsurface SM via a characteristic time length [29]. SWI is also used  
90 on analyses of statistical anomalies over several years to derive the Moisture Anomaly  
91 Index (MAI) [15], that has shown a good performance to provide quantitative informations  
92 on drought periods over the central region of Tunisia. The soil moisture deficit index  
93 (SMDI), developed by [30], was helpful for planning to mitigate drought impacts. This  
94 index is calculated from weekly percentage SM deficit or excess using long-term median,  
95 maximum and minimum soil water [30]. The main limitation of drought indices based  
96 on passive microwave data is their low resolution. Moreover, quantitative soil moisture  
97 estimation is still a challenge nowadays, especially under vegetation cover [12,31].

98 RS maps of Land Surface Temperature (LST) are very informative of water availabil-  
99 ity and form a good indicator of incipient droughts [32]. Indeed, water stress induces a  
100 stomatal closure that generates in turn an elevated canopy temperature [33–37]. Thermal  
101 channels of Landsat, AVHRR or MODIS were exploited to retrieve remotely sensed LST esti-  
102 mates. LST can be used as a signature of the land surface energy budget and in particular, to  
103 determine whether the dissipation of available energy is more into sensible (dry conditions)  
104 or/and latent heat (wet conditions) [38]. It also enables to compute evapotranspiration  
105 (ET) from latent heat computed as the residual of the surface energy budget [39,40]. ET  
106 is particularly informative about water stress conditions [40,41]. Indeed, the crop water  
107 requirements are adjusted to match the water losses from the actual evapotranspiration  
108 (ETa) [42]. Consequently, in water stress conditions, the plant reduces its transpiration in  
109 comparison with unstressed vegetation under the same atmospheric conditions [43]. Thus,  
110 the water stress intensity may be inferred from the ratio of actual and potential (unstressed)  
111 ET. Several drought indicators that integrate ET have been explored and showed improved  
112 drought monitoring. [44] proposed the so-called Water Deficit Index (WDI) based on the  
113 approach Vegetation Index/Temperature Trapezoid (VIT). WDI is estimated for each pixel  
114 to retrieve dry/wet conditions based on the fractional vegetation cover and the difference  
115 between LST and air temperature (Ta). A good example of TIR indices is the evaporative  
116 stress index (ESI) [35] that describes anomalies in the actual and the reference ET ratio. ESI  
117 was shown to provide early warning particularly for flash droughts that could produce  
118 damaging impacts over short time periods [5]. The stress index (SI) is defined as the ratio  
119 between actual and potential evapotranspiration rates.

120 In summary, several drought indices are commonly used for drought monitoring and  
121 impact assessment in agriculture. However, it is hard to assess their respective performance  
122 in deciphering stress/non stress and drought/non drought situations. An inter-comparison  
123 between the performance of these different drought indicators provided from different  
124 wave lengths, in terms of consistency, reliability and ability to detect incipient water plant  
125 stress is thus relevant. [19] evaluated VCI and the Temperature Condition Index (TCI)  
126 in different regions of the world. The analyses show that VCI has an excellent ability  
127 to detect each period of stress, the drought onset, intensity, duration, and impact on  
128 vegetation. The TCI provides additional information about vegetation stress through  
129 LST. [19] used VCI and TCI to develop the so-called Vegetation Health Index (VHI), a very  
130 known combined stress index and widely used for drought detection and assessment of  
131 drought severity and duration [45]. However, as a simple averaging is performed, there  
132 is no real theoretical support for the way to weight the relative impact of stress (through

133 LST) and vegetation development (through NDVI). For this reason, we suggest in this  
134 work to compare different indices provided from individual biophysical variables (1) the  
135 "Normalized difference vegetation index" (NDVI) from the solar (Visible/Near InfraRed  
136 spectrum), (2) "Soil Water Index" (SWI) from the microwave domains and, (3) a "Thermal  
137 stress Index" (SI) from Thermal InfraRed retrieved from a surface energy budget model.  
138 Analyses are carried out in the Kairouan area in central Tunisia which is subject to semi-  
139 arid climate. The thermal stress index used is based on ET simulated from a dual source  
140 energy balance model that provides robust estimates of ET when meteorological forcing  
141 and vegetation cover are accurately known [46]. A statistical downscaling method [47]  
142 combined with reanalysis data is used to generate surrogate series in order to extend the  
143 observation period. Then, simulated ET series from the downscaling method or from the  
144 unprocessed reanalyses are used to constrain the dual-source model Soil Plant Atmosphere  
145 and Remote Evapotranspiration (SPARSE) [48]. In this paper, we address 3 objectives:

146 (1) We first assess the consistency between those indices and a precipitation-based  
147 drought index called the "Uniformized Precipitation Index" (UPI) which is a variant of the  
148 SPI, at different time scales (annual and decadal variation) at regional scale. We assume that  
149 a robustness of the indices at very low resolution (12 Km) translates into a good accuracy  
150 at higher resolution (kilometric).

151 (2) We assess the reliability of very low (UPI, SWI) and low (SI, NDVI) resolution  
152 indices at higher resolution (kilometric) at local scale, by focusing on drought monitoring  
153 for pixels either with a high proportion of rainfed wheat or that correspond to the extra  
154 large aperture scintillometer (XLAS) transect;

155 (3) We compare SI when the energy balance model is either forced with the downscaled  
156 meteorological data or with the unprocessed reanalysis data, in order to test the applicability  
157 of SI with routinely available data at global scale.

## 158 2. Study area

159  
160 The study site, the Merguellil catchment (Figure 1) is located in central Tunisia, which  
161 is typical of semi-arid environments. The study area is influenced both by the Mediter-  
162 ranean climate (dry subhumid) and the pre-Saharan's climate (arid). It is characterised by  
163 the inter-annual irregularity of precipitation, with an average of annual rainfall of about  
164 300 mm per year, with a short rainy seasons mostly between September and May [49] and  
165 by a high evaporative demand of about 1600 mm per year.

166 The upstream and the downstream sub-catchments are separated by El Haouareb dam  
167 that protect the village from inundations and provides surface irrigation water for the plain.  
168 However, most of the water used for irrigation is extracted from ground water. Merguellil  
169 is emblematic of hydrological processes which have been profoundly modified by human  
170 activities since the end of the 1960s, by the inclusion of soil and water conservation works,  
171 the construction of large and small dams, and intensification of irrigated farming [49,50].  
172 As expected, agriculture is the main water consumer of available water resources in this  
173 region (around 80 of the available water%) [51]. Therefore, the aquifer system shows  
174 a considerable decrease over time and space due to over-exploitation [50]. This makes  
175 the Merguellil catchment an interesting case study to investigate the ability to monitor  
176 agricultural water management in the recent years. In this work, we are interested only by  
177 the lower sub-catchment (3000 km<sup>2</sup>) which forms an alluvial plain mainly flat with altitude

178 between 50 and 200 m. The plain incorporates mainly small cultivated areas [52], with  
 179 characteristic vegetation of semi-arid regions: rainfed agriculture (olive trees and cereals)  
 180 and vegetables (melons, chili's and tomatoes). The gauged network in the Merguellil plain  
 181 catchment has three stations (Fig. 1), installed since 2012 for the earliest ones.

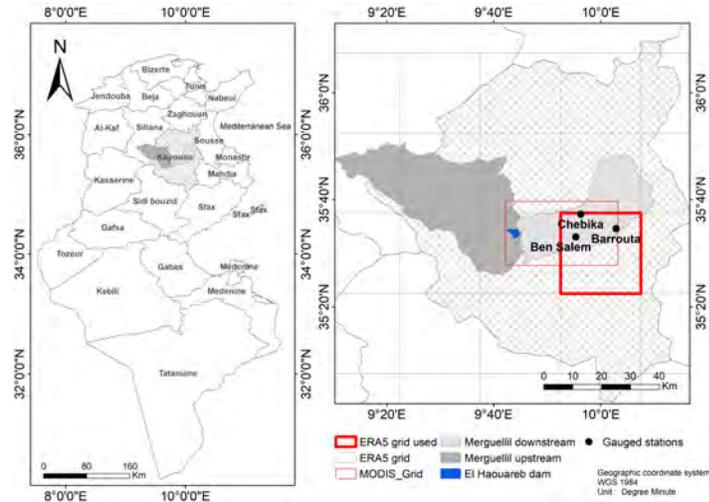


Figure 1. Merguellil catchment in central Tunisia.

### 182 3. Data and methodology

#### 183 3.1. Biophysical indices derivation

##### 184 3.1.1. NDVI

185 We used MOD13 version 6 product from the Moderate Resolution Imaging Spectro-  
 186 radiometer (MODIS) sensor on board the NASA Terra and Aqua Earth Observation System  
 187 satellites that provides each 16-day composite at 1 km of spatial resolution. Time series  
 188 are retrieved between 2000 and 2019 over our study area (see figure 1). NDVI uses the  
 189 normalized difference between near-infrared (NIR) and (RED) reflectances.

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}} \quad (1)$$

190 where  $\rho_{NIR}$  and  $\rho_{Red}$  are the Near-infrared (NIR) and red reflectances, respectively.  
 191 NDVI increases with the amount of healthy green photosynthetically active vegetation.  
 192 Uniformized NDVI measures reflect the “current” vegetation conditions according to a  
 193 longer-term historical average NDVI value over the same study area.

##### 194 3.1.2. SWI

195 The soil water index (SWI) data is derived from the surface soil moisture (SSM) using  
 196 an infiltration model to describe the relation between surface soil moisture and profile soil  
 197 moisture as a function of time.

$$SWI(t) = \frac{\sum SSM(t_i) e^{-(t-t_i)/T}}{\sum e^{-(t-t_i)/T}}, \quad (2)$$

198 where SSM is the surface soil moisture estimate from the ASCAT at time  $t_i$ . The  
 199 parameter T, called the characteristic time length, represents the time period to integrate  
 200 SSM data. T is the most important input parameter to derive the profile soil moisture

201 content from remotely sensed surface time series: a high T value describes a deeper soil  
 202 layer if the soil diffusivity is constant and the same T value describes different depths  
 203 for different soils (texture) [27]. [53] found in a semi-arid region characterized with an  
 204 heterogeneous texture distribution in the soil profiles as our study region, having surface  
 205 soils of sandy texture and a higher clay concentration to a depth of 2 m [54], that T=60  
 206 days is the best choice for the 50 – 100 cm soil profile which corresponds to the root  
 207 layer according to the predominantly agricultural use of our study region (olive trees and  
 208 cereals). Indeed, the upper part of the profiles promotes high infiltration rates and low  
 209 water retention forces [53].

210 The SSM is retrieved from scatterometer observations from the ASCAT instruments on  
 211 board the MetOP satellites, which measures radar backscatter at C-band in VV polarization,  
 212 with initial spatial resolution of 25 km, re-sampled at 12.5 km [55]. SWI product is retrieved  
 213 from the Copernicus Global Land service and provides global daily information since  
 214 January 2007 to present [56].

### 215 3.1.3. Thermal infrared stress index

216 We used a stress index (SI) obtained from TIR. The surface temperature (LST) provides  
 217 indirect estimates of water stress since it enters in the surface energy balance equations and  
 218 it is thus related to the evapotranspiration rate flux. The stress index is defined as the ratio  
 219 between actual ( $ET_a$ ) and potential evaporation rates ( $ET_p$ ) [40].

$$SI = \frac{ET_a}{ET_p} \quad (3)$$

220 If the actual evaporation value is close to the potential value, the stress index takes  
 221 values close to one that reflect unstressed conditions. However, if the actual evaporation  
 222 is low compared to its potential value, the stress index values may reach zero, which  
 223 represents fully stressed conditions.  $ET_a$  and  $ET_p$  are simulated from the energy balance  
 224 model SPARSE [48] using meteorological observations and remote sensing variables.

225 LST, viewing angle and emissivity data are retrieved from the daily 1-kilometer LST  
 226 product (MOD11A1) from MODIS VI products on board Terra and Aqua sensors. Quality  
 227 control (QC) assessment for LST and emissivity product, is also provided from the LST  
 228 product. For stress index retrieval, the Leaf Area Index (LAI) and albedo products are also  
 229 needed. We used 8-day of albedo series (MCD43A3) from MODIS at 500 m, resampled to a  
 230 spatial resolution of 1 km. Finally, the LAI is retrieved from the NDVI time series using an  
 231 empirical equation [57], defined in the following equation (eq 4):

$$LAI = -\frac{1}{k} \ln \frac{NDVI_{max} - NDVI_i}{NDVI_{max} - NDVI_{min}} \quad (4)$$

232 where k is an extinction factor, about 1.13,  $NDVI_{max} = 0.97$  and  $NDVI_{min} = 0.05$   
 233 corresponding to the NDVI value of a bare soil [46]. The region is characterized by tree-  
 234 dominant cover with interrow distance of 12 m corresponding to low LAI values (about 0.3  
 235 to 0.4) on an image [58].

### 236 3.1.4. UPI

237 We used the Climate Hazards group Infrared Precipitation with Stations (CHIRPS)  
 238 data which are available from 1981 till now. This product incorporates three types of

239 information: global climatologies, satellite estimates and in situ observations [59]. CHIRPS  
240 is available daily and at a spatial resolution of about 6 Km.

241 The CHIRPS precipitation time series is used to compute the Uniformized Precipitation  
242 Index UPI which is constructed following the same steps as the SPI except the last step  
243 that transforms into Normal standard quantiles [7]. SPI maps precipitation amounts to  
244 the [0,1] interval by using either the empirical or a parametric cumulative function. Then,  
245 a final transformation is applied to obtain values that are normally distributed. UPI is  
246 computed with a single transformation: the empirical cumulative function is used to map  
247 precipitation amounts derived from historical records, to uniform quantiles in the [0,1]  
248 interval, as expressed in equation 5:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{x_i \leq x\}}, \quad (5)$$

249 where  $1_{\{\cdot\}}$  is the indicator function,  $x_1, \dots, x_n$  are rainfall amounts aggregated at  
250 different time scales (years, months or decades) and  $\hat{F}_n(x)$  is the UPI value with  $\hat{F}_n(x) =$   
251  $P(X \leq x) \in [0, 1]$ , with  $X$  the random variable representing the rainfall amounts at a given  
252 time scale. In contrast, SPI is computed as  $\Phi^{-1}(\hat{F}_n(x))$  where  $\Phi^{-1}$  is the quantile function  
253 (the inverse of the cumulative function) of the standard normal distribution.

254 We believe the interpretation in terms of uniform quantiles which are simply probabil-  
255 ities of non-exceedance, is more straightforward than normal quantiles. For example, if  
256 UPI equals 0.1 on a given year, it means that 10% of the yearly precipitation amounts are  
257 inferior or equal to the amount observed that year. In other words, that year is among the  
258 10% driest years. The SPI value corresponding to 0.1 is approximately -1.28.

### 259 3.2. Thermal stress index derived from energy balance model

#### 260 3.2.1. SPARSE model

261 We rely on the dual source energy balance model SPARSE [48] to simulate evapotran-  
262 spiration. It estimates separately evaporation (E) and transpiration (T). This is particularly  
263 relevant for arid and semi-arid areas which are characterized by sparse crop canopy and  
264 by an uneven relative contribution of evaporation and transpiration [60].

265 The model can be run in two modes: retrieval and prescribed mode. In retrieval  
266 mode, the respective stress levels (between non evaporating/transpiring and fully evapo-  
267 rating/transpiring i.e., potential rates) are two unknowns which are determined from the  
268 single piece of information provided by  $T_{surf}$  by assuming at first that the vegetation is  
269 not stressed, that allows a simplification to solve the underdetermination problem.  $T_{surf}$  is  
270 used to estimate the latent heat component from the soil ( $LEs$ ), corresponding to the soil  
271 evaporation E. If the vegetation is suffering from water stress, the resulting  $LEs$  will de-  
272 crease to unrealistic levels (negative values). In that case, we assume that the soil surface is  
273 stressed and  $LEs$  is set to a minimum value close to zero. Then the energy budget equation  
274 is solved for the vegetation component of the latent heat flux ( $LEv$ ), which corresponds  
275 to the transpiration T. If  $LEv$  is also negative, fully stressed conditions are imposed for  
276 both soil and surface components [58]. The prescribed mode provides an estimate of the  
277 potential latent flux for the soil and the vegetation ( $LEs_{pot}$  and  $LEv_{pot}$  respectively). Water  
278 stress indice SI is then defined from the actual and unstressed evapotranspiration rates  
279 (potential) at the time of the satellite overpass.

SPARSE model is forced by a series of meteorological observations (air temperature, relative humidity, global radiation and wind speed) and remote sensing variables (NDVI, LAI, albedo and LST). Outputs are derived at meteorological time steps (half-hourly).

### 3.2.2. Meteorological forcing

It is important to consider a long meteorological series in order to perform a robust statistical characterization of drought periods. In addition, owing to the temporal and spatial scales of climate variability, evapotranspiration and water stress index must be monitored at high temporal (subdaily to daily scales) and spatial resolution. However, available gauged stations (see section 2) are scarce with short observation periods and numerous gaps, insufficient to perform a robust statistical analyses and characterization of drought periods. We rely therefore on surrogate meteorological series that extend in time the original series in order to be used to constrain the SPARSE energy balance model. In this work, we used two types of surrogate meteorological data:

1. Unprocessed reanalysis data ERA5 extracted at the grid cell closest to the region of interest (see figure 1): ERA5 reanalyses are available at 31 km of spatial resolution [61] from 1950 till now at hourly temporal scale [61]. Reanalysis series that correspond to the four meteorological variables required for the energy balance model to simulate the corresponding index  $SI_{ERA5}$  are: the incoming global solar radiation at the surface (bottom of atmosphere), wind speed at 10 m, air temperature at 2 m and the relative humidity that was derived from 2 m air temperature and 2 m dewpoint temperature ERA5 products, according to the procedures defined in [62]. The specific aim of using unprocessed reanalysis data for our study, is to assess its performance to constrain an energy balance model for regions with no gauged stations.
2. Simulated series from a Stochastic Weather Generator (SWG) called "MetGen" [47]: Its implementation is publicly and freely available as an R library. MetGen generates scenarios of meteorological variables at sub-daily temporal resolution in order to extend local observations in the past. It relies on low resolution ERA5 reanalysis data and exploits observations provided by three gauged stations located in our study region (see figure 1) to simulate regional climatic information. The corresponding index simulated using the SWG meteorological to constrain SPARSE model, is denoted  $SI_{SWG}$ .

Inter-comparison between indices will be computed at daily time scale for a long period in the past starting from 2000 to 2019. Characteristics of indices used are as follows, see Table 1.

### 3.3. Indices standardization

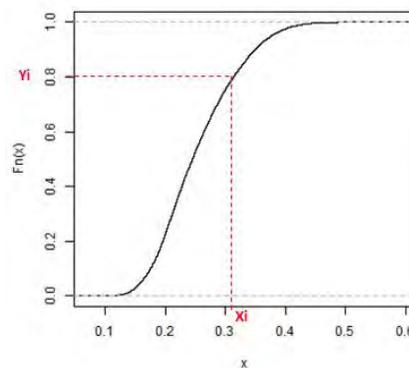
For added interpretability, the different indices are usually re-scaled or standardized in order (1) to be expressed as frequencies that can be easily interpreted [63] and (2) to characterize statistically the deviation (anomaly) of the indices values according to a long series recorded over the study period. For this aim, we standardize each index  $X$  with its empirical cumulative density function (ECDF). Let  $(x_1, \dots, x_n)$  be the values taken by the index  $X$  for a given data set. Then the ECDF is given by :

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{x_i \leq x\}}, \quad (6)$$

**Table 1:** Characteristics of indices used.

wave lengths	Visible/Near Infrared	Infrared + Visible + Meteo.		Microwave	Meteo. + Satellites data
Indices	NDVI	$SI_{SWG}$	$SI_{ERA5}$	SWI	UPI
Satellites	MODIS	MODIS	MODIS	ASCAT	CHIRPS
Model used	$\times$	SPARSE	SPARSE	$\times$	$\times$
Spatial resolution	1 Km	1 Km	1 Km	12.5 Km	6 Km
Temporal resolution	daily	daily	daily	daily	daily
Temporal availability	since 2000	since 2000	since 2000	since 2007	since 1981

321 where  $1_{\{x\}}$  is the indicator function,  $X_i$  is the indices value for a time step  $i$  and  $Y_i$  is the  
 322 standardized value of the index  $x$ . The ECDF standardizes by mapping a value  $x$  taken by  
 323  $X$  to its empirical non exceedance frequency, i.e., the percentage of occurrences of values  
 324 lesser than or equal to  $x$  [64]. This is illustrated in Fig. 2 : the ECDF maps the values  
 325  $x \in [0.1, 0.6]$  to  $[0, 1]$ . In particular, the value  $x_i = 0.31$  is mapped to  $y_i = 0.8$  which means  
 326 that 80 % of the values of the index  $X$  are expected to be below 0.31.

**Figure 2.** Empirical cumulative density function illustration.

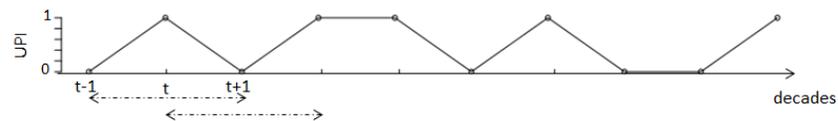
327 The advantages of the proposed standardization approach with the ECDF are that 1)  
 328 it does not rely on the assumption that the raw indices values follow a normal distribution  
 329 which is implicit in standardization approaches that rely on centering and scaling and 2) it  
 330 allows for flexibility in the choice of reference sample (depending on seasons for example),  
 331 and therefore can be used for short scales monitoring and 3) produces results that are  
 332 easily interpreted in terms of non-exceedance probabilities, as for UPI, especially when  
 333 simultaneous variables are analyzed [64]. Standardization using the empirical cumulative  
 334 distribution function is performed over the whole period (20 years). At annual scale,  
 335 standardization is performed over each year according to available values observed during  
 336 the whole study period. At decadal scale, standardization is performed according to season  
 337 (winter, autumn, summer and spring).

### 3.4. Evaluation of the different drought indices performance

#### 3.4.1. At regional scale

For comparison at a regional scale, all indices are aggregated at very low spatial resolution (about 12 kilometers). For this comparison, the CHIRPS precipitation time series is used as a reference to identify water stress conditions at different time scales: annual and decadal scales. In order to assess robustness and consistency between indices, we used for instance time series for visual interpretation and Kendall's  $\tau$  which is a rank-based correlation coefficient that measures the strength of the relationship between two indices.

To focus on consistency between precipitation index and other indices, we assess the ability of the different indices to identify correctly the stress periods defined according to UPI. We define stress periods based on UPI as follows. We assume that a drought event is a stress period with probability of occurrence of approximately 20% and that it lasts at least two decades. The threshold of 20% in terms of UPI corresponds to a threshold of about -0.84 in terms of SPI. The drought classes defined in [7] begin when SPI values fall below 0 (this corresponds to 50% in terms of UPI). In our study region, the threshold of 20% corresponds to precipitation amounts inferior to 200 mm, which is approximately the precipitation amount average at a growth season. We transform the decadal standardized indices into Boolean values according to the stress threshold: 0 to denote unstressed situation ( $> 0.2$ ) and 1 for stress conditions ( $< 0.2$ ). The second step is to check for each time step ( $t$ ) how many stress periods are recorded within the period between  $(t - 1)$  and  $(t + 1)$ , one decade being considered as a buffer (see figure 3).



**Figure 3.** Identification of stress periods using UPI index transformed into Boolean values.

This method allows us to take into account the memory effects for three decades. Moreover, the use of a moving average of UPI allows to take further consideration of memory effects and delayed responses of the different indices. We consider therefore, the use of a UPI moving average lagged up to 2, 3 and 4 decades. If the number of stress periods based on UPI or lagged moving averages UPI exceeds two of the three decades considered, we assign a drought condition for this time step. Then we assess the response stress/unstress of the other drought indices ( $SI_{SWG}$ ,  $SI_{ERA5}$ , SWI and NDVI).

#### 3.4.2. At local scale

For comparison at finer spatial resolution, a kilometric scale, we use local precise informations based on:

- **XLAS in-situ measurements** Sensible heat flux measurements using an extra-large aperture scintillometer (XLAS) are provided as part of the work of [58], for the period ranging between March 2013 and June 2015. The scintillometer (XLAS) was installed close to the Ben Salem village over a 4 km transect above mixed vegetation canopy: trees (mainly olive orchards) with some annual crops (cereals and market gardening) [58]. For our analyses, pixels enclosed in the mean XLAS are selected in order to

375 compare the different drought indices with the stress index derived from the sensible  
376 heat flux measurements, denoted  $SI_{XLAS}$ .

377 • **Historical rainfed areas selection** Rainfed crops are more sensitive to rainfall depletion  
378 and thus to drought. For our analyses, we identify historical rainfed wheat areas  
379 relying on non-irrigated cereal mask (see figure 4a), computed for the agricultural  
380 year 2011-2012, as part of the work computed by [65]. It is computed using an object-  
381 oriented classification technique basing on Spot image of 31 March 2012. We generate  
382 the percentage of non-irrigated cereal fields for this year, over each MODIS pixels,  
383 (see fig. 4b). Then, we select pixels that contain more than 40% of rainfed cereal cover.  
384 Rainfed cereal pixel selected are used as reference to locate non-irrigated cereal fields  
385 in precedent years, in order to assess the response of the different indices over a dry  
386 and a wet year.

387 In order to facilitate interpretation, we define four stress classes that characterize the  
388 severity of water stress identified using the different drought indices. The four stress  
389 classes are listed in table 2, "High stress", "Stress", "Moderate stress" and "No stress"  
390 according to the values of a given index.

**Table 2:** Drought indices classification.

Index values	Drought class
0-0.15	High stress
0.15-0.3	Stress
0.3-0.7	Moderate stress
0.7-1	No stress

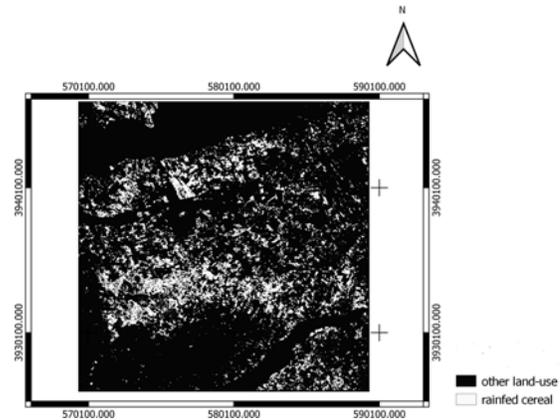
## 391 4. Results

### 392 4.1. Drought Indices inter-comparison at regional scale

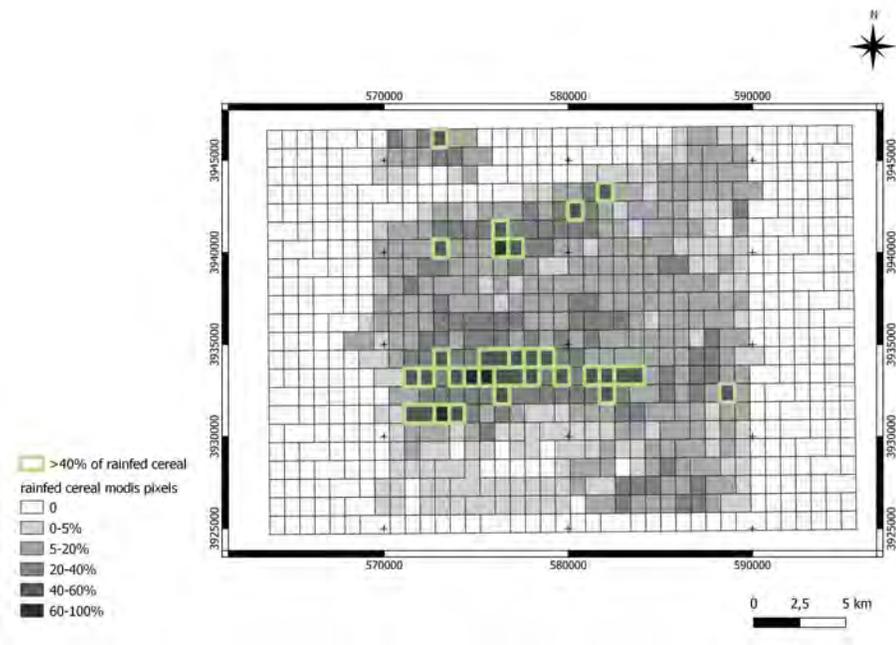
#### 393 4.1.1. Annual scale

394 Figure 5 presents the cumulative rainfall during the growth season (September to  
395 May) with the corresponding NDVI variation for each year (fig 5a), in order to characterize  
396 visually the inter-seasonal variation over the study period. Rainfall data are provided from  
397 CHIRPS precipitation time series. The red line in Fig. 5a corresponds to 200 mm/growth  
398 season of precipitation amount. We assume that a dry year presents a cumulative rainfall  
399 below 200 mm during a growth season. The figure 5b shows the time series of different  
400 standardized drought indices: the TIR SI simulated when the SPARSE model is constrained  
401 by the SWG or ERA5 meteorological data  $SI_{SWG}$  and  $SI_{ERA5}$ , the standardized NDVI index,  
402 the UPI and the standardized SWI which is available since 2007. The red line in Fig. 5b  
403 indicates the drought threshold, fixed in this study at 0.2 which corresponds to a drought  
404 event with an approximately 20% of non exceedance frequency. The stress threshold is  
405 necessary to discuss drought events, as well as their onset, duration and intensity. The  
406 farther below the red line a standardized index is, the worse the drought is.

407 Overall, we see that annual indices show a similar variation.  $SI_{SWG}$  and  $SI_{ERA5}$   
408 show similar inter-annual variability. Moreover, we observe that they reproduce UPI  
409 variability and succeed to identify wet and dry conditions simultaneously. Dry periods  
410 which correspond to abnormally dry situations that fall below the drought threshold, are  
411 well depicted also by a cumulative rainfall less than 200 mm during each growing season,  
412 and a low NDVI level. However, we observe in some instances that UPI is not very well



(a) Rainfed cereal mask, produced for the 2011-2012 agricultural season [65]



(b) Percentages of rainfed cereal in each Modis pixels of the 2011-2012 agricultural season

Figure 4. Identification of rainfed cereals in MODIS pixels over the 2011-2012 agricultural season.

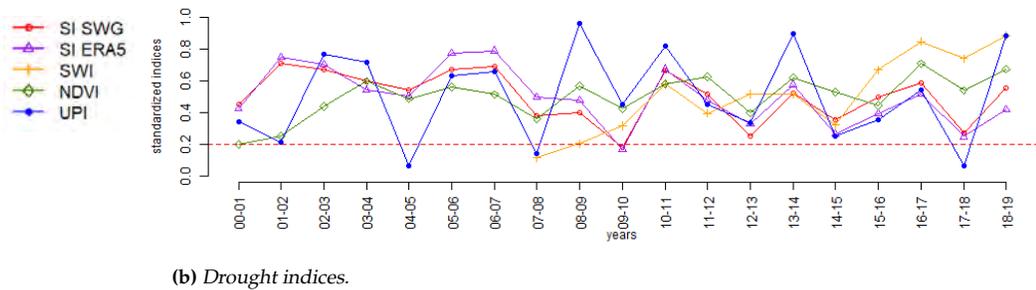
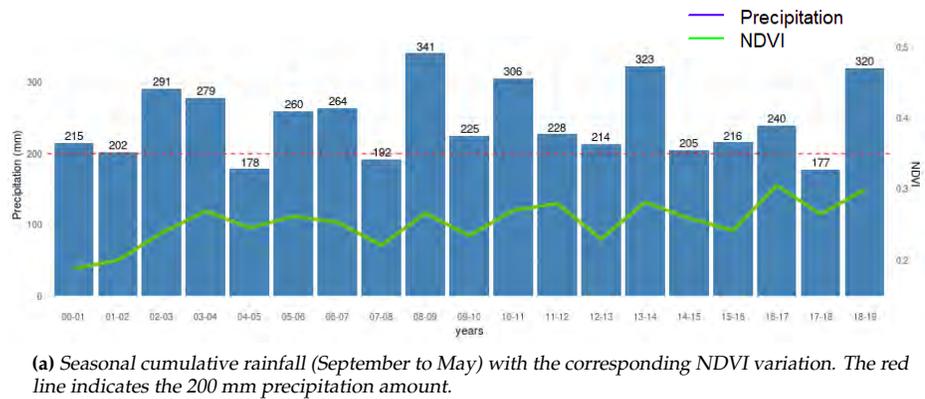


Figure 5. Time-series of seasonal variation of standardized drought indices cumulated in growth season.

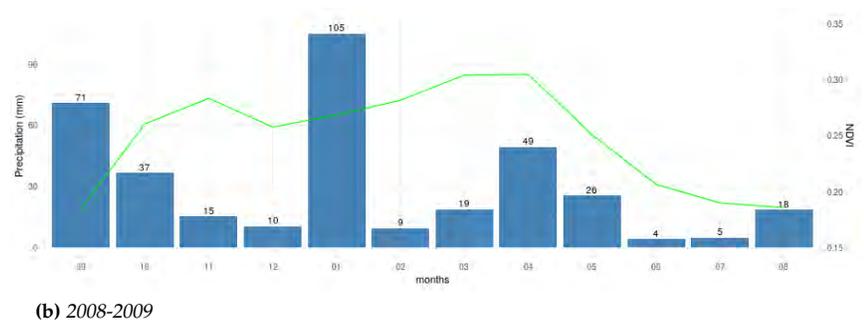


Figure 6. Precipitation and NDVI monthly variation over agronomic years (2004-2005) and (2008-2009).

413 correlated with other standardized indices. This is the case for 2004-2005 or 2008-2009  
414 where we observe that UPI is either above the unusual unstressed condition or below the  
415 drought threshold while other indices present similar stress level. To better explore the  
416 allocation of precipitation intensity along these years, we present in figure 6, the monthly  
417 cumulative precipitation and monthly NDVI variation over these agronomic years. There  
418 are in figure 6a, low cumulative rainfalls for several months. However, almost all the  
419 largest cumulative rainfall amounts are observed from September till February. Rainfall  
420 during this period was sufficient and coincide with the beginning of the growth season.  
421 For this reason, we observe a high level of NDVI in March and May that reflect a good  
422 development of vegetation, particularly cereal crops. Consequently, unstressed conditions  
423 were also shown by the rest of the standardized indices. Besides, we observe abnormally  
424 unstressed situation depicted by UPI during the agronomic year 2008-2009. However, the  
425 other standardized indices indicate a medium stress condition. Indeed, we observe very  
426 low rainfall amounts during the whole year, especially in November and December that  
427 correspond to seeding and early growth. Then, we observe an abnormally high rainfall  
428 amount that exceeds 100 mm/month in January. This high cumulative amount increases  
429 the annual cumulative rainfall but it is late for vegetation growth. Indeed, an adequate  
430 response of the stress is always dependent on receiving sufficient rainfall at the appropriate  
431 moments, with sufficient amounts. In contrast, the agronomic year 2017-2018 shows very  
432 low cumulative seasonal rainfall. The TIR SI values succeed to detect the stress period,  
433 by presenting low values. However, standardized SWI and NDVI show a medium stress  
434 condition. SWI and NDVI didn't succeed to detect the drought intensity for this year.

435 To assess consistency between the various indices at annual scale, we use Kendall's  $\tau$ .  
436 Kendall's  $\tau$  is suitable for non-gaussian distributions, as opposed to the Pearson correlation  
437 coefficient [66]. Positive values indicate that both indices tend to increase or decrease  
438 simultaneously while negative values indicate that they tend to vary in an opposite manner.  
439 A value near zero signals a lack of dependence. In figure 7, we see that all indices present  
440 positive correlation coefficients. All indices present high Kendall's coefficients with UPI  
441 that exceed 0.35, particularly  $SI_{ERA5}$  which presents Kendall's  $\tau$  about 0.46. However,  
442 SWI presents very low correlation coefficient with UPI. This is could be explained by the  
443 different responses of these indices to a rainfall supply. UPI presents an instantaneous  
444 response to rainfall variation. In contrast, SWI which presents the root zone soil moisture,  
445 presents a delayed response to rainfall supply. On the other hand, we observe that SWI  
446 presents the highest  $\tau$  value with NDVI, which reflect a very strong seasonal correlation  
447 between them. Also,  $SI_{SWG}$  is highly correlated with other indices. It presents a Kendall's  
448  $\tau$  about 0.58 with NDVI and with SWI and the highest correlation with  $SI_{ERA5}$ , with a  
449 Kendall's  $\tau$  about 0.64. However,  $SI_{ERA5}$  presents low correlation coefficients with NDVI  
450 and SWI.

#### 451 4.1.2. Decadal scale

452 We focus on a finer temporal scale, a decadal scale, in order to assess short term  
453 variations. Indeed, a short time scale (monthly or decadal) is more appropriate to provide  
454 information on the growing season development over a few months and therefore on the  
455 water supply, unlike the annual time scale [6].

456 In figure 8, we propose a comparison analysis carried out at a decadal scale of the  
457 cereal growing season (from September to May) selected from some dry and wet years.

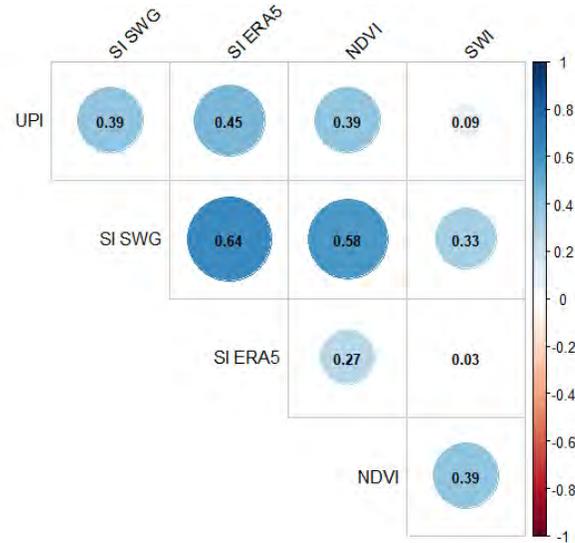


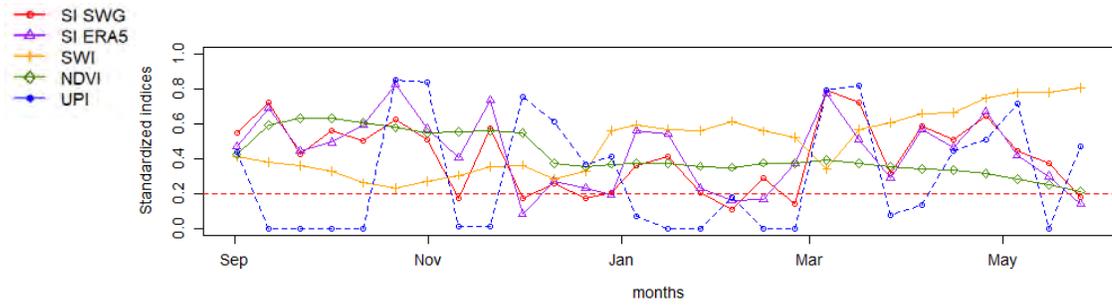
Figure 7. Seasonal Kendall's  $\tau$  correlation between drought indices.

458 We used the values of precipitation index UPI lagged up to five decades rather than the  
 459 instantaneous value, in order to be more correlated with the aggregated response of the  
 460 different indices. The agronomic year 2012-2013 (see figure 8a) is considered as a dry  
 461 year, based on the cumulative rainfall over the corresponding growth season (fig 5a). We  
 462 observe that the standardized NDVI index presents moderate drought conditions (50%  
 463 of probability of occurrence) at the beginning. Then, it decreases in December to 0.4  
 464 and then, it reaches 0.2 in the end of the growth season. However, the standardized TIR  
 465 indices show mainly similar variations, more correlated with lagged UPI variations. UPI  
 466 shows considerable fluctuations along the whole period. We observe very low values  
 467 during several decades in September and October and in January till March, that lead to  
 468 an important drops of  $SI_{SWG}$ ,  $SI_{ERA5}$  and an extended low values of SWI which fluctuate  
 469 around 0.2 in the beginning of the growth season.

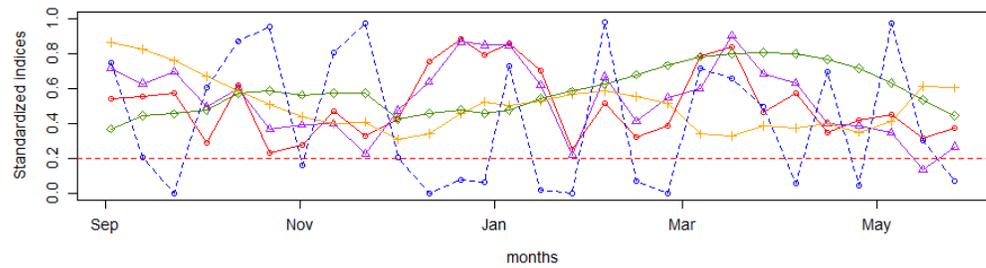
470 2013-2014 represents a typical wet growth season. In figure 8b, we observe that the  
 471 different drought indices are above the drought threshold over the whole period. The  
 472 standardized NDVI index present high values in spring that reflect a good vegetation  
 473 health and a maximum crop development during this period. High NDVI values are  
 474 observed after an important water supply provided during the growing season. UPI  
 475 presents abnormally high values that reach 100% of probability of non exceedance for  
 476 several decades.  $SI_{SWG}$ ,  $SI_{ERA5}$  succeed to reproduce the main fluctuation of the lagged  
 477 UPI: relevant increases in UPI values lead to an important elevation of both the TIR indices  
 478 and a relevant drop of UPI which falls below the stress threshold mainly in January and  
 479 February and induces an important rapid decline of both TIR SI values. SWI presents  
 480 however a more delayed responses to UPI variations. In fact, we observe a depression  
 481 of SWI in autumn and spring due to a successive drop fluctuations of UPI during these  
 482 periods.

483 In figure 8c, we present another dry growth season. The lack of precipitation observed  
 484 mainly at the beginning of the growth season induces a decrease of the standardized NDVI  
 485 and SWI indices at the end of the growing season. The rainfall shortage during this season  
 486 triggers a relevant depletion of the TIR indices especially in winter where we observe

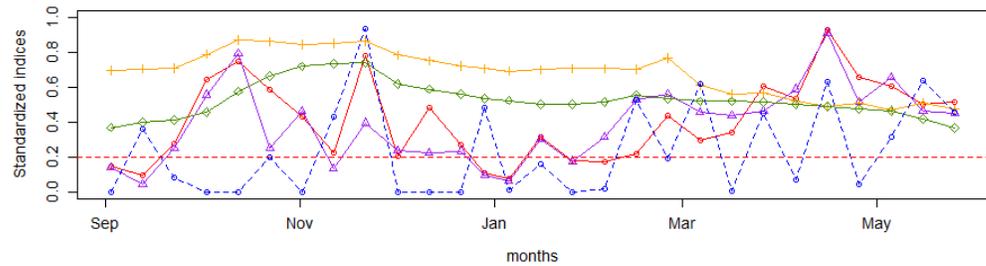
487  $SI_{SWG}$ ,  $SI_{ERA5}$  and UPI are below the stress threshold. However, each lagged UPI value  
 488 increases lead to an elevation on TIR indices.



(a) 2012-2013



(b) 2013-2014



(c) 2017-2018

**Figure 8.** Decadal times series of standardized stress indices derived from different wavelenghts over the growing season of two dry years (2012-2013) and (2017-2018), and a wet year (2013-2014).

489 A further analysis is carried out to assess more closely the consistency between the  
 490 different drought indices in identifying stress periods at the decadal timescale according  
 491 to UPI. As explained in section 3.4.1, we define stress periods according to the different  
 492 indices. Besides, we identify three cases:(1) the stress indices reproduce correctly the  
 493 drought condition identified with UPI (stress period or not stress period), flagged as  
 494 "Congruent with UPI", (2) the stress indices identify a stress period not confirmed by UPI,  
 495 called a "Not confirmed by UPI", and the last case, (3) the stress indices don't succeed to  
 496 identify a stress period detected by the UPI, called a "missed stress". The first case estimates  
 497 the consistency between the different drought indices and UPI to identify the same stress

498 conditions. Analyses are performed according to seasons, in order to assess drought indices  
499 responses through different seasons.

500 In figure 9, we present the probability to identify these three different cases, according  
501 to seasons and UPI values in figure 9a and lagged moving average UPI to 2, 3 and 4 decades  
502 respectively for figure 9b, 9c and 9d. Overall, we observe that the different indices show a  
503 good performance to identify the same stress condition in autumn, winter and especially  
504 in spring. However, in summer, the probability to detect the same stress conditions using  
505 the different stress indices is decreased at the expense of missed stress. Indeed, in summer  
506 the probability of rainfall events is extremely low in our study region. Therefore, drought  
507 periods identified by UPI increase, but they are not necessarily identified by other drought  
508 indices due to the supply of water provided from irrigation as an alternative water supply  
509 for this period. The NDVI shows high probability to detect stress status congruent with  
510 UPI, especially in spring where we have the maximum of vegetation development. The  
511 case of "congruent with UPI" could be related to a stress or an unstressed period. In this  
512 case, the high level of the "congruent with UPI" performed by NDVI is mainly related to a  
513 no stress period over the rainy season and particularly in spring.

514 Besides, we observe that  $SI_{SWG}$  and  $SI_{ERA5}$  show a good performance to detect  
515 correctly the stress conditions determined from the UPI and it present low probability to  
516 identify stress periods not confirmed by UPI especially using  $SI_{SWG}$ . This performance  
517 is maintained during the different season and even enhanced with the comparison with  
518 the lagged moving average of the UPI. On the other hand, we observe that, SWI shows  
519 better performance in winter and spring. Indeed, during this season, vegetation is well  
520 developed with deeper root zones, which could explain the high performance of SWI to  
521 identify a stress period or unstressed period according to UPI. As well, this performance is  
522 improved incrementally when we increase the lagged moving average value of UPI. The  
523 application of a moving average of 4 decades seems the best delayed period for adequate  
524 responses for all drought indices. Beyond this value too many memory effects are lost.

## 525 4.2. Drought Indices inter-comparison at local scale

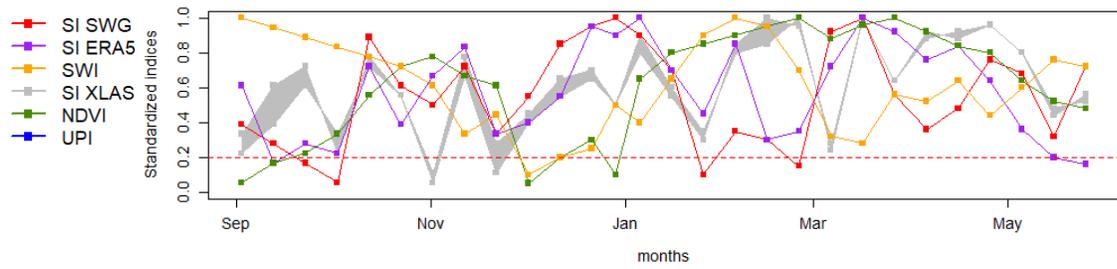
### 526 4.2.1. evaluation with XLAS in-situ measurements

527 In figure 10, we present decadal time series of the different standardized indices during  
528 two growing season where XLAS measurements are available at both local (figure 10a,  
529 10c) and regional scale (figure 10b, 10d). As mentioned above, we used for this comparison  
530 the lagged UPI (up at 5 decades) and  $SI_{XLAS}$  respectively as regional and local references.  
531 The aim of the comparison of indices at a local and a regional scale is to assess if an index  
532 found robust at low resolution (12 Km) maintains its performance at higher resolution  
533 (kilometric). For interpretation, we consider two groups of indices according to their spatial  
534 resolution: A group with low resolution (kilometric) ( $SI_{SWG}$ ,  $SI_{ERA5}$  and NDVI), and a  
535 second group with very low spatial resolution, dozen of kilometer (SWI and UPI).

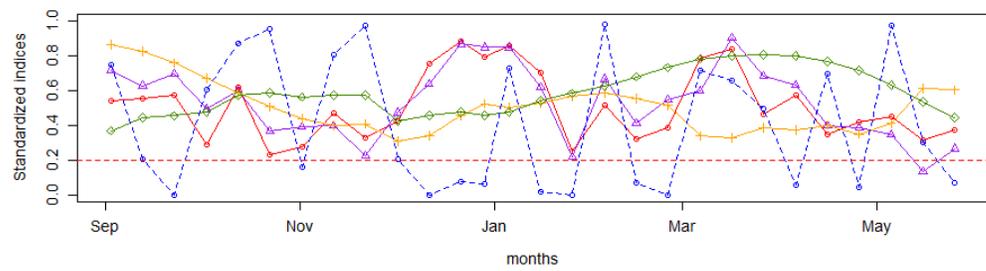
536 In figure 10a and 10b, we present a wet year (2013-2014). We see that indices  $SI_{SWG}$   
537 and  $SI_{ERA5}$  succeed to better reproduce  $SI_{XLAS}$  fluctuations in figure 10a. In addition, we  
538 observe that these indices, especially  $SI_{SWG}$ , follows more closely the  $SI_{XLAS}$  variations  
539 particularly during the rainfall season (from October to February) and the last decades of  
540 this growth season. Besides, we observe that stress periods defined by the  $SI_{XLAS}$  (under  
541 the threshold line), observed in October, December and February, are mostly isolated  
542 with  $SI_{SWG}$ . Moreover,  $SI_{SWG}$  succeed to identify stress conditions defined by  $SI_{XLAS}$



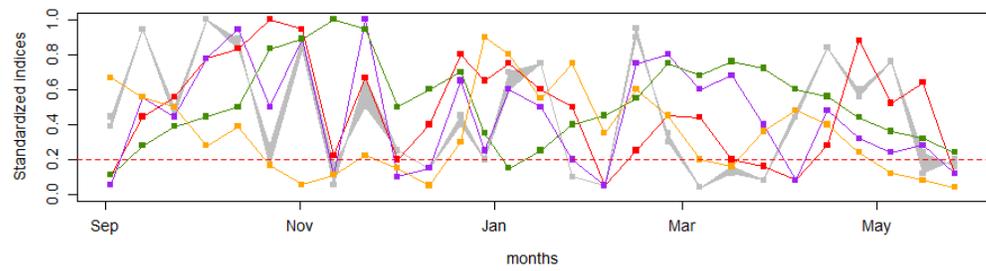
**Figure 9.** Probability of drought detection based on UPI and lagged UPI during the period between 2007 and 2019.



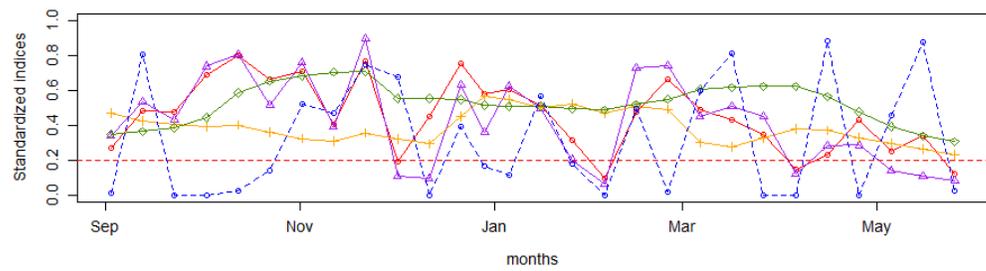
(a) Local scale (2013-2014)



(b) Regional scale (2013-2014)



(c) Local scale (2014-2015)



(d) Regional scale (2014-2015)

**Figure 10.** Decadal times series of standardized stress indices derived from different wave-lengths over the growing season of two years (2013-2014) and (2014-2015) at a local scale 10a, 10c, in comparison with  $SI_{XLAS}$  derived from XLAS measurements over pixels containing the scintillometer transect and at a regional scale 10b, 10d.

543 even using its aggregated values at the regional scale (see figure 10b). The NDVI, in spite  
544 of being defined among higher resolution indices, didn't succeed to reproduce  $SI_{XLAS}$   
545 decadal variations. However, it brings as expected, an overall delayed information about  
546 stress conditions. We observe for example in figure 10a, which presents a wet year, an  
547 elevation in NDVI values in winter and spring that correspond also to several elevated  
548 values observed by  $SI_{XLAS}$ . On the other hand, we observe a different behavior of the  
549 standardized SWI according to the different spatial scale. At a local scale (figure 10a),  
550 SWI didn't succeed to reproduce correctly  $SI_{XLAS}$  variations. However, at a regional  
551 scale (figure 10b), SWI succeed to identify dry or wet conditions, according to the lagged  
552 UPI fluctuations. Dry conditions were observed at the beginning of the growth season and  
553 February. Wet conditions were observed in November-December and the end of the growth  
554 season.

555 2014-2015 presents a relatively dry year. We observe that NDVI presents mainly  
556 moderate stress values in figure 10d. In figure 10c, we observe that TIR stress indices  
557 succeed to better reproduce  $SI_{XLAS}$  in this growth season, particularly using  $SI_{SWG}$ . Indeed,  
558  $SI_{SWG}$  succeed to identify a water stress period caught by  $SI_{XLAS}$  which persists for several  
559 decades during March, as well as, an unstressed period observed later in April and May,  
560 visible in the same figure. At local scale (figure 10c), SWI shows low values observed below  
561 the stress threshold during several decades in Autumn and March, which amply testify an  
562 overall dry situation during this growth season. Even at a regional scale (figure 10d), SWI  
563 shows moderate stress values. However, SWI was not able to identify with accuracy stress  
564 conditions in comparison with in-situ measurements. This could be explained by its very  
565 low spatial resolution, inadequate for local stress detection.

#### 566 4.2.2. Drought detection in rainfed areas

567 In figure 11, we present stress class maps of the different indices at the seasonal  
568 scale (growth season) of the year 2011-2012: fig. 11a for the standardized thermal index  
569 constrained by meteorological SWG data, fig. 11b for the standardized thermal index  
570 constrained by meteorological ERA5 data, fig. 11c for standardized NDVI index, fig. 11d  
571 for standardized SWI index and fig. 11e for standardized precipitation index. In addition,  
572 we present the percentage of stress classes identified by each index using extracted rainfed  
573 cereal pixels as explained in section 3.4.2 (see figure 11f). Combined results presented  
574 in figure 11f show that mostly all indices depicted a moderate stress class for this year  
575 with the highest percentage. However, we observe that the TIR stress indices and SWI  
576 show an elevated percentages of Stress class and even a very low percentage of a high  
577 stress class for  $SI_{SWG}$  and  $SI_{ERA5}$ . This finding is in agreement with the hydrological year  
578 characteristics presented in figure 5a. Indeed, this year characterized by a cumulative  
579 seasonal rainfall about 228 mm didn't belong to very wet or very dry year. Stress classes  
580 depicted by thermal indices and SWI could be more precise than NDVI characterized by  
581 a delayed response and than the UPI which is less informative when aggregated at low  
582 temporal scale (seasonal scale).

583 Then, the same analysis was carried for a dry year 2009-2010. In figure 12, we observe  
584 that the rainfed cereal areas selected correspond mainly to high stress or stress classes.  
585 More than 80% of rainfed cereal pixels indicate a high stress according to  $SI_{SWG}$  and  $SI_{ERA5}$   
586 maps. Using the SWI information, we identify also a high stress with more than 50%.  
587 However, using the NDVI, we identify mainly stress class and moderate stress as well

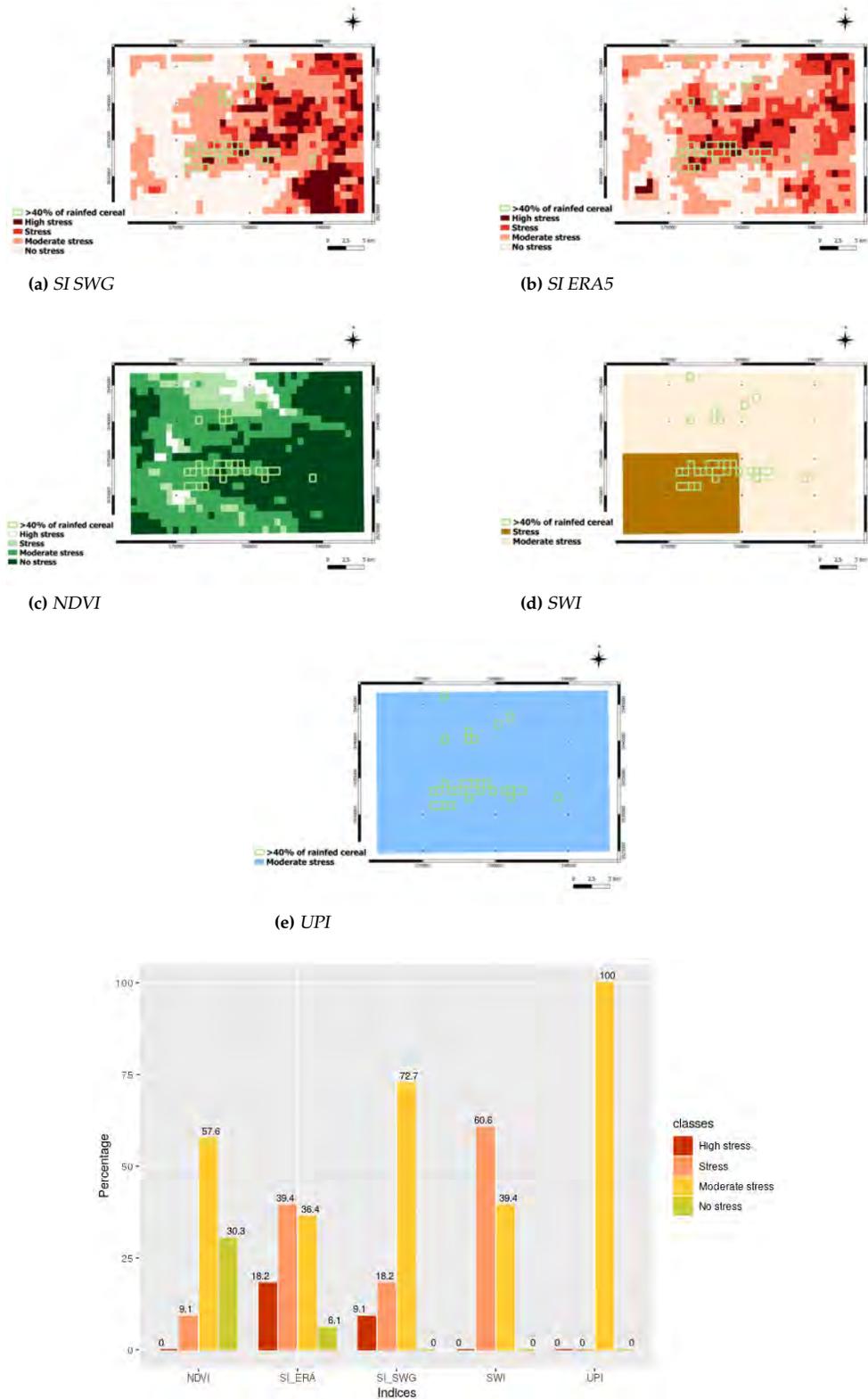
588 as the response of the UPI. Expected results are obtained mostly by the TIR indices and  
589 SWI. In fact, these indices are related directly with vegetation water requirement especially  
590 for the TIR that depends also on atmospheric demand. Precise information concerning  
591 vegetation and surface stress should be obtained from these indices. Due to its low spatial  
592 resolution, SWI might be less accurate for local analyses.

## 593 5. Discussion

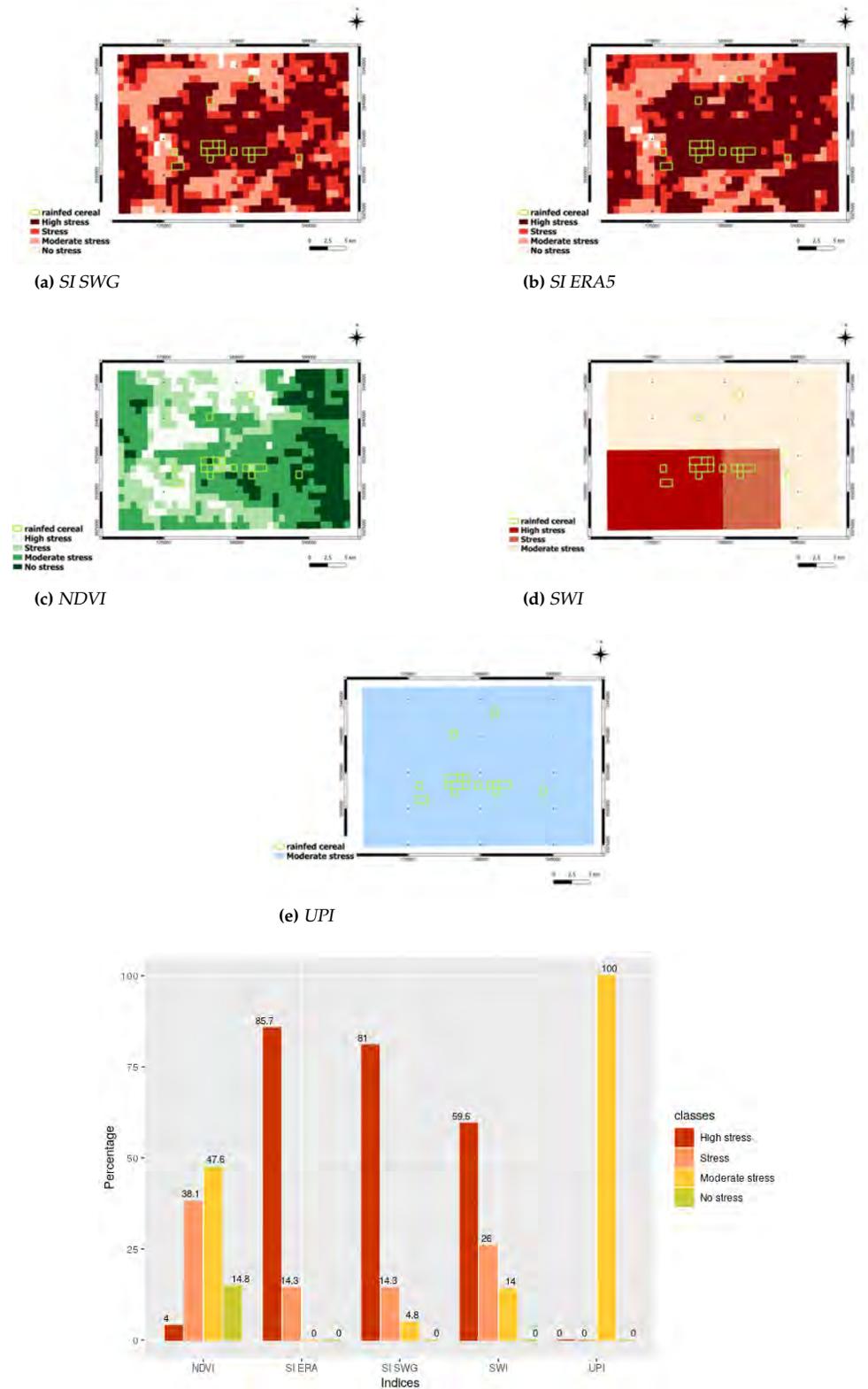
594 We focused on growing seasons to perform our analysis (from September to May).  
595 Indeed, droughts could be damaging for crops health in these periods of the year. Analyses  
596 are carried out at different spatial scales: regional and local scale. At the regional scale, UPI  
597 is considered as the reference index to characterize year-to-year water statement and to  
598 assess the response of the aggregated indices used at a large scale (12 km). Overall, we find  
599 that drought indices from the different wave lengths show mostly the same variations at  
600 the annual time scale. All the indices succeed to reproduce the UPI variations. We find  
601 also that, thermal stress indices provided from SPARSE model when constrained either  
602 from SWG or from ERA5 meteorological data,  $SI_{SWG}$  and  $SI_{ERA5}$  respectively, are very  
603 well correlated with UPI. However,  $SI_{ERA5}$  has a higher Kendall's  $\tau$  with UPI. Indeed,  
604  $SI_{ERA5}$  is simulated using a regional climatic information (ERA5 reanalyses) and local  
605 remote sensing information. The use of the large scale meteorological data to simulate  
606  $SI_{ERA5}$  could explain its high correlation with UPI provided also from gridded data-sets.  
607 Moreover, we observe a very strong seasonal correlation between  $SI_{SWG}$  and  $SI_{ERA5}$ . This  
608 correlation, measured by Kendall's coefficient, was conserved also at the decadal scale (see  
609 figure 8), due to their high spatial resolution. SWI and NDVI present also high correlation  
610 due to their aggregated response to a stress condition. Indeed, the NDVI presents a delayed  
611 response in comparison with the other indices, and shows less variability because of the  
612 time lag needed for vegetation to respond for a supply or a lack of water. On the other  
613 hand, SWI, due to the time needed for water supply to be infiltrated to the root zone soil,  
614 shows also a delayed response to water supply in comparison to the other indices.

615 We assess the consistency between drought indices and UPI as a reference stress index  
616 for regional analyses. We find that there is more consistency between UPI and  $SI_{SWG}$   
617 to identify the stress condition, especially in autumn, winter and spring. Moreover, this  
618 index presents the smallest probability to identify a stress period not confirmed by UPI.  
619 SWI shows more consistency with UPI in spring. Indeed, during this season, vegetation  
620 presents deeper root zones, which could explain the high performance of this index to  
621 identify a stress period or unstressed period according to UPI. Our analyses show also that,  
622 using a moving average of UPI improves the adequation of drought indices to the hydric  
623 status. In fact, all indices need a time lagged to respond for a water supply or shortage.  
624 This temporal lag depends obviously on the index used.

625 We can conclude therefore that TIR stress indices are more sensitive for a drought  
626 period. Indeed, a shortage of water supply will affect the soil (evaporation) and the vegeta-  
627 tion (root zone), inducing a stomatal closure and a surface temperature elevation detected  
628 by thermal sensors. However, SWI which carries information on the soil moisture in the  
629 root zone, didn't show immediate response as a result of a lack of precipitation. Indeed,  
630 the availability of some water storage in the profile soil after surface soil evaporation could  
631 delay the response of SWI to a water stress. For this aim, we need further analyses to  
632 identify the optimal parameter for the time infiltration (T) according to the characteristics



**Figure 11.** Stress classes for each drought index used over the agronomic year 2011-2012, 11a for the standardized thermal index constrained by meteorological SWG data, 11b for the standardized thermal index constrained by meteorological ERA5 data, 11c for standardized NDVI index 11d for standardized SWI index, 11e for standardized precipitation index and 11f for percentage of the stress class identified by each index.



**Figure 12.** Stress classes for each drought index used over the agronomic year 2009-2010, 12a for the standardized thermal index constrained by meteorological SWG data, 12b for the standardized thermal index constrained by meteorological ERA5 data, 12c for standardized NDVI index 12d for standardized SWI index, 12e for standardized precipitation index and 12f for percentage of the stress class identified by each index.

of our study region. Besides, satellite measurements to estimate soil moisture may not be sufficiently precise because of the low spatial resolution of the product used SSM (see section 3.1.2). However,  $SI_{ERA5}$  which is derived from very low resolution unprocessed ERA5 reanalyses, shows a good performance to identify water stress conditions since it is simulated at kilometric scale with precise satellite information. Analyses at the local scale strengthen our findings. Indeed, comparison of the different drought indices with the index derived from the XLAS measurements over the scintillometer transect show that the thermal stress indices are promising tools for drought identification at low spatial resolution scale (kilometric). In this case,  $SI_{SWG}$ , which incorporates information from the local meteorological station close to the XLAS transect, shows better performances than  $SI_{ERA5}$  and allows an accurate identification of the stress condition. Moreover, the identification of drought classes in non-irrigated cereal pixels shows therefore an ambiguous situations corresponding to  $SI_{ERA5}$ . This could be explained by the large spatial resolution of meteorological data (31 km [61]) used to constrain the SPARSE model in order to simulate  $SI_{ERA5}$ . For this task, although the SPARSE model allows retrieval of separate estimates of evaporation and transpiration, vegetation analyses seem to be insufficient to identify plant stress periods owing to the sparsity of the vegetation cover and the lack of consistent information about rainfed areas.

## 6. Conclusion

This study examined the consistency and the ability of different drought indices to identify stress periods. We rely on RS drought indicators provided from different wave lengths : the visible/near infrared, the thermal infrared and the microwave domains. These indicators provide information about vegetation health, water requirement and soil moisture that can help us to identify the plant hydric status. We use the SPARSE model, a dual source energy balance model [48], in order to retrieve estimates of evapotranspiration and water stress indices from the thermal infrared domain. SPARSE relies on satellite information and meteorological observation series to characterize vegetation cover and atmospheric demand. As far as we know, not many studies of drought indices comparison have been proposed in the literature. Our study show that, NDVI is very informative on the year-to-year water stress conditions. However, in spite of its low resolution (kilometric), its delayed response to water stress forms a major disadvantage. Thermal indices and SWI show consistent information at annual scale. Nevertheless,  $SI_{SWG}$  seems to be more precise for relevant water stress identification especially at a local scale. Indeed, SWI is not sufficiently reliable to identify with accuracy stress intensity at finer time scale due to it is large spatial scale. On the other hand, this work highlights the performance of large scale meteorological variables (reanalyses in our case) to identify periods of droughts. We proposed an efficient alternative when local meteorological data are not available particularly at regional analyses. Perspectives for this work include the simulation of a drought index simulated from transpiration only in order to focus on agronomic droughts. Indeed, information about vegetation water requirement rather than information provided from the "vegetation + soil" composite could be an efficient tool for drought management in semi-arid regions characterized by crops disparity. Moreover, adding mesoscale numerical weather prediction could provide efficient drought prediction. Lastly, the SWG-SPARSE tool will be tested for similar applications but in different climatic conditions (e.g. coastal area).

678 **Author Contributions:** writing original draft preparation, N.F.; data processing, N.F. and M.L.P.;  
 679 Software, N.F., M.L.P. and J.C.; data analysis, N.F. and G.B.; results interpretation N.F., G.B., Z.L.C.  
 680 and J.C.; Visualisation, G.B., and M.L.P.; writing review and editing, G.B., M.L.P. and J.C.; ideas and  
 681 discussions, J.C., M.L.P. and G.B.; data curation, Z.L.C., Z.K. and G.B.; funding acquisition, J.C., Z.L.C.  
 682 and G.B.; Conceptualisation, N.F. and G.B.; methodology, G.B. and N.F.; supervision, G.B.; validation,  
 683 G.B. All authors have read and agreed to the published version of the manuscript.

684 **Funding:** This work has been supported by the NAILA International Laboratory through a PhD  
 685 grant, as well as the PHC Utique French-Tunisian bilateral "Amande" project, the H2020 RISE mobility  
 686 "ACCWA" project, the WaterWorks2017 "FLUXMED", PRIMA 2018 "ALTOS" and ERANETMED  
 687 "CHAAMS" ERA-NET co-fund projects.

688 **Acknowledgments:** We extend our warm thanks to all the technical teams of the IRD, GREEN-TEAM,  
 689 CESBIO and INAT for their strong collaboration and support.

690 **Conflicts of Interest:** The authors declare no conflict of interest.

## 691 Appendix G. Post-processing of instantaneous evapotranspiration estimates

692 Evapotranspiration is simulated at the satellite overpass time and only during days  
 693 with clear sky. Some post-processing must be performed to reconstruct seasonal evapo-  
 694 transpiration useful for hydrological studies and particularly in agricultural applications.  
 695 The post-processing involves three steps: fusion, extrapolation and interpolation, that are  
 696 described below.

### 697 Appendix G.1. Extrapolation

698 ET derived from TIR RS data rely on once-a-day acquisitions. An extrapolation  
 699 algorithm is used to reconstruct its sub-daily variations using a method based on the  
 700 Evaporative Fraction (EF), well described in [67]. EF at satellite overpass  $t$  is defined as the  
 701 ratio between the Latent heat flux (LE) and the Available Energy (AE), which is derived  
 702 from the net radiation (Rn) and soil heat flux (G), as defined in the following equation:

$$EF_{obs} = \frac{LE}{AE_t} = \frac{LE}{Rn - G} \quad (A7)$$

703 According to [68], EF is relatively stable during the daylight hours on clear days. This  
 704 hypothesis allows to reconstruct the diurnal behavior (hourly or half-hourly time scale),  
 705 see equation A8, based on the empirical equation of EF defined by [69] (denoted  $EF_{sim}$ ),  
 706 see equation A9, and observed EF at satellite overpass (equation A7).

$$EF_{(30min)} = EF_{sim(30min)} * \frac{EF_{obs}}{EF_{sim}} \quad (A8)$$

707 where,

$$EF_{sim} = [1.2 - (0.4 * \frac{GR_t}{1000} + 0.5 * \frac{RH_t}{100})] \quad (A9)$$

708 GR and RH are respectively the global radiation and relative humidity for a time scale  
 709  $t$  (semi-hourly or satellite overpass).

710 Finally, daily evaporation ( $E_d$ ) can be simply obtained using the equation A10:

$$E_d = EF * AE_d, \quad (A10)$$

711 where  $AE_d$  is computed from an equation suggested by [70]: in where AE is assumed  
712 to present the same diurnal variation as the global radiation, see A11.

$$AE_d = Rg_d * \frac{AE_t}{Rg_t}, \quad (A11)$$

### 713 *Appendix G.2. TERRA ET and AQUA ET merge*

714 ET simulations are derived from TERRA and AQUA satellites. We choose to combine  
715 these two sources of data in order to increase the overall temporal availability. Results  
716 presented by [58] show that ET derived from these different sensors are potentially biased  
717 because of their different overpass time (10:30 am and 13:30 am for Terra and Aqua respec-  
718 tively) and that TERRA simulations are more correlated with ET derived from measured  
719 sensible heat flux (H) from an extra large aperture scintillometer (XLAS). Therefore, in  
720 order to take into consideration systematic differences in distributional properties, we  
721 perform a bias correction approach for the AQUA simulations (available since 2002) before  
722 merging. We considered CDF-t [71], a univariate bias correction method which allows  
723 non-linear corrections to reproduce the statistical distributions of the reference time series.  
724 Our aim is to keep ET simulations derived from TERRA RS (available since 2000) and to  
725 add bias corrected ET derived from AQUA only in days with no TERRA acquisition.

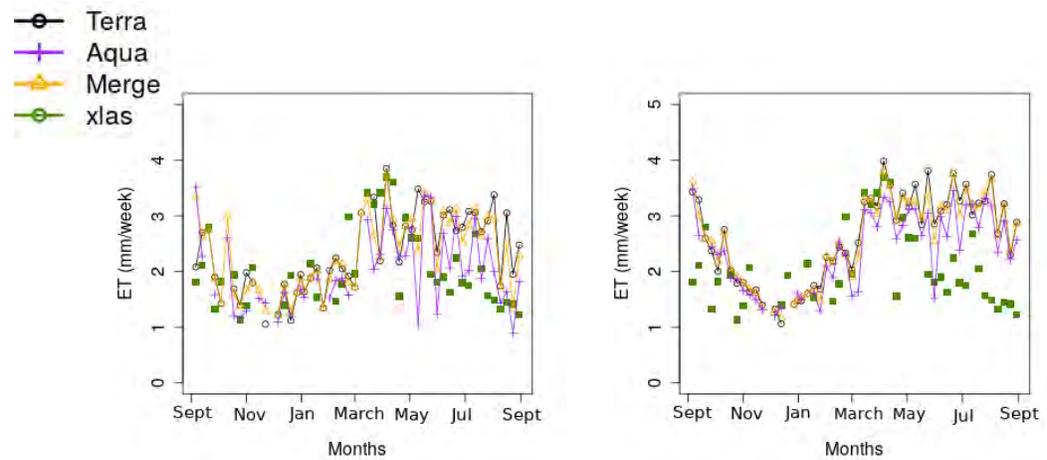
### 726 *Appendix G.3. Interpolation*

727 to derive seasonal evapotranspiration, we have to fill the gaps between satellite  
728 acquisitions in order to reconstruct days with missing ET data. Methods that rely on self  
729 preservation or known diurnal shape of the ratio of Evapotranspiration to a scale factor are  
730 usually used for this aim [67]. Indeed, this scale factor will monitor interpolation between  
731 two acquisition dates of successive images. In this work, we used the global radiation as  
732 an interpolation reference quantity. [41] have shown that it is the most robust and best  
733 performing scale factor for seasonal timescales.

### 734 **Appendix H. Daily ET simulations**

735 Figure A13 shows time series of the ET estimates derived from merging ET Terra  
736 and ET Aqua over the agronomic year 2013-2014 (from September to August), for which  
737 the residual ET derived from scintillometer (XLAS) measurements are available. We  
738 present separately ET simulations when the SPARSE model is constrained by the SWG  
739 meteorological data (see figure A13a) and when the SPARSE model is constrained by the  
740 ERA5 reanalyses meteorological data (see figure A13b). Overall, we observe in both figures  
741 that ET constrained from Aqua RS data (ET Aqua), is always biased in comparison with  
742 ET constrained from Terra RS data (ET Terra), and that ET Terra present higher values. In  
743 figure A13a, we observe that, in almost all cases, ET derived from XLAS measurements  
744 shows similar variation to ET Terra, excepted for summer periods. In addition, ET provided  
745 by the merging of Terra and Aqua reproduces more closely ET XLAS and succeeds to  
746 remove biases observed originally between ET Terra and Aqua. Using unprocessed ERA5  
747 reanalyses (figure A13b), ET from merging succeeds as well to reduce biases observed  
748 initially between ET Terra and Aqua. However, the three ET series simulated from ERA5  
749 meteorological data are not as close to the ET XLAS variations as the ET series simulated

750 from the SWG meteorological data. Indeed, ERA5 series are always biased in comparison  
751 with observations data.



(a) ET constrained by SWG meteorological surrogate series

(b) ET constrained by ERA5 meteorological surrogate series

**Figure A13.** Weekly times series of evapotranspiration simulated initially from TERRA and AQUA satellite data superposed with a time series computed from the merging (Merge) of TERRA and bias corrected AQUA series and ET from XLAS measurements for the agronomic year 2013-2014.

752

- 753 1. Wilhite, D.A.; Svoboda, M.D. Drought early warning systems in the context of drought preparedness and mitigation. *Early warning systems for drought preparedness and drought management* **2000**, pp. 1–21.
- 754 2. Schilling, J.; Hertig, E.; Trambly, Y.; Scheffran, J. Climate change vulnerability, water resources and social implications in North Africa. *Regional Environmental Change* **2020**, *20*, 1–12.
- 755 3. V., P. Zhai and A. Pirani and S. L. Connors and C. Péan, S. Berger and N. Caud, Y. Chen and L. Goldfarb and M. I. Gomis and M. Huang and K. Leitzell and E. Lonnoy and J.B.R. Matthews and T. K. Maycock and T. Waterfield and O. Yelekçi and R. Yu and B. Zhou (eds.). *Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2021.
- 756 4. Anderson, M.C.; Hain, C.; Wardlow, B.; Pimstein, A.; Mecikalski, J.R.; Kustas, W.P. Evaluation of drought indices based on thermal remote sensing of evapotranspiration over the continental United States. *Journal of Climate* **2011**, *24*, 2025–2044.
- 757 5. Otkin, J.A.; Anderson, M.C.; Hain, C.; Mladenova, I.E.; Basara, J.B.; Svoboda, M. Examining rapid onset drought development using the thermal infrared-based evaporative stress index. *Journal of Hydrometeorology* **2013**, *14*, 1057–1074.
- 758 6. Mishra, A.K.; Singh, V.P. A review of drought concepts. *Journal of hydrology* **2010**, *391*, 202–216.
- 759 7. McKee, T.B.; Doesken, N.J.; Kleist, J.; others. The relationship of drought frequency and duration to time scales. *Proceedings of the 8th Conference on Applied Climatology*. Boston, 1993, Vol. 17, pp. 179–183.
- 760 8. Mishra, A.; Desai, V. Drought forecasting using stochastic models. *Stochastic environmental research and risk assessment* **2005**, *19*, 326–339.
- 761
- 762
- 763
- 764
- 765
- 766
- 767
- 768
- 769
- 770
- 771
- 772
- 773
- 774
- 775

- 776 9. Mishra, A.; Singh, V.P. Analysis of drought severity-area-frequency curves using a general  
777 circulation model and scenario uncertainty. *Journal of Geophysical Research: Atmospheres* **2009**,  
778 *114*.
- 779 10. Zhang, J.; Xu, Y.; Yao, F.; Wang, P.; Guo, W.; Li, L.; Yang, L. Advances in estimation methods of  
780 vegetation water content based on optical remote sensing techniques. *Science China Technological*  
781 *Sciences* **2010**, *53*, 1159–1167.
- 782 11. AghaKouchak, A. Advancements in Satellite Remote Sensing for Drought Monitoring. In  
783 *Drought and Water Crises*; CRC Press, 2017; pp. 225–258.
- 784 12. Dorigo, W.; Wagner, W.; Albergel, C.; Albrecht, F.; Balsamo, G.; Brocca, L.; Chung, D.; Ertl, M.;  
785 Forkel, M.; Gruber, A.; others. ESA CCI Soil Moisture for improved Earth system understanding:  
786 State-of-the art and future directions. *Remote Sensing of Environment* **2017**, *203*, 185–215.
- 787 13. Wigneron, J.P.; Chanzy, A.; Calvet, J.C.; Bruguier, N. A simple algorithm to retrieve soil  
788 moisture and vegetation biomass using passive microwave measurements over crop fields.  
789 *Remote Sensing of Environment* **1995**, *51*, 331–341.
- 790 14. Zribi, M.; Nativel, S.; Le Page, M. Analysis of Agronomic Drought in a Highly Anthropogenic  
791 Context Based on Satellite Monitoring of Vegetation and Soil Moisture. *Remote Sensing* **2021**,  
792 *13*, 2698.
- 793 15. Amri, R.; Zribi, M.; Lili-Chabaane, Z.; Wagner, W.; Hasenauer, S. Analysis of C-band scatterom-  
794 eter moisture estimations derived over a semiarid region. *IEEE transactions on geoscience and*  
795 *remote sensing* **2012**, *50*, 2630–2638.
- 796 16. Zhang, D.; Tang, R.; Zhao, W.; Tang, B.; Wu, H.; Shao, K.; Li, Z.L. Surface soil water content  
797 estimation from thermal remote sensing based on the temporal variation of land surface  
798 temperature. *Remote Sensing* **2014**, *6*, 3170–3187.
- 799 17. Jiao, W.; Wang, L.; McCabe, M.F. Multi-sensor remote sensing for drought characterization:  
800 current status, opportunities and a roadmap for the future. *Remote Sensing of Environment* **2021**,  
801 *256*, 112313.
- 802 18. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation.  
803 *Remote sensing of Environment* **1979**, *8*, 127–150.
- 804 19. Kogan, F.N. Application of vegetation index and brightness temperature for drought detection.  
805 *Advances in space research* **1995**, *15*, 91–100.
- 806 20. Quiring, S.M.; Ganesh, S. Evaluating the utility of the Vegetation Condition Index (VCI) for  
807 monitoring meteorological drought in Texas. *Agricultural and Forest Meteorology* **2010**, *150*, 330–  
808 339.
- 809 21. Kogan, F.N. Operational space technology for global vegetation assessment. *Bulletin of the*  
810 *American meteorological society* **2001**, *82*, 1949–1964.
- 811 22. Amri, R.; Zribi, M.; Lili-Chabaane, Z.; Duchemin, B.; Gruhier, C.; Chehbouni, A. Analysis of  
812 vegetation behavior in a North African semi-arid region, using SPOT-VEGETATION NDVI  
813 data. *Remote sensing* **2011**, *3*, 2568–2590.
- 814 23. Farrar, T.; Nicholson, S.; Lare, A. The influence of soil type on the relationships between NDVI,  
815 rainfall, and soil moisture in semiarid Botswana. II. NDVI response to soil moisture. *Remote*  
816 *sensing of Environment* **1994**, *50*, 121–133.
- 817 24. Ozelkan, E.; Chen, G.; Ustundag, B.B. Multiscale object-based drought monitoring and compar-  
818 ison in rainfed and irrigated agriculture from Landsat 8 OLI imagery. *International Journal of*  
819 *Applied Earth Observation and Geoinformation* **2016**, *44*, 159–170.
- 820 25. AghaKouchak, A.; Farahmand, A.; Melton, F.; Teixeira, J.; Anderson, M.; Wardlow, B.D.; Hain,  
821 C. Remote sensing of drought: Progress, challenges and opportunities. *Reviews of Geophysics*  
822 **2015**, *53*, 452–480.
- 823 26. Paulik, C.; Dorigo, W.; Wagner, W.; Kidd, R. Validation of the ASCAT Soil Water Index using  
824 in situ data from the International Soil Moisture Network. *International journal of applied earth*  
825 *observation and geoinformation* **2014**, *30*, 1–8.

- 826 27. Paulik, C.; Naeimi, V.; Dorigo, W.; Wagner, W.; Kidd, R. A global validation of the ASCAT  
827 Soil Water Index (SWI) with in situ data from the International Soil Moisture Network. EGU  
828 General Assembly Conference Abstracts, 2012, p. 10189.
- 829 28. Baghdadi, N.; Cresson, R.; El Hajj, M.; Ludwig, R.; La Jeunesse, I. Soil parameters estima-  
830 tion over bare agriculture areas from C-band polarimetric SAR data using neural networks.  
831 *Hydrology & Earth System Sciences Discussions* **2012**, *9*.
- 832 29. Babaeian, E.; Sadeghi, M.; Jones, S.B.; Montzka, C.; Vereecken, H.; Tuller, M. Ground, proximal,  
833 and satellite remote sensing of soil moisture. *Reviews of Geophysics* **2019**, *57*, 530–616.
- 834 30. Narasimhan, B.; Srinivasan, R. Development and evaluation of Soil Moisture Deficit Index  
835 (SMDI) and Evapotranspiration Deficit Index (ETDI) for agricultural drought monitoring.  
836 *Agricultural and forest meteorology* **2005**, *133*, 69–88.
- 837 31. Zhang, N.; Hong, Y.; Qin, Q.; Liu, L. VSDI: a visible and shortwave infrared drought index for  
838 monitoring soil and vegetation moisture based on optical remote sensing. *International journal*  
839 *of remote sensing* **2013**, *34*, 4585–4609.
- 840 32. Jackson, R.D.; Idso, S.; Reginato, R.; Pinter Jr, P. Canopy temperature as a crop water stress  
841 indicator. *Water resources research* **1981**, *17*, 1133–1138.
- 842 33. Wang, X.; Zhao, C.; Guo, N.; Li, Y.; Jian, S.; Yu, K. Determining the canopy water stress for  
843 spring wheat using canopy hyperspectral reflectance data in loess plateau semiarid regions.  
844 *Spectroscopy Letters* **2015**, *48*, 492–498.
- 845 34. Jones, H.G.; Serraj, R.; Loveys, B.R.; Xiong, L.; Wheaton, A.; Price, A.H. Thermal infrared  
846 imaging of crop canopies for the remote diagnosis and quantification of plant responses to  
847 water stress in the field. *Functional Plant Biology* **2009**, *36*, 978–989.
- 848 35. Anderson, M.C.; Zolin, C.A.; Sentelhas, P.C.; Hain, C.R.; Semmens, K.; Yilmaz, M.T.; Gao, F.;  
849 Otkin, J.A.; Tetrault, R. The Evaporative Stress Index as an indicator of agricultural drought  
850 in Brazil: An assessment based on crop yield impacts. *Remote Sensing of Environment* **2016**,  
851 *174*, 82–99.
- 852 36. Norman, J.M.; Kustas, W.P.; Humes, K.S. Source approach for estimating soil and vegetation  
853 energy fluxes in observations of directional radiometric surface temperature. *Agricultural and*  
854 *Forest Meteorology* **1995**, *77*, 263–293.
- 855 37. Lili, Z.; Duchesne, J.; Nicolas, H.; Rivoal, R.; de BREGER, P. Détection infrarouge thermique des  
856 maladies du blé d’hiver 1. *Eppo Bulletin* **1991**, *21*, 659–672.
- 857 38. Sheffield, J.; Wood, E.F. *Drought: past problems and future scenarios*; Routledge, 2012.
- 858 39. Lagouarde, J.P.; Boulet, G. Energy balance of continental surfaces and the use of surface  
859 temperature. In *Land Surface Remote Sensing in Continental Hydrology*; Elsevier, 2016; pp. 323–  
860 361.
- 861 40. Boulet, G.; Chehbouni, A.; Gentine, P.; Duchemin, B.; Ezzahar, J.; Hadria, R. Monitoring water  
862 stress using time series of observed to unstressed surface temperature difference. *Agricultural*  
863 *and Forest Meteorology* **2007**, *146*, 159–172.
- 864 41. Delogu, E.; Olioso, A.; Allières, A.; Demarty, J.; Boulet, G. Evaluation of Multiple Methods for  
865 the Production of Continuous Evapotranspiration Estimates from TIR Remote Sensing. *Remote*  
866 *Sensing* **2021**, *13*, 1086.
- 867 42. Diarra, A.; Jarlan, L.; Er-Raki, S.; Le Page, M.; Aouade, G.; Tavernier, A.; Boulet, G.; Ezzahar,  
868 J.; Merlin, O.; Khabba, S. Performance of the two-source energy budget (TSEB) model for the  
869 monitoring of evapotranspiration over irrigated annual crops in North Africa. *Agricultural*  
870 *Water Management* **2017**, *193*, 71–88.
- 871 43. Moran, M.S. Thermal infrared measurement as an indicator of plant ecosystem health. In  
872 *Thermal remote sensing in land surface processes*; CRC Press, 2004; pp. 256–282.
- 873 44. Moran, M.; Clarke, T.; Inoue, Y.; Vidal, A. Estimating crop water deficit using the relation  
874 between surface-air temperature and spectral vegetation index. *Remote sensing of environment*  
875 **1994**, *49*, 246–263.

- 876 45. Cunha, A.P.; Zeri, M.; Deusdará Leal, K.; Costa, L.; Cuartas, L.A.; Marengo, J.A.; Tomasella, J.;  
877 Vieira, R.M.; Barbosa, A.A.; Cunningham, C.; others. Extreme drought events over Brazil from  
878 2011 to 2019. *Atmosphere* **2019**, *10*, 642.
- 879 46. Chirouze, J.; Boulet, G.; Jarlan, L.; Fieuzal, R.; Rodriguez, J.; Ezzahar, J.; Raki, S.E.; Bigeard, G.;  
880 Merlin, O.; Garatuza-Payan, J.; others. Intercomparison of four remote-sensing-based energy  
881 balance methods to retrieve surface evapotranspiration and water stress of irrigated fields in  
882 semi-arid climate. *Hydrology and Earth System Sciences Discussions* **2014**, pp. 1165–1188.
- 883 47. Farhani, N.; Carreau, J.; Boulet, G.; Kassouk, Z.; Mougenot, B.; Le Page, M.; Lili Chabaane,  
884 Z.; Zitouna, R. Scenarios of hydrometeorological variables based on auxiliary data for water  
885 stress retrieval in central Tunisia. 2020 Mediterranean and Middle-East Geoscience and Remote  
886 Sensing Symposium (M2GARSS). IEEE, 2020, pp. 293–296.
- 887 48. Boulet, G.; Mougenot, B.; Lhomme, J.; Fanise, P.; Lili-Chabaane, Z.; Oliosio, A.; Bahir, M.;  
888 Rivalland, V.; Jarlan, L.; Merlin, O.; others. The SPARSE model for the prediction of water  
889 stress and evapotranspiration components from thermal infra-red data and its evaluation over  
890 irrigated and rainfed wheat. *Hydrology and Earth System Sciences Discussions* **2015**, pp. 4653–4672.
- 891 49. Alazard, M.; Leduc, C.; Travi, Y.; Boulet, G.; Salem, A.B. Estimating evaporation in semi-arid  
892 areas facing data scarcity: Example of the El Haouareb dam (Merguellil catchment, Central  
893 Tunisia). *Journal of Hydrology: Regional Studies* **2015**, *3*, 265–284.
- 894 50. Massuel, S.; Riaux, J. Groundwater overexploitation: why is the red flag waved? Case study on  
895 the Kairouan plain aquifer (central Tunisia). *Hydrogeology Journal* **2017**, *25*, 1607–1620.
- 896 51. Leduc, C.; Ammar, S.B.; Favreau, G.; Beji, R.; Virrion, R.; Lacombe, G.; Tarhouni, J.; Aouadi,  
897 C.; Chelli, B.; Jebnoun, N.; others. Impacts of hydrological changes in the Mediterranean  
898 zone: environmental modifications and rural development in the Merguellil catchment, central  
899 Tunisia/ Un exemple d'évolution hydrologique en Méditerranée: impacts des modifications  
900 environnementales et du développement agricole dans le bassin-versant du Merguellil (Tunisie  
901 centrale). *Hydrological Sciences Journal/Journal des Sciences Hydrologiques* **2007**, *52*, 1162–1178.
- 902 52. Molle, F.; Wester, P. *River basin trajectories: societies, environments and development*; Vol. 8, IWMI,  
903 2009.
- 904 53. Ceballos, A.; Scipal, K.; Wagner, W.; Martínez-Fernández, J. Validation of ERS scatterometer-  
905 derived soil moisture data in the central part of the Duero Basin, Spain. *Hydrological Processes:  
906 An International Journal* **2005**, *19*, 1549–1566.
- 907 54. Kanzari, S.; Hachicha, M.; Bouhlila, R.; Battle-Sales, J. Characterization and modeling of water  
908 movement and salts transfer in a semi-arid region of Tunisia (Bou Hajla, Kairouan)—Salinization  
909 risk of soils and aquifers. *Computers and Electronics in Agriculture* **2012**, *86*, 34–42.
- 910 55. Brocca, L.; Ciabatta, L.; Moramarco, T.; Ponziani, F.; Berni, N.; Wagner, W. Use of satellite soil  
911 moisture products for the operational mitigation of landslides risk in central Italy. In *Satellite  
912 soil moisture retrieval*; Elsevier, 2016; pp. 231–247.
- 913 56. C. Paulik. *Copernicus Global Land Operations "Vegetation and Energy"*. TU Wien, 2017.
- 914 57. Clevers, J. Application of a weighted infrared-red vegetation index for estimating leaf area  
915 index by correcting for soil moisture. *Remote Sensing of Environment* **1989**, *29*, 25–37.
- 916 58. Saadi, S.; Boulet, G.; Bahir, M.; Brut, A.; Delogu, E.; Fanise, P.; Mougenot, B.; Simonneaux, V.;  
917 Lili Chabaane, Z. Assessment of actual evapotranspiration over a semi arid heterogeneous land  
918 surface by means of coupled low-resolution remote sensing data with an energy balance model:  
919 comparison to extra-large aperture scintillometer measurements. *Hydrology and Earth System  
920 Sciences* **2018**, *22*, 2187–2209.
- 921 59. Funk, C.; Peterson, P.; Landsfeld, M.; Pedreros, D.; Verdin, J.; Shukla, S.; Husak, G.; Rowland, J.;  
922 Harrison, L.; Hoell, A.; others. The climate hazards infrared precipitation with stations—a new  
923 environmental record for monitoring extremes. *Scientific data* **2015**, *2*, 1–21.

- 924 60. Massman, W. A surface energy balance method for partitioning evapotranspiration data into  
925 plant and soil components for a surface with partial canopy cover. *Water Resources Research*  
926 **1992**, *28*, 1723–1732.
- 927 61. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.;  
928 Peubey, C.; Radu, R.; Schepers, D.; Simmons, A.; Soci, C.; Abdalla, S.; Abellan, X.; Balsamo,  
929 G.; Bechtold, P.; Biavati, G.; Bidlot, J.; Bonavita, M.; De Chiara, G.; Dahlgren, P.; Dee, D.;  
930 Diamantakis, M.; Dragani, R.; Flemming, J.; Forbes, R.; Fuentes, M.; Geer, A.; Haimberger,  
931 L.; Healy, S.; Hogan, R.J.; Hólm, E.; Janisková, M.; Keeley, S.; Laloyaux, P.; Lopez, P.; Lupu,  
932 C.; Radnoti, G.; de Rosnay, P.; Rozum, I.; Vamborg, F.; Villaume, S.; Thépaut, J.N. The ERA5  
933 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **2020**, *146*, 1999–2049,  
934 [<https://rjmet.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3803>]. doi:10.1002/qj.3803.
- 935 62. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M.; others. Crop evapotranspiration-Guidelines for  
936 computing crop water requirements-FAO Irrigation and drainage paper 56. *Fao, Rome* **1998**,  
937 *300*, D05109.
- 938 63. Mega, N.; Medjerab, A. Statistical comparison between the standardized precipitation index  
939 and the standardized precipitation drought index. *Modeling Earth Systems and Environment* **2021**,  
940 *7*, 373–388.
- 941 64. Hoffman, R.N.; Boukabara, S.A.; Kumar, V.K.; Garrett, K.; Casey, S.P.; Atlas, R. An empirical  
942 cumulative density function approach to defining summary NWP forecast assessment metrics.  
943 *Monthly Weather Review* **2017**, *145*, 1427–1435.
- 944 65. Chahbi Bellakanji, A.; Zribi, M.; Lili-Chabaane, Z.; Mougnot, B. Forecasting of cereal yields in  
945 a semi-arid area using the simple algorithm for yield estimation (SAFY) agro-meteorological  
946 model combined with optical SPOT/HRV images. *Sensors* **2018**, *18*, 2138.
- 947 66. Joe, H. *Multivariate models and multivariate dependence concepts*; CRC Press, 1997.
- 948 67. Delogu, E.; Boulet, G.; Olioso, A.; Coudert, B.; Chirouze, J.; Ceschia, E.; Le Dantec, V.; Marloie,  
949 O.; Chehbouni, G.; Lagouarde, J.P. Reconstruction of temporal variations of evapotranspiration  
950 using instantaneous estimates at the time of satellite overpass. *Hydrology and earth system*  
951 *sciences* **2012**, *16*, 2995–3010.
- 952 68. Lhomme, J.P.; Elguero, E. Examination of evaporative fraction diurnal behaviour using a  
953 soil-vegetation model coupled with a mixed-layer model. *Hydrology and Earth System Sciences*  
954 **1999**, *3*, 259–270.
- 955 69. Hoedjes, J.; Chehbouni, A.; Jacob, F.; Ezzahar, J.; Boulet, G. Deriving daily evapotranspira-  
956 tion from remotely sensed instantaneous evaporative fraction over olive orchard in semi-arid  
957 Morocco. *Journal of Hydrology* **2008**, *354*, 53–64.
- 958 70. Jackson, R.D.; Hatfield, J.L.; Reginato, R.; Idso, S.; Pinter Jr, P. Estimation of daily evapotranspi-  
959 ration from one time-of-day measurements. *Agricultural Water Management* **1983**, *7*, 351–362.
- 960 71. Michelangeli, P.A.; Vrac, M.; Loukos, H. Probabilistic downscaling approaches: Application to  
961 wind cumulative distribution functions. *Geophysical Research Letters* **2009**, *36*.

## 4.4 Utilisation de l'indice thermique pour le stress de la végétation

L'évaluation des consommations réelles et le suivi de l'état hydrique des plantes passent notamment par la quantification de l'évapotranspiration (ET). L'évapotranspiration est un processus qui comprend la quantité d'eau transpirée par les plantes et aussi la quantité d'eau qui s'est évaporée directement de la surface du sol sans avoir été absorbée par les végétaux. Elle comprend aussi souvent la quantité d'eau interceptée par le feuillage qui s'est évaporée sans pénétrer dans le sol ou dans la plante (Doorenbos & Pruitt, 1997) (voir figure 4.7). Il est clair qu'une bonne partie de l'eau qui s'évapore sans pénétrer dans la plante est considérée comme une perte. Un bon développement de la végétation et un bon rendement sont toujours reliés à la disponibilité de l'eau avec des quantités adéquates au niveau de la zone racinaire (Passioura, 1983). La transpiration (T) forme un bon indicateur sur l'état hydrique de la végétation et par conséquent sur l'efficacité de l'utilisation de l'eau. Par ailleurs, une estimation de la composante de transpiration permet une estimation plus efficace des besoins réels de la végétation. Concrètement, la transpiration forme la quantité d'eau perdue à travers les feuilles du couvert végétal, plus précisément à partir des ouvertures stomatiques, lorsque la végétation est en bonne santé c'est-à-dire, un bon état hydrique. La quantité d'eau transpirée dépend de la teneur en eau dans le sol, des conditions météorologiques, et aussi des propriétés du sol (Mishra & Singh, 2010; Allen *et al.*, 1998). Une estimation de la composante de transpiration est possible par le biais des modèles à double source (Norman *et al.*, 1995).

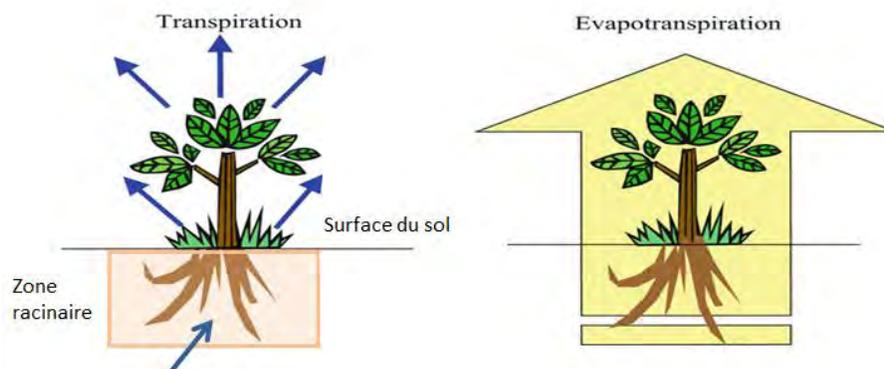


Figure 4.7 – Schématisation de l'évapotranspiration au niveau du composite "sol-plante" et la transpiration issue uniquement de la plante.

L'utilisation des modèles de bilan d'énergie à double source offre l'avantage d'estimer chacune des deux composantes «transpiration des plantes» et «évaporation du sol» (Lagouarde & Boulet, 2016). Cette estimation séparée est très utile pour estimer les flux d'énergie au niveau des zones arides et semi-arides caractérisées par un couvert végétal épars, et par une contribution inégale de ces deux composantes dans le processus d'évapotranspiration (Norman *et al.*, 1995). En particulier, la connaissance de la contribution de la plante dans le processus d'évapotranspiration au niveau de ces régions, est très informative par rapport à l'estimation

des besoins de la végétation, réelle et efficace en eau. Par ailleurs, l'estimation de T est très utile pour éviter les phénomènes de sécheresse de la végétation et comprendre au mieux le mécanisme de leurs formations qui présente des conditions de déclenchement différentes de celles présentées au niveau de la surface (Guan & Wilson, 2009).

#### 4.4.1 Simulation de la transpiration

Le modèle de bilan d'énergie à double source SPARSE utilise la température radiative de surface (LST) afin de dissocier une source virtuelle «sol» de l'évaporation, en intégrant les températures du sol ( $T_s$ ), et une autre source virtuelle «végétation» pour la transpiration, en intégrant les températures des feuilles/tiges du couvert ( $T_v$ ). Ce modèle interprète la température de brillance sur la base de la quantité globale relative de sol et de la végétation observée par le capteur. Par ailleurs, les températures  $t_s$  et  $t_v$  sont utilisées comme deux variables clés pour résoudre les deux bilans correspondant au sol et à la végétation. Pour l'estimation de flux de chaleur latente du sol  $LE_s$  (évaporation) et de la végétation  $LE_v$  (transpiration). SPARSE se base sur les mêmes hypothèses proposées par Norman *et al.* (1995), pour résoudre les deux bilans du sol et de la végétation en utilisant une seule information de température (LST) et éviter par conséquent le problème d'indétermination. SPARSE peut être aussi résolu dans des conditions prescrites, c'est-à-dire, conditions pour un niveau de stress donné, et donc pour des conditions comprises entre absence de stress (conditions potentielles) à stress total (absence d'évaporation ou de transpiration). Le stress est défini pour SPARSE à l'aide des coefficients d'efficacité, qui pondèrent le flux exprimé sans stress, et donc correspondent grossièrement au rapport entre une Evaporation (Transpiration) réelle et Evaporation (Transpiration) potentielle. Ils jouent un rôle équivalent à la résistance de surface (évaporation) ou un facteur de stress de la résistance stomatique (transpiration). Les valeurs de ces coefficients d'efficacité,  $\beta_s$  et  $\beta_v$ , respectivement pour le sol et la végétation, varient donc entre 0 et 1.

Classiquement, une évapotranspiration potentielle est calculée en supposant que la surface est suffisamment humide en surface et dans la zone racinaire pour qu'à la fois le sol évapore au taux maximal ( $\beta_s = 1$ ) et la plante transpire au taux maximal ( $\beta_v = 1$ ). Pourtant, toutes les combinaisons de  $\beta_s$  et  $\beta_v$  sont théoriquement possibles, et même si la transpiration maximale est obtenue avec  $\beta_v = 1$ , le flux correspondant, du fait des interactions avec le sol et l'atmosphère commune dans le cadre du schéma double-source, peut varier de façon significative en raison de ce que l'on pourrait appeler «effet de micro-oasis». En effet, lorsque le sol s'assèche en surface après une pluie par exemple, la fraction d'énergie disponible non utilisée par l'évaporation est en partie convertie en chaleur sensible supplémentaire et donc contribue à réchauffer et assécher l'air au sein du couvert. Cela produit une augmentation de la transpiration potentielle dès lors que la zone racinaire est bien alimentée en eau. Aussi comment devons-nous définir la transpiration potentielle ? Celle d'un sol sec intercalaire ou bien d'un sol humide en surface ? Si l'on souhaite utiliser la notion de transpiration potentielle pour détecter le stress hydrique, la première option semble être la plus pertinente, et cela renvoie à l'hypothèse de Norman : si la plante commence à souffrir de stress hydrique, il y a de fortes chances que le sol soit déjà très sec en surface ( $\beta_s = 0$ ). La différence entre les valeurs de transpiration potentielle avec  $\beta_s = 0$  ou  $\beta_s = 1$  est illustrée sur la figure 4.8. Dans notre cas, nous pouvons

avoir près de 0.2 mm/ jour de différence. Ce graphique montre la variation décadaire de la transpiration réelle et potentielle calculée. Nous avons choisi une année sèche (2009-2010), et un pixel contenant un pourcentage élevé de céréale pluviale afin d'avoir un signal détectable de stress hydrique de la végétation. Pour ce faire, nous avons utilisé l'information issue de la carte de mask de céréales irrigués 2011-2012, élaborée par Chahbi *et al.* (2014). Nous avons alors repéré des pixels ayant plus de 60% de couverture en céréales en sec pour cette année. Ensuite, nous avons vérifié si ces pixels correspondent aussi à des pixels du céréale pendant l'année 2009-2010. Ainsi, nous avons présenté le cumul de précipitations et la variation de NDVI pour la même période, pour faciliter l'interprétation de la variation de transpiration. Au niveau du graphique 4.8a la transpiration potentielle est calculée en supposant que la surface ainsi que la zone racinaire sont suffisamment humides, pour effectuer une évaporation et une transpiration à un taux maximal  $\beta_s = 1$  et  $\beta_v = 1$ . Cependant, d'après ce graphique, nous observons que la transpiration réelle est plus forte que la transpiration potentielle. Ce fait nous ramène à confronter le problème d'effet de micro-oasis et avoir par conséquent des indices de stress négatifs. Nous avons alors reproduit la transpiration potentielle en se basant sur l'hypothèse qu'un stress de la plante arrive à un sol très sec donc  $\beta_s = 0$  et  $\beta_v = 1$  pour rester dans les conditions potentielles de transpiration de la plante. D'après le graphique 4.8b, nous observons que les résultats deviennent plus cohérents avec la théorie, avec un taux de transpiration potentielle plus fort qu'un taux de transpiration réelle.

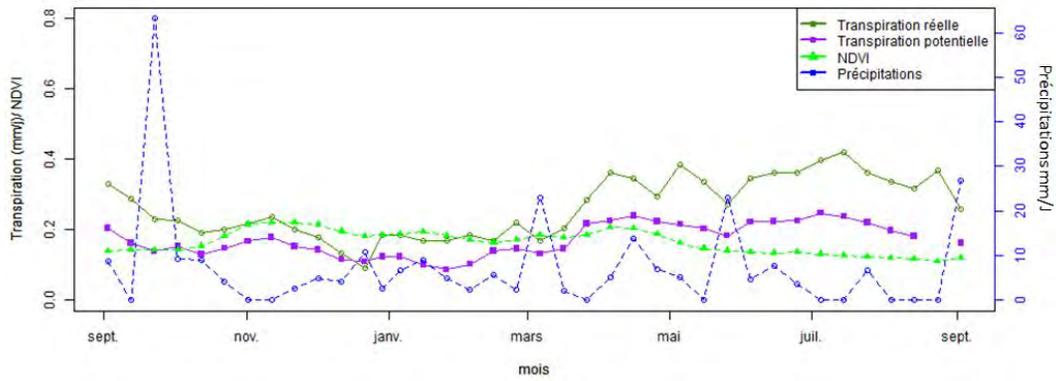
Le graphique 4.8b montre des valeurs très faibles de la transpiration réelle et potentielle ne dépassant pas 0.6mm/j. Un manque de précipitations durant la période entre octobre et février a engendré une diminution de la transpiration réelle qui montre une valeur très basse en fin Décembre, de l'ordre de 0.18mm/j. L'augmentation de la transpiration en été est due à l'enregistrement des faibles apports en précipitations mais qui étaient suffisants pour augmenter le taux de transpiration. Le NDVI à son tour présente des valeurs très faibles pour toute l'année. Durant le mois de Mars, période de développement maximal de céréales, nous observons une augmentation très légère de NDVI. Ceci, montre que le pixel choisi ne contient pas un pourcentage assez suffisant pour avoir une réponse nette des cultures de céréales et pouvoir interpréter la réponse de la transpiration pour des conditions de stress. Il s'agit peut être d'une occupation du sol d'arboriculture majoritairement.

#### 4.4.2 Indice de stress végétation

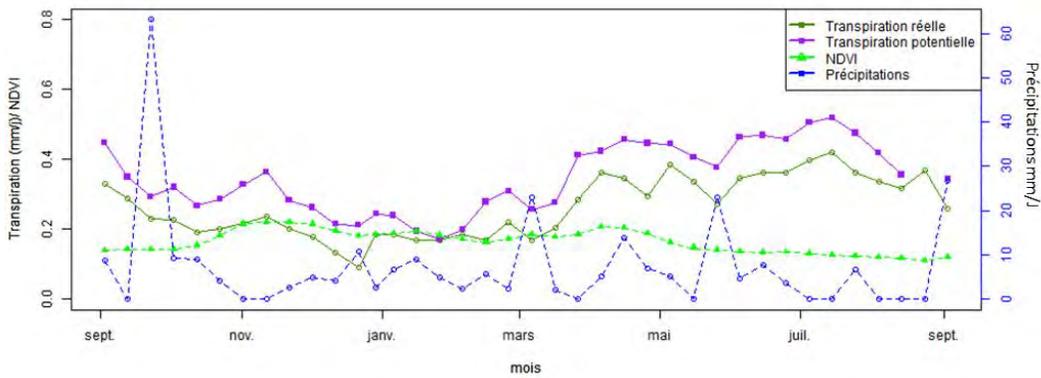
Pour mieux explorer dans le stress de la végétation, nous avons élaboré un indice thermique calculé seulement à partir de la transpiration. Cet indice ( $SI_{veg}$ ) est simulé à partir du modèle de bilan d'énergie à double source SPARSE en se basant sur la composante de flux de chaleur latente de la végétation (LEv), comme présenté au niveau de l'équation 4.7.

$$SI_{veg} = 1 - \frac{LEv}{LEv_{pot}}. \quad (4.7)$$

Au niveau de la figure 4.9, nous avons présenté l'indice de stress total issu de l'évapotranspiration, l'indice de stress de la végétation calculé à partir de la transpiration et l'indice standardisé des précipitations, pour la même année sèche (2009-2010) et sur le pixel choisi de céréale.



(a) la transpiration potentielle avec  $\beta_s=1$  et  $\beta_v=1$



(b) la transpiration potentielle avec  $\beta_s=0$  et  $\beta_v=1$

Figure 4.8 – Variation décadaire de la moyenne de transpiration journalière simulée sur un pixel MODIS contenant un pourcentage élevé de céréale sec.

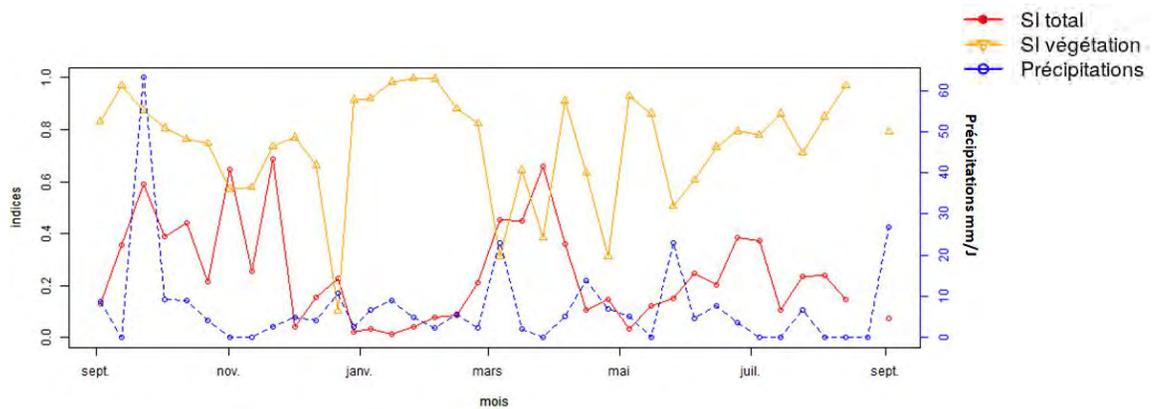


Figure 4.9 – Variation décadaire de l'indice de stress calculé à partir de l'évapotranspiration et l'indice calculé seulement à partir de la composante de transpiration, tous les deux simulés sur un pixel MODIS contenant un pourcentage élevé de céréale sec durant l'année 2009-2010.

Au niveau de ce graphique (4.9), nous observons que l'indice SI total montre une variation en cohérence avec la variation de UPI. Un apport important en précipitations au début de l'année a engendré des valeurs élevées de cet indice. Puis, une baisse de précipitations en automne est suivie d'une diminution remarquable de SI, qui reprend des valeurs moyennes avec des apports de précipitations variables. Pour l'indice thermique de la végétation (issu de la transpiration), nous observons des valeurs proches de 1 ce qui indique une situation de non stress et d'humidité assez suffisante au niveau de la zone racinaire de la plante. Cependant, nous observons en janvier une chute remarquable de cet indice qui coïncide avec la chute de la transpiration réelle observée au niveau du graphique 4.8b, ce qui témoigne d'une situation de stress pendant cette période, précédée par un déficit de précipitations sévère. Nous observons aussi en Mai, une valeur faible de cet indice avec une probabilité de non dépassement de 20%. La diminution de l'indice de stress végétation est effectuée suite à la diminution de la transpiration réelle par rapport à son taux potentiel pour cette période (voir fig. 4.8b), ce qui indique par ailleurs une situation de stress.

## 4.5 Comparaison SI et ESI

La variabilité de l'évapotranspiration à la fois à l'échelle spatiale et temporelle est due à la variabilité de l'humidité du sol, la couverture végétale et les conditions atmosphérique (la température de l'air, vitesse du vent et déficit de pression de vapeur) (Anderson *et al.*, 2007). L'évapotranspiration potentielle estime la quantité d'eau susceptible d'être perdue dans les mêmes conditions quand l'eau n'est pas un facteur limitant (Brochet, 1977). Par ailleurs cette information joue un rôle très important pour exprimer un stress relatif à la variabilité de l'évapotranspiration réelle par rapport à une condition potentielle. Dans la littérature, plusieurs indices de stress hydriques ont été développés en se basant sur ce principe, notamment CWSI (Jackson *et al.*, 1977), WDI (Moran *et al.*, 1994), ESI (Anderson *et al.*, 2007) et notre indice SI (voir la section 2.1.1). Les deux indices SI et ESI sont simulés à partir des modèles d'énergie à double source, ALEXI et SPARSE respectivement. Ces deux indices présentent alors l'avantage d'estimer avec précision le stress grâce à leur simulation à partir des modèles à double source (simulation séparée à partir des composantes du sol et de la végétation), et de pouvoir refléter directement une estimation du stress et non pas une variable qui lui est relié indirectement.

Les conditions potentielles d'évapotranspiration sont exprimées d'une façon différente entre ces deux indices. ESI est calculé à partir d'une évapotranspiration de référence ( $ET_0$ ). En revanche, SI est estimé à partir de l'évapotranspiration potentielle ( $ET_p$ ).  $ET_0$  est définie comme le flux d'ET d'une culture hypothétique de référence, très proche de l'ET d'une grande surface de gazon de hauteur uniforme, en croissance active, avec un approvisionnement en eau convenable (Allen *et al.*, 1998).  $ET_0$  est calculée par la méthode FAO dérivée de la formule de Penman-Monteith (voir équation 4.8). Cette équation est exprimée seulement en fonction des paramètres climatiques (température de l'air, vitesse du vent, rayonnement...).

$$ET_0 = \frac{0.408\Delta(Rn - G) + \gamma \frac{900}{T+273} U_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)} \quad (4.8)$$

Avec,  $(Rn-G)$  : énergie disponible ( $MJm^{-2}J^1$ ),  $\delta$  : pente de la relation entre la pression de vapeur de saturation et la température ( $KPa/^\circ C$ ),  $T$  : température de l'air moyenne journalière mesurée à 2 m ( $^\circ C$ ),  $U_2$  : vitesse du vent mesurée à 2 m ( $m/s$ ),  $\gamma$  : constante psychrométrique ( $KPa/^\circ C$ ),  $e_a - e_s$  : déficit de pression de vapeur saturante ( $KPa$ ).

Une évapotranspiration réelle est exprimée en fonction de  $ET_0$  avec l'application d'un coefficient cultural ( $k_c$ ) qui dépend essentiellement des caractéristiques propres de la culture (voir équation 4.9). L'évapotranspiration est dite potentielle lorsque l'eau ne forme pas un facteur limitant au niveau de l'absorption racinaire, la régulation stomatique est minimale et l'évapotranspiration est maximale. Par ailleurs,  $ET_p$  dépend à la fois du type de culture, de son stade végétatif, de la disponibilité en eau et des conditions climatiques.  $ET_p$  permet d'estimer la demande en eau des cultures en fonction des conditions climatiques, contrairement à  $ET_0$  qui est fonction de la demande évaporative de l'atmosphère. On peut donc supposer que normaliser par  $ET_p$  va isoler le stress hydrique lui-même de l'impact du stress sur un moindre développement végétal (déjà pris en compte dans  $ET_p$  à travers le LAI), tandis que normaliser par  $ET_0$  va combiner les deux aspects, la référence étant la demande climatique.

$$ET_r = K_c ET_0 \quad (4.9)$$

Au niveau de la figure 4.10, nous avons présenté  $ET_p$  et  $ET_0$  moyennée sur toute la plaine.  $ET_p$  est simulée à partir de SPARSE et  $ET_0$  est calculée à partir de la formule de Penman-Monteith, présentée au niveau de l'équation 4.8. Ces deux variables sont issues d'un forçage climatique de la projection du générateur stochastique depuis l'année 2000. Au niveau de ce graphique, nous observons une variation saisonnière très semblable entre ces deux variables. Cependant,  $ET_0$  présente des valeurs plus élevées qu'une  $ET_p$  dans la plupart du temps, compte tenu des faibles NDVI sur la zone, sauf pour quelques années comme l'année 2008 et 2016, où les deux  $ET$  sont bien superposées. En effet,  $ET_p$  est essentiellement une fonction de la demande évaporative de l'atmosphère comme  $ET_0$ , mais les propriétés des cultures qui montrent des variations très importantes en termes du types de culture et leur stade de développement montre la différence entre les deux indices d'une année à une autre.

Au niveau de la figure 4.11a, nous avons présenté la variation décadaire des deux indices SI et ESI calculés respectivement à partir de  $ET_p$  et  $ET_0$ , pour l'année 2012-2013. Ces deux indices sont aussi issus d'un forçage climatique de la projection du générateur stochastique. Nous observons une variation très corrélée entre ces deux indices, mais qui varie en fonction des saisons. Pour mieux comprendre la variation inter-annuelle observée au niveau des variables  $ET_0$  et  $ET_p$  (graphique 4.10), et la variation inter-saisonnière de ces indices, nous avons présenté au niveau du graphique 4.11b, les données climatiques décennales standardisées pour cette année : la température de l'air à partir de laquelle nous allons déduire la variation de l'humidité de l'air (anti-corrélées), la vitesse du vent et le rayonnement global.

Au niveau du graphique 4.11a, SI indique moins de stress par rapport à ESI pendant la saison de croissance. Cependant, en été c'est ESI qui a indiqué moins de stress. En effet, pendant la saison de croissance, les valeurs élevées de SI par rapport à ESI indiquent que pour cette période, la demande climatique, présentée par  $ET_0$ , a été supérieure par rapport à la demande des cultures durant cette période. Nous observons par ailleurs au niveau du graphique 4.11b, des valeurs élevées de la vitesse du vent, de la température de l'air, essen-

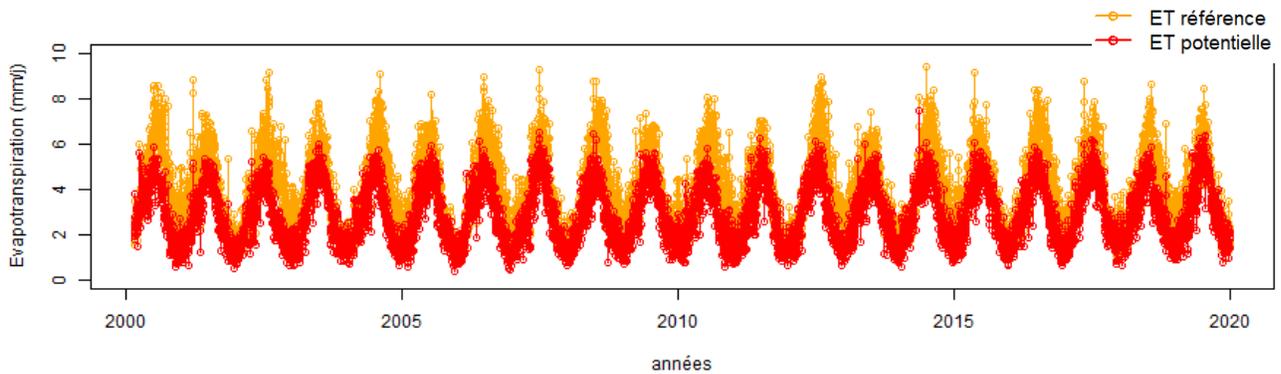


Figure 4.10 – Comparaison de l'évapotranspiration de référence et l'évapotranspiration potentielle sur toute la période d'étude, toutes les deux simulées à partir d'un forçage climatique issu du générateur stochastique, représentant la plaine.

tiellement en Janvier et Février, d'où l'observation d'une diminution très importante de ESI durant cette période. D'autre part, cette année a montré des valeurs de NDVI très faibles (voir figure 7 dans l'article de comparaison des indices, section 4.3.2), ce qui engendre un coefficient cultural plus faible pour cette année, et des valeurs de ETp faibles pour cette saison. En été, nous observons une augmentation parallèle de SI et ESI pour indiquer une situation de non stress, avec des valeurs ESI supérieures à SI. Cette augmentation témoigne de l'existence des conditions d'humidité de l'air et du sol suffisantes pour subvenir aux besoins de la demande climatique importante et la demande culturelle (ETp), qui était plus élevée en cette période. En effet, d'après la figure 4.11b, nous observons des températures de l'air anormalement faibles pour une période d'été, notamment en Juin et des taux d'ensoleillement (rayonnement global) moyens et faibles. Ces conditions climatiques favorisent une diminution de la demande évaporative, les taux d'évapotranspiration réelles arrivent à compenser les demandes atmosphériques quantifiées par  $ET_0$  et la demande culturelle pour cette période (ETp). Finalement, cette comparaison nous permet de conclure que ces deux indices SI et ESI présentent une variation très similaire, ce qui induit un suivi des périodes de sécheresse comparable, avec des légères différences en termes d'intensité engendrées par la considération de la demande atmosphérique seule ou la demande des cultures en fonction des conditions atmosphériques.

## 4.6 Conclusion partielle et synthèse

La détection du stress hydrique est d'importance primordiale à la fois pour le développement agricole et pour la gestion des ressources en eau. Elle permet en outre d'identifier d'une façon précoce une période de sécheresse, d'atténuer ses dégâts et de prévenir les décideurs de son intensité et de sa répartition spatiale. Les données de télédétection et notamment les indices de stress qui en dérivent sur différentes longueurs d'onde jouent un rôle important pour la

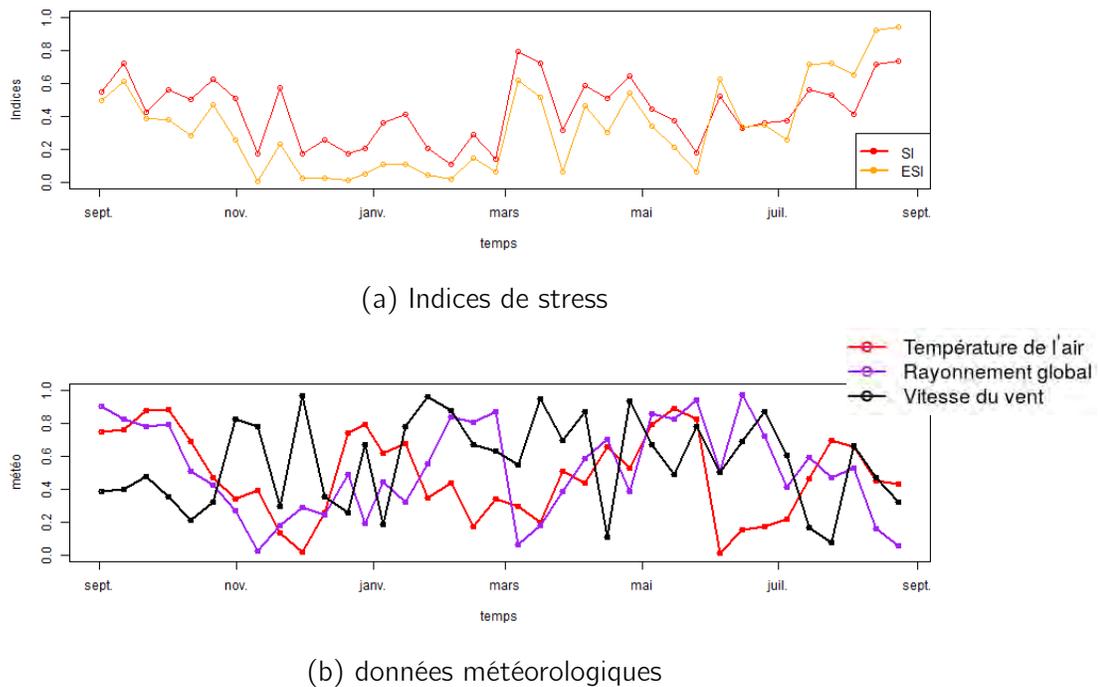


Figure 4.11 – Variation décadaire des indices standardisés SI et ESI et les données météorologiques standardisées durant l'année agricole 2012-2013.

caractérisation de l'état hydrique des surfaces. La première section de ce chapitre comporte une évaluation des méthodes d'interpolation de l'évapotranspiration permettant de reconstruire une information saisonnière, utile pour les études de sécheresse et l'élaboration d'un indice de stress thermique. Nous avons comparé deux types d'interpolation : une méthode classiquement utilisée dans la littérature, basée sur le rayonnement global comme support d'interpolation, et une deuxième méthode, inspirée par la méthode utilisée dans Delogu *et al.* (2021), basée sur l'indice des pluies antérieures (API). La méthode d'API a montré une très bonne performance en comparaison avec les données de mesure de scintillomètre XLAS. En effet, cette méthode reproduit au mieux la variation de l'évapotranspiration en fonction de l'information issue des précipitations qui à son tour, jouent un rôle très important sur les conditions hydriques de la surface. Ceci est d'autant plus important, au niveau des régions semi-arides, caractérisées par un dessèchement du sol très rapide après un événement pluvial, et une élévation rapide et brusque de ET. Cependant, la méthode de API nécessite encore un travail de paramétrage (choix de  $API_{max}$  par exemple).

La deuxième section de ce chapitre comporte une évaluation de la cohérence et de la pertinence de trois indices de stress : un indice thermique, le NDVI issu du visible/proche infra-rouge et le SWI issu du micro-onde. L'indice thermique (SI) est simulé à partir du modèle de bilan d'énergie à double source SPARSE en utilisant deux types de forçage climatique, à partir des données météorologiques issues du générateur stochastique  $SI_{SWG}$  (voir section 3.2.2) et à partir des données brutes des réanalyses ERA5  $SI_{ERA5}$ . A une échelle régionale, la comparaison

de ces différents indices est effectuée en se basant sur un indice standardisé de précipitations UPI, utilisé comme référence pour l'identification des périodes de déficit hydrique. Cette comparaison est effectuée sur différentes échelles temporelles, (1) une échelle saisonnière qui couvre la période de croissance de la végétation (de Septembre jusqu'au Mai) pour voir la variabilité inter-annuelle de ces indices en fonction du statut hydrique identifié par UPI pour chaque saison, et (2) une échelle décadaire pour voir l'évolution de ces indices en réponse à une disponibilité limitée ou suffisante des besoins en eau sur une plus courte période. Une étude des indices de stress sur une échelle temporelle fine nous permet par ailleurs d'analyser le mécanisme de déclenchement des périodes de stress.

Ce travail nous a permis de voir qu'à une échelle annuelle, tous les indices de stress montrent une variation inter-saisonnière similaire à celle de UPI. Toutefois, nous observons parfois une incohérence entre l'indice UPI et le reste des indices standardisés. Ce fait a mis l'accent sur le besoin de la végétation d'un apport d'eau adéquat durant une période sensible de son développement, ce qui engendre une différence entre le cumul de précipitations saisonnières et la réponse de la végétation reflétée par le biais des indices de stress. Sur une échelle décadaire, nous observons que le temps nécessaire pour répondre à un déficit hydrique ou un apport d'eau est différent d'un indice à un autre. Le NDVI montre moins de variabilité et une réponse très tardive par rapport à la variation de UPI. Toutefois, cet indice arrive à caractériser correctement le statut hydrique de la saison en se basant sur le seuil de stress que nous avons défini (20% des valeurs les plus faibles pour chaque indice). D'autre part, on observe une variation similaire des 3 indices  $SI_{SWG}$ ,  $SI_{ERA5}$  et SWI. Cependant, lors d'un déficit hydrique les indices thermiques réagissent rapidement par une diminution de leurs valeurs. En effet, les indices thermiques calculés à partir de l'évapotranspiration illustrent l'état hydrique à la fois du sol et de la végétation. Un déficit hydrique engendre un dessèchement rapide de la surface du sol, ce qui fait varier l'indice thermique d'une façon précoce. Selon l'hypothèse de Norman la plante commence à souffrir de stress hydrique après le dessèchement total du sol. SWI qui illustre l'état hydrique au niveau de la zone racinaire, va réagir alors après le dessèchement de la surface du sol.

D'autre part, nous avons évalué la probabilité de chaque indice à détecter une période de stress identifiée par UPI ou sa moyenne mobile calculée sur différentes fenêtres temporelles (2, 3 et 4 décades). Cette analyse effectuée en fonction des saisons, nous a permis de voir une meilleure cohérence entre UPI et  $SI_{SWG}$  notamment en automne, hiver et printemps. En particulier,  $SI_{SWG}$  a présenté la probabilité la plus faible de détecter un stress hydrique non identifié par UPI, contrairement à  $SI_{ERA5}$  ayant une probabilité élevée à détecter une fausse période de stress. SWI présente une meilleure cohérence avec UPI au printemps. En effet, durant cette saison, la végétation montre un développement maximal avec des racines bien profondes dans le sol, ce qui explique peut-être la performance de cet indice à bien identifier le statut hydrique en fonction des apports de précipitations. L'application des moyennes mobiles sur l'indice standardisé de précipitations a amélioré sa cohérence avec les autres indices. Cela témoigne d'un délai spécifique pour chaque indice afin de répondre à un stress hydrique. Ce décalage est approximativement, de l'ordre de 2 décades pour les indices thermiques, 3 décades pour SWI et à partir de 4 décades pour NDVI. En effet, le NDVI reflète une sénescence précoce engendrée par un déficit hydrique marqué dans le sol qui peut être identifié par ailleurs d'une

façon précoce par SWI ou les indices thermiques qui sont en liaison directe avec la teneur en eau dans le sol.

A l'échelle locale, nous avons identifié la zone qui correspond au transect XLAS afin de pouvoir comparer les différents indices avec des données de vérité terrain, issues des mesures de scintillomètre XLAS. Cette analyse a montré que l'indice thermique suit au mieux les variations de l'indice XLAS, notamment l'indice  $SI_{SWG}$ , qui arrive à identifier dans la plupart du temps les périodes de stress. En effet,  $SI_{SWG}$  est simulé à partir d'un forçage météorologique simulé sur une échelle locale. Tandis que,  $SI_{ERA5}$  est simulé à partir d'un forçage climatique d'ERA5 ayant une résolution spatiale très faible. SWI n'est pas montré précis pour cette analyse à une échelle locale. Cette déficience peut être expliquée par sa résolution spatiale faible et par ailleurs son incapacité à déterminer un état hydrique sur une échelle plus fine. D'autre part, une analyse des indices de stress sur des parcelles de céréales pluviales est très utile pour avoir un signal de réponse à un stress détectable par les indices. Pour ce faire, nous avons identifié un pixel contenant un pourcentage élevé de céréales pluviales durant une année humide et une année sèche. Cette analyse a montré que l'indice  $SI_{SWG}$ ,  $SI_{ERA5}$  et SWI ont été les plus performants à identifier la classe de stress pour chaque année.

Finalement, nous avons évalué la réponse de la transpiration réelle et potentielle au niveau des pixels contenant un pourcentage élevé de céréales en sec. La transpiration potentielle est simulée par SPARSE à l'aide des coefficients d'efficacité  $\beta_s$  et  $\beta_v$ , respectivement pour le sol et la végétation. Nous avons testé deux hypothèses pour pondérer les coefficients  $\beta_s$  et  $\beta_v$  : Un sol bien humide en surface et alors une évaporation potentielle  $\beta_s = 1$  et une transpiration potentielle  $\beta_v = 1$ , ou bien un sol sec intercalaire et par ailleurs, un sol sec  $\beta_s = 0$  et une transpiration maximale  $\beta_v = 1$ . La simulation de la transpiration en se basant sur ces deux hypothèses a montré que la deuxième hypothèse était la plus cohérente, et plus réaliste pour éviter les effets de micro-oasis et des valeurs d'indice de stress négatifs. Toutefois, pour avoir des meilleurs résultats, il faut pallier les valeurs des coefficients d'efficacité pour s'approcher au maximum du comportement réaliste de l'écosystème. Ainsi, cette étude nécessite une information précise concernant les parcelles de céréale pluviale, et aussi des étendus importants de ce type d'occupation des sols afin d'avoir un signal détectable de stress hydrique de la végétation.

La région méditerranéenne a été classée parmi les zones à risque en termes de changements climatiques par le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC). En effet, la plupart des pays de cette région, définie comme un «hot-spot» de crise de l'eau, souffrent d'une élévation de la température, d'une pénurie d'eau croissante, de la fréquence accrue des périodes de sécheresses ainsi que de l'intensification des phénomènes de désertification. Ces contraintes climatiques, combinées à une forte croissance démographique, laissent présager une augmentation de la contrainte anthropique induite sur les disponibilités limitées des ressources hydriques. Par ailleurs, ces régions sont dans le besoin impératif de lutter contre les conséquences du changement climatique en adoptant des mesures d'atténuation et notamment d'adaptation des bonnes pratiques de gestion de ces ressources. Le bassin versant de Merguellil situé au centre de la Tunisie et qui fait l'objet de notre étude, présente un modèle typique des zones semi-arides de la région de la Méditerranée. Ce bassin versant est caractérisé par une forte variabilité climatique inter-annuelle, des faibles apports en précipitations et une sur-exploitation des ressources en eau disponibles. Ces régions doivent avoir une stratégie adéquate d'exploitation de ces ressources afin de garantir une disponibilité durable pour les différents usages. Pour ce faire une étude des périodes de sécheresses s'avère indispensable pour comprendre le mécanisme de leur génération, leur récurrence et leur intensité. Une meilleure compréhension de cet aléa naturel nous permettra d'atténuer ses dégâts par la définition des politiques de mitigation, essentiellement au niveau du secteur agricole qui forme le consommateur primordial des ressources hydriques au niveau de ces régions (environ 80%). Des indices pour quantifier et identifier les périodes de sécheresse/stress hydrique sont mises en œuvre. Ces indices qui se basent sur des informations de l'état de surface, reflètent le statut hydrique de la zone étudiée. Dans ce contexte, on se base alors sur des indices de stress qui utilisent les données satellitaires. En effet, grâce à leur accessibilité et leur emprise spatiale et temporelle, les données de télédétection constituent un outil d'intérêt pour décrire les variations spatio-temporelles des variables de surface : le développement de la végétation à partir du visible et du proche infrarouge, l'humidité de surface à partir du micro-onde actif ou passif et la température de surface à partir du thermique infrarouge.

En particulier, l'évapotranspiration constitue en ce sens une variable clé pour la compréhension des interactions surface-atmosphère à différentes échelles spatio-temporelles. Ceci est d'autant plus vrai au niveau des régions semi-arides, très exposées aux aléas climatiques, dû à son rôle important dans le cycle de l'eau et de l'énergie de la surface. En période de déficit hydrique, la plante réduit son évapotranspiration réelle (ET<sub>r</sub>) par rapport au taux d'évapotranspiration potentielle (ET<sub>p</sub>). Cette réduction est imposée par la fermeture des stomates, ce qui engendre une élévation de la température de surface (LST) détectée par les capteurs thermiques. Le rapport entre ET<sub>r</sub> et ET<sub>p</sub> est par ailleurs très informatif sur l'existence

d'un stress hydrique. L'évapotranspiration et l'indice de stress thermique peuvent être simulés d'une façon précise à partir des modèles de bilan d'énergie à double source. Ces modèles nécessitent à la fois des données satellitaires (température de surface, NDVI, albédo et LAI) et des données météorologiques (température de l'air, humidité relative de l'air, vitesse du vent et rayonnement global). Toutefois, les données météorologiques disponibles à l'échelle locale ne sont pas suffisantes pour une étude des périodes de sécheresse (profondeur temporelle courte et présence des lacunes au niveau des séries d'observations). En effet, une étude de sécheresse nécessite une base de données assez longue, ayant une résolution temporelle fine pour des analyses robustes pour la détection des anomalies. Le premier objectif de la thèse est alors de prolonger les données météorologiques disponibles. Nous avons eu recours aux données de réanalyses qui présentent une estimation du climat observé à grande échelle, du fait qu'elles reposent sur des observations réelles du système climatique. Ces données sont disponibles depuis 1950, avec une résolution spatiale de dizaine de kilomètres et un pas de temps infra-journalier. Toutefois, ces données restent inadéquates pour les utiliser directement au niveau des modèles d'impacts, car elles présentent dans la plupart du temps des biais systématiques par rapport aux observations, due à la différence de résolution spatiale. Une étape intermédiaire, nommée descente d'échelle, doit être mise en œuvre. Cette approche consiste à dériver l'information climatique à haute résolution spatiale nécessaire aux modèles d'impact à partir des variables à basse résolution (réanalyses).

Dans ce contexte, nous avons développé un générateur stochastique de conditions météorologiques "MetGen", inspiré des travaux de Chandler (2020). MetGen est basé sur les données de réanalyses ERA5 (Hersbach, 2016), pour effectuer à la fois le comblement des données météorologiques existantes et générer des scénarios des différentes variables sur une longue période dans le passé. Il simule ces scénarios en se basant sur des modèles linéaires généralisés (GLM) en tirant profit des informations climatiques issues des différentes stations de mesure ayant une proximité géographique. Cela permet d'augmenter la base de données disponible et de fournir une information généralisée sur la zone d'étude. MetGen est un outil régional et multi-variables. Il nous permet de simuler les variables nécessaires pour forcer le modèle de bilan d'énergie (température de l'air, humidité relative de l'air, vitesse du vent, et rayonnement global), ainsi que d'autres variables très utilisées pour les études de changement climatiques (précipitations et pression atmosphérique). La simulation est effectuée en choisissant une loi de distribution appropriée et un jeu de covariables adéquats pour chaque variable météorologique d'une façon séparée. Grâce à ses covariables, nous avons pu adapter le générateur au contexte de l'estimation du stress hydrique en régions semi-arides, ce qui présente la valeur ajoutée de notre travail par rapport aux travaux de Chandler (2020). En effet, l'introduction des cycles saisonniers et diurnes ainsi que les données de ré-analyses, exploitées en tant que covariables, permettent de guider les simulations à une résolution infra-journalière (horaire ou demi-horaire). D'autres covariables sont utilisées afin d'introduire les dépendances inter-variables initialement observées, informations géographiques, et la persistance temporelle. La validation de notre méthode statistique est effectuée pendant une période en validation croisée qui couvre la période d'observation (2011-2016). Cette évaluation est réalisée en termes d'extension temporelle et de gap-filling, en effectuant 50 répliques. D'autre part, la méthode de générateur stochastique est comparée avec une autre méthode de descente d'échelle, la

méthode de correction de biais dans sa forme univariée en utilisant la fonction CDFt (Michelangeli *et al.*, 2009) et multivariée en utilisant la fonction MBCn (Cannon, 2018). Pour ces méthodes, nous avons travaillé plutôt sur les anomalies en enlevant les fluctuations systématiques des cycles diurnes et saisonniers. En effet, pour une saison donnée, nous observons toujours le même cycle. Ces scénarios climatiques issus de ces différentes sources (MetGen, CDFt, MBCn et ERA5 brutes) sont évalués en termes de leur capacité à bien reproduire les caractéristiques des variables climatiques observées in-situ, ainsi que leur capacité à reproduire la simulation de l'évapotranspiration et du stress hydrique avec un modèle de bilan d'énergie à double source SPARSE (Boulet *et al.*, 2015).

Les analyses ont montré que l'utilisation directe des données des réanalyses peut engendrer des valeurs biaisées par rapport aux observations. Ce biais est dû à un mélange d'effet d'échelle et des erreurs de modélisation météorologique. MetGen et les méthodes de corrections de biais ont réussi à corriger ce biais et à reproduire une variabilité très proche de celle des observations. Nous avons trouvé une performance très comparable entre ces deux méthodes, essentiellement en termes de reproduction de la distribution des variables météorologiques observées. La dépendance inter-variable est très bien reproduite par ces deux méthodes. Pour MetGen, ces relations sont bien conservées entre les variables grâce à la définition d'un schéma d'interdépendance illustrées par le graphique 3.1. Ces relations d'interdépendance sont inspirées par les liens entre les variables pré-définis pour l'application en RGLimclim au Royaume-Uni (Chandler, 2015). D'autre part, MetGen a réussi à bien simuler les variables au niveau des différentes stations et à tenir en compte de l'hétérogénéité observée au sein de ces stations, bien que nous n'ayons pas introduit un mécanisme pour tenir en compte de la dépendance inter-site. Un bon exemple, le pouvoir du générateur à simuler correctement la vitesse du vent au niveau de l'une des stations entourées par des brises-vent (station de Chebika). En effet, l'implication d'une covariable pour spécifier ce critère et différencier par conséquent les caractéristiques de cette station par rapport aux autres, nous a permis de simuler cette variable correctement au niveau des différents sites de mesure. Certes, la méthode de générateur stochastique nécessite plus de travail pour l'ajustement des variables, le choix de probabilité de distribution et des covariables à introduire par rapport aux méthodes de correction de biais, mais ce mécanisme de reconstitution des modèles statistiques offre plus de flexibilité pour la modélisation des variables météorologiques et permet de reproduire la variabilité saisonnière et inter-annuelle d'une façon plus réaliste grâce à l'introduction des différentes covariables. Ces derniers agissent sur le comportement des variables à l'échelle temporelle et spatiale, ce qui est très utile pour une longue période de simulation. Cependant, les méthodes de corrections de biais sont fortement reliées aux modèles climatiques et aux fluctuations observées pendant les périodes disponibles d'observation. Une bonne reproduction des variabilités saisonnières et diurnes par les méthodes de correction de biais revient au fait de conserver les cycles des observations, supposés constants pour chaque saison. Dans ce cas, la profondeur de la période de calibration est de grande importance pour une simulation assez fiable et assez robuste en utilisant les méthodes de correction de biais. D'autre part, le générateur stochastique offre l'avantage de simuler d'une façon stochastique différents scénarios climatiques, en fonction de la loi de distribution préalablement choisie pour chaque variable. En se basant sur ce critère, nous avons présenté que les quantiles faibles 2.5% et les

quantiles élevés 97.5% de l'indice de stress et l'évapotranspiration pour donner une estimation sur l'ordre des grandeurs de leurs simulations effectuées par SPARSE en utilisant les différentes réalisations de MetGen. La comparaison des différentes simulations de SPARSE en utilisant les différentes séries répliquées générées à partir de MetGen (seulement les quantiles 2.5 et 97.5 sont présentés), des deux méthodes de corrections de biais et ERA5 brutes montrent des différences notables. Cependant, les simulations de MetGen ont montré une capacité à bien reproduire les faibles valeurs de l'indice de stress thermiques qui sont très informatives pour la détection d'un début de stress hydrique. Le reste des méthodes sous-estiment les faibles valeurs de stress. Les fortes valeurs de l'indice de stress sont bien reproduites en utilisant MetGen et la méthode de correction de biais multivariée.

Suite à la validation de MetGen, nous avons utilisé une réalisation de cette méthode afin de simuler des variables météorologiques sur toute la période d'étude (2000-2019). Cette série prolongée dans le passé est utilisée pour contraindre le modèle SPARSE et simuler l'évapotranspiration et l'indice de stress qui en dérive. L'élaboration de l'information saisonnière à partir des données discontinues de l'évapotranspiration a été abordée au début de ce travail. Les techniques d'interpolation de ET font l'objet de plusieurs travaux dans la littérature. Cependant, l'utilisation de l'information issue des précipitations représente un des développements innovants de ce travail en continuité avec les travaux de Delogu *et al.* (2021). Cette innovation consiste à bien diriger les variations de ET, fortement dépendantes des précipitations pendant les jours pluvieux. En effet, dues à la présence de l'ennuage, les taux de ET vont être baissés énormément. Une interpolation classique va se baser sur ces faibles valeurs de ET, se contenter seulement de suivre les variations de la variable utilisée comme support pour l'interpolation ( $R_g$ ,  $R_c$ , etc.). Cependant, en réalité, les jours pluvieux font varier d'une façon très remarquable les valeurs de ET notamment au niveau des zones semi-arides, caractérisées par une demande évaporative conséquente pour les jours de ciel clair suivant l'épisode pluvieux. Il est important alors, d'introduire ces variations, et de forcer l'évolution de ET pour que ça soit plus réaliste au contexte de notre zone d'étude et aux variations des précipitations qui forment un facteur primordial dans la réponse des surfaces en termes de ET.

Pour l'indice thermique élaboré, nous avons gardé pour les comparaisons, la série ERA5 brute, pour évaluer la valeur ajoutée d'une méthode de descente d'échelle pour les modèles d'impact. Ces indices thermiques simulés à partir de MetGen  $SI_{SWG}$  ou à partir des données d'ERA5 brutes  $SI_{ERA5}$  sont comparés avec d'autres indices issus de différentes longueurs d'onde, le NDVI issu du visible/Proche infrarouge et SWI issu du micro-onde. Ces deux derniers indices sont considérés comme des indices de sécheresse agronomique vue leur liaison directe avec l'activité photosynthétique (pour le NDVI) et la teneur en eau au niveau de la zone racinaire (pour le SWI). Le deuxième objectif de la thèse est alors de comparer ces indices en termes de pertinence, de réalisme, de cohérence et de précocité en nous basant sur un indice de précipitations UPI en tant que référence pour une analyse à l'échelle régionale. Pour ce faire, nous avons effectué une standardisation qui utilise la fonction de répartition empirique de chaque indice afin de l'interpréter en termes de probabilité de non dépassement. A l'échelle régionale, nous avons présenté des séries chronologiques de ces indices afin d'évaluer leur comportement sur différentes échelles temporelles (saisonnière et décadaire). L'échelle sai-

sonnière (de Septembre jusqu'au Mai) est très représentative des variations inter-saisonnières du statut hydrique durant la période d'étude. Cependant, les analyses à une échelle plus fine (décadaire), illustrent au mieux le mécanisme de l'évolution des périodes de sécheresse et la réponse des indices pour ces périodes de déficit hydrique. Pour ces analyses, la définition d'un seuil de stress est de grande importance pour évaluer à la fois l'existence et l'intensité d'une période de stress. Nous avons défini alors un seuil de stress, de façon analogue aux classes de stress défini par SPI, qui correspond dans notre cas aux 20% des valeurs les plus faibles pour chaque indice.

Ces analyses ont montré qu'à l'échelle annuelle, les indices présentent des variations similaires et une cohérence pour la caractérisation des années (humide/sec) en se basant sur le seuil de stress fixé. A une échelle décadaire, nous observons une différence très nette au niveau de la réponse des différents indices. Alors que le NDVI montre une réponse tardive par rapport aux variations de UPI, les indices thermiques et SWI sont des indices précurseurs pour l'identification d'une période de stress. La sensibilité de ces indices liés à l'humidité du sol pour un déficit hydrique en comparaison à un indice basé sur le développement du couvert végétal a été prévisible. En effet, la réponse de la végétation pour un stress hydrique est effectuée généralement sur des plus longues amplitudes. Ainsi,  $SI_{SWG}$ ,  $SI_{ERA5}$  et SWI ont présenté une probabilité élevée pour identifier correctement le statut hydrique en concordance avec UPI, notamment en période de l'automne, hiver et printemps. Cependant, en été nous observons plus de dispersion en termes de concordance avec la réponse de UPI. En effet, les pluies sont plus efficaces en période de croissance végétative, d'où la bonne corrélation entre ces indices et l'indice de pluie en cette période. Cependant, en été tous les indices ne sont plus en concordance avec UPI, du fait de l'introduction de l'irrigation pour subvenir aux besoins de la végétation vue la rareté des précipitations dans notre zone d'étude en cette période de l'année. D'autre part, la comparaison de ces indices avec la moyenne mobile de UPI calculée sur différentes fenêtres temporelles a montré l'importance de tenir en compte des effets mémoires et d'appliquer un délai pour avoir la bonne réponse des indices.

Une analyse à l'échelle locale est aussi primordiale afin de voir si les indices gardent leurs performances en passant à une échelle spatiale plus fine, à l'échelle du kilomètre. Pour ces analyses, nous avons isolé quelques pixels qui correspondent au transect de XLAS ou bien qui correspondent à des parcelles de céréales pluviales. En effet, la sécheresse agronomique nous demande de regarder les cultures herbacées pluviales, les plus sensibles à la sécheresse. Toutefois, le contexte irrigué est aussi impacté par une sécheresse météorologique, qui induit une réduction d'évaporation du sol. La sécheresse hydrologique n'est pas visible dans notre cas d'étude grâce à la présence d'un aquifère profond bien accessible contrairement à d'autres zones dépendant des apports de surface (barrages etc) et qui sont donc plus sensibles aux variations climatiques. Par ailleurs, une étude sur les zones de céréales en sec est particulièrement informative sur la réponse de ces indices pour des périodes de sécheresse. Malgré une performance très comparable des deux indices thermiques  $SI_{SWG}$  et  $SI_{ERA5}$ , ces analyses nous ont permis d'approuver la performance de l'indice  $SI_{SWG}$  pour l'identification des périodes de stress. Ce fait témoigne de l'importance de l'utilisation des données météorologiques assez fines pour les modèles d'impacts.

Enfin, la dernière étape de ces travaux a consisté en un exercice d'évaluation de l'indice de

stress de la composante "végétation" (et non de la surface comme précédemment), calculé à partir de la transpiration. Pour ce faire, nous étions appelés à revoir la méthode de calcul de la transpiration potentielle par le modèle SPARSE. En effet, ce modèle se base sur la définition des coefficients d'efficacité évaporative pour le sol et pour la végétation afin de simuler un état de stress. Théoriquement, pour reproduire une condition potentielle de transpiration, ces deux coefficients doivent être mis à une valeur maximale. Cependant, cette réflexion nous a mené à une situation irréaliste, avec une  $E_{Tr}$  plus forte que l' $E_{Tp}$  et un indice de stress qui devrait être négatif. Ce constat nous a poussé à revoir le comportement de ces deux compartiments "sol" et "végétation" afin de reproduire des réponses de transpiration plus réalistes. L'hypothèse de Norman qui suppose que si la plante commence à souffrir de stress hydrique, le sol doit être déjà très sec en surface, était de grande utilité. Nous avons appliqué par ailleurs cette théorie, en appliquant un coefficient  $\beta_s$  du sol minimal (égal à 0) et un coefficient de végétation  $\beta_v$  maximal (égal à 1) afin d'exprimer un état potentiel de transpiration plus adéquat. Cette hypothèse a permis d'avoir des résultats plus cohérents.

Comme n'importe quel exercice de recherche, ce travail présente un certain nombre de limites, qu'elles soient inhérentes à la précision et à la représentativité des résultats ou à la représentation des processus. L'élaboration d'une station représentative de toute la plaine afin d'avoir une information généralisée sur les conditions météorologiques de la zone d'étude était de grande utilité pour simuler de l'information météorologique assez longue et continue pour l'utiliser comme forçage climatique pour le modèle SPARSE. Cependant, avoir une seule information météorologique pour toute la zone, bien qu'elle présente une similarité des propriétés climatiques, reste insuffisante pour reproduire avec précision les conditions de stress à l'échelle kilométrique. En effet, la variation observée des valeurs de stress simulées par SPARSE revient aux informations satellitaires qui englobent la température de surface, l'état de développement de la végétation et le type d'occupation du sol issus des variables utilisées comme entrées du modèle LST, albédo, NDVI et LAI. Ces variations issues de l'information satellitaire sont très informatives pour la simulation des conditions de stress. Mais avoir aussi une information météorologique à une échelle kilométrique sur toute la zone d'étude sera de grande utilité pour le modèle d'impact qui est très sensible aux variations météorologiques. Un bon exemple, SPARSE est très sensible pour la hauteur de mesure du vent. De même, cette variable qui est caractérisée par sa turbulence et sa variabilité doit être simulée à une échelle spatiale très fine. D'autre part, toujours dans le cadre de la simulation des données météorologiques, nous avons défini des relations d'inter-dépendance des variables (présentée par le graphique 3.1) qui, à notre avis, présentent des relations physiques cohérentes et robustes. Cependant, nous n'avons pas testé l'incidence du choix du graphe de dépendance pour induire la dépendance entre les variables. D'autre part, nous n'avons pas réussi à bien reproduire les précipitations en termes d'occurrence et d'intensité. Cette variable n'était pas utilisée pour contraindre le modèle de bilan d'énergie, mais elle présente l'une des variables les plus importantes dans l'étude des changements climatiques. Nous avons utilisé une loi de probabilité gamma pour modéliser son intensité. Cette loi de probabilité peut bien s'ajuster avec les faibles valeurs d'intensité de la pluie, mais elle peut engendrer des écarts dès que les valeurs de la pluie passent au delà d'une certaine valeur qui diffère selon les caractéristiques de la zone. Par ailleurs, l'utilisation de gamma pour modéliser les précipitations peut engendrer

des sous-estimations des évènements extrêmes de pluie notamment au niveau des zones semi-arides caractérisées à la fois par les faibles fréquences des précipitations et par leurs fortes amplitudes. Ce travail consiste aussi à la détection du stress en fixant des seuils pour définir les périodes de stress. Nous avons ajusté ce seuil en équivalence au seuil des classes de stress défini par l'indice SPI (16%). D'autre part, pour la définition d'une période de stress nous avons identifié qu'il s'agit d'une période de stress pour au moins deux décades en choisissant une fenêtre glissante de 3 décades. Le choix des seuils ainsi que la définition de la période de stress fait l'objet d'un choix subjectif, tout comme pour d'autres travaux sur la détection du stress, on cite à titre d'exemple Farahmand *et al.* (2015) qui ont choisi pour leur travail un seuil de 30% d'occurrence d'un évènement de stress. Ces choix sont différents d'un travail à un autre, ce qui traduit la complexité de l'exercice de la détection de stress.

Malgré ces limites, l'approche proposée dans ce travail pour l'extension des données météorologiques et pour l'élaboration d'un indice de stress robuste pour la détection du stress d'une manière précoce a montré plusieurs avantages dans son application. Toutefois, cette approche laisse en ce sens envisager certaines perspectives quant à son utilisation dans le futur, ainsi que certaines questions scientifiques abordées au long de ce travail auxquelles il faut répondre.

Le travail mené au début de la thèse pour prolonger les données météorologiques à l'échelle temporelle doit être complété par une extension spatiale, afin d'avoir des données d'évapotranspiration et un indice de stress spatialisés sur toute la zone d'étude. Ce travail de spatialisations peut être effectué à l'aide d'un modèle qui tient en compte des propriétés de surface et de sa variabilité physique (relief, occupation du sol), comme le modèle régional Weather Research and Forecasting WRF. Ce modèle climatique est utilisé à la fois en recherche et en prévision opérationnelle du temps. Nous disposons des simulations de WRF sur 3 ans (de 2014 jusqu'au 2016), présentant une résolution spatiale de trois kilomètres et une résolution temporelle demi-horaire. Dans ce contexte, nous pouvons exploiter la structure spatiale des simulations WRF pour interpoler spatialement les séries aux stations afin d'avoir des données qui couvrent la totalité de la période d'étude (2000-2019). D'autre part, une étude de la sensibilité de SPARSE par rapport à ses variables d'entrées satellitaires et météorologiques peut être aussi de grande utilité pour améliorer la précision des simulations.

Les résultats générés par l'interpolation de ET en utilisant la méthode de pluie antérieure API, en tant que support d'interpolation se sont révélés pertinents en comparaison avec les données reconstituées à partir des mesures du scintillomètre à large ouverture (XLAS). Ce fait nous incite à mieux travailler cette méthode, largement sous-exploitée dans la littérature malgré sa performance. Cette approche peut être par ailleurs une candidate intéressante pour une méthode opérationnelle d'interpolation pour les futures missions spatiales de haute résolution telles que TRISHNA ou LSTM pour lesquelles on souhaite un produit ET journalier. Par ailleurs, améliorer les données de précipitations simulées par MetGen, paraît aussi inévitable pour pouvoir les utiliser en tant que support pour cette méthode. Nous pouvons choisir une procédure plus complexe comme la méthode utilisée dans les travaux de Carreau & Vrac (2011).

Notre travail a pour objectif d'étudier la sécheresse agronomique. Les indices utilisés sont considérés comme des indices agronomiques vu leur liaison directe avec la végétation (SI et

NDVI) et la teneur en eau au niveau de la zone racinaire (SWI). L'avantage de l'utilisation des modèles à double source est la simulation d'un état de stress plus précis notamment au niveau des zones semi-arides caractérisées par un couvert végétal dispersé et une contribution inégale des composantes d'évaporation et de transpiration, qui varient à la fois dans le temps et l'espace. Vu l'intérêt à détecter spécifiquement le stress de la végétation, nous avons élaboré à la fin de ce travail un indice de stress de la végétation. À nos connaissances, un indice de stress thermique basé seulement sur la composante de la végétation, n'est jamais abordé dans la littérature. Les résultats présentés nous laissent réfléchir à des études complémentaires afin de dépasser la difficulté de cibler le stress pour les couverts épars. En effet, la manière de définir les coefficients d'efficience pour SPARSE afin de simuler une transpiration potentielle peut jouer un rôle très important pour une simulation plus réaliste de cet indice. D'autre part, l'identification des parcelles de céréales en sec existantes au niveau de la région durant notre période d'étude sera de grande utilité pour avoir un signal de stress détectable et des réponses des indices plus cohérentes. Nous proposons alors, comme continuité de ce travail, d'établir un historique des parcelles pluviales au niveau de notre zone d'étude. Ce travail peut être effectué à l'aide des images SPOT. Ainsi donc, une étude de l'indice de stress de la végétation peut être accomplie.

- AghaKouchak, Amir. 2017. Advancements in Satellite Remote Sensing for Drought Monitoring. *Pages 225–258 of : Drought and Water Crises*. CRC Press.
- Alazard, Marina, Leduc, Christian, Travi, Y, Boulet, Gilles, & Salem, A Ben Salem. 2015. Estimating evaporation in semi-arid areas facing data scarcity : Example of the El Haouareb dam (Merguellil catchment, Central Tunisia). *Journal of Hydrology : Regional Studies*, **3**, 265–284.
- Alboghday, Mohamed, & El-Hendawy, Salah. 2016. Economic impacts of climate change and variability on agricultural production in the Middle East and North Africa region. *International Journal of Climate Change Strategies and Management*.
- Allen, Richard G., Pereira, Luis S, Raes, Drik., & Smith, Martin. 1998. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *FAO, Rome*, **300**(9), D05109.
- Allen, Richard G, Tasumi, Masahiro, & Trezza, Ricardo. 2007. Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC)—Model. *Journal of irrigation and drainage engineering*, **133**(4), 380–394.
- Alley, William M. 1984. The Palmer drought severity index : limitations and assumptions. *Journal of Applied Meteorology and Climatology*, **23**(7), 1100–1109.
- Amigues, Jean-Pierre, Debaeke, Philippe, Itier, Bernard, Lemaire, Gilles, *et al.* 2006. *Sécheresse et agriculture. Réduire la vulnérabilité de l'agriculture à un risque accru de manque d'eau. Expertise scientifique collective. Rapport*. Ph.D. thesis, INRA.
- Amri, Rim, Zribi, Mehrez, Lili-Chabaane, Zohra, Duchemin, Benoit, Gruhier, Claire, & Chehbouni, Abdelghani. 2011. Analysis of vegetation behavior in a North African semi-arid region, using SPOT-VEGETATION NDVI data. *Remote sensing*, **3**(12), 2568–2590.
- Amri, Rim, Zribi, Mehrez, Lili-Chabaane, Zohra, Wagner, Wolfgang, & Hasenauer, Stefan. 2012. Analysis of C-band scatterometer moisture estimations derived over a semiarid region. *IEEE transactions on geoscience and remote sensing*, **50**(7), 2630–2638.
- Anderson, Martha, & Kustas, William. 2008. Thermal remote sensing of drought and evapotranspiration. *Eos, Transactions American Geophysical Union*, **89**(26), 233–234.
- Anderson, Martha C, Norman, John M, Mecikalski, John R, Otkin, Jason A, & Kustas, William P. 2007. A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing : 1. Model formulation. *Journal of Geophysical Research : Atmospheres*, **112**(D10).

- Anderson, Martha C, Hain, Christopher, Wardlow, Brian, Pimstein, Agustin, Mecikalski, John R, & Kustas, William P. 2011a. Evaluation of drought indices based on thermal remote sensing of evapotranspiration over the continental United States. *Journal of Climate*, **24**(8), 2025–2044.
- Anderson, Martha C, Kustas, William P, Norman, John M, Hain, Christopher R, Mecikalski, John R, Schultz, L, Gonzalez-Dugo, Maria Pat, Cammalleri, C, d’Urso, Guido, Pimstein, A, *et al.* 2011b. Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery. *Hydrology and Earth System Sciences*, **15**(1), 223–239.
- Anderson, Martha C, Zolin, Cornelio A, Sentelhas, Paulo C, Hain, Christopher R, Semmens, Kathryn, Yilmaz, M Tugrul, Gao, Feng, Otkin, Jason A, & Tetrault, Robert. 2016. The Evaporative Stress Index as an indicator of agricultural drought in Brazil : An assessment based on crop yield impacts. *Remote Sensing of Environment*, **174**, 82–99.
- AQUASTAT, FAO. 2015. *Aquastat*.
- Arnell, NW, & Reynard, NS. 1996. The effects of climate change due to global warming on river flows in Great Britain. *Journal of hydrology*, **183**(3-4), 397–424.
- Asrar, Ghassem R, Lucas, Paul, van Vuuren, Detlef, Pereira, Laura, Vervoort, Joost, Bhargava, Rohan, *et al.* 2019. Outlooks in geo-6-global environment outlook (geo-6) : Healthy planet, healthy people chapter 19. *Global Environment Outlook (GEO-6) : Healthy Planet, Healthy People*.
- Ayar, Pradeebane Vaittinada, Vrac, Mathieu, Bastin, Sophie, Carreau, Julie, Déqué, Michel, & Gallardo, Clemente. 2016. Intercomparison of statistical and dynamical downscaling models under the EURO-and MED-CORDEX initiative framework : present climate evaluations. *Climate dynamics*, **46**(3-4), 1301–1329.
- Babaeian, Ebrahim, Sadeghi, Morteza, Jones, Scott B, Montzka, Carsten, Vereecken, Harry, & Tuller, Markus. 2019. Ground, proximal, and satellite remote sensing of soil moisture. *Reviews of Geophysics*, **57**(2), 530–616.
- Baccour, H., Feki, H., Slimani, M., & Cudennec, C. 2012. Interpolation de l’évapotranspiration de référence en Tunisie par la méthode de krigeage ordinaire. *Science et changements planétaires/Sécheresse*, **23**(2), 121–132.
- Baghdadi, N, Cresson, R, El Hajj, M, Ludwig, R, & La Jeunesse, I. 2012. Soil parameters estimation over bare agriculture areas from C-band polarimetric SAR data using neural networks. *Hydrology & Earth System Sciences Discussions*, **9**(3).
- Balsamo, Gianpaolo. 2012. *Apport de la modélisation et de l’assimilation de données pour la caractérisation des surfaces continentales en prévision numérique du temps*. Ph.D. thesis, Université Paul Sabatier-Toulouse III.

- Bannari, Abdou, Morin, D, Bonn, F, & Huete, AjRsr. 1995. A review of vegetation indices. *Remote sensing reviews*, **13**(1-2), 95–120.
- Baret, Frédéric, Guyot, G, Bégué, Agnès, Maurel, P, & Podaire, A. 1986. Etude de la complémentarité du moyen infrarouge avec le visible et le proche infrarouge pour le suivi de la végétation. In : 3. *International Colloquium*, vol. 247. ASE.
- Bastiaanssen, Wim GM, Menenti, Massimo, Feddes, RA, & Holtslag, AAM. 1998. A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation. *Journal of hydrology*, **212**, 198–212.
- Bausch, Walter C. 1993. Soil background effects on reflectance-based crop coefficients for corn. *Remote Sensing of Environment*, **46**(2), 213–222.
- Bento, Virgílio A, Gouveia, Célia M, DaCamara, Carlos C, & Trigo, Isabel F. 2018. A climatological assessment of drought impact on vegetation health index. *Agricultural and Forest Meteorology*, **259**, 286–295.
- Berliner, P, Oosterhuis, DM, & Green, GIC. 1984. Evaluation of the infrared thermometer as a crop stress detector. *Agricultural and Forest meteorology*, **31**(3-4), 219–230.
- Besbes, Mustapha, Chahed, Jamel, Shayeb, Hedi, & Hamdane, Abdelkader. 2013. L'eau en Tunisie. *Techniques de l'ingénieur Technologies de l'eau*.
- Besbes, Mustapha, Chahed, Jamel, & Hamdane, Abdelkader. 2014. *Sécurité hydrique de la Tunisie*. L'HARMATTAN.
- Bhuiyan, C, Singh, RP, & Kogan, FN. 2006. Monitoring drought dynamics in the Aravalli region (India) using different indices based on ground and remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, **8**(4), 289–302.
- Boe, Julien. 2007. *Changement global et cycle hydrologique : Une étude de régionalisation sur la France*. Ph.D. thesis, Université Paul Sabatier-Toulouse III.
- Boulet, Gilles, Olioso, Albert, Ceschia, Eric, Marloie, Olivier, Coudert, Benoît, Rivalland, Vincent, Chirouze, Jonas, & Chehbouni, Ghani. 2012. An empirical expression to relate aerodynamic and surface temperatures for use within single-source energy balance models. *Agricultural and Forest Meteorology*, **161**, 148–155.
- Boulet, Gilles., Mougénot, Bernard., Lhomme, Jean-Pierre., Fanise, Pascal., Lili-Chabaane, Zohra., Olioso, Albert., Bahir, Malik., Rivalland, Vincent., Jarlan, Lionel., Merlin, Olivier., et al. 2015. The SPARSE model for the prediction of water stress and evapotranspiration components from thermal infra-red data and its evaluation over irrigated and rainfed wheat. *Hydrology and Earth System Sciences Discussions*, 4653–4672.
- Brochet, PATRICK. 1977. La sécheresse 1976 en France : aspects climatologiques et conséquences/The 1976 drought in France : climatological aspects and consequences. *Hydrological Sciences Journal*, **22**(3), 393–411.

- Burgin, Mariko S, Colliander, Andreas, Njoku, Eni G, Chan, Steven K, Cabot, Francois, Kerr, Yann H, Bindlish, Rajat, Jackson, Thomas J, Entekhabi, Dara, & Yueh, Simon H. 2017. A comparative study of the SMAP passive soil moisture product with existing satellite-based soil moisture products. *IEEE Transactions on Geoscience and Remote Sensing*, **55**(5), 2959–2971.
- Cannon, Alex J. 2018. Multivariate quantile mapping bias correction : an N-dimensional probability density function transform for climate model simulations of multiple variables. *Climate dynamics*, **50**(1), 31–49.
- Carlson, Toby. 2007. An overview of the "triangle method" for estimating surface evapotranspiration and soil moisture from satellite imagery. *Sensors*, **7**(8), 1612–1629.
- Carreau, Julie, & Vrac, Mathieu. 2011. Stochastic downscaling of precipitation with neural network conditional mixture models. *Water Resources Research*, **47**(10).
- Carter, Gary M, Dallavalle, J Paul, & Glahn, Harry R. 1989. Statistical forecasts based on the National Meteorological Center's numerical weather prediction system. *Weather and Forecasting*, **4**(3), 401–412.
- Chahbi, Aicha, Zribi, Mehrez, Lili-Chabaane, Zohra, Duchemin, Benoit, Shabou, Marouen, Mougenot, Bernard, & Boulet, Gilles. 2014. Estimation of the dynamics and yields of cereals in a semi-arid area using remote sensing and the SAFY growth model. *International Journal of Remote Sensing*, **35**(3), 1004–1028.
- Chahbi, Aicha, Zribi, Mehrez, Lili-Chabaane, Zohra, & Mougenot, Bernard. 2015. Forecasting of cereals yields in a semi-arid area using the agrometeorological model «SAFY» combined to optical SPOT/HRV images. *Page 963729 of : Remote Sensing for Agriculture, Ecosystems, and Hydrology XVII*, vol. 9637. International Society for Optics and Photonics.
- Chahbi, Aicha, Zribi, Mehrez, & Lili-Chabaane, Zohra. 2016. Remote sensing and modelling of vegetation dynamics for early estimation and spatial analysis of grain yields in semiarid context in central Tunisia. *Pages EPSC2016–8761 of : EGU General Assembly Conference Abstracts*.
- Chahbi Bellakanji, Aicha, Zribi, Mehrez, Lili-Chabaane, Zohra, & Mougenot, Bernard. 2018. Forecasting of cereal yields in a semi-arid area using the simple algorithm for yield estimation (SAFY) agro-meteorological model combined with optical SPOT/HRV images. *Sensors*, **18**(7), 2138.
- Chandler, Richard. 2015. A multisite, multivariate daily weather generator based on Generalized Linear Models. *User guide : R package*.
- Chandler, Richard. E. 2020. Multisite, multivariate weather generation based on generalised linear models. *Environmental Modelling & Software*, **134**, 104867.

- Chirouze, J., Boulet, G., Jarlan, L., Fieuzal, R., Rodriguez, J.C., Ezzahar, J., Raki, S. Er., Bigeard, G., Merlin, O., Garatuza-Payan, J., *et al.* 2014. Intercomparison of four remote-sensing-based energy balance methods to retrieve surface evapotranspiration and water stress of irrigated fields in semi-arid climate. *Hydrology and Earth System Sciences Discussions*, 1165–1188.
- Clevers, JGPW. 1989. Application of a weighted infrared-red vegetation index for estimating leaf area index by correcting for soil moisture. *Remote Sensing of Environment*, **29**(1), 25–37.
- Daki, Yousra, Zahour, Ghalem, Lachgar, Rahal, & El Hadi, Hassan. 2016. Caractérisation de la sécheresse climatique du bassin versant d'oum errbia (Maroc) par le biais de l'indice de précipitation standardisé (SPI). *European Scientific Journal*, **12**(14).
- Dash, Prasanjit, Göttsche, Frank-M, Olesen, Folke-S, & Fischer, Herbert. 2001. Retrieval of land surface temperature and emissivity from satellite data : physics, theoretical limitations and current methods. *Journal of the Indian Society of Remote Sensing*, **29**(1), 23–30.
- De Oliveira, A Costa, Marini, N, & Farias, DR. 2014. Climate change : New breeding pressures and goals. *Encyclopedia of Agriculture and Food Systems*, 284–293.
- Delogu, Emilie, Olioso, Albert, Alliès, Aubin, Demarty, Jérôme, & Boulet, Gilles. 2021. Evaluation of Multiple Methods for the Production of Continuous Evapotranspiration Estimates from TIR Remote Sensing. *Remote Sensing*, **13**(6), 1086.
- Diarra, A, Jarlan, Lionel, Er-Raki, S, Le Page, M, Aouade, G, Tavernier, A, Boulet, Gilles, Ezzahar, J, Merlin, Olivier, & Khabba, S. 2017. Performance of the two-source energy budget (TSEB) model for the monitoring of evapotranspiration over irrigated annual crops in North Africa. *Agricultural Water Management*, **193**, 71–88.
- Doorenbos, J, & Pruitt, Wo. 1997. La petite irrigation dans les zones arides Principes et options. Collection FAO : développement 2. Rome. Italie.
- Dorigo, Wouter, Wagner, Wolfgang, Albergel, Clement, Albrecht, Franziska, Balsamo, Gianpaolo, Brocca, Luca, Chung, Daniel, Ertl, Martin, Forkel, Matthias, Gruber, Alexander, *et al.* 2017. ESA CCI Soil Moisture for improved Earth system understanding : State-of-the art and future directions. *Remote Sensing of Environment*, **203**, 185–215.
- Entekhabi, Dara, Reichle, Rolf H, Koster, Randal D, & Crow, Wade T. 2010. Performance metrics for soil moisture retrievals and application requirements. *Journal of Hydrometeorology*, **11**(3), 832–840.
- Fahad, Shah, Bajwa, Ali A, Nazir, Usman, Anjum, Shakeel A, Farooq, Ayesha, Zohaib, Ali, Sadia, Sehrish, Nasim, Wajid, Adkins, Steve, Saud, Shah, *et al.* 2017. Crop production under drought and heat stress : plant responses and management options. *Frontiers in plant science*, **8**, 1147.

- FAO, Rome. 2011. Prospects for food, nutrition, agriculture and major commodity groups. *World agriculture : towards, 2030*, 2050.
- Farahmand, Alireza, AghaKouchak, Amir, & Teixeira, Joao. 2015. A vantage from space can detect earlier drought onset : An approach using relative humidity. *Scientific reports*, **5**(1), 1–6.
- Fathalli, Bilel, Pohl, Benjamin, Castel, Thierry, & Safi, Mohamed Jomâa. 2019. Errors and uncertainties in regional climate simulations of rainfall variability over Tunisia : a multi-model and multi-member approach. *Climate Dynamics*, **52**(1), 335–361.
- Fisher, Joshua B, Melton, Forrest, Middleton, Elizabeth, Hain, Christopher, Anderson, Martha, Allen, Richard, McCabe, Matthew F, Hook, Simon, Baldocchi, Dennis, Townsend, Philip A, *et al.* 2017. The future of evapotranspiration : Global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources. *Water Resources Research*, **53**(4), 2618–2626.
- Friston, Karl John, Frith, Christopher Donald, & Frackowiak, RSJ. 1993. Principal component analysis learning algorithms : a neurobiological analysis. *Proceedings of the Royal Society of London. Series B : Biological Sciences*, **254**(1339), 47–54.
- Gallego-Elvira, Belen, Oliosio, Albert, Mira, Maria, Reyes-Castillo, Sergio, Boulet, Gilles, Marloie, Olivier, Garrigues, Sébastien, Courault, Dominique, Weiss, Marie, Chauvelon, Philippe, *et al.* 2013. EVASPA (EVapotranspiration assessment from SPACe) tool : An overview. *Procedia Environmental Sciences*, **19**, 303–310.
- Gentine, Pierre, Entekhabi, Dara, Chehbouni, Abdelghani, Boulet, Gilles, & Duchemin, Benoît. 2007. Analysis of evaporative fraction diurnal behaviour. *Agricultural and forest meteorology*, **143**(1-2), 13–29.
- Gerber, Nicolas, & Mirzabaev, Alisher. 2017. Benefits of Action and Costs of Inaction : Drought Mitigation and Preparedness—A Literature Review. Integrated Drought Management Programme Working Paper No. 1. *World Meteorological Organization, Geneva, Switzerland and Global Water Partnership, Stockholm, Sweden.*
- Giglio, Louis, Csiszar, Ivan, & Justice, Christopher O. 2006. Global distribution and seasonality of active fires as observed with the Terra and Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) sensors. *Journal of geophysical research : Biogeosciences*, **111**(G2).
- Granger, Clive, & Lin, Jin-Lung. 1994. Using the mutual information coefficient to identify lags in nonlinear models. *Journal of time series analysis*, **15**(4), 371–384.
- Grouillet, B, Ruelland, D, Ayar, PV, & Vrac, M. 2015. Sensitivity analysis of runoff modeling to statistical downscaling models in the western Mediterranean. *Hydrology & Earth System Sciences Discussions*, **12**(10).

- Guan, Huade, & Wilson, John L. 2009. A hybrid dual-source model for potential evaporation and transpiration partitioning. *Journal of Hydrology*, **377**(3-4), 405–416.
- Guo, Yiping, & Senior, Matthew J. 2006. Climate model simulation of point rainfall frequency characteristics. *Journal of Hydrologic Engineering*, **11**(6), 547–554.
- Guttman, Nathaniel B. 1998. Comparing the palmer drought index and the standardized precipitation index 1. *JAWRA Journal of the American Water Resources Association*, **34**(1), 113–121.
- Guyot, Lionel. 1964. Etude pédologique du Djebel Fadeloun (Bled Sisseb). *ORSTOM*, 39 p. *multigr.*
- Herrera, Edgar, Ouarda, Taha, & Bobée, Bernard. 2006. Méthodes de désagrégation appliquées aux modèles du climat global Atmosphere-Océan (MCGAO). *Revue des sciences de l'eau/Journal of Water Science*, **19**(4), 297–312.
- Hersbach, H., de Rosnay, P., Bell, B., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Alonso-Balmaseda, M., Balsamo, G., Bechtold, P., Berrisford, P., Bidlot, J.-R., de Boissés, E., Bonavita, M., Browne, P., Buizza, R., Dahlgren, P., Dee, D., Dragani, R., Diamantakis, M., Flemming, J., Forbes, R., Geer, A. J., Haiden, T., Hólm, E., Haimberger, L., Hogan, R., Horányi, A., Janiskova, M., Laloyaux, P., Lopez, P., Muñoz-Sabater, J., Peubey, C., Radu, R., Richardson, D., Thépaut, J. N., Vitart, F., Yang, X., Zsótér, E., & Zuo, H. 2018. *Operational global reanalysis : progress, future directions and synergies with NWP*. ERA Report Series.
- Hersbach, Hans. 2016. The ERA5 Atmospheric Reanalysis. *Pages NG33D-01 of : AGU fall meeting abstracts*, vol. 2016.
- Hirabayashi, Yukiko, Mahendran, Roobavannan, Koirala, Sujana, Konoshima, Lisako, Yamazaki, Dai, Watanabe, Satoshi, Kim, Hyungjun, & Kanae, Shinjiro. 2013. Global flood risk under climate change. *Nature climate change*, **3**(9), 816–821.
- Hoffmann, Lars, Günther, Gebhard, Li, Dan, Stein, Olaf, Wu, Xue, Griessbach, Sabine, Heng, Yi, Konopka, Paul, Müller, Rolf, Vogel, Bärbel, *et al.* 2019. From ERA-Interim to ERA5 : the considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations. *Atmospheric Chemistry and Physics*, **19**(5), 3097–3124.
- Huete, AR, Jackson, RD, & Post, DF. 1985. Spectral response of a plant canopy with different soil backgrounds. *Remote sensing of environment*, **17**(1), 37–53.
- Idso, SB, Jackson, RD, Ehler, WL, & Mitchell, ST. 1977. Infrared emittance determination of leaves. *Ecology*, **50**, 899–902.
- IPCC. 2021. Climate Change 2021 : The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change ; Technical Summary. *Climate Change 2021*.

- Jackson, R. D., Idso, SB., Reginato, RJ., & Pinter Jr, PJ. 1981. Canopy temperature as a crop water stress indicator. *Water resources research*, **17**(4), 1133–1138.
- Jackson, R D, Hatfield, J L, Reginato, RJ, Idso, SB, & Pinter Jr, PJ. 1983. Estimation of daily evapotranspiration from one time-of-day measurements. *Agricultural Water Management*, **7**(1-3), 351–362.
- Jackson, Ray D, Reginato, RJ, & Idso, SerwoodB. 1977. Wheat canopy temperature : a practical tool for evaluating water requirements. *Water resources research*, **13**(3), 651–656.
- Jiao, Wenzhe, Wang, Lixin, & McCabe, Matthew F. 2021. Multi-sensor remote sensing for drought characterization : current status, opportunities and a roadmap for the future. *Remote Sensing of Environment*, **256**, 112313.
- Jiménez-Donaire, María del Pilar, Tarquis, Ana, & Giráldez, Juan Vicente. 2020. Evaluation of a combined drought indicator and its potential for agricultural drought prediction in southern Spain. *Natural Hazards and Earth System Sciences*, **20**(1), 21–33.
- Justice, Christopher O, Vermote, Eric, Townshend, John RG, Defries, Ruth, Roy, David P, Hall, Dorothy K, Salomonson, Vincent V, Privette, Jeffrey L, Riggs, George, Strahler, Alan, *et al.* 1998. The Moderate Resolution Imaging Spectroradiometer (MODIS) : Land remote sensing for global change research. *IEEE transactions on geoscience and remote sensing*, **36**(4), 1228–1249.
- Kassouk, Z, Mabrouki, R, Zayani, H, Mougnot, B, & Lili-chabaane, Z. 2018. Cartographie multi temporelle des occupations du sol par télédétection et SIG : Cas de la plaine de Merguellil (Tunisie). *In : International Scientific Symposium INAT - 120th Anniversary, 17-19 octobre 2018.*
- Kennedy, PJ. 1989. Monitoring the phenology of Tunisian grazing lands. *International Journal of Remote Sensing*, **10**(4-5), 835–845.
- Kerr, Yann H, Waldteufel, Philippe, Wigneron, J-P, Martinuzzi, JAMJ, Font, Jordi, & Berger, Michael. 2001. Soil moisture retrieval from space : The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE transactions on Geoscience and remote sensing*, **39**(8), 1729–1735.
- Klein, William H. 1982. Statistical weather forecasting on different time scales. *Bulletin of the American Meteorological Society*, 170–177.
- Kogan, F, Stark, R, Gitelson, A, Jargalsaikhan, L, Dugrajav, C, & Tsooj, S. 2004. Derivation of pasture biomass in Mongolia from AVHRR-based vegetation health indices. *International Journal of Remote Sensing*, **25**(14), 2889–2896.
- Kogan, Felix N. 1990. Remote sensing of weather impacts on vegetation in non-homogeneous areas. *International Journal of Remote Sensing*, **11**(8), 1405–1419.

- Kogan, Felix N. 1995. Application of vegetation index and brightness temperature for drought detection. *Advances in space research*, **15**(11), 91–100.
- Kogan, Felix N. 1997. Global drought watch from space. *Bulletin of the American Meteorological Society*, **78**(4), 621–636.
- Kogan, Felix N. 2001. Operational space technology for global vegetation assessment. *Bulletin of the American meteorological society*, **82**(9), 1949–1964.
- Kustas, William P, & Norman, John M. 1999. Evaluation of soil and vegetation heat flux predictions using a simple two-source model with radiometric temperatures for partial canopy cover. *Agricultural and Forest Meteorology*, **94**(1), 13–29.
- Lagouarde, Jean-Pierre, & Boulet, Gilles. 2016. Energy balance of continental surfaces and the use of surface temperature. *Pages 323–361 of : Land Surface Remote Sensing in Continental Hydrology*. Elsevier.
- Le Goulven, Patrick, Leduc, Christian, Bachta, Mohamed Salah, & Poussin, Jean-Christophe. 2009. Sharing scarce resources in a Mediterranean river basin : Wadi Merguellil in Central Tunisia. *River Basin Trajectories : Societies, Environments and Development*, Wallingford, CABI, 147–170.
- Le Page, Michel, & Zribi, Mehrez. 2019. Analysis and predictability of drought in Northwest Africa using optical and microwave satellite remote sensing products. *Scientific reports*, **9**(1), 1–13.
- Leduc, C., Ammar, S. Ben, Favreau, G., Beji, R., Virrion, R, Lacombe, G., Tarhouni, J, Aouadi, C, Chelli, B.Z., Jebnoun, N, *et al.* 2007. Impacts of hydrological changes in the Mediterranean zone : environmental modifications and rural development in the Merguellil catchment, central Tunisia/ Un exemple d'évolution hydrologique en Méditerranée : impacts des modifications environnementales et du développement agricole dans le bassin-versant du Merguellil (Tunisie centrale). *Hydrological Sciences Journal/Journal des Sciences Hydrologiques*, **52**(6), 1162–1178.
- Lili, Z, Duchesne, J, Nicolas, Hervé, Rivoal, Roger, & de BREGER, P. 1991. Détection infrarouge thermique des maladies du blé d'hiver 1. *Eppo Bulletin*, **21**(3), 659–672.
- Liu, Xiaoli, Coulibaly, P, & Evora, N. 2008. Comparison of data-driven methods for downscaling ensemble weather forecasts. *Hydrology and Earth System Sciences*, **12**(2), 615–624.
- Lohou, F, Kergoat, Laurent, Guichard, F, Boone, Aaron, Cappelaere, Bernard, Cohard, J-M, Demarty, Jérôme, Galle, Sylvie, Grippa, Manuela, Peugeot, Christophe, *et al.* 2014. Surface response to rain events throughout the West African monsoon. *Atmospheric Chemistry and Physics*, **14**(8), 3883–3898.
- Loudière, Daniel, & Gourbesville, Philippe. 2020. *Rapport mondial des Nations Unies sur la mise en valeur des ressources en eau 2020-L'eau et les changements climatiques*.

- Luquet, Delphine. 2002. *Suivi de l'état hydrique des plantes par infrarouge thermique : analyse expérimentale et modélisation 3D de la variabilité des températures au sein d'une culture en rang de cotonniers*. Ph.D. thesis, INA-PG.
- Luquet, Delphine, Vidal, Alain, Dauzat, Jean, Bégué, Agnès, Olios, Albert, & Clouvel, Pascal. 2004. Using directional TIR measurements and 3D simulations to assess the limitations and opportunities of water stress indices. *Remote Sensing of Environment*, **90**(1), 53–62.
- Lutz, Wolfgang, Sanderson, Warren, & Scherbov, Sergei. 2001. The end of world population growth. *Nature*, **412**(6846), 543–545.
- Maraun, Douglas, Wetterhall, F, Ireson, AM, Chandler, RE, Kendon, EJ, Widmann, Martin, Brienen, S, Rust, HW, Sauter, T, Themeßl, M, *et al.* 2010. Precipitation downscaling under climate change : Recent developments to bridge the gap between dynamical models and the end user. *Reviews of geophysics*, **48**(3).
- Massuel, Sylvain, & Riaux, Jeanne. 2017. Groundwater overexploitation : why is the red flag waved ? Case study on the Kairouan plain aquifer (central Tunisia). *Hydrogeology Journal*, **25**(6), 1607–1620.
- McKee, Thomas B, Doesken, Nolan J, Kleist, John, *et al.* 1993. The relationship of drought frequency and duration to time scales. *Pages 179–183 of : Proceedings of the 8th Conference on Applied Climatology*, vol. 17. Boston.
- Mega, Nabil, & Medjerab, Abderrahmane. 2021. Statistical comparison between the standardized precipitation index and the standardized precipitation drought index. *Modeling Earth Systems and Environment*, **7**(1), 373–388.
- Michelangeli, P-A, Vrac, Matthieu, & Loukos, Harilaos. 2009. Probabilistic downscaling approaches : Application to wind cumulative distribution functions. *Geophysical Research Letters*, **36**(11).
- Min, Seung-Ki, Zhang, Xuebin, Zwiers, Francis W, & Hegerl, Gabriele C. 2011. Human contribution to more-intense precipitation extremes. *Nature*, **470**(7334), 378–381.
- Ministère de l'Environnement et du Développement Durable, ANPE. 2005. *Gestion Durable Des Ressources en Eaux*.
- Mishra, Ashok K, & Singh, Vijay P. 2010. A review of drought concepts. *Journal of hydrology*, **391**(1-2), 202–216.
- Mohanty, Binayak P, Cosh, Michael H, Lakshmi, Venkat, & Montzka, Carsten. 2017. Soil moisture remote sensing : State-of-the-science. *Vadose Zone Journal*, **16**(1), 1–9.
- Monteith, JL. 1965. Light distribution and photosynthesis in field crops. *Annals of Botany*, **29**(1), 17–37.

- Monteny, Bruno, & Lhomme, Jean-Paul. 1980. Système d'analyse des relations entre le couvert végétal et les paramètres climatiques. *Actes du colloque d'Abijan informatique et biosphere*.
- Monteny, Bruno A. 1987. *Contribution à l'étude des interactions végétation-atmosphère en milieu tropical humide : importance du rôle du système forestier dans le recyclage des eaux de pluie*. Ph.D. thesis, Paris 11.
- Moran, M Susan. 2004. Thermal infrared measurement as an indicator of plant ecosystem health. *Pages 256–282 of : Thermal remote sensing in land surface processes*. CRC Press.
- Moran, MS, Clarke, TR, Inoue, Yoshio, & Vidal, A. 1994. Estimating crop water deficit using the relation between surface-air temperature and spectral vegetation index. *Remote sensing of environment*, **49**(3), 246–263.
- Moran, MS, Rahman, AF, Washburne, JC, Goodrich, DC, Wertz, MA, & Kustas, WP. 1996. Combining the Penman-Monteith equation with measurements of surface temperature and reflectance to estimate evaporation rates of semiarid grassland. *Agricultural and forest Meteorology*, **80**(2-4), 87–109.
- Mu, Qiaozhen, Zhao, Maosheng, Kimball, John S, McDowell, Nathan G, & Running, Steven W. 2013. A remotely sensed global terrestrial drought severity index. *Bulletin of the American Meteorological Society*, **94**(1), 83–98.
- Narasimhan, B, & Srinivasan, R. 2005. Development and evaluation of Soil Moisture Deficit Index (SMDI) and Evapotranspiration Deficit Index (ETDI) for agricultural drought monitoring. *Agricultural and forest meteorology*, **133**(1-4), 69–88.
- Niemeyer, Stefan, *et al.* 2008. New drought indices. *Options Méditerranéennes. Série A : Séminaires Méditerranéens*, **80**, 267–274.
- Norman, J. M, Kustas, W. P, & Humes, K. S. 1995. Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature. *Agricultural and Forest Meteorology*, **77**(3-4), 263–293.
- Olukayode Oladipo, E. 1985. A comparative performance analysis of three meteorological drought indices. *Journal of Climatology*, **5**(6), 655–664.
- OMM. 2010. *Guide des instruments et des méthodes d'observation météorologiques*. OMM 8, Geneve.
- OMM N°1090, MONDIALE ORGANISATION MÉTÉOROLOGIQUE Temps Climat Eau. 2012. *Guide d'utilisation de l'indice de précipitations normalisé*.
- Otkin, Jason A, Anderson, Martha C, Hain, Christopher, Mladenova, Iliana E, Basara, Jeffrey B, & Svoboda, Mark. 2013. Examining rapid onset drought development using the thermal infrared-based evaporative stress index. *Journal of Hydrometeorology*, **14**(4), 1057–1074.

- Palmer, Wayne C. 1965. *Meteorological drought*. Vol. 30. US Department of Commerce, Weather Bureau.
- Park, Seo-Yeon, Sur, Chanyang, Kim, Jong-Suk, & Lee, Joo-Heon. 2018. Evaluation of multi-sensor satellite data for monitoring different drought impacts. *Stochastic environmental research and risk assessment*, **32**(9), 2551–2563.
- Passioura, JB. 1983. Roots and drought resistance. *Pages 265–280 of : Developments in Agricultural and Managed Forest Ecology*, vol. 12. Elsevier.
- Paulik, Christoph, Naeimi, V, Dorigo, W, Wagner, W, & Kidd, R. 2012. A global validation of the ASCAT Soil Water Index (SWI) with in situ data from the International Soil Moisture Network. *Page 10189 of : EGU General Assembly Conference Abstracts*.
- Paulik, Christoph, Dorigo, Wouter, Wagner, Wolfgang, & Kidd, Richard. 2014. Validation of the ASCAT Soil Water Index using in situ data from the International Soil Moisture Network. *International journal of applied earth observation and geoinformation*, **30**, 1–8.
- Pellarin, Thierry, Calvet, Jean-Christophe, & Wagner, Wolfgang. 2006. Evaluation of ERS scatterometer soil moisture products over a half-degree region in southwestern France. *Geophysical Research Letters*, **33**(17).
- Péridy, Nicolas, Brunetto, Marc, & Ghoneim, Ahmed. 2012. The Economic Costs of Climate Change in MENA countries : A Micro-Spatial Quantitative Assessment and a Survey of Policies. *Femise, Marseille*.
- Quiring, Steven M, & Ganesh, Srinivasan. 2010. Evaluating the utility of the Vegetation Condition Index (VCI) for monitoring meteorological drought in Texas. *Agricultural and Forest Meteorology*, **150**(3), 330–339.
- Rajasekaran, eswar, Sekhar, Muddu, Bhattacharya, Bimal K, & Bandyopadhyay, Soumya. 2017. Spatial disaggregation of latent heat flux using contextual models over India. *Remote Sensing*, **9**(9), 949.
- Richardson, Clarence W. 1981. Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water resources research*, **17**(1), 182–190.
- Roerink, GJ, Su, Zhongbo, & Menenti, M. 2000. S-SEBI : A simple remote sensing algorithm to estimate the surface energy balance. *Physics and Chemistry of the Earth, Part B : Hydrology, Oceans and Atmosphere*, **25**(2), 147–157.
- Saadi, S., Boulet, G., Bahir, M., Brut, A., Delogu, E., Fanise, P., Mougnot, B., Simonneaux, V., & Chabaane, Z. 2018. Assessment of actual evapotranspiration over a semi arid heterogeneous land surface by means of coupled low-resolution remote sensing data with an energy balance model : comparison to extra-large aperture scintillometer measurements. *Hydrology and Earth System Sciences*, **22**(4), 2187–2209.

- Schilling, Janpeter, Hertig, Elke, Trambly, Yves, & Scheffran, Jürgen. 2020. Climate change vulnerability, water resources and social implications in North Africa. *Regional Environmental Change*, **20**(1), 1–12.
- Schmitz, Christoph, Lotze-Campen, Hermann, Gerten, Dieter, Dietrich, Jan Philipp, Bodirsky, Benjamin, Biewald, Anne, & Popp, Alexander. 2013. Blue water scarcity and the economic impacts of future agricultural trade and demand. *Water Resources Research*, **49**(6), 3601–3617.
- Seiler, RA, Kogan, F, & Wei, Guo. 2000. Monitoring weather impact and crop yield from NOAA AVHRR data in Argentina. *Advances in Space Research*, **26**(7), 1177–1185.
- Shabou, Marouen, Mougenot, Bernard, Lili Chabaane, Zohra, Walter, Christian, Boulet, Gilles, Benaissa, Nadhira, & Zribi, Mehrez. 2015. Soil clay content mapping using a time series of Landsat TM data in semi-arid lands. *Remote sensing*, **7**(5), 6059–6078.
- Shafer, BA, & Dezman, LE. 1982. Development of surface water supply index (SWSI) to assess the severity of drought condition in snowpack runoff areas. proceeding of the western snow conference.
- Singh, Ramesh P, Roy, Sudipa, & Kogan, F. 2003. Vegetation and temperature condition indices from NOAA AVHRR data for drought monitoring over India. *International journal of remote sensing*, **24**(22), 4393–4402.
- Slimani, Mohamed, Cudennec, Christophe, & Feki, Haifa. 2007. Structure du gradient pluviométrique de la transition Méditerranée–Sahara en Tunisie : déterminants géographiques et saisonnalité/Structure of the rainfall gradient in the Mediterranean–Sahara transition in Tunisia : geographical determinants and seasonality. *Hydrological Sciences Journal/Journal des Sciences Hydrologiques*, **52**(6), 1088–1102.
- Smith, William K, Dannenberg, Matthew P, Yan, Dong, Herrmann, Stefanie, Barnes, Mal-lory L, Barron-Gafford, Greg A, Biederman, Joel A, Ferrenberg, Scott, Fox, Andrew M, Hudson, Amy, *et al.* 2019. Remote sensing of dryland ecosystem structure and function : Progress, challenges, and opportunities. *Remote Sensing of Environment*, **233**, 111401.
- Stahle, David W, Fye, Falko K, Cook, Edward R, & Griffin, R Daniel. 2007. Tree-ring re-constructed megadroughts over North America since AD 1300. *Climatic change*, **83**(1), 133–149.
- Su, Zhongbo. 2002. The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. *Hydrology and earth system sciences*, **6**(1), 85–100.
- Sur, Chanyang, Park, Seo-Yeon, Kim, Tae-Woong, & Lee, Joo-Heon. 2019. Remote sensing-based agricultural drought monitoring using hydrometeorological variables. *KSCE Journal of Civil Engineering*, **23**(12), 5244–5256.
- Svoboda, M., & Fuchs, B.A. 2016. *Manuel des indicateurs et indices de sécheresse*.

- Thiemeßl, Matthias, & Gobiet, Andreas. 2010. Empirical-statistical downscaling and model error correction of daily temperature and precipitation from regional climate simulations and the effects on climate change signals. *Page 9664 of : EGU General Assembly Conference Abstracts*, vol. 12.
- Thornthwaite, Charles Warren. 1948. An approach toward a rational classification of climate. *Geographical review*, **38**(1), 55–94.
- Tomlinson, Charlie J, Chapman, Lee, Thornes, John E, & Baker, Christopher. 2011. Remote sensing land surface temperature for meteorology and climatology : A review. *Meteorological Applications*, **18**(3), 296–306.
- Tramblay, Yves, & Somot, Samuel. 2018. Future evolution of extreme precipitation in the Mediterranean. *Climatic Change*, **151**(2), 289–302.
- Tramblay, Yves, Koutroulis, Aristedeis, Samaniego, Luis, Vicente-Serrano, Sergio M, Volaire, Florence, Boone, Aaron, Le Page, Michel, Llasat, Maria Carmen, Albergel, Clement, Burak, Selmin, *et al.* 2020. Challenges for drought assessment in the Mediterranean region under future climate scenarios. *Earth-Science Reviews*, 103348.
- Trenberth, Kevin E, Dai, Aiguo, Van Der Schrier, Gerard, Jones, Philip D, Barichivich, Jonathan, Briffa, Keith R, & Sheffield, Justin. 2014. Global warming and changes in drought. *Nature Climate Change*, **4**(1), 17–22.
- Trigui, Ahmed. 1987. *Relations entre le climat, le sol et la production de la variété d'olivier Chemlali (Olea europea L.) dans la région de Sfax (Tunisie) : contribution à l'étude bioclimatique de l'espèce par l'analyse des propriétés biophysiques, structurales et spectrales des feuilles et des arbres*. Ph.D. thesis, Montpellier 2.
- Tucker, Compton J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, **8**(2), 127–150.
- Turco, Marco, Llasat, Maria C, Herrera, Sixto, & Gutiérrez, José Manuel. 2017. Bias correction and downscaling of future RCM precipitation projections using a MOS-Analog technique. *Journal of Geophysical Research : Atmospheres*, **122**(5), 2631–2648.
- UNCCD. 2017. *Global Land Outlook*. first edition edn. Germany : Secretariat of the United Nations Convention to Combat Desertification.
- Unganai, Leonard S, & Kogan, Felix N. 1998. Drought monitoring and corn yield estimation in Southern Africa from AVHRR data. *Remote sensing of environment*, **63**(3), 219–232.
- Valipour, Mohammad. 2015. Temperature analysis of reference evapotranspiration models. *Meteorological Applications*, **22**(3), 385–394.
- Van Loon, Anne F, & Van Lanen, Henny AJ. 2013. Making the distinction between water scarcity and drought using an observation-modeling framework. *Water Resources Research*, **49**(3), 1483–1502.

- Van Rooy, MP. 1965. A rainfall anomaly index independent of time and space, *Notos. WEATHER BUREAU SOUTH AFRICA*, **14**, 43–48.
- Van Zyl, Jakob J, Njoku, Eni G, & Jackson, Thomas J. 2003. Quantitative analysis of SMEX'02 AIRSAR data for soil moisture inversion. *Pages 404–406 of : IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No. 03CH37477)*, vol. 1. IEEE.
- Vicente-Serrano, Sergio M. 2007. Evaluating the impact of drought using remote sensing in a Mediterranean, semi-arid region. *Natural Hazards*, **40**(1), 173–208.
- Vicente-Serrano, Sergio M, McVicar, Tim R, Miralles, Diego G, Yang, Yuting, & Tomas-Burguera, Miquel. 2020. Unraveling the influence of atmospheric evaporative demand on drought and its response to climate change. *Wiley Interdisciplinary Reviews : Climate Change*, **11**(2), e632.
- Vogt, J, Naumann, Gustavo, Masante, Dario, Spinoni, Jonathan, Cammalleri, Carmelo, Erian, Wadid, Pischke, Frederik, Pulwarty, Roger, & Barbosa, Paulo. 2018. Drought risk assessment and management. *A Conceptual Framework*.
- Von Storch, Hans, Zorita, Eduardo, & Cubasch, Ulrich. 1993. Downscaling of global climate change estimates to regional scales : an application to Iberian rainfall in wintertime. *Journal of Climate*, **6**(6), 1161–1171.
- Vrac, Mathieu, & Friederichs, Petra. 2015. Multivariate intervariable, spatial, and temporal bias correction. *Journal of Climate*, **28**(1), 218–237.
- Wackernagel, H. 1995. Multivariate geostatistics. Springer-Verlag, New York. *Multivariate geostatistics. Springer-Verlag, New York*.
- Wagner, Wolfgang, Lemoine, Guido, & Rott, Helmut. 1999. A method for estimating soil moisture from ERS scatterometer and soil data. *Remote sensing of environment*, **70**(2), 191–207.
- Wang, Qunming, Shi, Wenzhong, Atkinson, Peter M, & Zhao, Yuanling. 2015. Downscaling MODIS images with area-to-point regression kriging. *Remote Sensing of Environment*, **166**, 191–204.
- Westermann, Sebastian, Langer, Moritz, & Boike, Julia. 2011. Spatial and temporal variations of summer surface temperatures of high arctic tundra on Svalbard implications for MODIS LST based permafrost monitoring. *Remote Sensing of Environment*, **115**(3), 908–922.
- Wigneron, J-P, Jackson, TJ, O'Neill, P, De Lannoy, Gabrielle, de Rosnay, Patricia, Walker, JP, Ferrazzoli, P, Mironov, V, Bircher, Simone, Grant, JP, *et al.* 2017. Modelling the passive microwave signature from land surfaces : A review of recent results and application to the L-band SMOS & SMAP soil moisture retrieval algorithms. *Remote Sensing of Environment*, **192**, 238–262.

- Wilby, Robert L, & Wigley, Thomas ML. 1997a. Downscaling general circulation model output : a review of methods and limitations. *Progress in physical geography*, **21**(4), 530–548.
- Wilby, Robert L, & Wigley, Thomas ML. 1997b. Downscaling general circulation model output : a review of methods and limitations. *Progress in physical geography*, **21**(4), 530–548.
- Wilby, Robert L, Wigley, TML, Conway, D, Jones, PD, Hewitson, BC, Main, J, & Wilks, DS. 1998. Statistical downscaling of general circulation model output : A comparison of methods. *Water resources research*, **34**(11), 2995–3008.
- Wilby, Robert L, Hay, Lauren E, Gutowski Jr, William J, Arritt, Raymond W, Takle, Eugene S, Pan, Zaitao, Leavesley, George H, & Clark, Martyn P. 2000. Hydrological responses to dynamically and statistically downscaled climate model output. *Geophysical Research Letters*, **27**(8), 1199–1202.
- Wilby, Robert L, Dawson, Christian W, & Barrow, Elaine M. 2002. SDSM—a decision support tool for the assessment of regional climate change impacts. *Environmental Modelling & Software*, **17**(2), 145–157.
- Wilby, Robert L, Charles, Steven P, Zorita, Eduardo, Timbal, Bertrand, Whetton, Penny, & Mearns, Linda O. 2004. Guidelines for use of climate scenarios developed from statistical downscaling methods. *Supporting material of the Intergovernmental Panel on Climate Change, available from the DDC of IPCC TGCI*, **27**.
- Wilhite, Donald A, & Glantz, Michael H. 1985. Understanding : the drought phenomenon : the role of definitions. *Water international*, **10**(3), 111–120.
- Wilks, Daniel S. 1989. Conditioning stochastic daily precipitation models on total monthly precipitation. *Water Resources Research*, **25**(6), 1429–1439.
- Wilks, Daniel S, & Wilby, Robert L. 1999. The weather generation game : a review of stochastic weather models. *Progress in physical geography*, **23**(3), 329–357.
- Willems, Patrick, & Vrac, M. 2011. Statistical precipitation downscaling for small-scale hydrological impact investigations of climate change. *Journal of hydrology*, **402**(3-4), 193–205.
- Xi, Zhimin, Pan, Hao, & Yang, Ren-Jye. 2017. Time dependent model bias correction for model based reliability analysis. *Structural Safety*, **66**, 74–83.
- Xu, C-Y, & Singh, V P. 2001. Evaluation and generalization of temperature-based methods for calculating evaporation. *Hydrological processes*, **15**(2), 305–319.
- Zargar, Amin, Sadiq, Rehan, Naser, Bahman, & Khan, Faisal I. 2011. A review of drought indices. *Environmental Reviews*, **19**, 333–349.

- Zhang, JiaHua, Xu, Yun, Yao, FengMei, Wang, PeiJuan, Guo, WenJuan, Li, Li, & Yang, LiMin. 2010. Advances in estimation methods of vegetation water content based on optical remote sensing techniques. *Science China Technological Sciences*, **53**(5), 1159–1167.
- Zhang, Ning, Hong, Yang, Qin, Qiming, & Liu, Lu. 2013. VSDI : a visible and shortwave infrared drought index for monitoring soil and vegetation moisture based on optical remote sensing. *International journal of remote sensing*, **34**(13), 4585–4609.
- Zhang, X-C. 2013. Verifying a temporal disaggregation method for generating daily precipitation of potentially non-stationary climate change for site-specific impact assessment. *International journal of climatology*, **33**(2), 326–342.
- Zorita, Eduardo, & von Storch, Hans. 1997. *A survey of statistical downscaling techniques*.
- Zorita, Eduardo, & Von Storch, Hans. 1999. The analog method as a simple statistical downscaling technique : Comparison with more complicated methods. *Journal of climate*, **12**(8), 2474–2489.
- Zouabi, Oussama, & Peridy, Nicolas. 2015. Direct and indirect effects of climate on agriculture : an application of a spatial panel data analysis to Tunisia. *Climatic change*, **133**(2), 301–320.
- Zribi, Mehrez, Nativel, Simon, & Le Page, Michel. 2021. Analysis of Agronomic Drought in a Highly Anthropogenic Context Based on Satellite Monitoring of Vegetation and Soil Moisture. *Remote Sensing*, **13**(14), 2698.