

# Université d'Angers

Master Mathématiques et Applications

**M2 DATA SCIENCE**

Année académique 2024-2025

## Rapport de stage

---

Générateurs stochastiques de météo pour  
évaluer l'impact du changement climatique sur  
les pommes et les vignes

---

Organisme d'accueil : UMR MISTEA - INRAE Centre  
Occitanie-Montpellier

*Étudiant :*

Arnaud GONIN

*Tuteur université :*

Gilles STUPFLER<sup>1</sup>

*Tuteurs de stage :*

David MÉTIVIER<sup>2</sup>

Bénédicte FONTEZ<sup>3</sup>

*Collaboratrices :*

Anne PELLEGRINO<sup>4</sup>

Isabelle FARRERA<sup>5</sup>

# Engagement de non plagiat

Je, soussigné, déclare être pleinement conscient que le plagiat de documents ou d'une partie d'un document publiés sur toutes formes de support, y compris internet, constitue une violation des droits d'auteur ainsi qu'une fraude caractérisée. En conséquence, je m'engage à citer toutes les sources que j'ai utilisées pour écrire ce rapport.

## Résumé

Les modèles phénologiques ont pour but de prédire les dates de débourrement annuelles des plantes. Dans le cas des pommes et des vignes, ces modèles sont dépendants des températures journalières. En particulier, des scénarios extrêmes de températures peuvent avoir des conséquences graves sur la santé de ces plantes et leur reproduction. Ces scénarios extrêmes sont dus à une variabilité significative des séries de températures journalières enregistrées. Cependant, les modèles climatiques classiques, tels que ceux utilisés dans les rapports du GIEC, reproduisent mal cette variabilité, et sont extrêmement coûteux numériquement.

Pour résoudre ces problèmes, nous allons nous servir des générateurs stochastiques de météo. Ce sont des modèles génératifs entraînés sur des données météorologiques, capables de simuler rapidement des séries de températures de même loi, avec la variabilité qui permet de former des scénarios extrêmes.

Ainsi, dans ce rapport, nous expliquerons le fonctionnement des générateurs stochastiques de météo. Ensuite, nous présenterons les modèles phénologiques et les résultats qu'ils donnent sur différentes zones géographiques. Enfin, nous verrons ce que donnent les modèles phénologiques appliqués à des séries de températures simulées avec des générateurs de météo entraînés sur des projections de modèles climatiques. Comme ces séries de températures auront la même loi que les projections, les modèles phénologiques nous donneront les distributions empiriques des scénarios phénologiques futures. Nous verrons alors que le réchauffement climatique peut avoir un impact dramatique sur la phénologie des pommes et des vignes dans les années à venir.



# Remerciements

Tout d'abord, j'aimerais remercier toutes les personnes qui m'ont encadré, aidé et accompagné au cours de ce stage :

Parmi elles, je remercie en premier mon encadrant principal David MÉTIVIER. C'est grâce à lui que j'ai pu intégrer l'équipe de l'UMR Mistea pour réaliser ce stage qui fut grandement formateur sur le plan académique, professionnel et humain. À partir du moment où je l'ai contacté pour exprimer mon intérêt pour l'offre de stage, il a toujours été disponible pour répondre à mes questions, me donner des conseils et m'accompagner dans chacune des difficultés que j'ai eu à surmonter. Il m'a même aidé dans des épreuves n'ayant pas un lien direct avec le sujet du stage comme ma recherche d'appartement à Montpellier. David MÉTIVIER est un encadrant de stage que je recommande vivement pour sa bienveillance et ses conseils riches et exigeants qui m'ont permis de grandement développer mes compétences scientifiques.

Je remercie aussi ma deuxième encadrante Bénédicte FONTEZ. Sa gentillesse et sa pédagogie m'ont permis par exemple de bien comprendre les modèles phénologiques pour ce projet et de faire en sorte que ma présentation de 20 minutes en anglais devant le consortium internationale NewClim se soit déroulée avec succès.

Je remercie également Anne PELLEGRINO et Isabelle FARRERA pour leurs explications sur la phénologie des pommes et des vignes et toutes les ressources pertinentes qu'elles m'ont partagées. Cela m'a permis d'implémenter correctement les modèles phénologiques qui occupent une place importante de mon stage.

Je remercie de même la doctorante Caroline COGNOT pour avoir accepté de prendre une partie de son temps pour m'aider à comprendre certains détails des générateurs de météo et de la régression LOESS.

Je remercie tous les membres du personnels de MISTEA qui m'ont aidé dans les aspects administratifs et logistiques du stage comme Laetitia PORCHER et Malika NASSIF.

Et enfin, je remercie toutes les personnes que j'ai côtoyées durant ce stage et avec lesquelles j'ai passé des bons moments, notamment lors des évènements organisés par l'unité ou les quelques sorties à l'extérieur du laboratoire.

*Merci !*

# Table des matières

<b>Introduction</b>	<b>1</b>
<b>1 Enjeux et démarche scientifiques du stage</b>	<b>3</b>
1.1 Contexte . . . . .	3
1.1.1 Organisme . . . . .	3
1.1.2 Financement et intérêt du stage . . . . .	3
1.1.3 Langage de programmation utilisé : Julia . . . . .	3
1.2 État de l’art . . . . .	4
1.2.1 Effet du réchauffement climatique sur les vignes . . . . .	4
1.2.2 Effet du réchauffement climatique sur les pommes . . . . .	5
1.3 Objectifs . . . . .	5
1.4 Notations . . . . .	6
<b>2 Générateurs stochastiques de météo</b>	<b>8</b>
2.1 Principe . . . . .	8
2.2 Objectifs et état de l’art pour le générateur de météo. . . . .	9
2.3 Décomposition d’une série de températures . . . . .	10
2.3.1 Tendance . . . . .	11
2.3.2 Saisonnalité . . . . .	12
2.4 Modèles pour la partie résiduelle . . . . .	12
2.4.1 Modèle AR . . . . .	12
2.4.2 Modèle AR mensuel . . . . .	14
2.4.3 Discussion sur l’estimation mensuel . . . . .	14
2.5 Simulation de séries . . . . .	16
2.6 Sélection du modèle . . . . .	16
2.6.1 Critères quantitatifs . . . . .	17
2.6.2 Critères qualitatifs . . . . .	18
<b>3 Modèles phénologiques</b>	<b>19</b>
3.1 Principe . . . . .	19
3.2 Modèle pour les vignes (BRIN) . . . . .	21
3.3 Modèle pour les pommes (F1 Gold 1) . . . . .	22
<b>4 Mises en œuvre et résultats</b>	<b>23</b>
4.1 Données . . . . .	23
4.1.1 Séries de températures enregistrées : données ECA&D . . . . .	23
4.1.2 Projections climatiques : données du portail DRIAS . . . . .	23
4.2 Sélection des générateurs de météo . . . . .	24
4.2.1 Sélection des modèles univariés . . . . .	24
4.2.2 Sélection des modèles multivariés . . . . .	28
4.3 Modèles phénologiques appliqués aux données enregistrées . . . . .	29
4.4 Modèles phénologiques appliqués aux données simulées avec générateurs de météo entraînés sur températures enregistrées . . . . .	31
4.5 Modèles phénologiques appliqués aux données simulées avec générateurs de météo entraînés sur projections climatiques . . . . .	34
4.5.1 Scénario rcp4.5 . . . . .	35
4.5.2 Scénario rcp8.5 . . . . .	37

<b>Conclusion</b>	<b>40</b>
<b>Table des annexes</b>	<b>41</b>
<b>Annexe A Tables des BIC et AIC</b>	<b>41</b>
A.1 Tables des meilleurs modèles selon le critère BIC . . . . .	41
A.2 Tables des meilleurs modèles selon le critère AIC . . . . .	42
<b>Annexe B Résultats d'estimations du générateur de séries appariées de type TN et TX sur les températures enregistrées à Nantes</b>	<b>43</b>
B.1 Résultats pour la série de températures minimales journalières (TN) . . . . .	43
B.2 Résultats pour la série de températures maximales journalières (TX) . . . . .	45
<b>Annexe C Résultats des modèles phénologiques appliquées aux simulations de générateurs météo</b>	<b>47</b>
C.1 Générateurs de météo entraîné sur les données de Montpellier-Aéroport . . . . .	47
C.2 Générateurs de météo entraîné sur les données de Nantes-Bougenais . . . . .	48
<b>Annexe D Histogrammes des prédictions des modèles phénologiques appliquées aux simulations de générateurs de météo entraînés sur projections climatiques.</b>	<b>49</b>
<b>Annexe E Lien du dépôt github du projet</b>	<b>51</b>

# Introduction

Évaluer et prévenir tous les risques dus au réchauffement climatique est un enjeu actuel majeur. Cela a été une partie importante du travail du groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) depuis 1988 dont le dernier rapport d'évaluation date de 2023. Parmi les risques énoncées, on trouve l'augmentation de la fréquence des événements climatiques extrêmes ([McSweeney, 2021](#)), ces derniers pouvant avoir des effets dramatiques sur l'agriculture.

L'objectif du stage est d'évaluer les risques liés aux réchauffement climatiques sur les pommes et les vignes. Pour cela nous nous intéresserons à leur phénologie, c'est-à-dire à leurs changements saisonniers qui dépendent de la température et qui sont donc sensibles aux effets du réchauffement climatiques. Nous utiliserons des modèles capables de prédire des dates importantes de leur cycle phénologique telle que la date de débourrement à partir des données de températures journalières. Nous utiliserons ceux présentés dans [Legave et al. \(2013\)](#) pour les pommes et dans [García de Cortázar-Atauri et al. \(2009\)](#) pour les vignes.

Ainsi, la première idée que l'on pourrait avoir serait de prédire les dates phénologiques à partir de ces modèles en mettant en entrée des données de températures enregistrées ou des projections futures de températures issues de modèles climatiques, qui sont en grande partie basées sur la physique. Ces modèles climatiques sont par exemple utilisés par le GIEC pour établir leurs quatre scénarios possibles du réchauffement de la planète pour les années à venir. Cependant, l'utilisation de ces données de températures présente des problèmes importants :

- Pour les données de températures enregistrées, par station, d'un point de vu probabiliste nous avons l'équivalent de seulement une réalisation d'une série de température avec possiblement quelque événements extrêmes qui se sont réalisés. Cela n'est pas suffisamment représentatif de la loi d'une telle série et il aurait été nécessaire d'obtenir plusieurs séries de température enregistrées du même endroit, ce qui est logiquement impossible. L'intérêt d'avoir un grand nombre de séries de températures est de pouvoir connaître leur loi empirique afin de comprendre mieux l'évolution des cycles phénologiques au cours des dernières années.
- Les simulations de températures de modèles climatiques sont très coûteux numériquement, ce qui fait que leur nombre est trop faible pour avoir un échantillon satisfaisant de la loi correspondante. Par exemple, la Trajectoire de Réchauffement de référence pour l'Adaptation au Changement Climatique (TRACC, [Rousset \(2025\)](#)), une étude sur les conséquences du changement climatique avec des propositions d'adaptation, utilise seulement 17 simulations par scénario du GIEC pour avoir un intervalle confiance. De plus, les modèles climatiques ne prennent pas suffisamment en compte la variabilité des températures, qui sont à l'origine des événements extrêmes comme les canicules.

La solution envisagée pour palier ces deux problèmes est l'utilisation des générateurs stochastiques de météo (stochastic weather generators). Un générateur de météo est un modèle entraîné sur une série de températures capable de simuler des séries avec la même loi de probabilité. Ces modèles sont capables de générer des séries de températures pour un coût numérique nettement inférieur aux modèles climatiques et leur stochasticité permet de prendre en compte la variabilité des températures lors de l'apprentissage et de la reproduire lors des simulations ([Ailliot et al., 2015](#)). Générer beaucoup de simulations permettra alors d'obtenir l'évolution des distributions empiriques des températures et d'avoir une vision de tous les scénarios extrêmes possibles. Enfin, en mettant tous les scénarios de températures générées en entrée des modèles phénologiques, on obtient tous les scénarios possibles d'évolution des cycles phénologiques des



pommes et des vignes dans un contexte de réchauffement climatique.

Ainsi, dans ce rapport on présentera d'abord tous les enjeux et la démarche scientifiques du stage, ensuite on expliquera en détail ce que sont les générateurs stochastiques de météo et les modèles phénologiques et enfin on présentera les résultats de ces modèles et de leur couplage afin de mieux connaître les effets potentiels du réchauffement climatique sur les pommes et les vignes.

# 1 Enjeux et démarche scientifiques du stage

## 1.1 Contexte

### 1.1.1 Organisme

Le stage a été réalisé au sein de l'Unité Mixte de Recherche de Mathématiques, Informatique et Statistique pour l'Environnement et l'Agronomie (UMR [MISTEA](#)), affiliée à l'INRAE (département MathNum), à l'Institut Agro de Montpellier (département SABP) et à l'université de Montpellier au sein du pôle MIPS (Mathématiques, Informatique, Physique, Systèmes). Cette unité de recherche est divisée en quatre axes :

- L'axe [Probabilité et Statistique](#) auquel mes encadrants et moi sommes affiliés. Les principaux objets d'étude de cet axe sont :
  - L'apprentissage machine adapté aux spécificités des données d'agriculture numérique, de phénotypage et d'agroécologie.
  - L'analyse de sensibilité des modèles agronomiques.
  - Et la conception et analyse d'algorithmes stochastiques en écologie des populations et écologie microbienne.
- L'axe [Systèmes Dynamiques](#), travaillant d'une part sur la modélisation, simulation et analyse des systèmes biologiques et d'autre part sur l'optimisation et le contrôle optimal pour le génie des procédés, la gestion des ressources renouvelables et la préservation de l'environnement.
- L'axe [Informatique](#), faisant des recherches appliquées à l'agronomie et l'environnement sur la gestion et l'analyse de données, l'ingénierie des connaissances, le web sémantique, les ontologies, les graphes de connaissances et la programmation par contraintes.
- Et enfin une cellule d'ingénierie travaillant sur la suite logicielle [OpenSILEX](#), qui consiste en la mise en place de systèmes d'information dédiés à l'exploitation des données à grandes échelles en agronomie.

Avoir intégré cette unité de recherche a été enrichissant, notamment en assistant tout au long du stage à de nombreuses présentations (au moins une par semaine sauf en Juillet et en Août) comme le séminaire des doctorants et post-doctorants qui a lieu tous les mois. J'ai eu aussi l'occasion de faire une présentation de mes travaux devant mes collègues lors de la journée des stagiaires au début du mois de Juillet.

### 1.1.2 Financement et intérêt du stage

Ce stage a été financé par le [métaprogramme CLIMAE](#), un programme national de l'INRAE d'aide à la transition écologique de l'agriculture et des systèmes forestiers pour leur adaptation au climat futur et la réduction des émissions de gaz à effet de serre. En plus de ce métaprogramme, les travaux réalisés ont aussi intéressé le consortium scientifique international [Newclim](#), dans leur objectif d'avoir des projections régionales du climat et de recherche de stratégies d'adaptation. En particulier, leur souhait est de voir la réalisation d'un jeu sérieux qui utilise les modèles développés au cours du stage. Une présentation de 20 minutes en anglais a été faite devant quelque membres de Newclim en Juin pour montrer le travail réalisé et toutes les perspectives d'améliorations.

### 1.1.3 Langage de programmation utilisé : Julia

Le langage de programmation utilisé au cours de ce stage est [Julia](#). Il s'agit d'un langage utilisé pour le calcul scientifique avec une syntaxe aussi simple que R et Python et une rapidité comparable au C grâce au fait que c'est un langage compilé ([JuliaLang.org contributors](#),

2024). De plus, ce langage permet de facilement surcharger des fonctions pour qu'elles puissent prendre en entrée des objets de différents types, en particulier ceux issus de packages (cette fonctionnalité est appelée multi-dispatching). Ces deux atouts permettent aux utilisateurs de ce langage de créer aisément de nouveaux programmes utilisant des concepts de différentes branches scientifiques tout en gardant une grande rapidité. Par exemple, si on veut implémenter des équations différentielles neuronales (Chen et al., 2018), cela est faisable en installant un package d'équations différentielles et un package de deep learning, tous les deux codés entièrement en Julia. En effet, le package d'équations différentielles ne sera pas un problème pour la différentiation automatique grâce au multi-dispatching et à l'absence d'autres langages dans les deux packages. Implémenter des équations différentielles neuronales aurait été beaucoup plus compliqué en Python à cause de l'absence d'interopérabilité entre les packages d'équations différentielles et les packages de deep learning. L'absence de multi-dispatching et le fait que les principaux modules de python utilisent très souvent d'autres langages (par exemple C++) rend difficile la réalisation de nouveaux packages à partir de différents packages préexistants. Les utilisateurs de Julia sont majoritairement des ingénieurs, scientifiques et étudiants dans les domaines de la science des données, la modélisation numérique, les statistiques et l'apprentissage automatique (Andrew Claster, 2023). Ce langage a été choisi pour ce stage car c'est le langage de calcul scientifique le mieux maîtrisé par mon encadrant David Métivier. Par exemple, j'ai dû quelque fois utiliser son package Julia [StochasticWeatherGenerators.jl](#). L'emploi de Julia est d'autant plus utile pour raccourcir le temps d'entraînement des générateurs météo qui peut être assez long, notamment lorsqu'il y a un très grand nombre d'hyperparamètres à tester (exemple à la section 4.2.1). En effet, il a été nécessaire d'utiliser des boucles "for" pour implémenter les modèles et les programmes auraient par conséquent été beaucoup plus long avec R ou Python.

## 1.2 État de l'art

Dans cette section nous présenterons certains travaux qui ont eu pour but de décrire les effets du réchauffement climatique sur les pommes et les vignes et de proposer des solutions d'adaptation.

### 1.2.1 Effet du réchauffement climatique sur les vignes

- Dans Caffarra et al. (2012), on décrit une étude qui a été faite sur l'impact du réchauffement climatique sur les interactions entre les vignes et les insectes ravageurs. Par exemple, cet article montre que l'augmentation moyenne des températures crée une désynchronisation entre la période de résistance de la plante et l'arrivée des larves de ces insectes. Ainsi, la plante devient plus vulnérable aux ravageurs, ce qui a des conséquences néfastes sur les rendements.
- Dans Sgubin et al. (2018) on cherche à expliquer l'impact du réchauffement climatique à partir des risques de gel post-débourrement. Cette étude utilise des projections de modèles climatiques. On y montre que ce risque dépend beaucoup du modèle phénologique utilisé et de la position géographique.
- Van Leeuwen et al. (2019) présente les nécessités d'adaptation des vignobles en terme de méthodes de plantation et de récolte dues à l'augmentation des températures et des sécheresses.
- Dans la même ligne, Parker et al. (2020) présente des modèles pour la quantité de sucre et d'autres critères de la vigne afin de donner des méthodes d'adaptation pour la préservation de la qualité du vin face à la hausse des températures.

### 1.2.2 Effet du réchauffement climatique sur les pommes

- [Funes et al. \(2016\)](#) montre que l'augmentation des températures réduit la réponse au besoin de refroidissement "chilling" des pommes. Ce refroidissement est nécessaire pour arrêter la période de dormance. Ainsi le réchauffement climatique rend certaines espèces de pommes incapables de survivre à l'endroit où elles étaient cultivées. De plus, cette étude a montré que le réchauffement climatique est à l'origine d'un changement de la date de floraison et d'autres désordres.
- [Sugiura et al. \(2013\)](#) présente les effets du réchauffement climatique sur le goût et la texture des pommes. Il s'agit par exemple d'une baisse de l'acidité, de la fermeté et de la quantité en eau de ces fruits. Cette étude montre que les pommes commercialisées sont sujettes à une baisse de qualité sur le long terme.
- [Legave et al. \(2015\)](#) montre que les effets du réchauffement climatique sur les dates de floraison varient en fonction de la zone géographique. Par exemple, les floraisons sont plus précoces dans la majorité des régions européennes et au Maroc alors qu'elles restent stationnaires dans certaines régions du Brésil.
- Dans [Kőrösi et al. \(2018\)](#) on présente une expérience dans laquelle on décale artificiellement les cycles phénologiques en plaçant les pommiers dans des serres ou des zones refroidies puis en les installant dans des vergers. L'objectif est de connaître les changements des interactions avec les pollinisateurs et les autres arthropodes. Le résultat est que la répartition de ces espèces n'est plus la même pour les pommiers désynchronisés, ce qui indique un risque pour la santé de l'arbre et sa reproduction.

## 1.3 Objectifs

Comme énoncé dans l'introduction, l'objectif principal du stage est d'utiliser les générateurs stochastiques de météo et les modèles phénologiques pour obtenir une distribution des scénarios possibles d'effets du réchauffement climatique sur les pommes et les vignes. Cela peut se décomposer en trois étapes importantes :

### 1. Implémentation des générateurs stochastiques de météo et choix du modèle :

Le premier objectif est d'implémenter les générateurs de météo en Julia. Il en existe plusieurs catégories, et pour un modèle on a un choix d'hyperparamètres à faire. Ainsi, le choix du meilleur modèle est aussi un objectif important. Par exemple, le modèle choisi devra avant tout être capable de simuler des séries avec les mêmes propriétés que la série de température d'entraînement (par exemple, la série d'entraînement doit être dans l'intervalle des minimums et maximums des séries générés, elles doivent avoir des variances mensuelles proches, etc.). Des critères statistiques comme l'AIC et le BIC seront aussi pris en compte.

### 2. Implémentation des modèles phénologiques :

Le deuxième objectif principal sera également l'implémentation des modèles phénologiques en Julia. Contrairement aux générateurs de météo, pour ce projet nous utiliserons des modèles déjà ajustés avec leurs paramètres proposés dans les articles [Legave et al. \(2013\)](#) et [García de Cortázar-Atauri et al. \(2009\)](#).

### 3. Analyse des résultats et conclusion :

Le dernier objectif sera d'utiliser ces modèles afin de tirer des conclusions sur l'impact

du réchauffement climatique sur les pommes et les vignes. Cela consistera majoritairement à appliquer les modèles phénologiques aux simulations des générateurs de météo afin d'obtenir une distribution empirique des dates phénologiques prédites. Dans un premier temps, on entraînera le générateur de météo sur des températures enregistrées pour faire une analyse de l'impact du réchauffement climatique sur les pommes et les vignes jusqu'à aujourd'hui. Nous entraînerons aussi le générateur météo sur des projections par des modèles climatiques, ce qui nous permettra d'avoir un bien plus grand ensemble de simulations de températures dans le futur. Ces données sont présentées dans la section 4.1. En y appliquant les modèles phénologiques, ces simulations nous permettront d'avoir un échantillon des possibilités d'évolution des cycles phénologiques des pommes et des vignes en incluant leur variabilité, et par extension d'évaluer les risques possibles dans le futur.

Comme dit précédemment, une des ambitions de ce stage est de pouvoir développer des outils pour la mise en place d'un jeu sérieux de sensibilisation au changement climatique.

## 1.4 Notations

Voici la liste des notations utilisées dans ce rapport :

### Notations mathématiques et statistiques :

- $(Y_t)_{1 \leq t \leq N} = (Y_1, Y_2, \dots, Y_N)$  une série temporelle d'indices allant de 1 à  $N$ . Pour simplifier on pourra noter  $Y = (Y_t)_{1 \leq t \leq N}$  si cela n'engendre pas d'ambiguïté.
- $\hat{Y}_t$  une prédiction ou une simulation d'un modèle pour l'indice  $t$ .
- $M^T$  la transposée de la matrice  $M$
- $|x|$  la valeur absolue de  $x$ .
- $\mathcal{M}_l(\mathbb{R})$  l'espace des matrices de taille  $l \times l$  à coefficients réels.
- $\mathcal{N}(\mu, \Gamma)$  la loi normale d'espérance  $\mu$  et de matrice de variance-covariance  $\Gamma$ .
- $\theta$  le paramètre d'un modèle.
- $\hat{\theta}$  l'estimateur de  $\theta$ .
- $f_\theta((Y_t)_{1 \leq t \leq N} | (Z_t)_{1 \leq t \leq M}) = f_\theta(Y_1, Y_2, \dots, Y_N | Z_1, Z_2, \dots, Z_M)$  la densité de  $(Y_t)_{1 \leq t \leq N}$  sachant  $(Z_t)_{1 \leq t \leq M}$  selon le modèle avec le paramètre  $\theta$ .
- $L((Y_t)_{t \in I}; (Z_s)_{s \in J}, \theta)$  la vraisemblance du modèle avec le paramètre  $\theta$  sur la série  $(Y_t)_{t \in I}$  conditionnellement à la série  $(Z_s)_{s \in J}$ .

### Notations adaptées au contexte météorologique :

- On suppose que les séries de températures journalières sont indexées entre 1 (le premier jour de la série) et  $N$  (le dernier jour). Dans ce contexte l'indice  $t$  représente toujours une date.
- $m(t)$  le numéro du mois associé à la date de  $t$ . Par exemple si  $t = 500$  correspond à un 22 septembre, alors  $m(500) = 9$ .
- $TN_t$ ,  $TG_t$  et  $TX_t$  respectivement la température minimale, moyenne et maximale du jour  $t$ .
- $T_t$  est, en fonction du contexte précisé, une de ces valeurs ou un vecteur contenant deux de ces valeurs ou les trois à la fois (par exemple  $T_t = (TN_t, TX_t)$ ). Dans la section 2.3  $T_t$  est en dimension 1 pour tout  $t$  entre 1 et  $N$  tandis qu'à partir de 2.4 la dimension peut être 1, 2 ou 3.

- $d$  est la dimension de  $T_t$  pour tout  $t$  entre 1 et  $N$ .
- $X_{\text{TN}_t}$ ,  $X_{\text{TG}_t}$  et  $X_{\text{TX}_t}$  respectivement les parties résiduelles de  $\text{TN}_t$ ,  $\text{TG}_t$  et  $\text{TX}_t$  après la décomposition décrite à la section 2.3.
- $X_t$  est la partie résiduelle selon la décomposition décrite à la section 2.3 de  $T_t$  si  $T_t$  est en dimension 1.

Si  $T_t$  est en dimension supérieure à 1, alors  $X_t$  est le vecteur des parties résiduelles des composantes de  $T_t$ . Par exemple si  $T_t = (\text{TN}_t, \text{TX}_t)$  alors  $X_t = (X_{\text{TN}_t}, X_{\text{TX}_t})$

## 2 Générateurs stochastiques de météo

### 2.1 Principe

Un générateur stochastique de météo est un modèle génératif entraîné sur des séries de températures ayant pour but de générer des séries avec les mêmes caractéristiques statistiques. Ce principe est illustré sur la figure 1. La série d'entraînement et les séries générées doivent par exemple à avoir des tendances, des saisonnalités, des variances et des autocorrélations similaires. Dans notre modèle on suppose que les séries de températures sont des séries de variables aléatoires. Le rôle d'un générateur stochastique de météo est donc d'apprendre à générer des séries de températures de même loi que leur série d'entraînement.

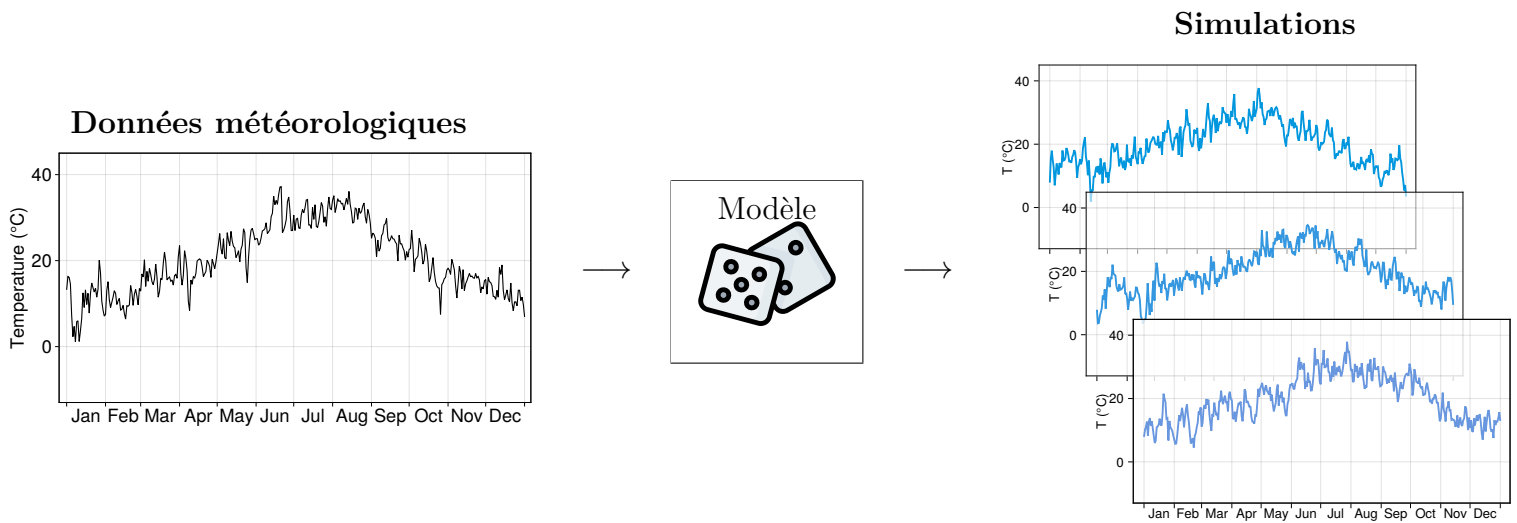


FIGURE 1 – Schéma qui représente le principe de fonctionnement des générateurs stochastiques de météo. À gauche il s'agit d'une série de températures enregistrées par une station météorologique, en l'occurrence les températures maximales journalières de Montpellier en 2003, et à droite il s'agit des séries générées pour cette même période avec un générateur de météo entraîné sur la série de gauche. Le modèle de ce générateur de météo sera détaillé à partir de la section 2.3. Pour des raisons pédagogiques on a représenté des séries de un an, mais en réalité ce type de modèle considère des séries de plusieurs années.

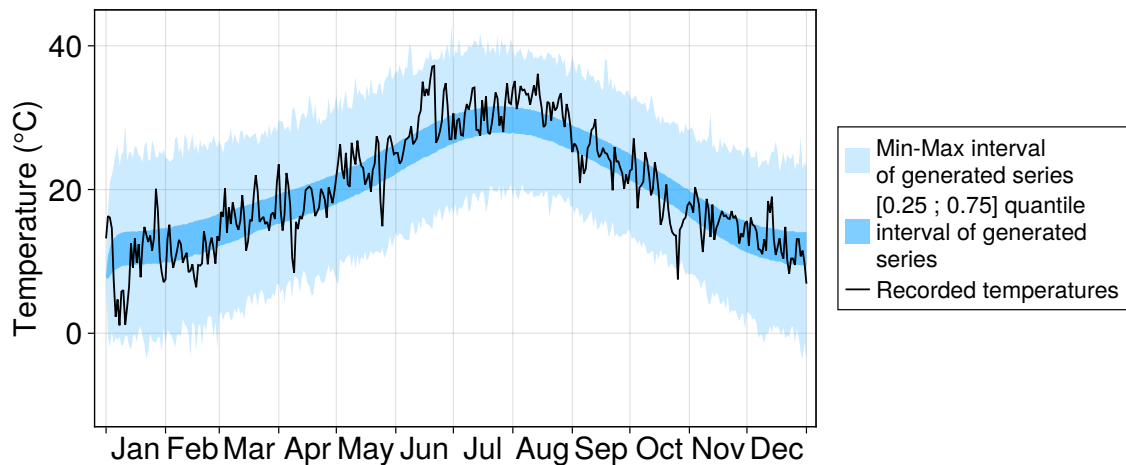


FIGURE 2 – Températures maximales journalières de Montpellier en 2003 et intervalles minimum-maximum et interquartiles des températures de 5000 séries générées pour chaque jour.

Générer un grand nombre de simulations permet alors d’obtenir une distribution empirique de la loi de la série d’entraînement. En effet, sur la figure 2 on peut voir l’étendu des températures simulées à chaque jour de l’année ainsi que l’intervalle interquartile, représentatif de cette loi. Cela est utile pour savoir quels événements extrêmes sont possibles. Dans notre illustration on voit que toutes les valeurs observées sont incluses dans la distribution générée, même les plus extrêmes.

## 2.2 Objectifs et état de l’art pour le générateur de météo.

Pour pouvoir générer des séries de températures correctes et utilisables par les modèles phénologiques, il faut que les générateurs stochastiques de météo puissent :

- Être capables de générer des couples de séries de températures minimales et maximales journalières (TN, TX) et de températures minimales et moyennes (TN, TG) en respectant les corrélations entre les deux séries dans les deux cas, pour les raisons exprimées à la section 4.2.2.
- Être capables de reconstituer l’évolution moyenne et l’évolution de la variance de la série au cours des années.
- Être capables de reconstituer les effets périodiques des saisons sur la valeur moyenne et la variance.
- Mettre une part d’aléatoire réaliste dans la génération des séries pour éviter que toutes les séries générées ne soient identiques. Il faut en particulier donner une variance correcte aux séries générées et qu’elle soient proches de la série d’entraînement tout en étant capables d’atteindre des valeurs extrêmes plausibles que la série d’entraînement n’a pas forcément atteintes.
- Respecter les propriétés climatiques, comme la mémoire des températures journalières des jours précédents.

Nous noterons que nous ne considérerons pas de corrélation spatiale entre les températures pour chaque station. Nous travaillerons sur des modèles qui considèrent un seul site géographique à la fois.

Les modèles présentés dans [Richardson \(1981\)](#) et bien développés depuis par exemple par [Parlange and Katz \(2000\)](#) consistent à retirer les saisonnalités en moyenne et en variance puis à appliquer un modèle AR(1) multivarié ([Indjehagopian, 1979](#)) sur les résidus. L’avantage à utiliser des tels modèles est qu’ils considèrent les corrélations entre les types de températures, par exemple entre les températures minimales et maximales journalières. Ces modèles considèrent aussi la pluie mais dans ce stage on ne la prendra pas en compte car elle n’est pas utilisée par nos modèles phénologiques. Le fait que les modèles autorégressifs ne soient qu’à l’ordre 1 nous a paru être une limite pour considérer la mémoire sur plusieurs jours donc a décidé d’essayer des modèles à des ordres supérieures (voir définition 2.3). Ensuite comme le font [Cognot et al. \(2025\)](#) et [Parey et al. \(2014\)](#), on a décidé d’estimer aussi les tendances dans les parties déterministes, ce qui rend le modèle beaucoup plus réaliste dans un contexte d’augmentation moyenne des températures due au réchauffement climatique. Cela donne la décomposition des séries de températures de l’équation 1. Néanmoins, il n’est pas forcément pertinent d’utiliser les mêmes paramètres du modèle autorégressif sur chaque mois à cause des changements saisonniers du comportement des températures. Par exemple, les articles [Ailliot et al. \(2009\)](#), [Evin et al. \(2018\)](#) et [Evin \(2018\)](#) présentent des générateurs de météo pour lesquelles les paramètres sont estimés mois par mois. Cela est fait en concaténant tous les mois de Janvier de la série, tous les mois de Février, etc., puis en estimant le modèle sur chaque concaténation. Ainsi, nous aussi on ajuste les paramètres mois par mois mais la méthode pour faire cela est discutée dans la section 2.4.3. Pour simuler des scénarios extrêmes à partir de la partie résiduelle, dans le



domaine de la finance les modèles de type ARCH Engle (1982) considère une variance non constante pour pouvoir simuler des périodes de crises économiques. Cependant, dû à la modélisation de la volatilité, ce modèle peut considérer des changements trop brutaux qui peuvent devenir irréalistes dans le cas des températures. Enfin, il existe encore une approche différente du modèle inspiré de Richardson (1981) pour les générateurs de météo qui consiste en l'utilisation de champs gaussiens. Obakrim et al. (2025) détaille comment les générateurs de météo utilise cet outil et des articles qu'on a énoncés comme Cognot et al. (2025) et Parey et al. (2014) utilisent ce type de générateur. Les champs gaussiens sont en revanche adaptés à des générateurs multi-site, c'est-à-dire ceux qui considèrent une corrélation spatiale entre les séries de chaque station. Par conséquent, nous maintiendrons l'utilisation de modèles autorégressifs pour la modélisation uni-site des températures.

## 2.3 Décomposition d'une série de températures

Dans un premier temps, le modèle du générateur de météo implique une décomposition additive-multiplicative des séries temporelles (Cognot et al., 2025, Parey et al., 2014) :

$$T_t = M_t + S_t + \sigma_{M_t}\sigma_{S_t}X_t \quad (1)$$

où :

- $T_t$  est la température enregistrée
- $M_t$  est la tendance additive
- $S_t$  est la saisonnalité additive
- $\sigma_{M_t}$  est la tendance multiplicative
- $\sigma_{S_t}$  est la saisonnalité multiplicative
- $X_t$  est la partie résiduelle, contenant une partie stochastique (2.4).

L'objectif de cette décomposition est d'obtenir une partie stationnaire (ici  $X_t$ ) afin d'y appliquer des modèles statistiques adaptés (voir à la section 2.4). De plus, les parties déterministes additives ( $M_t+S_t$ ) et multiplicatives ( $\sigma_{M_t}\sigma_{S_t}$ ) représentent respectivement la moyenne et l'écart-type de la température au jour  $t$ . Voici la démarche pour estimer chacune de ces parties :

1. On commence d'abord par estimer la tendance additive sur toute la série  $T_t$  via une méthode appelée Locally Estimated Scatterplot Smoothing abrégée LOESS (Cleveland, 1979). L'usage de cette méthode est brièvement expliqué dans la sous-section suivante. Cela donne un estimateur de la tendance additive  $\hat{M}_t$ .
2. Ensuite on estime la saisonnalité additive sur  $T_t - \hat{M}_t$  via une régression par moindres carrés pour obtenir une saisonnalité  $\hat{S}_t$  sous forme de polynôme trigonométrique. Cette estimation est expliquée de manière plus détaillée dans 2.3.2.
3. On estime le carré de la tendance multiplicative sur  $(T_t - \hat{M}_t - \hat{S}_t)^2$ , toujours avec une régression LOESS pour obtenir  $\hat{\sigma}_{M_t}^2$ .
4. On estime la saisonnalité multiplicative sur  $(T_t - \hat{M}_t - \hat{S}_t)^2 / \hat{\sigma}_{M_t}^2$  avec de même une régression linéaire pour nous donner le terme  $\hat{\sigma}_{S_t}^2$  sous forme de polynôme trigonométrique.
5. Et enfin, les résidus que nous considérerons se définissent par :

$$X_t = \frac{T_t - \hat{M}_t - \hat{S}_t}{\hat{\sigma}_{M_t}\hat{\sigma}_{S_t}} \quad (2)$$

. Ils existent plusieurs modèles pour la partie résiduelle qui seront présentés à la section 2.4. On entraîne l'un de ces modèles sur cette partie, ce qui nous permettra de simuler

une partie résiduelle la plus similaire possible et d'y rajouter ensuite les tendances et saisonnalités estimées afin d'obtenir une nouvelle série de températures (cela sera expliqué en détail dans la section 2.5).

### 2.3.1 Tendances

Lors de l'estimation du modèle, nous calculons les tendances additives et multiplicatives à l'aide d'une méthode de régression locale appelée LOESS, initialement présentée par [Cleveland \(1979\)](#). Le principe est le suivant : on considère  $x_1, x_2, \dots, x_n$  les abscisses et  $y_1, y_2, \dots, y_n$  les ordonnées des points sur lesquels on veut faire une régression. Pour chaque valeur  $x_i$ , on réalise une régression pondérée (weighted regression) d'un polynôme de degré  $D$  sur l'ensemble des points, ce qui nous donne une prédiction  $\hat{y}_i$  qui sera retenue comme prédiction de la LOESS au point  $x_i$ . Cela se formalise de la manière suivante :

**Définition 2.1** (Régression LOESS). *Soit  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  un ensemble de points,  $h$  un réel positif,  $D$  un entier et  $W$  une fonction réelle avec les propriétés suivantes :*

- $W(x) > 0$  pour  $|x| < 1$  ;
- $W(-x) = W(x)$  ;
- $W(x)$  est décroissante pour  $x \geq 0$  ;
- $W(x) = 0$  pour  $|x| \geq 1$ .

Pour  $i$  entre 1 et  $n$  on pose  $w_i(x) = W\left(\frac{x_i - x}{h}\right)$ . On appelle régression LOESS un algorithme qui résout, pour tout entier  $j$  entre 1 et  $n$ , le problème d'optimisation suivant :

$$\min_{\beta_{0,j}, \dots, \beta_{D,j}} \sum_{i=1}^n w_i(x_j) \left( y_i - \sum_{d=0}^D \beta_{d,j} x_i^d \right)^2 \quad (3)$$

ce qui correspond à  $n$  problèmes d'optimisation. Les prédictions de la régression LOESS sont calculées avec cette formule :

$$\hat{y}_j = \sum_{d=0}^D \hat{\beta}_{d,j} x_j^d \quad (4)$$

où les valeurs  $(\hat{\beta}_{d,j})_{d,j}$  sont les solutions calculées des problèmes d'optimisation (3).

Pour réaliser cette régression, nous utilisons le package [Loess.jl](#) qui applique une méthode de calcul présentée dans [Cleveland and Grosse \(1991\)](#).

L'hyperparamètre  $h$  correspond à un facteur d'adoucissement : plus  $h$  est élevé, plus la courbe de régression est lisse. Ainsi on peut choisir  $h$  en fonction du niveau de lissage que l'on souhaite avoir. Cependant, il existe une méthode de sélection de ce paramètre appelée "Modified Partitioned Cross-Validation" (MPCV) développée dans la thèse de [Hoang \(2010\)](#), basée par exemple sur le découpage en plusieurs groupes de la série de données et le calcul de scores de validation croisée. La mise en place de cette méthode de sélection est très complexe et il n'existe pas de code disponible en open source pour l'utiliser. La valeur que nous avons choisie est une valeur qui nous a été recommandée directement par l'autrice principale de [Cognot et al. \(2025\)](#) et qui vaut  $0.08 \times N$ . Pour le choix de la valeur  $D$ , nous le prendrons égale à 1 comme ce qui est fait dans les articles de [Cognot et al. \(2025\)](#) et [Parey et al. \(2014\)](#), ce qui correspond à une régression linéaire locale.

### 2.3.2 Saisonnalité

Pour estimer les saisonnalités additives et multiplicatives, on considérera qu'elles ont une forme de polynôme trigonométrique centré comme ce qui est recommandé dans les articles de [Cognot et al. \(2025\)](#) et [Parey et al. \(2014\)](#). C'est-à-dire qu'elles suivent cette équation :

$$S_t^* = c_0 + \sum_{j=1}^q \alpha_j \cos(\omega jt) + \beta_j \sin(\omega jt) \quad (5)$$

avec  $\omega = 2\pi/365.25$ ,  $q$  l'ordre et  $\alpha_1, \alpha_2, \dots, \alpha_q, \beta_1, \beta_2, \dots, \beta_q$  les coefficients à estimer.

L'estimation de la tendance additive a fait que l'estimation de la saisonnalité additive a eu lieu sur un jeu de données centré, ce qui fait que pour  $S_t^* = S_t$  on a  $c_0 = 0$ . De la même manière, l'estimation de la tendance multiplicative a fait que l'estimation de la saisonnalité multiplicative a eu lieu sur un jeu de données réduit, ce qui donne  $c_0 = 1$  pour  $S_t^* = \sigma_{S_t}$ . Les coefficients sont estimés à l'aide de la méthode des moindres carrés. Pour estimer les saisonnalités, on doit alors choisir leur ordre. On notera  $q$  l'ordre de la saisonnalité additive et  $\sigma_q$  l'ordre de la saisonnalité multiplicative. Le choix de ces entiers sera expliqué dans la section 2.6.

## 2.4 Modèles pour la partie résiduelle

Maintenant il nous reste à modéliser la partie résiduelle  $X$ , qui est censé être stationnaire. Il est important de pouvoir bien la modéliser, car c'est elle qui contient de l'aléatoire et qui est à l'origine de propriétés importantes comme les autocorrélations. Pour cette partie ils existent quelque modèles et nous allons en présenter certains dans cette partie.

On rappelle qu'à partir de cette section  $T_t$  et  $X_t$  pourront être en dimension supérieure à 1 comme cela a été expliqué dans la section 1.4.

### 2.4.1 Modèle AR

L'un des pionniers de la recherche sur les générateurs stochastiques de météo, [Richardson \(1981\)](#), propose un modèle relativement simple pour une série de température qui a été centrée et réduite, ce qui correspond à notre partie résiduelle :

**Définition 2.2** (Modèle de génération multivarié de Richardson). *Soit  $(X_t)_{1 \leq t \leq N}$  la série des  $d$  parties résiduelles appariées (i.e  $X_t \in \mathbb{R}^d \forall t$ ). Le modèle de génération multivarié de Richardson est le modèle où pour  $t$  entre 2 et  $N$ ,  $X_t$  suit cette équation :*

$$X_t = AX_{t-1} + \Sigma \varepsilon \quad (6)$$

avec  $\varepsilon \sim \mathcal{N}(0, I_d)$ ,  $A \in \mathcal{M}_d(\mathbb{R})$  une matrice de paramètres à estimer et  $\Sigma \in \mathcal{M}_d(\mathbb{R})$  également une matrice de paramètres.

Cela correspond donc à un modèle AR(1) multivarié ([Indjehagopian, 1979](#)). On peut interpréter cette formule de la manière suivante :

- $AX_{t-1}$  correspond à la mémoire du jour précédent. C'est-à-dire l'effet des températures au jour  $t - 1$  sur celles du jour  $t$ .
- $\Sigma \varepsilon$  correspond à un effet aléatoire sur les températures au jour  $t$ , de matrice de variance-covariance  $\Sigma \Sigma^T$ .

Cependant, une limite de ce modèle est qu'il ne considère que la mémoire du jour précédent. On peut alors considérer le modèle AR( $p$ ) multivarié qui prendra en compte la mémoire sur plusieurs jours :

**Définition 2.3** (Modèle AR( $p$ )). Soit  $(X_t)_{1 \leq t \leq N}$  la série des  $d$  parties résiduelles appariées et  $p$  un entier strictement positif. Le modèle autorégressif multivarié d'ordre  $p$  (noté AR( $p$ )) est le modèle où pour  $t$  entre  $p+1$  et  $N$ ,  $X_t$  suit cette équation :

$$X_t = \sum_{i=1}^p A_i X_{t-i} + \Sigma \varepsilon \quad (7)$$

avec  $\varepsilon \sim \mathcal{N}(0, I_d)$ ,  $(A_i)_{1 \leq i \leq p}$  une suite de matrices de dimension  $d \times d$  de paramètres à estimer et  $\Sigma \in \mathcal{M}_d(\mathbb{R})$  également une matrice de paramètres.

Pour la suite on notera  $\theta = (A_1, A_2, \dots, A_p, \Sigma)$  l'ensemble des paramètres de ce modèle. Pour pouvoir estimer les paramètres, il faut comprendre qu'elle est la loi de  $X$ . Dans un premier temps, l'équation (7) nous dit que pour tout  $t > p$ , la loi de  $X_t$  conditionnellement à  $X_{t-1}, X_{t-2}, \dots, X_{t-p}$  est  $\mathcal{N}(\sum_{i=1}^p A_i X_{t-i}, \Sigma \Sigma^T)$ . Ainsi, on peut calculer la vraisemblance de la loi de  $(X_t)_{1 \leq t \leq N}$  conditionnellement aux  $p$  premiers termes, et estimer l'estimateur du maximum de vraisemblance  $\hat{\theta}$ . Soit  $f_\theta$  la fonction de densité de la variable aléatoire qu'elle prend en entrée et  $L_\theta = L((X_t)_{p+1 < t \leq N}; (X_t)_{1 \leq t \leq p}, \theta)$  la vraisemblance du modèle (conditionné à  $X_1, \dots, X_p$ ) :

$$\begin{aligned} L_\theta &= f_\theta(X_N, X_{N-1}, \dots, X_{p+1} | X_p, X_{p-1}, \dots, X_1) \\ &= f_\theta(X_N | X_{N-1}, X_{N-2}, \dots, X_1) f_\theta(X_{N-1}, \dots, X_{p+1} | X_p, X_{p-1}, \dots, X_1) \\ &= f_\theta(X_N | X_{N-1}, X_{N-2}, \dots, X_{N-p}) f_\theta(X_{N-1}, \dots, X_{p+1} | X_p, X_{p-1}, \dots, X_1) \end{aligned}$$

Par indépendance conditionnelle de  $X_N$  avec  $X_{N-p-1}, X_{N-p-2}, \dots, X_1$ . En réitérant  $N - p - 1$  fois cette manipulation sur  $f_\theta(X_{N-1}, \dots, X_{p+1} | X_p, \dots, X_2, X_1)$  on obtient :

$$\begin{aligned} L_\theta &= \prod_{k=p+1}^N f_\theta(X_k | X_{k-1}, \dots, X_{k-p}) \\ &= \prod_{k=p+1}^N \frac{1}{\sqrt{(2\pi)^d \det(\Sigma \Sigma^T)}} \exp \left( -\frac{1}{2} (X_k - \sum_{i=1}^p A_i X_{k-i})^T (\Sigma \Sigma^T)^{-1} (X_k - \sum_{i=1}^p A_i X_{k-i}) \right) \\ &= \frac{1}{\sqrt{(2\pi)^d \det(\Sigma \Sigma^T)}^{N-p}} \exp \left( \sum_{k=p+1}^N -\frac{1}{2} (X_k - \sum_{i=1}^p A_i X_{k-i})^T (\Sigma \Sigma^T)^{-1} (X_k - \sum_{i=1}^p A_i X_{k-i}) \right) \end{aligned}$$

Ainsi pour trouver l'estimateur du maximum de vraisemblance il suffit d'appliquer un algorithme d'optimisation sur la log-vraisemblance :

$$\log(L_\theta) = -\frac{1}{2} \left( (N-p)(d \log(2\pi) + \log(\det(\Sigma \Sigma^T))) + \sum_{k=p+1}^N (X_k - \mu_k)^T (\Sigma \Sigma^T)^{-1} (X_k - \mu_k) \right) \quad (8)$$

avec  $\mu_k = \sum_{i=1}^p A_i X_{k-i}$ .

Une limite de ce modèle est que les coefficients estimés restent constants pendant toute la période de la série, ce qui n'est pas forcément réaliste. En particulier, il se peut que l'effet des températures des jours précédents sur ceux d'un jour considéré soit différent pour chaque saison de l'année. En d'autres termes, ce modèle nécessite l'hypothèse que  $X$  est stationnaire exactement tout le temps, ce qu'on peut difficilement garantir dus aux évolutions qui peuvent exister au cours d'une année. Cela nous amène à modifier légèrement ce modèle dans la section suivante.

### 2.4.2 Modèle AR mensuel

Une variante du modèle précédent est celui où les paramètres à estimer varient en fonction du mois de l'année. Ce modèle est adapté à une hypothèse de stationnarité par mois, c'est-à-dire qu'on suppose  $X_t$  de même loi sur tous les mois de Janvier, de même loi sur tous les mois de Février, etc. Avoir des paramètres différents pour chaque mois se concrétise par une multiplication par 12 de la dimension de  $\theta$  :

$$\theta = (A_{1,1}, A_{2,1}, \dots, A_{p,1}, \Sigma_1, A_{1,2}, A_{2,2}, \dots, A_{p,2}, \Sigma_2, \dots, A_{1,12}, A_{2,12}, \dots, A_{p,12}, \Sigma_{12}).$$

Voici comment on définit ce nouveau modèle :

**Définition 2.4** (Modèle AR( $p$ ) mensuel). *Soit  $(X_t)_{1 \leq t \leq N}$  la série des  $d$  parties résiduelles appariées et  $p$  un entier strictement positif. Le modèle autorégressif multivarié mensuel d'ordre  $p$  (noté AR( $p$ )) est le modèle où pour  $t$  entre  $p+1$  et  $N$ ,  $X_t$  suit cette équation :*

$$X_t = \sum_{i=1}^p A_{i,m(t)} X_{t-i} + \Sigma_{m(t)} \varepsilon \quad (9)$$

avec  $\varepsilon \sim \mathcal{N}(0, I_d)$ ,  $m(t)$  le numéro du mois de la date  $t$ ,  $(A_{i,j})_{\substack{1 \leq i \leq p \\ 1 \leq j \leq 12}}$  une suite de matrices de dimension  $d \times d$  de paramètres à estimer et  $(\Sigma_j)_{1 \leq j \leq 12}$  également une suite de matrices de paramètres de même dimension.

L'équation (9) nous donne pour  $t$  entre  $p+1$  et  $N$  la loi conditionnelle de  $X_t$  :  $X_t | \{X_{t-1}, X_{t-2}, \dots, X_{t-p}\} \sim \mathcal{N}(\sum_{i=1}^p A_{i,m(t)} X_{t-i}, \Sigma_{m(t)} \Sigma_{m(t)}^T)$  avec  $m(t)$  le numéro du mois de la date  $t$ . Avec la même hypothèse que  $X_t$  est indépendant de  $X_{t-j}$  pour  $j > p$ , en reprenant les 4 premières lignes de calculs qui nous ont mené à connaître la vraisemblance du modèle AR( $p$ ) multivarié (section 2.4.1), on obtient cette vraisemblance pour le modèle AR( $p$ ) mensuel multivarié :

$$L_\theta = \prod_{k=p+1}^N \frac{1}{\sqrt{(2\pi)^d \det(\Sigma_{m(k)} \Sigma_{m(k)}^T)}} \exp \left( -\frac{1}{2} (X_k - \sum_{i=1}^p A_{i,m(k)} X_{k-i})^T (\Sigma_{m(k)} \Sigma_{m(k)}^T)^{-1} (X_k - \sum_{i=1}^p A_{i,m(k)} X_{k-i}) \right) \quad (10)$$

. Le passage au logarithme donne :

$$\log(L_\theta) = -\frac{1}{2} \sum_{k=p+1}^N d \log(2\pi) + \log(\det(\Sigma_{m(k)} \Sigma_{m(k)}^T)) + (X_k - \mu_k)^T (\Sigma_{m(k)} \Sigma_{m(k)}^T)^{-1} (X_k - \mu_k) \quad (11)$$

avec  $\mu_k = \sum_{i=1}^p A_{i,m(k)} X_{k-i}$ .

Ainsi, il suffit d'utiliser un algorithme d'optimisation sur cette valeur pour calculer l'estimateur du maximum de vraisemblance  $\hat{\theta}$ .

### 2.4.3 Discussion sur l'estimation mensuel

Pour estimer un modèle AR mensuel, au cours du stage d'autres méthodes ont été réfléchies :

1. On peut estimer un modèle AR standard sur chaque mois de la série, puis calculer la moyenne ou la médiane de chaque paramètre pour chaque mois de l'année. Par exemple, pour une série de 30 ans qui commence un 1<sup>er</sup> Janvier et fini un 31 Décembre, on va avoir 30 mois de Janvier donc 30 estimateurs de  $A_{1,1}$ . L'estimateur  $\hat{A}_{1,1}$  que l'on retient pour le modèle mensuel est la moyenne ou la médiane des 30 estimateurs de  $A_{1,1}$ .
2. On peut concaténer tous les mois de Janvier ensemble et faire l'estimation d'un AR standard dessus pour obtenir les paramètres du mois de Janvier, puis faire la même chose avec les 11 autres mois de l'année.

3. Et enfin, pour estimer les paramètres du mois de Janvier par exemple, on peut calculer les vraisemblances des modèles AR sur chaque mois de Janvier, puis faire l'optimisation sur la somme de toutes ces vraisemblances. On fait la même chose avec les 11 autres mois de l'année pour obtenir l'estimation complète des paramètres.

Nous considérerons la méthode qui consiste à optimiser la log vraisemblance (11) énoncée dans la section précédente comme étant la méthode 4. Pour avoir une idée de quelle est la meilleure méthode parmi les quatre, on a généré une série univariée ( $d = 1$ ) de températures résiduelles jouet de 100 ans ( $N = 36524$ ) selon un AR(2) mensuel dont on a choisi les paramètres. Dans ce cas les paramètres dans l'équation (9) seront des réels. Ces paramètres sont détaillés dans la table 1 et on montre avec le calcul des racines  $r_1$  et  $r_2$  des polynômes caractéristiques que les séries sont bien stationnaires pour chaque mois, en remarquant que ces racines sont toutes strictement supérieures à 1 en valeur absolue (voir le cours de séries chronologiques [Stupfler and Proia \(2024\)](#)).

	janv.	févr.	mars	avr.	mai	juin	juill.	août	sept.	oct.	nov.	déc
$A_1$	0.5	-0.7	0.1	0.6	-0.2	0.05	-0.3	0.5	0.3	-0.4	0.3	0.5
$A_2$	0.2	0.1	0.8	0.2	0.7	0.9	0.2	0.4	0.1	0.3	0.1	0.2
$\Sigma$	1.5	2.0	4.5	7.0	8.0	3.0	4.5	1.0	7.0	2.5	3.0	6.0
$r_1$	-3.81	-1.22	-1.18	-4.19	-1.06	-1.08	-1.61	-2.33	-5.0	-1.28	-5.0	-3.81
$r_2$	1.31	8.22	1.06	1.19	1.35	1.03	3.11	1.08	2.0	2.61	2.0	1.31

TABLE 1 – Paramètres et racines du polynôme caractéristique du modèle AR(2) mensuel jouet.

Pour pouvoir mesurer la qualité d'estimation, on calcule la moyenne sur les 12 mois des erreurs relatives entre le paramètre estimé et le vrai paramètre qui a servi pour la simulation. Par exemple, pour mesurer la qualité d'estimation de  $(\hat{A}_{1,j})_{1 \leq j \leq 12}$ , on calcule  $\frac{1}{12} \sum_{j=1}^{12} \left| \frac{\hat{A}_{1,j} - A_{1,j}}{A_{1,j}} \right|$ . Les résultats sont présentés dans la table 2. On voit que pour ce modèle et la série jouet

Paramètre	Estimation	Sur	Avec somme des	Avec
		concaténation	vraisemblances par	vraisemblance
		mensuel (2.)	mois (3.)	totale (4.)
$A_1$		15.08	4.71	3.96
$A_2$		17.95	8.04	8.64
$\Sigma$		10.48	1.31	1.14

Paramètre	Estimation	Par moyennes des	Par médianes des
		paramètres estimés sur	paramètres estimés sur
		chaque mois (1.)	chaque mois (1.)
$A_1$		6.21	7.21
$A_2$		26.48	25.45
$\Sigma$		4.21	4.38

TABLE 2 – Moyenne sur les 12 mois des erreurs relatives d'estimation en pourcentage pour chaque méthode et chaque paramètre.

la meilleure méthode d'estimation est la 4., suivi de près par la méthode 3. La méthode 3. est meilleure que la méthode 4. seulement pour l'estimation de  $A_2$ . Il peut exister plusieurs explications des meilleurs résultats de la méthode 4. comparés à ceux des trois autres :

- Comme on considère des modèles AR conditionnels aux  $p$  premières valeurs des séries, utiliser ces modèles sur chaque mois soit pour estimer les paramètres (méthode 1.) soit

pour juste calculer la vraisemblance (méthode 3.) crée une perte d'information en ne prenant pas en compte les  $p$  premières températures journalières de chaque mois, tandis que la méthode 4. est conditionnée seulement par les  $p$  premières valeurs de toute la série.

- Concaténer tous les mois de Janvier par exemple revient à considérer que le 31 Janvier de la première année est suivi par le 1er Janvier de l'année suivante, ce qui est faux. Or c'est ce que fait la méthode 2. pour les 12 mois de l'année sur la série. Comme les modèles autorégressifs se servent de la mémoire des jours précédents pour calculer la température au jour suivant, il est important d'avoir une continuité réaliste entre les valeurs de la série sur laquelle on fait notre estimation.

Ainsi nous avons retenu la méthode 4. pour estimer les modèles AR mensuels.

## 2.5 Simulation de séries

Une fois que l'on a estimé :

- Les tendances et les saisonnalités de  $T_t$  ou de toutes les composantes de  $T_t$  si  $d > 1$ ,
- Les paramètres du modèle de la partie résiduelle,

Voici comment on simule une nouvelle série de températures  $(\hat{T}_t)_{1 \leq t \leq N}$  :

1. On définit d'abord les conditions initiales :  $\hat{X}_1 = X_1, \hat{X}_2 = X_2, \dots, \hat{X}_p = X_p$ .
2. On simule  $\varepsilon_t \sim \mathcal{N}(0, I_d)$  pour tout  $t > p$ .
3. On utilise le modèle de la partie résiduelle pour simuler de manière itérative la partie résiduelle de la nouvelle série. Dans notre cas, comme on a choisi un modèle AR( $p$ ) mensuel, il suffit d'appliquer, pour  $t$  entre  $p+1$  et  $N$ , cette formule de manière récursive :

$$\hat{X}_t = \sum_{i=1}^p \hat{A}_{i,m(t)} \hat{X}_{t-i} + \hat{\Sigma}_{m(t)} \varepsilon_t \quad (12)$$

avec  $(\hat{A}_{i,j})_{\substack{1 \leq i \leq p \\ 1 \leq j \leq 12}}$  et  $(\hat{\Sigma}_j)_{1 \leq j \leq 12}$  les matrices des paramètres estimées en maximisant la log-vraisemblance exprimée dans (11).

4. Une fois la partie stochastique générée, il suffit de rajouter les tendances et saisonnalités qui ont été estimées précédemment (voir section 2.3) pour obtenir  $\hat{T}_t$ . Par exemple, pour  $d = 1$ , cela revient à calculer  $\hat{T}_t = \hat{M}_t + \hat{S}_t + \hat{\sigma}_{M_t} \hat{\sigma}_{S_t} \hat{X}_t$ . Si  $d > 1$ , on fait ce même calcul pour chaque composante de  $\hat{X}_t$ .

Dans le cas où l'on veut générer plusieurs séries simultanément ( $d \leq 2$ ), par exemple des séries de températures minimales (TN) et maximales (TX) journalières, cet algorithme nous donne aucune garantie que  $TX > TN$ . Une solution serait de vérifier à chaque calcul de  $\hat{X}_t$  si, en rajoutant les tendances et saisonnalités juste en  $t$ ,  $\hat{T}_t = (\hat{TN}, \hat{TX})$  respecte bien cette propriété. Si ce n'est pas le cas, on resample  $\varepsilon_t$ , on calcule  $\hat{X}_t$  avec cette valeur, on calcule encore  $\hat{T}_t$  et on vérifie à nouveau que l'inégalité est bien respectée. Le cas échéant on retient  $\hat{T}_t$  et sinon on réitère jusqu'à ce que la propriété soit satisfaite.

## 2.6 Sélection du modèle

Pour entraîner un générateur stochastique de météo, on doit :

- Choisir l'ordre de la saisonnalité additive  $q$  et de la saisonnalité multiplicative  $\sigma_q$  (2.3.2).

— Choisir l'ordre  $p$  du modèle AR mensuel utilisé pour la partie résiduelle (2.4.1).

Pour simplifier la modélisation, on a décidé de prendre la même valeur pour l'ordre des saisonnalités additive et multiplicative (i.e  $q = \sigma_q$ ), ce qui nous donne un total de 2 hyperparamètres à choisir :  $p$  et  $q$ .

Ainsi, pour choisir quel est le meilleur modèle pour un certain ensemble de couples d'hyperparamètres on va se baser sur :

- Des critères quantitatifs : Critère d'information d'Akaike (abrégié "AIC", [Akaike \(1973\)](#)) et le critère d'information bayésien (abrégié "BIC", [Schwarz \(1978\)](#)) des modèles.
- Des critères qualitatifs : Est-ce que les séries générées ont bien les mêmes propriétés que la série d'entraînement (même variance, mêmes autocorrélations, réalisation de scénarios extrêmes, etc.).

### 2.6.1 Critères quantitatifs

Nous pouvons nous servir des critères AIC et BIC pour mesurer la qualité du modèle. D'abord, on peut s'appuyer sur les critères AIC et BIC des modèles de régression qui ont servi à l'estimation des saisonnalités (2.3.2) puis sur ceux du modèle autorégressif (2.4.2) et choisir quel couple d'hyperparamètres minimise tous ces critères. Cependant il peut être très difficile de trouver un modèle qui en minimise autant à la fois. Par exemple, un modèle pourra être très bon pour minimiser l'AIC des modèles des saisonnalités mais mauvais pour minimiser l'AIC du modèle des résidus, tandis qu'un autre modèle pourra être bon dans les cas inverses. Néanmoins, dans le cas  $d = 1$ , on peut calculer l'AIC ou le BIC du modèle complet en calculant sa vraisemblance. En injectant l'équation (9) dans (1), on obtient :

$$T_t = M_t + S_t + \sigma_{M_t} \sigma_{S_t} \left( \sum_{i=1}^p A_{i,m(t)} X_{t-i} \right) + \sigma_{M_t} \sigma_{S_t} \Sigma_{m(t)} \varepsilon \quad (13)$$

. On voit que, conditionnellement à  $X_{t-1}, X_{t-2}, \dots, X_{t-p}$  donc conditionnellement à  $T_{t-1}, T_{t-2}, \dots, T_{t-p}$  par transformation linéaire,  $T_t$  suit une loi normale d'espérance  $\mu_t = M_t + S_t + \sigma_{M_t} \sigma_{S_t} \left( \sum_{i=1}^p A_{i,m(t)} X_{t-i} \right)$  et de variance  $\gamma_t = (\sigma_{M_t} \sigma_{S_t} \Sigma_{m(t)})^2$ . Avec les mêmes manipulations faites pour montrer la vraisemblance du modèle AR multivarié dans la section 2.4.1, on obtient la vraisemblance suivante :

$$L((T_t)_{p+1 \leq t \leq N}; (T_t)_{1 \leq t \leq p}, \theta) = \prod_{k=p+1}^N \frac{1}{\sqrt{(2\pi)\gamma_k}} \exp\left(-\frac{(T_k - \mu_k)^2}{2\gamma_k}\right) \quad (14)$$

. Le passage au logarithme donne :

$$\log(L((T_t)_{p+1 \leq t \leq N}; (T_t)_{1 \leq t \leq p}, \theta)) = -\frac{1}{2} \sum_{k=p+1}^N \log(2\pi) + \log(\gamma_k) + \frac{(T_k - \mu_k)^2}{\gamma_k} \quad (15)$$

. Toujours dans le cas  $d = 1$ , on a  $12(p+1)$  paramètres pour la partie AR mensuel et  $2(2q+1)$  paramètres pour les saisonnalités additive et multiplicative, ce qui donne un total de  $12(p+1)+2(2q+1)$  paramètres pour le modèle complet. En notant  $L_\theta = L((T_t)_{p+1 \leq t \leq N}; (T_t)_{1 \leq t \leq p}, \theta)$ , on obtient :

$$\begin{cases} \text{AIC} = -2 \log(L_\theta) + 2(12(p+1) + 2(2q+1)) \\ \text{BIC} = -2 \log(L_\theta) + (12(p+1) + 2(2q+1)) \log(N) \end{cases} \quad (16)$$

. On a ainsi deux critères quantitatifs évaluant la qualité globale du modèle, ce qui nous permet de plus facilement choisir le meilleur couple d'hyperparamètres. Pour les modèles multivariés, c'est-à-dire ceux modélisant simultanément plusieurs types de températures avec  $d > 1$ , on se limitera à regarder quels étaient les meilleurs modèles univariés sur chacun de ses types de



températures et à sélectionner les hyperparamètres les plus proches des leurs. Par exemple, si pour une série de  $TN_t$  on a choisi  $p = 2$  et  $k = 5$  et pour une série de  $TX_t$  on a  $p = 2$  et  $k = 6$ , on pourra choisir  $p = 2$  et  $k = 6$  pour le modèle de  $(TN_t, TX_t)_{1 \leq t \leq N}$  quitte à avoir un léger sur-apprentissage. On se permet de faire ce choix pour plusieurs raisons :

- Les modèles multivariés sont très coûteux à entraîner, il aurait fallu énormément de temps (à l'ordre de plusieurs jours avec les moyens à disposition pour le stage) pour entraîner tout un panel de modèles avec par exemple  $p$  qui varie entre 1 et 8 et  $q$  qui varie entre 1 et 10.
- Pour une même station on a trouvé souvent des hyperparamètres assez proches pour chaque type de température (voir table 5) ce qui suggère un comportement assez similaire.

### 2.6.2 Critères qualitatifs

Enfin, le plus important pour un modèle de générateur de météo est qu'il puisse simuler des séries de températures avec la même loi que la série d'entraînement. Ainsi, une grande part de l'évaluation des modèles se fera de manière graphique. On pourra observer :

- Pour chaque jour de l'année, l'étendu et les quartiles des valeurs simulées et les comparer avec la valeur moyenne de la série d'entraînement.
- Les autocorrélations et autocorrélations partielles des séries d'entraînement et des séries simulées.
- Les maximums, moyennes et variances mensuels des séries générées et de la série d'entraînement.
- Pour les séries multivariées, les corrélations croisées entre les composantes. Par exemple, pour  $T_t = (TN_t, TX_t)$ , il s'agira d'étudier les corrélations entre  $TN_t$  et  $TX_{t+i}$  pour  $i$  allant de  $-4$  à  $4$  (ou un autre entier strictement positif). Ces corrélations croisées sont par exemple représentées dans [Richardson \(1981\)](#).

Avec ces résultats, on pourra savoir si les séries générées sont capables d'atteindre des valeurs extrêmes comme le fait la série d'entraînement.

### 3 Modèles phénologiques

Nous allons maintenant présenter le deuxième type de modèle utilisé pour notre étude : les modèles phénologiques. Chez les pommiers et les vignes à chaque printemps a lieu l'apparition des bourgeons, des pièces à l'origine des fruits et donc fondamentales pour la reproduction de la plante. Les modèles phénologiques que nous utiliserons ont pour but de prédire la date de débourrement, c'est-à-dire la date d'apparition de ces bourgeons, en fonction des températures journalières. Dans un contexte de réchauffement climatique, ces dates annuelles de débourrement risquent donc d'être modifiées. Voici deux exemples de risques que de tels changements peuvent amener :

- Une désynchronisation des dates de débourrement des pommiers de plusieurs espèces peut être néfaste pour leur reproduction. Cela peut avoir lieu si les effets du réchauffement climatique n'ont pas la même intensité sur toutes les zones géographiques (Kőrösi et al., 2018, Legave et al., 2015).
- Une date de débourrement trop avancée peut rendre le bourgeon vulnérable à un risque de gel (García de Cortázar-Atauri et al., 2009).

Par conséquent, avoir un modèle capable de prédire la date de débourrement d'une plante nous permet d'évaluer la présence potentielle de ces risques et donc de mesurer l'impact du réchauffement climatique sur les pommiers et les vignes à partir d'une série de températures donnée. Dans cette section on présentera comment fonctionne les modèles phénologiques que nous utiliserons et nous détaillerons les spécificités du modèle utilisé pour les pommes et de celui utilisé pour les vignes.

#### 3.1 Principe

Du début de l'automne jusqu'au débourrement au printemps, la plante vit deux phases successives :

- La première est appelée endo-dormance : la plante perd ses feuilles et dans notre modèle elle accumule des quantités appelées unités de "refroidissement" ou "chilling units". Comme leur nom l'indique, ces unités dépendent de la température journalière et en principe plus il fait froid, plus cette valeur est élevée. Lorsque la plante a accumulé une certaine quantité d'unités de refroidissement ( $C$  dans l'équation (17)), la période d'endo-dormance se termine et la seconde phase se déclenche.
- Cette seconde phase est appelée éco-dormance : maintenant le modèle considère l'accumulation d'unités de "réchauffement" ou "heating units". Elles peuvent être aussi appelées "forcing units". Ces unités dépendent aussi des températures journalières et généralement plus il fait chaud, plus ces unités sont élevées. Lorsque le cumul des unités de réchauffement atteint un certain seuil ( $H$  dans l'équation (18)), le débourrement a lieu.

Il est important de noter que la rupture d'endo-dormance, c'est-à-dire le changement entre les deux phases, n'est pas visible à l'œil nu. On peut illustrer ce cycle phénologique avec la figure 3. Le modèle phénologique utilisé ici (F1 Gold 1, voir section 3.3) prend en entrée seulement la série de température journalière moyenne  $(TG_t)_{1 \leq t \leq N}$ . Les températures mises en entrées sont celles de Bonn (origine des données présentée à la section 4.1.1). Voici comment interpréter ce graphe :

- La ligne verticale bleue représente le début de la période d'endo-dormance. À partir de cette date, on a modélisé l'accumulation des unités de refroidissement avec l'aire en bleue.
- Lorsque cette accumulation a atteint un certain seuil ( $C$  dans l'équation (17)), la rupture d'endo-dormance représenté par la ligne verticale orange a lieu. On prend en compte maintenant l'accumulation des unités de réchauffement qui est représentée par l'aire orange.

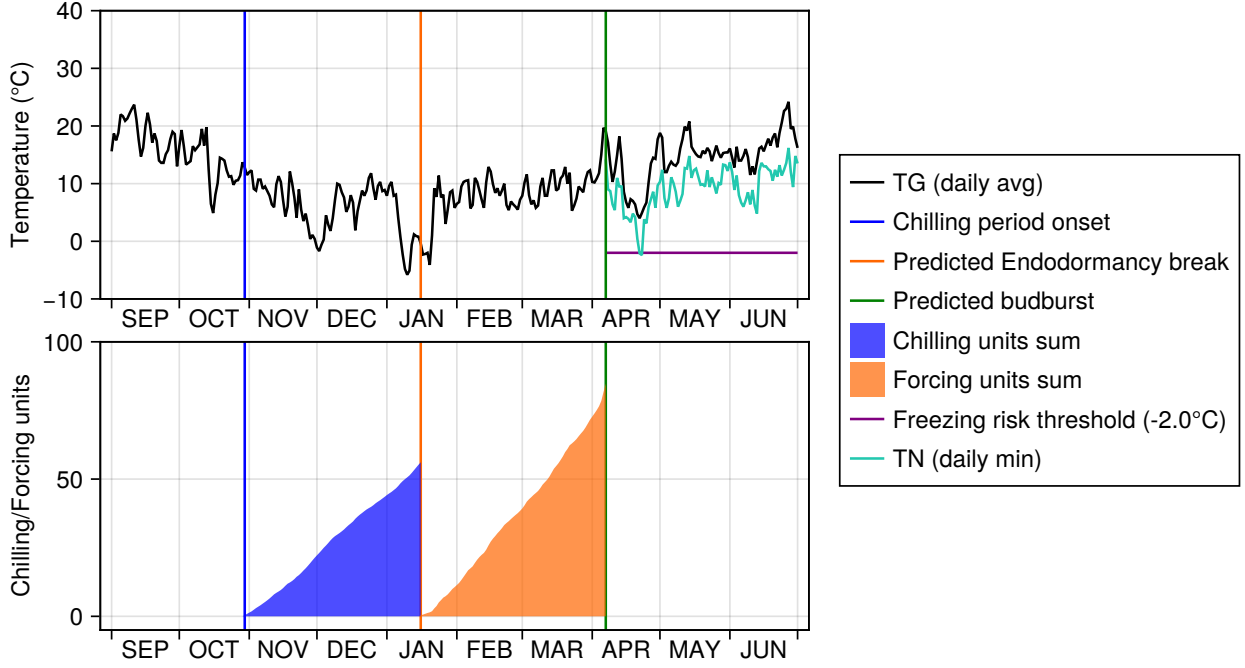


FIGURE 3 – Dates phénologiques prédites par le modèle pour les pommes (3.3) pour l'année 2023-2024 à Bonn.

- Une fois que cette dernière accumulation a atteint un certain niveau ( $H$  dans l'équation (18)), on obtient le débourrement qui est représenté par la ligne verticale verte.
- Après le débourrement, on a affiché les températures minimales journalières en cyan pour voir s'il y a un risque de gel après débourrement. Pour cette année, les températures journalières minimales sont descendues en dessous de  $-2^{\circ}\text{C}$  vers fin Avril alors que le modèle a prédit le débourrement vers début Avril. Cela constitue donc un risque de gel pour un bourgeon qui est par conséquent vulnérable.

Les modèles phénologiques pour les pommiers et les vignes se formalisent tous les deux de la manière suivante :

**Définition 3.1** (Modèle phénologique de pommier ou de vigne). *Un modèle phénologique de pommier ou de vigne est un modèle de prédiction de la date de rupture d'endo-dormance  $t_{eb}$  et de débourrement  $t_{bb}$  en fonction des températures journalières  $(T_t)_{1 \leq t \leq N}$  selon ces équations :*

$$t_{eb} = \min\left\{M, \sum_{t=\text{CPO}}^M F_c(T_t) > C\right\} \quad (17)$$

,

$$t_{bb} = \min\left\{M, \sum_{t=t_{eb}}^M F_h(T_t) > H\right\} \quad (18)$$

avec :

CPO : Début de période d'endo-dormance (Chilling period onset)

$C, H$  : Respectivement les quantités d'unités de refroidissement (Chilling) et de réchauffement (Heating) requises.

$F_c, F_h$  : Respectivement les fonctions de refroidissement et de réchauffement.

$F_c$  et  $F_h$  sont les fonctions qui renvoient la quantité d'unités de refroidissement et de ré-

chauffement en fonction des températures journalières. On notera que :

$$T_t = \begin{cases} \text{TG}_t & \text{Pour le modèle de pommier.} \\ (\text{TN}_t, \text{TX}_t) & \text{Pour la fonction de refroidissement } F_c \text{ du modèle de vigne.} \\ (\text{TN}_t, \text{TX}_t, \text{TN}_{t+1}) & \text{Pour la fonction de réchauffement } F_h \text{ du modèle de vigne.} \end{cases} \quad (19)$$

. Les valeurs CPO,  $C$  et  $H$  et les fonctions  $F_c$  et  $F_h$  sont propres aux modèles spécifiques aux pommiers et aux vignes et seront détaillées dans les sections 3.3 et 3.2.

### 3.2 Modèle pour les vignes (BRIN)

Pour les vignes nous utiliserons le modèle BRIN présenté dans [García de Cortázar-Atauri et al. \(2009\)](#). La fonction de refroidissement  $F_c$  proposée est donnée par :

$$F_c(\text{TN}_t, \text{TX}_t) = Q_{10c}^{-\frac{\text{TN}_t}{10}} + Q_{10c}^{-\frac{\text{TX}_t}{10}} \quad (20)$$

avec  $Q_{10c}$  un paramètre donné dans la table 3. On remarque ici que comme  $Q_{10c}$  est positif les applications partielles de  $F_c$  en  $\text{TN}_t$  et  $\text{TX}_t$  sont décroissantes : plus il fait froid, plus on accumule des unités de refroidissement.

La fonction de réchauffement  $F_h$  donne la quantité d'unités de réchauffement en fonction de  $\text{TN}_t, \text{TX}_t$  à partir de plusieurs calculs intermédiaires. D'abord, on interpole linéairement les températures horaires  $T^*(h, t)$  en considérant que  $T^*(0, t) = \text{TN}_t$ ,  $T^*(12, t) = \text{TX}_t$  et que  $T^*(24, t) = \text{TN}_{t+1}$  :

$$T^*(h, t) = \begin{cases} \text{TN}_t + h \left( \frac{\text{TX}_t - \text{TN}_t}{12} \right) & \text{si } h \leq 12 \\ \text{TX}_t - (h - 12) \left( \frac{\text{TX}_t - \text{TN}_{t+1}}{12} \right) & \text{si } h > 12 \end{cases} \quad (21)$$

. Ensuite on applique un ajustement sur  $T^*(h, t)$  donné par :

$$T(h, t) = \begin{cases} 0 & \text{if } T^*(h, t) < T_{0Bc} \\ T^*(h, t) - T_{0Bc} & \text{if } T_{0Bc} \leq T^*(h, t) \leq T_{MBc} \\ T_{MBc} - T_{0Bc} & \text{if } T_{MBc} < T^*(h, t) \end{cases} \quad (22)$$

où  $T_{0Bc}$  est un paramètre donné dans la table 3 et  $T_{MBc}$  est fixée à 25°C.

Enfin, la fonction de réchauffement est donnée par la somme sur une journée de ces températures horaires ajustées :

$$F_h(\text{TN}_t, \text{TX}_t, \text{TN}_{t+1}) = \sum_{h=1}^{24} T(h, n) \quad (23)$$

. On remarque ici que plus les températures horaires interpolées sont élevées, plus on accumule des unités de réchauffement. Voici l'ensemble des paramètres donnés dans l'article :

Paramètre	CPO	$C$	$H$	$Q_{10c}$	$T_{0Bc}$
Valeur	1 <sup>er</sup> Août	119.0 unités	13236 C°	2.17	8.19 C°

TABLE 3 – Paramètres donnés dans l'article [García de Cortázar-Atauri et al. \(2009\)](#) pour le modèle BRIN

### 3.3 Modèle pour les pommes (F1 Gold 1)

Pour les pommes nous utiliserons le modèle F1 Gold 1 présenté dans [Legave et al. \(2013\)](#). C'est un modèle spécifique à la variété Golden Delicious. La fonction de refroidissement  $F_c$  proposée est une fonction dite "triangulaire" donnée par :

$$F_c(T) = \begin{cases} 1 - (|T - T_c|/l_c) & \text{si } T \in (T_c - l_c, T_c + l_c) \\ 0 & \text{sinon} \end{cases} \quad (24)$$

avec  $T_c$  et  $l_c$  des paramètres donnés dans la table 4. Cette formule implique que la température maximisant la quantité d'unité de refroidissement est  $T_c$ , et que cette quantité décroît lorsque  $T$  s'éloigne de  $T_c$ . Comme ce n'est pas une fonction décroissante, cela signifie que des températures trop froides ne vont pas avancer la fin de la période d'endo-dormance mais la retarder.

La fonction de réchauffement  $F_h$  est une fonction exponentielle :

$$F_h(T) = \exp(T/T_h) \quad (25)$$

avec  $T_h$  un paramètre donné dans la table 4. Cette fois-ci il s'agit d'une fonction strictement croissante donc plus il fait chaud, plus le débourrement va être avancé. Voici l'ensemble des paramètres proposés dans l'article :

Paramètre	CPO	$C$	$H$	$T_c$	$l_c$	$T_h$
Valeur	30 Octobre	56.0 unités	83.58 unités	1.1°C	20.°C	9.0°C

TABLE 4 – Paramètres donnés dans l'article [Legave et al. \(2013\)](#) pour le modèle F1 Gold 1

On remarque que  $T_c$  correspond à une température hivernale. Comme c'est la valeur qui maximise  $F_c$ , cela montre que des températures basses vont généralement accélérer l'accumulation des unités de refroidissement tant qu'elles ne sont pas extrêmes.

## 4 Mises en œuvre et résultats

### 4.1 Données

#### 4.1.1 Séries de températures enregistrées : données ECA&D

Nous entraînerons nos modèles de générateurs stochastiques de météo et appliquerons nos modèles phénologiques d'abord sur des données de températures enregistrées par des stations météorologiques. Nous les avons collectées sur le site de l'European Climate Assessment and Dataset (abrégé ECA&D) : <https://www.ecad.eu/dailydata/predefinedseries.php>. Ce site contient une base de données des températures journalières des stations météorologiques européennes. Pour chacune des stations que nous avons sélectionnées nous avons collecté les températures minimales, moyennes et maximales journalières (pour rappel leurs notations respectives sont  $TN_t$ ,  $TG_t$  et  $TX_t$ ). Voici les stations que nous avons sélectionnées :

- **Montpellier-Aéroport, France** : pour la période du 01/01/1946 au 31/12/2022
- **Nantes-Bougenais, France** : pour la période du 01/01/1992 au 31/01/2025
- **Koln-Bonn, Allemagne** : pour la période du 01/11/1976 au 28/02/2025

S'il y a des valeurs manquantes, on a retiré de la série ce qu'il y a avant la dernière valeur manquante pour avoir une continuité. Si une valeur se trouve vers la fin de la série, on a retiré aussi ce qu'il y a après. Cela explique les différences entre les périodes de chaque station. On a choisi ces stations pour les raisons suivantes :

- Montpellier est un site étudié dans [García de Cortázar-Atauri et al. \(2009\)](#) et proche de celui de Nîmes lui-même étudié dans [Legave et al. \(2013\)](#). Cet endroit est un bon exemple de climat méditerranéen.
- Nantes est proche d'Angers, ce dernier étant un site étudié dans les deux derniers articles cités. Cela correspond à un bon exemple de climat océanique. La raison pour laquelle on n'a pas choisi Angers est que la série de Nantes contient moins de valeurs manquantes.
- Bonn est un site présenté dans [Legave et al. \(2013\)](#) est constitué un bon exemple de climat continental.

#### 4.1.2 Projections climatiques : données du portail DRIAS

Nous entraînerons aussi le générateur de météo sur des projections de modèles climatiques. Les séries récupérées sont issues du portail DRIAS *Les futurs du climat* (DRIAS pour "Donner accès aux scénarios climatiques Régionalisés français pour l'Impact et l'Adaptation de nos Sociétés et environnement") développé par Météo France pour rendre accessible les données issues des modèles numériques de simulation climatique : <https://www.drias-climat.fr/>. Le modèle de projection climatique a comme nom CNRM-ALADIN63 - CNRM-CERFACS-CNRM-CM5 (RCM - GCM). On a choisi la sortie du modèle pour la localisation  $47.173^\circ, -1.621930^\circ$ , le point disponible le plus proche de la station météorologique de Nantes-Bougenais pour avoir une continuité avec les résultats issus de la série de température enregistrée de cette station. Pour ce point nous avons deux simulations correspondant chacune aux deux scénarios rcp4.5 et rcp8.5 du GIEC (rcp pour "Representative Concentration Pathway", voir description sur le site du portail [DRIAS](#)). Ces simulations climatiques contiennent les données des températures minimales, moyennes et maximales journalières et s'étendent du 1<sup>er</sup> août 2006 au 31 décembre 2100.

## 4.2 Sélection des générateurs de météo

Nous allons maintenant présenter comment les modèles de générateurs de météo ont été sélectionnés selon la méthode décrite dans la section 2.6.

### 4.2.1 Sélection des modèles univariés

Sur les séries de températures enregistrées minimales, moyennes et maximales journalières de Bonn, Nantes et Montpellier, nous avons entraîné les modèles univariés de générateurs météo pour l'ordre  $p$  de la partie AR qui varie entre 1 et 8 et l'ordre  $q$  de la saisonnalité qui varie entre 1 et 10. Cela donne un total de  $3 \times 3 \times 8 \times 10 = 720$  modèles entraînés univariés. Lors des estimations, les vraisemblances ont été maximisées via un algorithme d'optimisation nommé **BFGS**. Comme il peut être très fastidieux de regarder les graphes des résultats de 720 modèles, on a décidé de d'abord trier quels modèles ont les meilleurs critères AIC ou BIC (ceux donnés par l'équation 16) puis de vérifier s'ils sont satisfaisants sur les autres critères observables. Sur les tableaux des résultats dans l'annexe A, on remarque que les modèles qui optimisent le critère BIC ont des valeurs de  $q$  plus faibles que ceux qui optimisent le critère AIC. Cela rend le critère BIC moins fiable étant donné qu'il est nécessaire que la saisonnalité estimée soit réaliste (voir graphe B. de la figure 4). Voici alors les couples d'hyperparamètres qui maximisent le critère AIC :

	TN		TG		TX	
	p	q	p	q	p	q
Montpellier	6	2	7	3	7	4
Nantes	3	2	3	2	3	2
Bonn	5	2	3	1	4	2

TABLE 5 – Table des meilleurs couples d'hyperparamètres selon le critère AIC pour les modèles entraînés sur les températures enregistrées aux stations de Montpellier-Aéroport, Nantes-Bougenais et Koln-Bonn.

On peut analyser maintenant les résultats de ces modèles de manières graphiques. On va montrer ici les résultat du modèle pour les températures maximales journalières de Nantes (avec  $p = 3$  et  $q = 2$ ) sur la figure 4. Les autres résultats seront disponibles dans le dossier AR/Best\_results du dépôt github donné dans l'annexe E. Pour chacun de ces graphes on a généré 1000 simulations avec le générateur de météo. Voici comment interpréter chacun des graphes de la figure 4 :

- Le graphe A. représente en bleue la saisonnalité estimée et en jaune la moyenne pour chaque jour de l'année de toutes les températures enregistrées. Par exemple, pour le 1<sup>er</sup> Janvier, on a calculé la moyenne des températures des 33 1<sup>er</sup> Janvier de la série. Le fait que les deux courbes soient superposées nous montre que les températures simulées sont centrées autour de la saisonnalité.
- Le graphe B. compare des statistiques journalières, c'est-à-dire pour chaque jour de l'année, de la série enregistrée comme la moyenne en orange et la température maximale atteinte en rouge avec celles des séries simulées comme l'étendue et l'intervalle interquartile représentés respectivement par les bandes bleue clair et bleue foncé. On voit d'abord que la moyenne annuelle de la série enregistrée est proche de la saisonnalité estimée, ce qui est un point positif pour l'estimation de cette saisonnalité. On voit aussi que la bande bleue foncé correspondant à la moitié des températures simulées est aussi centrée autour de cette saisonnalité et donc centrée autour de la moyenne des températures enregistrées, ce qui est satisfaisant pour montrer le réalisme des séries simulées. Enfin, on voit que pour

chaque jour de l'année, le générateur de météo a été capable de simuler une température supérieure à ce qui a été enregistré. Par exemple, le générateur a été capable de générer une température de 15 Juillet supérieure aux 33 températures maximales journalière de 15 Juillet de la série enregistrée à la station de Nantes. Cela montre une capacité du générateur de météo à générer des températures extrêmes, ce qui est l'un de nos objectifs recherchés.

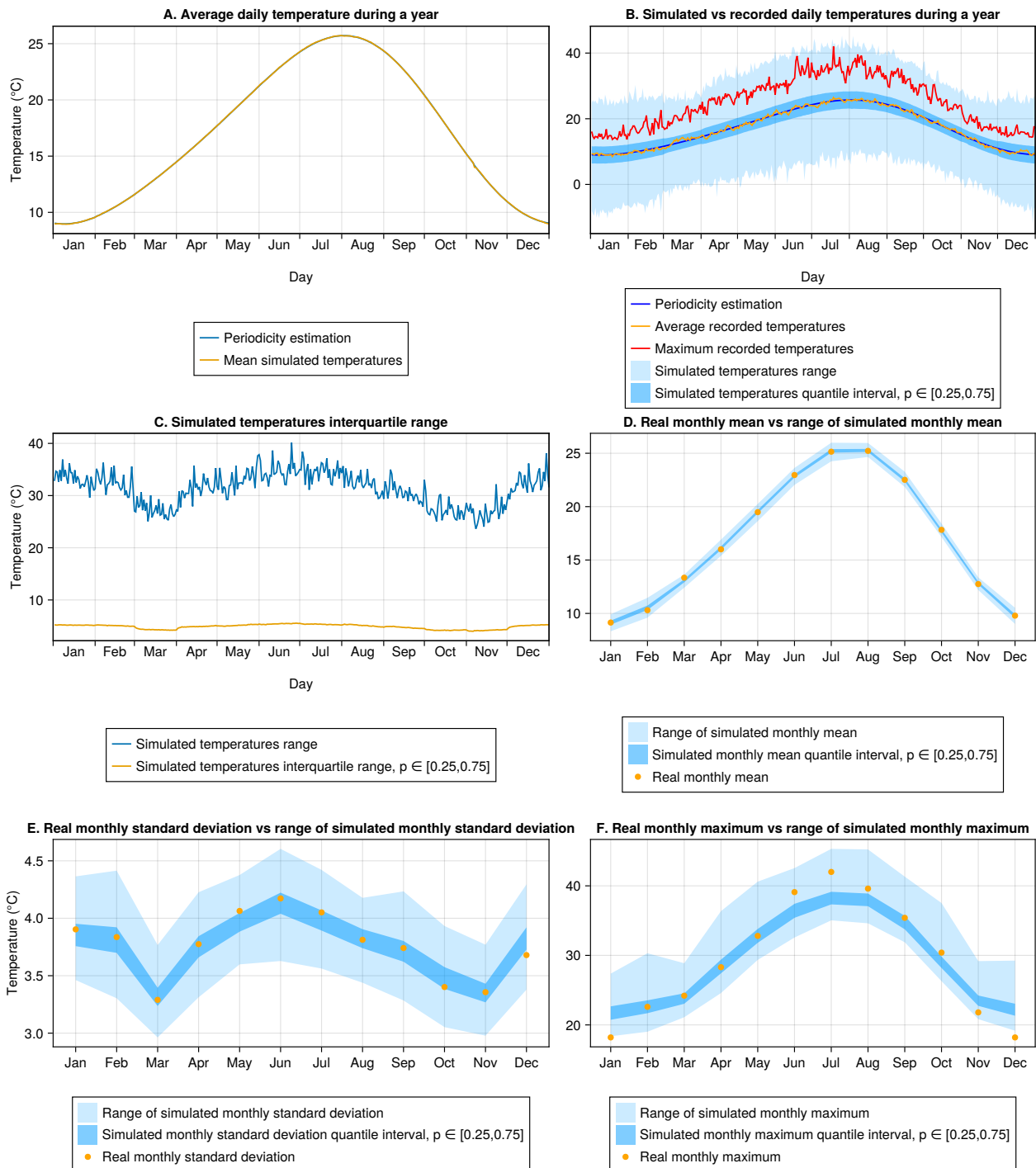


FIGURE 4 – Différentes comparaisons entre la série de températures maximales journalières enregistrées à la station Nantes-Bougenais et 1000 simulations du générateur de météo entraîné sur cette série avec  $p = 3$  et  $q = 2$ .

— Le graphe C. représente en ordonnée les largeurs des bandes de l'étendue et de l'intervalle



interquartile du graphe B. L'intérêt de ce graphe est surtout de voir que l'étendue des valeurs simulées varie au cours de l'année. En l'occurrence, ici cette étendue est plus élevée en hiver et en été et moins élevée lors des demi-saisons.

— Pour les moyennes, écarts-type et maximums sur chaque mois de l'année, respectivement les graphes D., E. et F. compare les données de la série enregistrée avec les étendues et quartiles des données des séries simulées. Par exemple, sur le graphe D., pour le mois de Janvier, le point jaune correspond à la moyenne de toutes les températures de tous les mois de Janvier de la série enregistrée. On a calculé cette moyenne aussi pour chaque simulation ce qui nous donne une distribution empirique de cette moyenne dont on montre l'étendue et les quartiles avec les bandes. Pour les moyennes et écarts-type, on voit que les données simulées restent centrées autour des données de la série enregistrée, ce qui montre que le générateur de météo est capable de simuler des séries proches en moyenne et en écart-type de la série d'entraînement. En revanche, pour le maximum mensuel, on remarque deux choses :

- Pour les mois de Décembre et Janvier, toutes les séries simulées ont un maximum de température supérieur à celui de la série d'entraînement. Cela signifie que le modèle sur-estime la présence et l'intensité de températures anormalement chaudes en période hivernale.
- En revanche, pour la période de Juin à Août, les maximums des séries simulées ne sont plus centrés autour de ceux de la série enregistrée mais plus de 75% d'entre eux sont en dessous. Il existe quand même des simulations qui ont des températures supérieures, mais cela montre quand même que le modèle sous-estime légèrement la présence d'extrêmes de chaleur en été.

Ces limites ne sont pas néanmoins dramatiques.

Ainsi, la figure 4 nous montre globalement un modèle génératif de qualité pour imiter des séries de températures maximales journalières. On va aussi comparer les ACF mensuelles des résidus des séries générées avec ceux des résidus de la série d'entraînement sur la figure 5. Pour faire ces graphes, pour la série d'entraînement et les séries simulées, on a calculé les ACF sur chaque mois de la série puis fait la moyenne pour chaque mois de l'année. Par exemple, les ACF affichées pour le mois de Janvier correspondent aux moyennes des ACF des 33 mois de Janvier de la série. On voit que, pour chaque mois de l'année et chaque ordre, globalement les ACF moyens des séries simulées sont proches de l'ACF moyenne de la série d'entraînement. Cela s'explique par le fait que pour chaque mois l'ACF de la série d'entraînement à une allure correspondant à un  $MA(\infty)$ , ce qui s'accorde donc avec le choix d'un modèle autorégressif pour modéliser les résidus. Et enfin, on a fait de même avec les PACF sur la figure 6. De mêmes, on voit que les PACF moyens des résidus des séries simulées sont proches de la PACF moyenne des résidus de la série d'entraînement. On remarque par exemple que la PACF à l'ordre 1 pour chaque mois est bien plus élevée que pour les autres ordres, mais que les PACF à l'ordre 2 et 3 peuvent varier d'un mois à l'autre. Cela est un peu visible à l'ordre 3, entre les mois de Mai et Juin pour lesquels cette valeur est un peu en dessous de 0 et les mois de Juillet et Août où cette valeur est très proche de 0. Cela est plus visible à l'ordre 2, où par exemple au mois d'Août la PACF à l'ordre 2 est bien inférieure à 0 tandis qu'au mois de Décembre elle est très proche de 0. Cela justifie donc l'utilisation d'un modèle autorégressif avec des paramètres différents pour chaque mois de l'année, même si on peut légèrement remettre en question la valeur  $p = 3$  choisie selon le critère AIC qui semble trop élevée au vu de ces résultats.

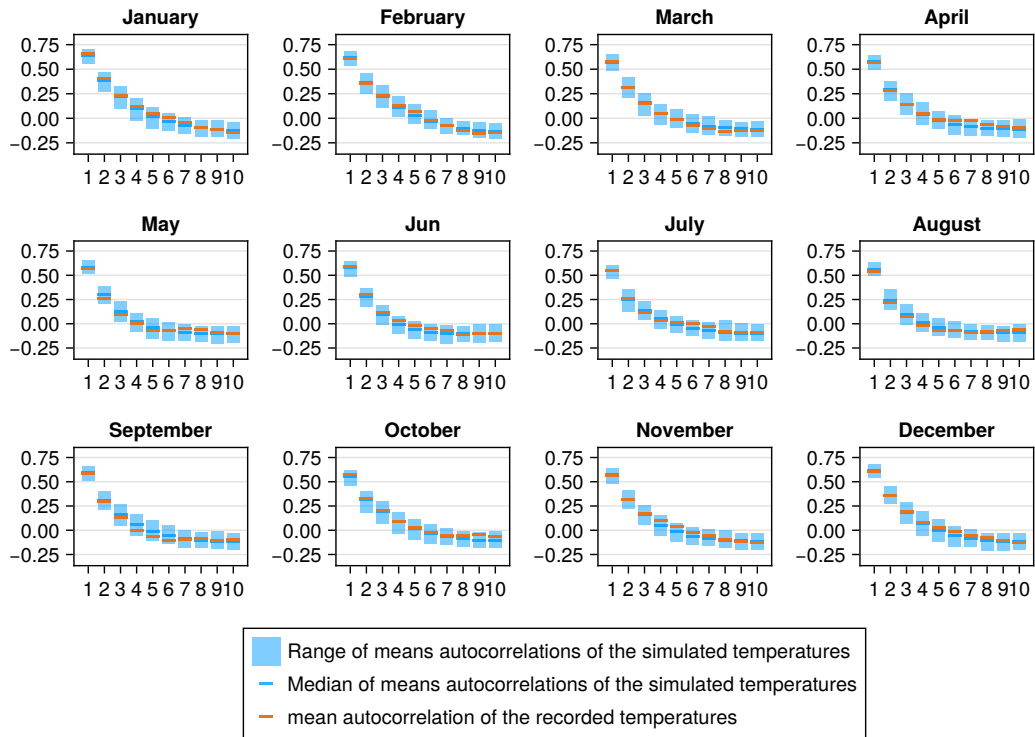


FIGURE 5 – Comparaison entre les ACF mensuelles des résidus de la série de températures maximales journalières enregistrées à Nantes-Bougenais et celles des résidus des 1000 simulations du générateur de météo entraîné sur cette série avec  $p = 3$  et  $q = 2$ .

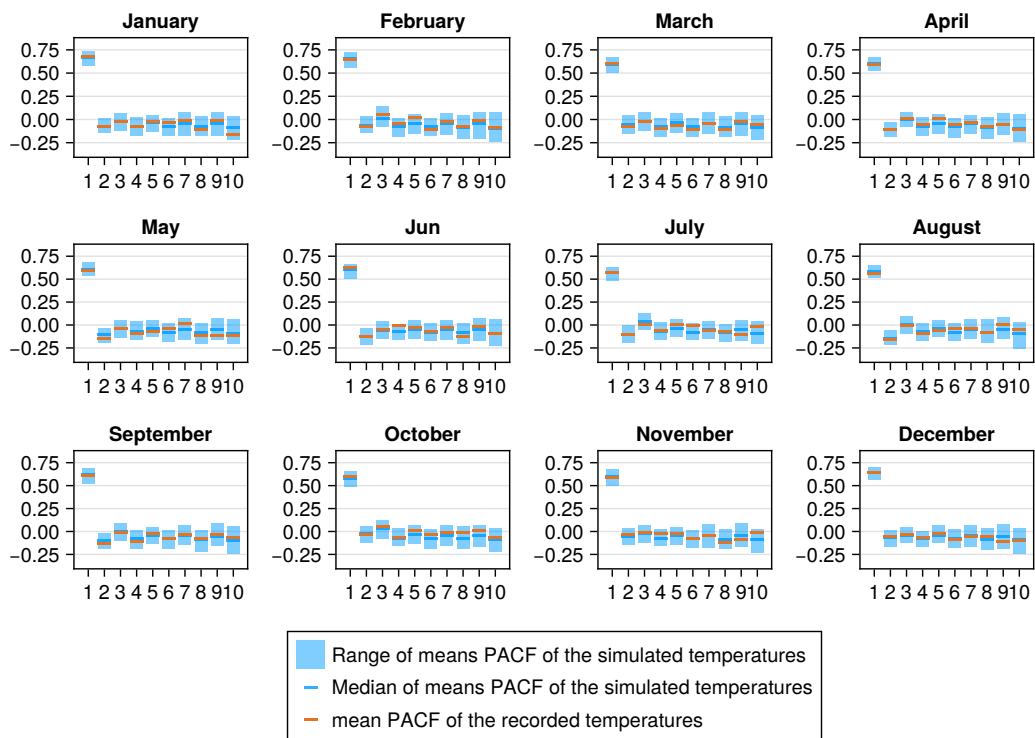


FIGURE 6 – Comparaison entre les PACF mensuelles des résidus de la série de températures maximales journalières enregistrées à Nantes-Bougenais et celles des résidus des 1000 simulations du générateur de météo entraîné sur cette série avec  $p = 3$  et  $q = 2$ .

## 4.2.2 Sélection des modèles multivariés

Pour pouvoir utiliser les modèles phénologiques sur des simulations de générateurs stochastiques de météo, selon la liste des entrées de modèles phénologiques (19), il faut que ce générateur soit un modèle multivarié :

- Pour les modèles de pommiers, on aurait pu se servir de modèles capables de générer seulement des séries univariées pour la température moyenne mais nous souhaitons aussi mesurer les risques de gel qui dépendent alors aussi des températures minimales journalières. Ainsi il nous faut un modèle capable de générer des séries de  $(TN_t, TG_t)$ .
- Pour les modèles de vignes, il nous faut un générateur capable de faire des simulations de séries de  $(TN_t, TX_t)$

Comme annoncé à la fin de la section 2.6.1, nous choisirons les modèles multivariés selon la sélection des modèles univariés via la table 5 mais aussi selon certains résultats vu sur les graphes. Pour certains modèles, on a par exemple baissé la valeur de  $p$  au vu des autocorrélations partielles (6) et augmenté la valeur de  $q$  pour avoir une saisonnalité plus fine. Voici la liste des modèles multivariés que nous avons retenus :

	TN et TG		TN et TX	
	p	q	p	q
Montpellier	5	2	6	5
Nantes	3	2	3	2
Bonn	5	2	5	3

TABLE 6 – Table des choix de couples d’hyperparamètres pour les modèles entraînés sur les températures enregistrées aux stations de Montpellier-Aéroport, Nantes-Bougenais et Koln-Bonn.

Tous les résultats d’estimation sont disponibles dans le dossier AR/Results\_multiple du dépôt github indiqué dans l’annexe E. Dans l’annexe sont mis à disposition les résultats d’estimation du modèle de TN et TX de Nantes Bougenais dans la section B sur des critères concernant les deux séries individuellement. Les résultats sont similaires à ceux du modèle univarié présenté dans la section 4.2.1, ce qui montre que le modèle multivarié est aussi efficace pour simuler des séries de températures. La seule différence est présente dans les graphes A. des figures 25 et 28. On voit que la moyenne des températures simulées est très légèrement supérieure à la saisonnalité au mois de décembre et janvier. Ce phénomène est apparu lorsque nous avons mis en place la correction dans l’algorithme de simulations expliquée dans le dernier paragraphe de la section 2.5 pour éviter d’avoir des jours  $t$  tels que  $TN_t > TX_t$ . Néanmoins un intérêt important d’utiliser un modèle multivarié au lieu d’utiliser un modèle univarié sur plusieurs séries de manière indépendante est de pouvoir reconstituer les corrélations entre les résidus des séries. En particulier, nous allons regarder les corrélations croisées entre les résidus des séries de températures minimales et des séries de températures maximales. Pour rappel, cela correspond aux corrélations entre  $X_{TN_t}$  et  $X_{TX_{t+s}}$  pour  $s$  allant de -4 à 4 (on aurait pu prendre un autre entier). Les comparaisons entre le couple de séries appariées d’entraînement et tous les couples de séries appariées simulées pour ce critère sont visibles sur la figure 7. On voit que les corrélations croisées pour les couples de séries simulées sont proches de celles associées au couple d’entraînement. Cela montre que ce critère est bien pris en compte par le générateur de météo, ce qui le rend encore plus fiable dans notre objectif. Ainsi on gardera ce modèle pour la suite.

On tient à préciser que la même vérification a été faite sur les 5 autres modèles bivariés, ce qui nous a permis de valider chacun d’entre eux pour la suite du projet. Ils ont été utilisés pour générer des simulations sur lesquelles on a appliqué les modèles phénologiques comme cela est présenté dans la section 4.4.

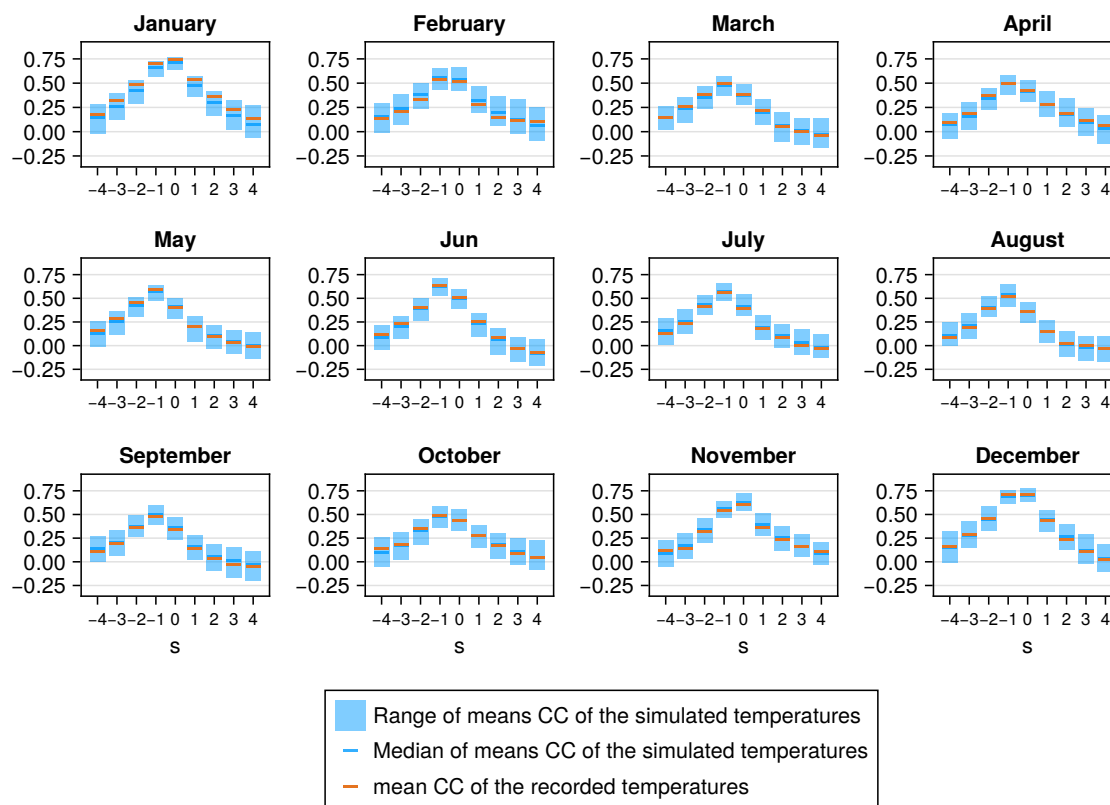


FIGURE 7 – Comparaison entre les corrélations mensuelles entre  $X_{TN_t}$  et  $X_{TX_{t+s}}$  issues du couple (TN, TX) enregistrée à Nantes-Bougenais en orange et celles issues des 1000 simulations du générateur de météo entraîné sur ce couple en bleu avec  $p = 3$  et  $q = 2$ .

### 4.3 Modèles phénologiques appliqués aux données enregistrées

Nous allons ici analyser les sorties des modèles phénologiques avec d'abord en entrée les températures enregistrées aux stations énoncées à la section 4.1.1. La figure 8 présente les résultats du modèle phénologique de vigne.

Cette figure se lit de bas en haut. Les trois courbes du bas correspondent aux dates annuelles de rupture d'endo-dormance et les trois du haut aux dates annuelles de débourrement. On remarque deux tendances :

- D'abord, on voit que les dates de ruptures d'endo-dormance sont de plus en plus tardives. En effet, une augmentation moyenne des températures fait que le seuil d'unités de refroidissement (voir définition 3.1) pour engendrer ce changement de phase est atteint plus lentement.
- Ensuite, on voit que les dates de débourrement ("Budburst") sont en revanche de plus en plus avancées. Cela s'explique toujours par l'augmentation des températures en moyenne qui fait que les unités de réchauffement sont accumulées plus rapidement pour atteindre le seuil qui déclenche le débourrement.

Nous avons aussi utilisé le modèle de pomme décrit dans la section 3.3 pour prédire les dates phénologiques annuelles en fonction des températures enregistrées aux mêmes stations. Ces résultats sont visibles sur la figure 9. On peut voir des résultats similaires aux résultats du modèle de vigne : le réchauffement climatique retarde les dates de ruptures d'endo-dormance et avance les dates de débourrement. En revanche, une différence importante est que pour le modèle de pomme les dates de débourrement sont généralement plus avancées (entre fin Mars

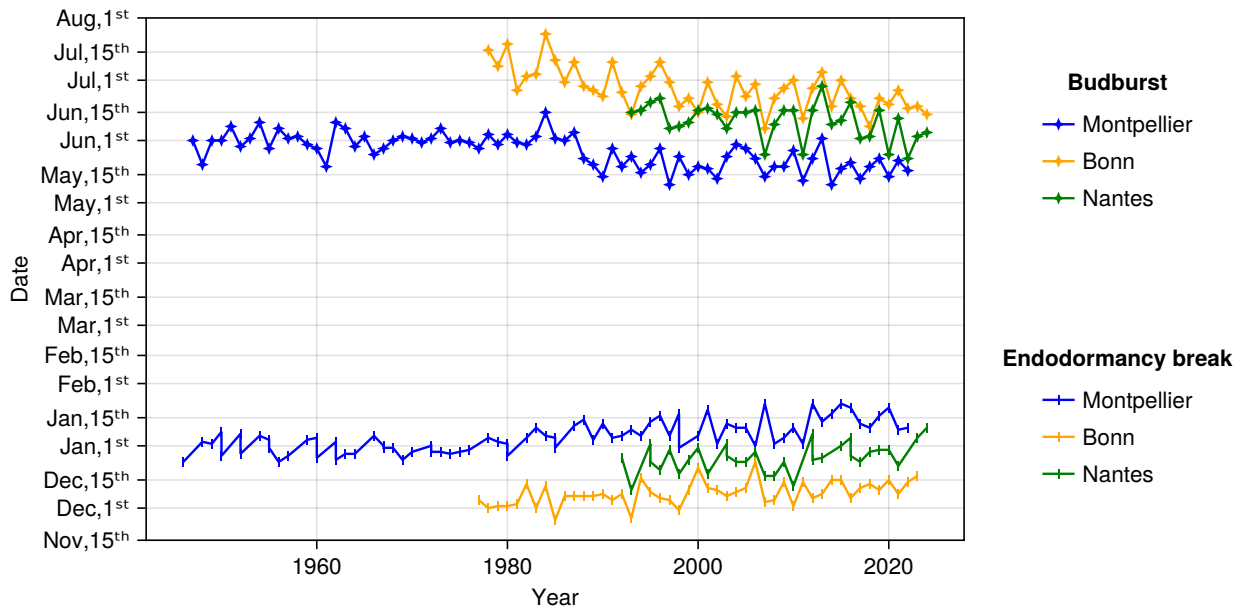


FIGURE 8 – Prédiction du modèle phénologique de vigne de la section 3.2 sur les stations de Montpellier-Aéroport, Nantes-Bougenais et Koln-Bonn.

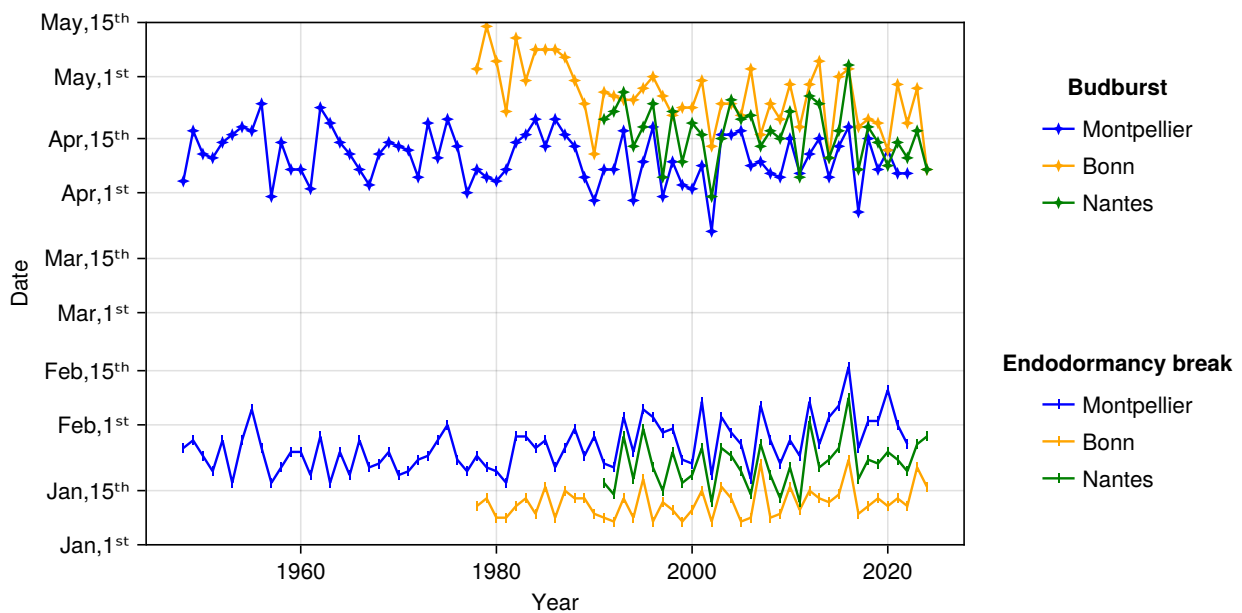


FIGURE 9 – Prédiction du modèle phénologique de pomme de la section 3.3 sur les stations de Montpellier-Aéroport, Nantes-Bougenais et Koln-Bonn.

et début Mai) que pour le modèle de vignes (entre début Mai et fin Juillet). Ainsi, comme les dates de débourrement sont de plus en plus avancées vers l'hiver, on peut questionner l'augmentation d'un risque de gel post-débourrement. Nous avons alors compté, pour les trois stations, le nombre de jours avec la température minimale  $TN_t$  inférieure à  $-2^{\circ}\text{C}$  dans les trois mois après les débournements de chaque année. On en a compté aucune pour les stations de Montpellier et Nantes, mais en revanche on en a compté plusieurs pour la station de Bonn dont les résultats sont dans la figure 10.

On voit que cet évènement devient fréquent à partir de 2014, même s'il est apparu deux fois en 1981. On peut alors émettre l'hypothèse que l'augmentation moyenne des températures serait

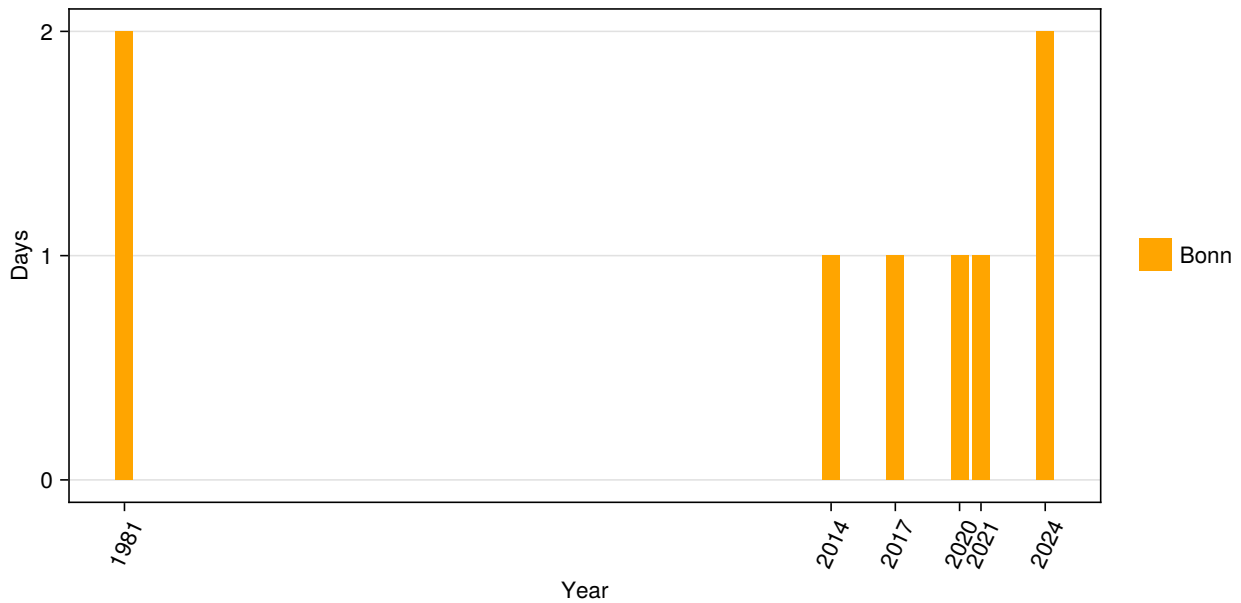


FIGURE 10 – Nombre de jours avec la température minimale journalière inférieure à  $-2^{\circ}\text{C}$  après la date de débournement prédite par le modèle phénologique de pomme 3.3 pour chaque année de la série de températures enregistrées à Koln-Bonn.

à l'origine de la multiplication du gel post-débournement dans des zones de climats continentaux comme à Bonn. En effet, on peut qualifier de tels évènements d'extrêmes et il est attendu qu'ils soient de plus en plus fréquents à cause du réchauffement climatique (McSweeney, 2021).

#### 4.4 Modèles phénologiques appliqués aux données simulées avec générateurs de météo entraînés sur températures enregistrées

Nous allons maintenant appliquer les modèles phénologiques sur un grand nombre de simulations des générateurs de météo afin d'avoir une distribution empirique des dates phénologiques. Ici nous analyserons les résultats sur les simulations des générateurs entraînés sur la série enregistrée à Koln-Bonn mais pour les autres stations les résultats seront disponibles dans la section C de l'annexe. Nous utiliserons les générateurs dont les paramètres sont présentés dans la table 6. D'abord, pour le modèle phénologique de vigne, nous avons utilisé le générateur de météo entraîné sur le couple (TN, TX) enregistré à la station de Koln-Bonn avec les paramètres  $p = 5$  et  $q = 3$ . On a généré 1000 simulations avec ce générateur sur lesquelles on a utilisé le modèle de vigne. Les résultats sont dans la figure 11.

Cette fois-ci, au lieu d'obtenir seulement les résultats d'une série de température, on a toute une distribution empirique des dates phénologiques pour chaque année. De cette distribution on a affiché l'intervalle interquartile avec les bandes foncées et l'étendue des valeurs avec les bandes claires. D'abord, on voit que le générateur de météo reste fiable car les dates phénologiques prédites sur les températures simulées restent relativement centrées autour des dates phénologiques prédites sur les températures enregistrées. On voit tout de même que les dates phénologiques des températures enregistrées sortent souvent de l'intervalle interquartile, ce qui montre une variabilité importante. Mais ce qui est intéressant c'est que l'on peut observer un bien plus grand échantillon de scénarios phénologiques pour la vigne grâce aux simulations, ce qui nous permet de voir quels évènements phénologiques extrêmes aurait pu arriver. Pour une année fixée, on peut mieux représenter la distribution des dates phénologiques des simulations avec l'histogramme de la figure 12.

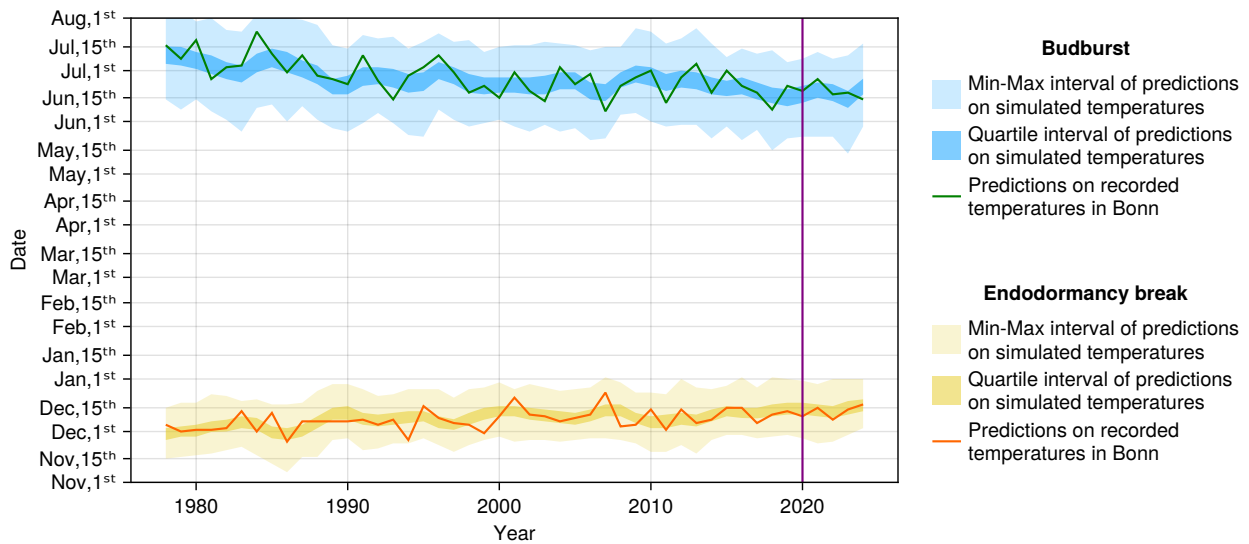


FIGURE 11 – Prédications du modèle phénologique de vigne de la section 3.2 sur le couple (TN, TX) enregistré à la station de Koln-Bonn et sur 1000 simulations du générateur météo entraîné sur ce couple avec  $p = 5$  et  $q = 3$ . La ligne verticale violette est un repère pour comprendre la figure 12.

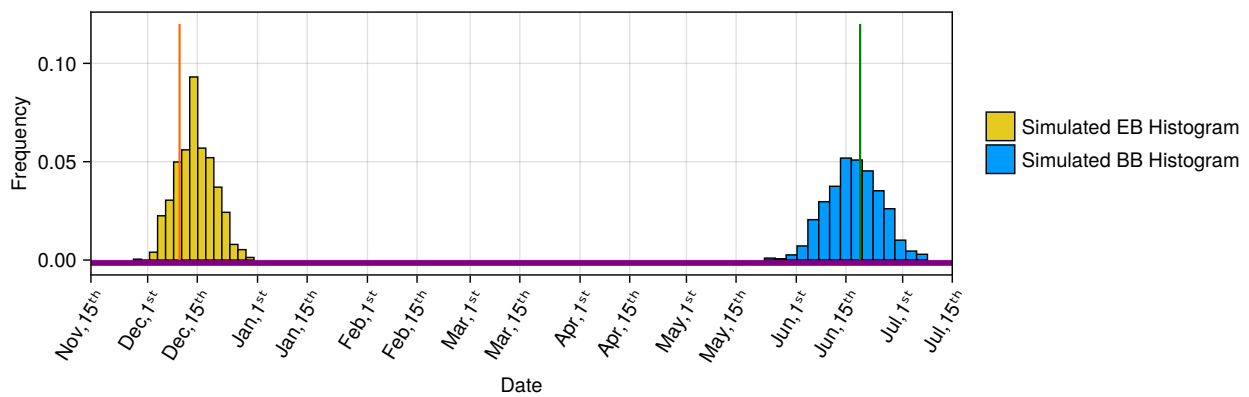


FIGURE 12 – Histogramme des prédictions pour l’année 2020 du modèle phénologique de vigne de la section 3.2 sur 1000 simulations du générateur météo entraîné sur le couple (TN, TX) enregistré à la station de Koln-Bonn avec  $p = 5$  et  $q = 3$ .

Ces histogrammes peuvent nous laisser supposer que, pour chaque année, les dates de ruptures d’endo-dormance et de débourrement suivent une loi ressemblant à la loi normale. Avoir des statistiques détaillées sur ces dates peuvent être utiles aux agronomes.

Nous avons fait un travail similaire pour le modèle phénologique de pommes : pour cela on a généré 1000 simulations avec le générateur de météo entraîné sur le couple (TN, TG) enregistré à la station de Koln-Bonn avec les paramètres  $p = 5$  et  $q = 2$ . Les résultats du modèle de pomme de la section 3.3 sur ces simulations sont visibles sur la figure 13. De même, on voit que le générateur de météo fonctionne aussi bien pour le modèle de pomme car les échantillons des dates phénologiques issus des simulations sont centrés autour des dates phénologiques associées aux températures enregistrées. Grâce aux générateurs de météo, on a maintenant aussi une distribution des dates phénologiques des pommes, ce qui a la même utilité que pour les vignes. On voit aussi, via la figure 14, que les distributions annuelles des dates phénologiques de pommes ont une forme similaire à celles des vignes.

On voit que, pour les modèles de pomme et de vigne appliqués aux simulations, les dates

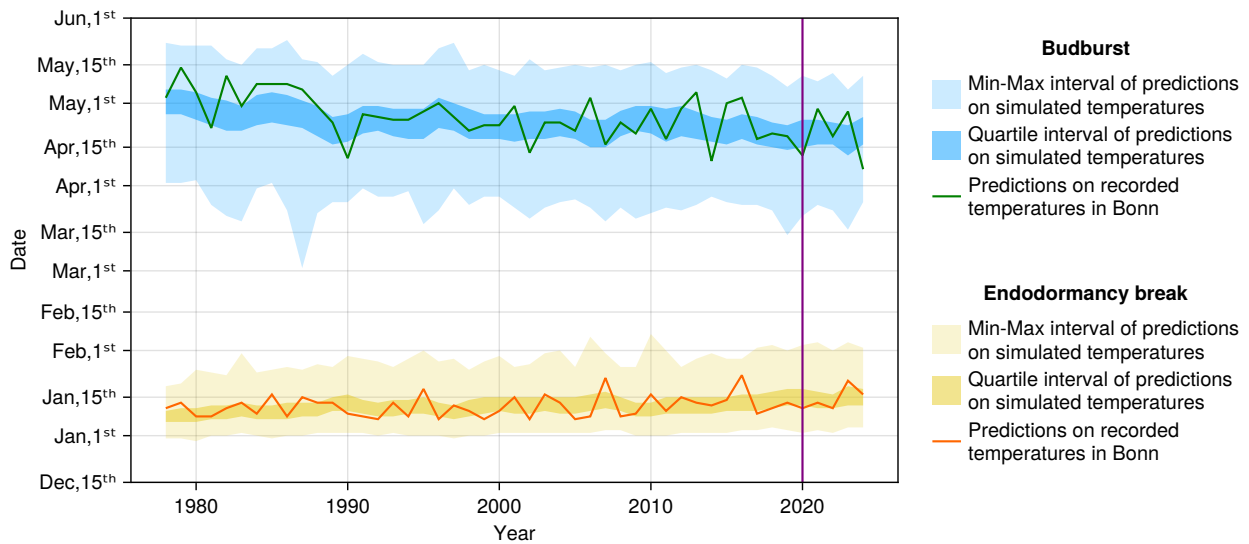


FIGURE 13 – Prédications du modèle phénologique de pomme de la section 3.3 sur le couple (TN, TG) enregistré à la station de Koln-Bonn et sur 1000 simulations du générateur météo entraîné sur ce couple avec  $p = 5$  et  $q = 2$ . La ligne verticale violette est un repère pour comprendre la figure 12.

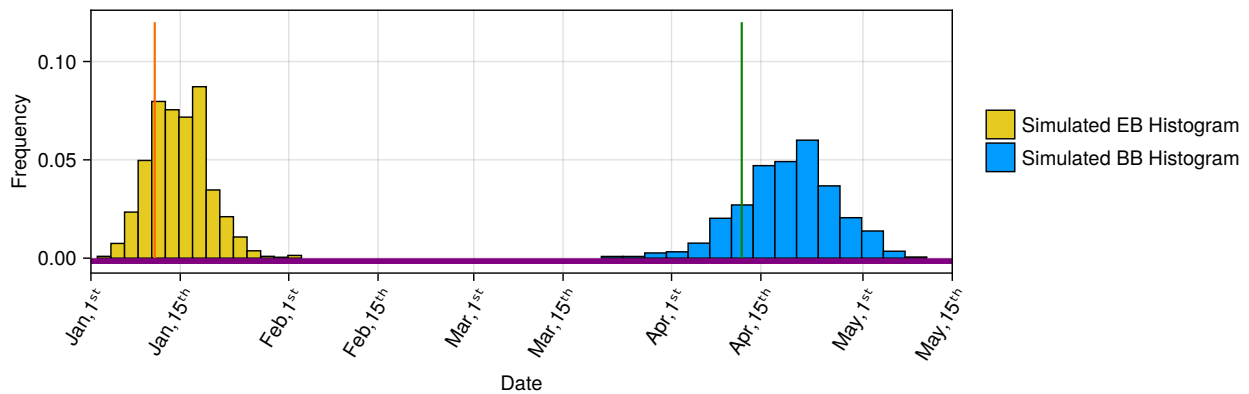


FIGURE 14 – Histogramme des prédictions pour l’année 2020 du modèle phénologique de pomme de la section 3.3 sur 1000 simulations du générateur météo entraîné sur le couple (TN, TG) enregistré à la station de Koln-Bonn avec  $p = 5$  et  $q = 2$ .

de ruptures d’endo-dormance ont tendance à reculer et les dates de débourrement à s’avancer, comme cela a été le cas lorsqu’ils sont appliqués aux températures enregistrées. Cela pourrait indiquer que le réchauffement climatique a eu un impact pas seulement sur la phénologie expliquée par les séries de températures enregistrées, mais sur la phénologie issue de la loi de ces séries telle que on l’a modélisée avec le générateur de météo. Ces phénomènes sont aussi visibles sur les résultats pour les villes de Montpellier et Nantes dans l’annexe C. Grâce aux simulations des séries de températures minimales journalières, on peut voir la distribution des risques de gel chez les pommiers avec la figure 15.

Ces résultats contredisent l’hypothèse que nous avons émises lors de l’interprétation de la figure 10. En effet, on pensait qu’avec les années le réchauffement climatique allait augmenter la fréquence de gel post-débourrement avec des débournements avancés. On pourrait expliquer cette baisse moyenne de ce risque par le fait que l’augmentation moyenne des températures pour les jours après débourrement est suffisamment élevée pour empêcher d’atteindre des seuils critiques comme celui de  $-2^{\circ}\text{C}$ , même si le débourrement est de plus en plus proche de l’hiver avec les années. Néanmoins, on a remarqué que, en fonction des années, la fréquence de ces gels



a une allure très ressemblante à celles des saisonnalités multiplicatives présentées dans la figure 16.

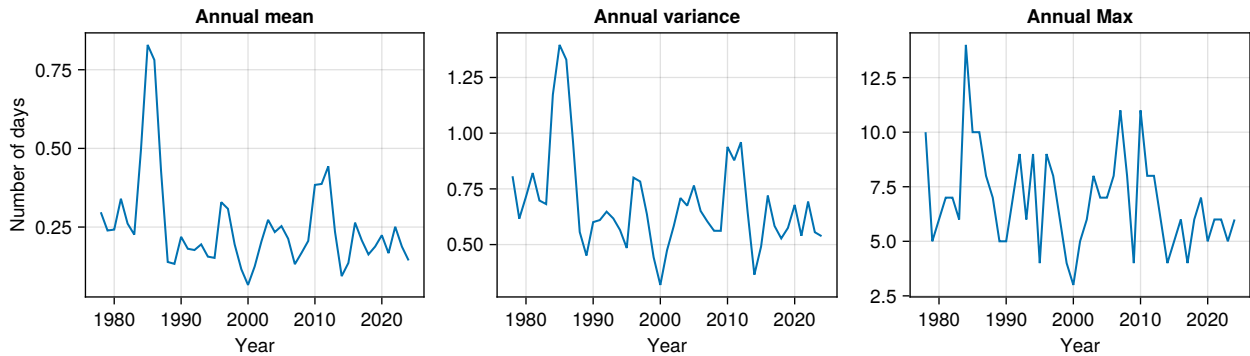


FIGURE 15 – Moyenne, variance et maximum (de gauche à droite) des nombres de jours avec  $TN_t$  inférieure à  $-2^\circ\text{C}$  après débourrement selon les prédictions du modèle phénologique de pomme de la section 3.3 sur les 1000 simulations du générateur météo entraîné sur le couple  $(TN, TG)$  enregistré à la station de Koln-Bonn avec  $p = 5$  et  $q = 2$ .

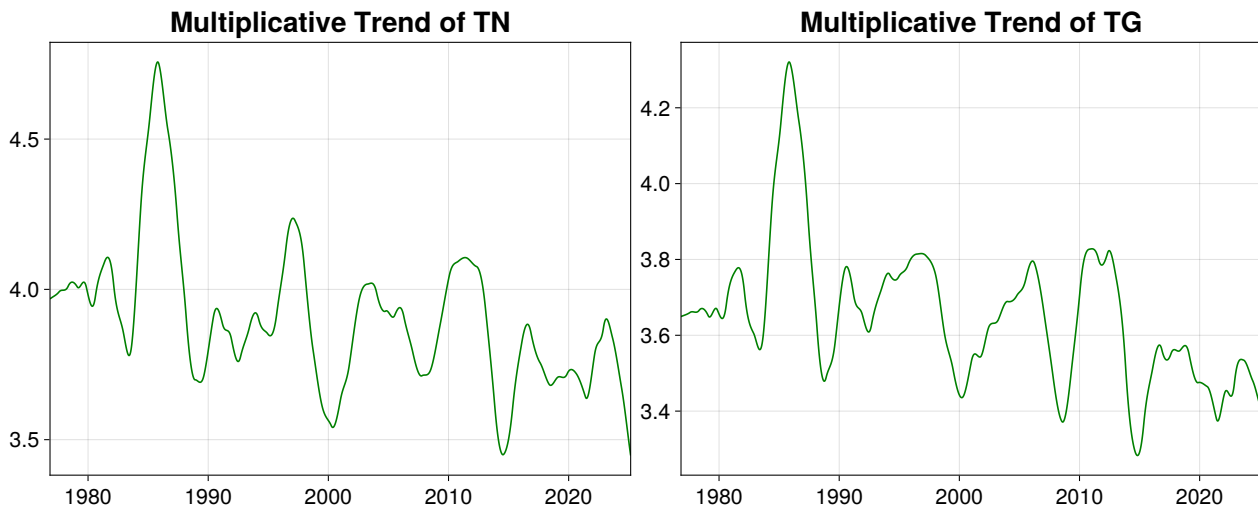


FIGURE 16 – Tendances multiplicatives estimées sur le couple  $(TN, TG)$  enregistré à la station de Koln-Bonn avec  $q = 2$ .

Cela indique que le risque de gel n'est pas principalement influencé par la température mais par la variance des températures.

#### 4.5 Modèles phénologiques appliqués aux données simulées avec générateurs de météo entraînés sur projections climatiques

Dans cette dernière partie, nous allons maintenant utiliser les modèles phénologiques sur les données des projections climatiques futures 4.1.2 et sur les générateurs de météo entraînés sur ces projections. Pour rappel, il s'agit de projections des températures à des coordonnées proches de la station de Nantes-Bougenais. Nous pourrons ainsi avoir une distribution des dates phénologiques futures selon les scénarios rep4.5 et rep8.5 du GIEC.

### 4.5.1 Scénario rcp4.5

D'abord, nous avons dans les figures 17 et 18, respectivement pour les pommes et les vignes, les prédictions des modèles phénologiques sur les températures enregistrées et les projections de températures selon le scénario rcp4.5 à Nantes.

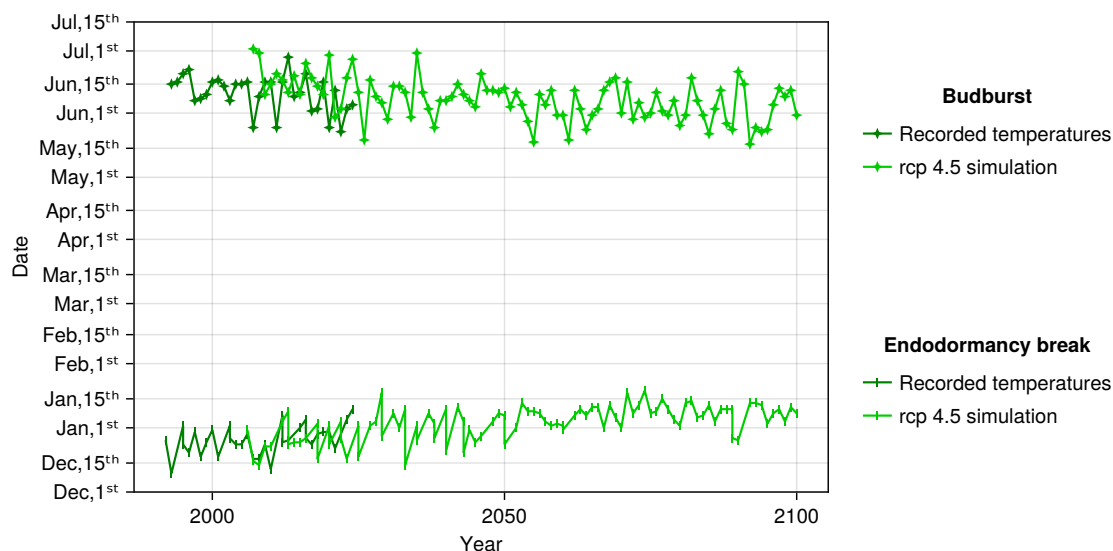


FIGURE 17 – Prédiction du modèle phénologique de vigne de la section 3.2 sur le couple (TN, TX) enregistré à la station de Nantes-Bougenais et les projections de températures aux coordonnées 7.173°,-1.621930° selon le scénario rcp4.5.

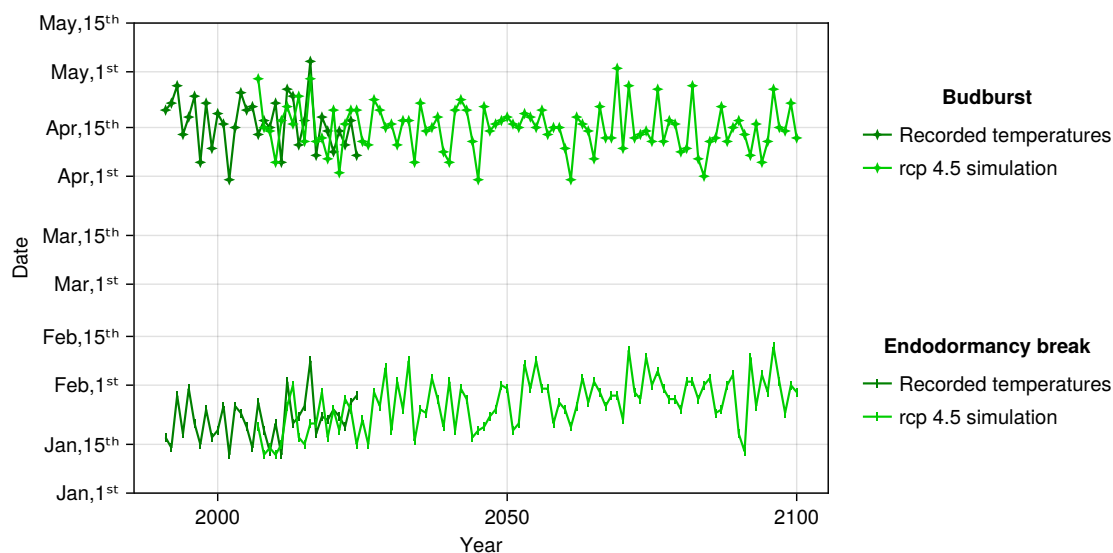


FIGURE 18 – Prédiction du modèle phénologique de pomme de la section 3.3 sur les températures moyennes journalières enregistrées à la station de Nantes-Bougenais et les projections de températures aux coordonnées 7.173°,- 1.621930° selon le scénario rcp4.5.

On voit toujours le même phénomène que sur les figures précédentes : les dates de rupture d'endo-dormance ont tendance à être de plus en plus tardives et les dates de débourrement de plus en plus tôt, même si pour le scénario rcp4.5 ces évolutions sont assez légères et finissent souvent par se stabiliser à partir de 2050.

Sur les figures 19 et 20, on a affiché les prédictions des modèles phénologiques sur les générateurs de météo entraînés sur les températures enregistrées et sur les projections climatiques.

On voit un phénomène similaire à ce qui est visible dans les figures 18 et 17 : des ruptures d'endo-dormance de plus en plus tardives et des débourrements de plus en plus en tôt mais avec une stabilisation à partir de 2050. Sur la figure 20 correspondant au modèle de pomme, on voit que les prédictions des fins d'endo-dormance sur les simulations du générateur entraîné sur les données enregistrées montent plus vite que celles du générateur entraîné sur les projections,

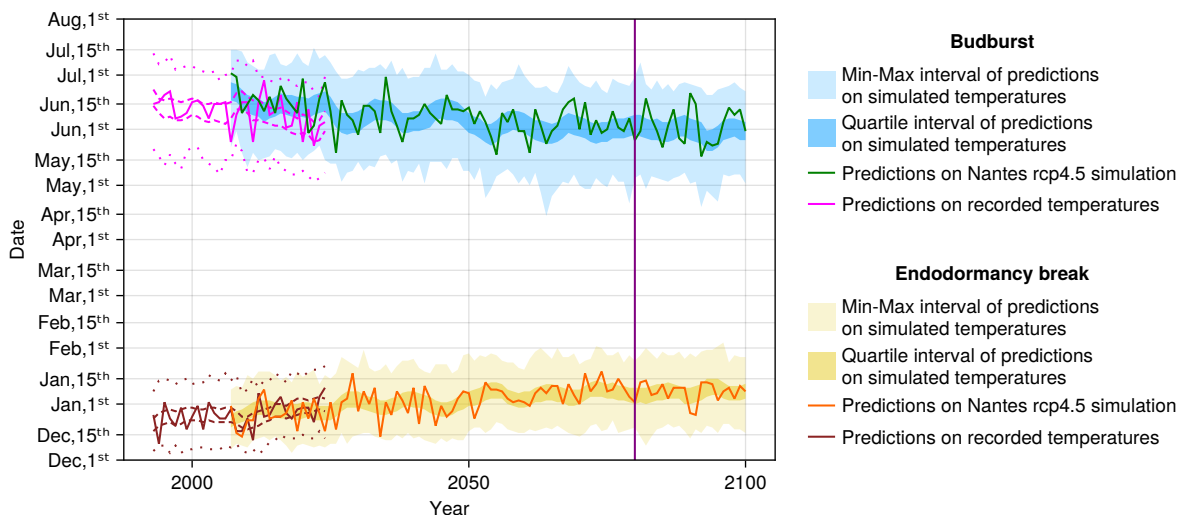


FIGURE 19 – Prédications du modèle phénologique de vigne de la section 3.2 sur :

- Le couple (TN, TX) enregistré à la station de Nantes-Bougenais et sur 1000 simulations du générateur météo entraîné sur ce couple avec  $p = 3$  et  $q = 2$  indiqué dans la légende par "Predictions on recorded temperatures".
- Le couple (TN, TX) de la projection climatique aux coordonnées  $7.173^\circ, -1.621930^\circ$  selon le scénario rcp4.5 et sur 1000 simulations du générateur météo entraîné sur ce couple avec  $p = 3$  et  $q = 2$ .

La ligne verticale violette est un repère pour comprendre la figure 35 dans l'annexe D.

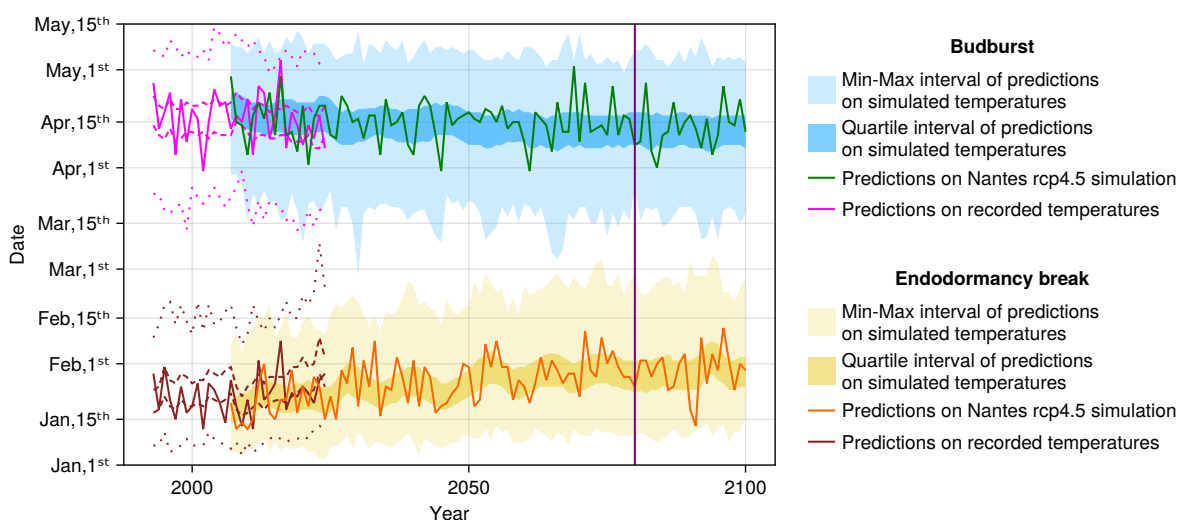


FIGURE 20 – Prédications du modèle phénologique de pomme de la section 3.3 sur :

- Le couple (TN, TG) enregistré à la station de Nantes-Bougenais et sur 1000 simulations du générateur météo entraîné sur ce couple avec  $p = 3$  et  $q = 2$  indiqué dans la légende par "Predictions on recorded temperatures".
- Le couple (TN, TG) de la projection climatique aux coordonnées  $7.173^\circ, -1.621930^\circ$  selon le scénario rcp4.5 et sur 1000 simulations du générateur météo entraîné sur ce couple avec  $p = 3$  et  $q = 2$ .

La ligne verticale violette est un repère pour comprendre la figure 36 dans l'annexe D.

comme si le scénario rcp4.5, commençant en 2006, était plus optimiste que les vrais évènements réalisés entre 2006 et aujourd'hui. Nous pouvons voir sur les figures 35 et 36 de l'annexe D que les distributions annuelles des dates phénologiques prédites sur les simulations ont toujours une forme proche de la loi normale, comme dans le cas où les générateurs sont entraînés sur des données enregistrées. Avoir ces distributions empiriques permet d'avoir l'ensemble des scénarios possibles pour la phénologie des plantes, au lieu de se limiter à une seule série de prédictions pour la projection climatique (correspondant aux courbes orange et verte).

#### 4.5.2 Scénario rcp8.5

Nous allons faire la même étude en considérant cette fois-ci la projection du scénario plus pessimiste rcp8.5. Comme dans la section précédente, on a d'abord appliqué les modèles phénologiques sur les projections de températures. Les résultats sont visibles dans les figures 21 et 22.

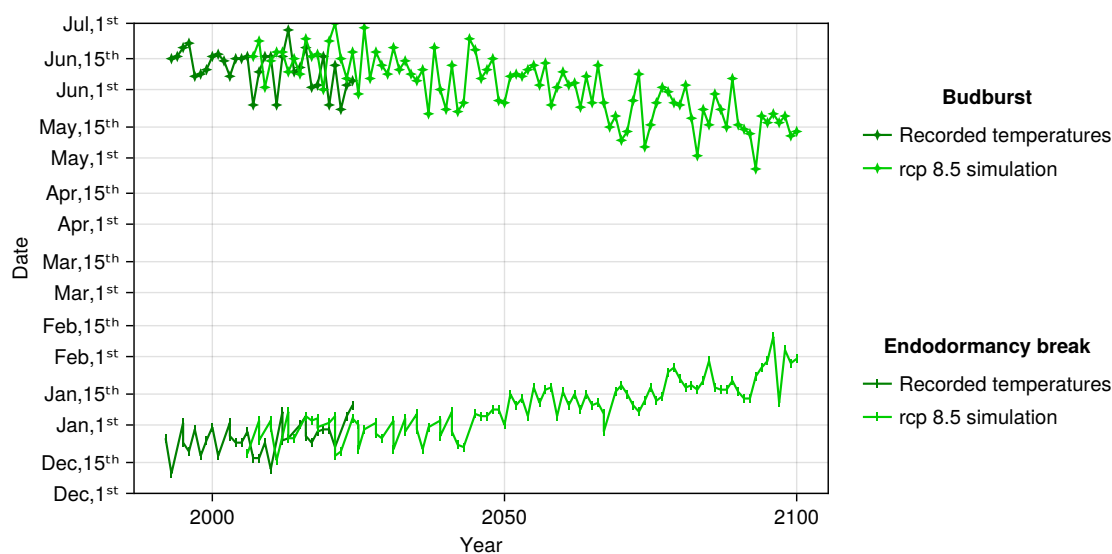


FIGURE 21 – Prédiction du modèle phénologique de vigne de la section 3.2 sur le couple (TN, TX) enregistré à la station de Nantes-Bougenais et les projections de températures aux coordonnées  $7.173^{\circ}$ ,  $-1.621930^{\circ}$  selon le scénario rcp8.5.

Pour le modèle de vigne, dans la figure 21, on remarque que le phénomène d'avancée des débourrements et de recul des ruptures d'endo-dormance est encore plus marqué que pour le scénario rcp4.5 (voir figure 17). En particulier, on ne voit plus de stabilisation à partir de 2050 : après cette année, les dates phénologiques continuent toujours d'évoluer en moyenne. En revanche, pour le modèle de pomme, on peut voir une différence sur la figure 22. Les dates de ruptures d'endo-dormance ont tendance à se retarder de manière intense jusqu'à la fin de la série, tandis que les dates de débourrement décroissent légèrement jusqu'en 2050 mais restent stables ensuite comme dans la figure 18 associée au scénario rcp4.5. L'hypothèse que nous avons développée pour expliquer ce résultat étonnant est la présence d'un phénomène de compensation entre :

- Le retardement des dates de fin d'endo-dormance tellement important que cela décale significativement la période d'accumulation des unités de réchauffement, ce qui implique en théorie des dates de débourrement plus tardives.
- L'augmentation moyenne des températures qui raccourcit cette même période et qui par conséquent avance les dates de débourrement.

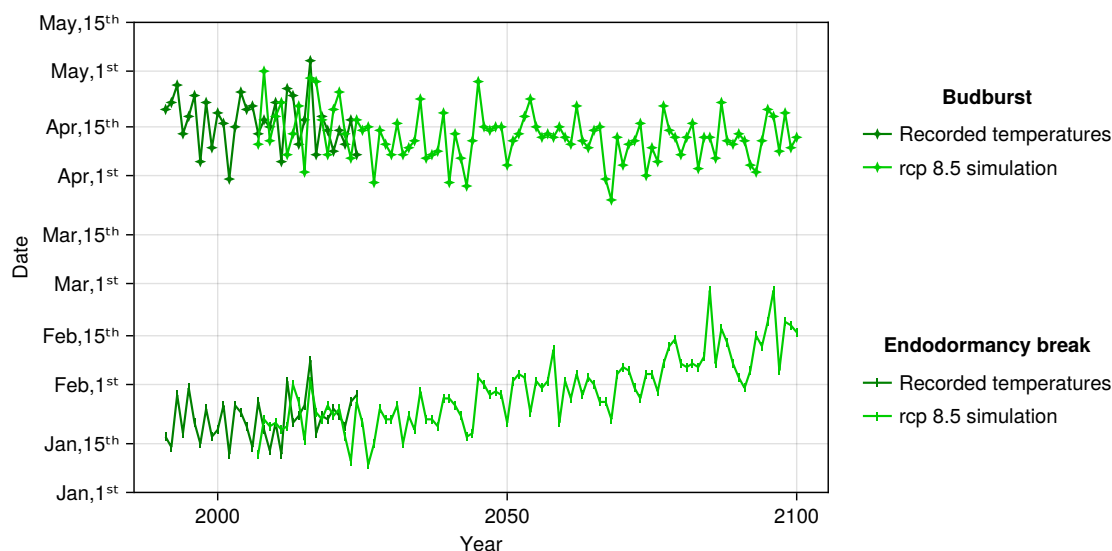


FIGURE 22 – Prédiction du modèle phénologique de pomme de la section 3.3 sur les températures moyennes journalières enregistrées de la station de Nantes-Bougenais et les projections de températures aux coordonnées  $7.173^\circ, -1.621930^\circ$  selon le scénario rcp8.5.

Même si on peut penser que la stabilité des dates de débournement ne présente pas de problème, il peut être intéressant d'étudier les conséquences d'une période d'éco-dormance trop courte (pour rappel il s'agit de la période entre les deux dates phénologiques étudiées, voir section 3.1). Cette recherche n'a pas pu être faite pendant le stage, mais on peut supposer que cela a des conséquences néfastes sur la santé et la reproduction de la plante. Pour connaître la distribution des scénarios phénologiques possibles dans l'avenir selon le scénario climatique rcp8.5, on a entraîné nos générateurs de météo sur les projections climatiques selon ce scénario. Les résultats sont visibles dans les figures 23 et 24. Sur les simulations du générateur,

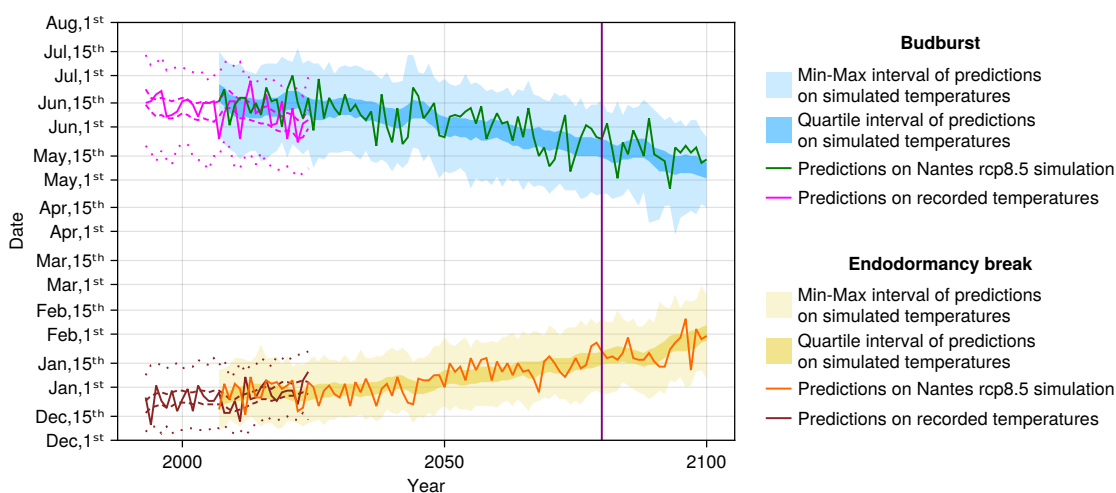


FIGURE 23 – Prédications du modèle phénologique de vigne de la section 3.2 sur :

- Le couple (TN, TX) enregistré à la station de Nantes-Bougenais et sur 1000 simulations du générateur météo entraîné sur ce couple avec  $p = 3$  et  $q = 2$  indiqué dans la légende par "Predictions on recorded temperatures".
- Le couple (TN, TX) de la projection climatique aux coordonnées  $7.173^\circ, -1.621930^\circ$  selon le scénario rcp8.5 et sur 1000 simulations du générateur météo entraîné sur ce couple avec  $p = 3$  et  $q = 2$ .

La ligne verticale violette est un repère pour comprendre la figure 37 dans l'annexe D.

on remarque les mêmes phénomènes d'évolution des dates phénologiques que sur les projections climatiques. Ces simulations nous permettent néanmoins d'avoir l'ensemble des scénarios phénologiques possibles selon le scénario rcp8.5, en particulier les plus catastrophiques pour le modèle de pomme visibles sur la figure 24. Il s'agit de la seule figure où les supports des ruptures d'endo-dormance et des débournements s'intersectent. On explique cela par l'existence de simulations où les ruptures d'endo-dormance ont lieu après les dates de débournement d'autres simulations. Cela est dû au raccourcissement de la période d'éco-dormance mais aussi à une augmentation visible des étendues des dates phénologiques à partir des années 2090. On tient à noter que, pour une année, la forme des distributions semble peu changer pour ces modèles comme cela est illustré pour l'exemple de 2080 sur les figures 37 et 38 de l'annexe D.

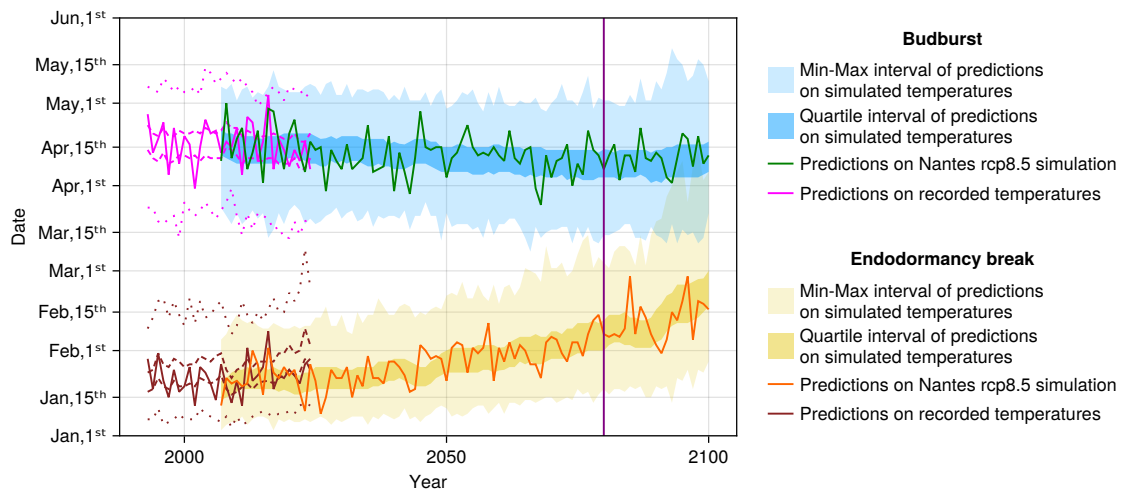


FIGURE 24 – Prédications du modèle phénologique de pomme de la section 3.3 sur :

- Le couple (TN, TG) enregistré à la station de Nantes-Bougenais et sur 1000 simulations du générateur météo entraîné sur ce couple avec  $p = 3$  et  $q = 2$  indiqué dans la légende par "Predictions on recorded temperatures".

- Le couple (TN, TG) de la projection climatique au coordonnées  $7.173^\circ, -1.621930^\circ$  selon le scénario rcp8.5 et sur 1000 simulations du générateur météo entraîné sur ce couple avec  $p = 3$  et  $q = 2$ .

La ligne verticale violette est un repère pour comprendre la figure 38 dans l'annexe D.

## Conclusion

En conclusion, on a montré que les générateurs stochastiques de météo sont très efficaces pour obtenir des échantillons de taille importantes de séries de températures réalistes sur la période d'entraînement. Lorsque les générateurs sont entraînés sur une série de températures donnée, ils permettent alors d'obtenir la distribution empirique de la loi de cette série. Ainsi, ce sont des outils puissants pour connaître l'évolution des lois des séries de températures enregistrées et des projections climatiques futures, au lieu de se limiter à trop peu de réalisations qui ne sont pas suffisamment représentatives de leur loi. En particulier, les générateurs de météo simulent des échantillons avec une variabilité pouvant donner des scénarios extrêmes.

Couplés avec les modèles phénologiques, ces générateurs nous ont permis d'obtenir les distributions empiriques des phénologies des pommes et des vignes sur les sites géographiques de Montpellier, Nantes et Bonn. Nous pouvons étudier ces distributions sur les périodes des séries de températures à notre disposition, c'est-à-dire sur les 30 à 70 dernières années (les dates précises sont disponibles à la section 4.1.1). Cela nous a permis de mettre en valeur un retardement en moyenne des dates de rupture d'endo-dormance et une avancée en moyenne des dates de débourrement sur ces périodes. Les changements de ces dates phénologiques sont les résultats directs de l'augmentation des températures moyennes due au réchauffement climatique. Cet effet est expliqué au début de la section 4.3. L'utilisation des générateurs de météo nous a permis d'observer, pour le site de Bonn qui a un climat continental, une corrélation entre les risques de gel post-débourrement et la variance tendancielle des températures sur les figures 15 et 16. Ce résultat contredit l'hypothèse que l'augmentation moyenne des températures augmente systématiquement le risque de gel.

Enfin, on a entraîné les générateurs de météo sur deux projections climatiques futures des températures à Nantes des scénarios rcp4.5 et rcp8.5, pour la période de 2006 à 2100. Utiliser ces générateurs nous permet de combler le manque de variabilité des projections climatiques et donc d'avoir plus d'évènements extrêmes dans l'échantillon des simulations utilisées. De même, nous les avons couplés avec les modèles phénologiques pour obtenir une distribution empirique des dates phénologiques futures. Pour le scénario rcp4.5, on a une stabilisation des phénologies à partir de 2050. En revanche, pour le scénario rcp8.5, on a des résultats qui semblent alarmants. En effet, on peut voir sur les figures 23 et 24 que les ruptures d'endo-dormance deviennent extrêmement tardives et, spécialement pour les vignes, que les débournements deviennent extrêmement avancés. Par exemple, sur la droite de la figure 24, le générateur de météo nous donne des simulations où, vers 2100, la fin de la période d'endo-dormance a lieu pendant la période correspondant normalement au débourrement.

Comme cela a été montré dans divers travaux (énoncés à la section 1.2), avoir des dates phénologiques décalées peuvent engendrer des risques importants pour la santé et la reproduction des pommiers et des vignes. Il y a par exemple la perturbation des interactions entre ces plantes et les espèces animales telles que les pollinisateurs et les ravageurs. Le couplage des générateurs stochastiques de météo avec les modèles phénologiques est donc un outil puissant de projection de tous les scénarios possibles de risques chez les pommiers et les vignes. Ainsi, il nous permet alors de mieux connaître les effets potentiels du réchauffement climatique sur la production de ces fruits. Cela en fait un dispositif efficace pour l'aide à l'adaptation agricole et pour la sensibilisation aux enjeux environnementaux du XXI<sup>ème</sup> siècle.

# Annexes

## Annexe A Tables des BIC et AIC

### A.1 Tables des meilleurs modèles selon le critère BIC

Montpellier :

TN				TG				TX			
p	q	AIC	BIC	p	q	AIC	BIC	p	q	AIC	BIC
3	2	133433.94	133912.11	3	2	108439.01	108916.3	1	2	132030.13	132311.38
2	2	133544.2	133923.44	3	3	108419.05	108929.26	1	3	131998.66	132313.0
1	2	133653.09	133933.39	2	2	108574.05	108952.59	2	2	131945.39	132325.91
3	3	133424.23	133935.38	2	3	108548.83	108960.29	2	3	131918.97	132332.57
2	3	133531.71	133943.92	3	4	108422.67	108965.79	1	4	131995.31	132342.73

TABLE 7 – Tables des 5 meilleurs couples d’hyperparamètres selon le critère BIC pour les modèles entraînés sur les températures enregistrées à la station de Montpellier-Aéroport dans l’ordre décroissant.

Nantes :

TN				TG				TX			
p	q	AIC	BIC	p	q	AIC	BIC	p	q	AIC	BIC
1	1	58669.67	58891.66	3	2	51754.4	52186.85	1	2	74610.99	74872.45
1	2	58649.24	58900.83	2	2	51844.29	52187.26	1	3	74612.64	74904.86
1	3	58654.23	58935.42	3	1	51789.56	52192.18	1	1	74708.07	74938.77
1	4	58664.45	58975.23	2	1	51888.75	52201.9	1	4	74622.17	74945.16
2	1	58666.64	58977.43	3	3	51758.29	52220.56	2	2	74597.5	74951.25

TABLE 8 – Tables des 5 meilleurs couples d’hyperparamètres selon le critère BIC pour les modèles entraînés sur les températures enregistrées à la station de Nantes-Bouguenais dans l’ordre décroissant.

Bonn :

TN				TG				TX			
p	q	AIC	BIC	p	q	AIC	BIC	p	q	AIC	BIC
1	2	86303.71	86568.18	3	1	77871.76	78291.81	1	2	120677.45	120953.28
1	1	86335.1	86568.45	2	1	77984.75	78311.45	2	2	120582.18	120955.36
2	2	86223.58	86581.4	3	2	77877.54	78328.7	1	1	120727.34	120970.72
2	1	86260.78	86587.48	2	2	77986.06	78343.87	2	1	120638.45	120979.18
1	3	86314.85	86610.44	3	3	77874.91	78357.18	1	3	120681.59	120989.88

TABLE 9 – Tables des 5 meilleurs couples d’hyperparamètres selon le critère BIC pour les modèles entraînés sur les températures enregistrées à la station de Koln-Bonn dans l’ordre décroissant.



## A.2 Tables des meilleurs modèles selon le critère AIC

Montpellier :

TN				TG				TX			
p	q	AIC	BIC	p	q	AIC	BIC	p	q	AIC	BIC
6	5	133383.62	134257.53	7	3	108401.77	109306.98	7	4	131851.2	132794.22
5	5	133384.31	134159.29	8	3	108403.46	109407.42	7	3	131852.09	132762.02
4	5	133385.57	134061.61	7	4	108405.39	109343.52	6	4	131861.1	132704.85
8	5	133392.81	134464.58	6	3	108406.76	109213.22	6	3	131861.9	132672.56
7	5	133392.98	134365.81	5	3	108407.07	109114.78	7	5	131862.12	132838.22

TABLE 10 – Tables des 5 meilleurs couples d’hyperparamètres selon le critère AIC pour les modèles entraînés sur les températures enregistrées à la station de Montpellier-Aéroport dans l’ordre décroissant.

Nantes :

TN				TG				TX			
p	q	AIC	BIC	p	q	AIC	BIC	p	q	AIC	BIC
3	2	58643.02	59072.2	3	2	51754.4	52186.85	3	2	74580.46	75026.49
4	2	58643.39	59161.37	4	2	51756.07	52277.99	3	3	74581.66	75058.45
6	2	58644.9	59340.47	3	5	51757.33	52279.24	6	2	74583.28	75306.16
5	2	58647.18	59253.96	3	3	51758.29	52220.56	6	3	74584.25	75337.88
3	3	58648.04	59106.82	4	5	51759.32	52370.71	3	5	74589.64	75127.95

TABLE 11 – Tables des 5 meilleurs couples d’hyperparamètres selon le critère AIC pour les modèles entraînés sur les températures enregistrées à la station de Nantes-Bouguenais dans l’ordre décroissant.

Bonn :

TN				TG				TX			
p	q	AIC	BIC	p	q	AIC	BIC	p	q	AIC	BIC
5	2	86177.19	86815.03	3	1	77871.76	78291.81	4	2	120534.42	121102.31
3	2	86177.69	86628.85	3	3	77874.91	78357.18	5	2	120534.56	121199.8
6	2	86181.15	86912.34	3	2	77877.54	78328.7	8	2	120537.11	121494.41
4	2	86183.04	86727.54	3	4	77879.51	78392.9	7	2	120537.33	121397.27
5	3	86188.66	86857.62	4	1	77880.47	78393.86	4	3	120539.02	121139.35

TABLE 12 – Tables des 5 meilleurs couples d’hyperparamètres selon le critère AIC pour les modèles entraînés sur les températures enregistrées à la station de Koln-Bonn dans l’ordre décroissant.

# Annexe B Résultats d'estimations du générateur de séries appariées de type TN et TX sur les températures enregistrées à Nantes

## B.1 Résultats pour la série de températures minimales journalières (TN)

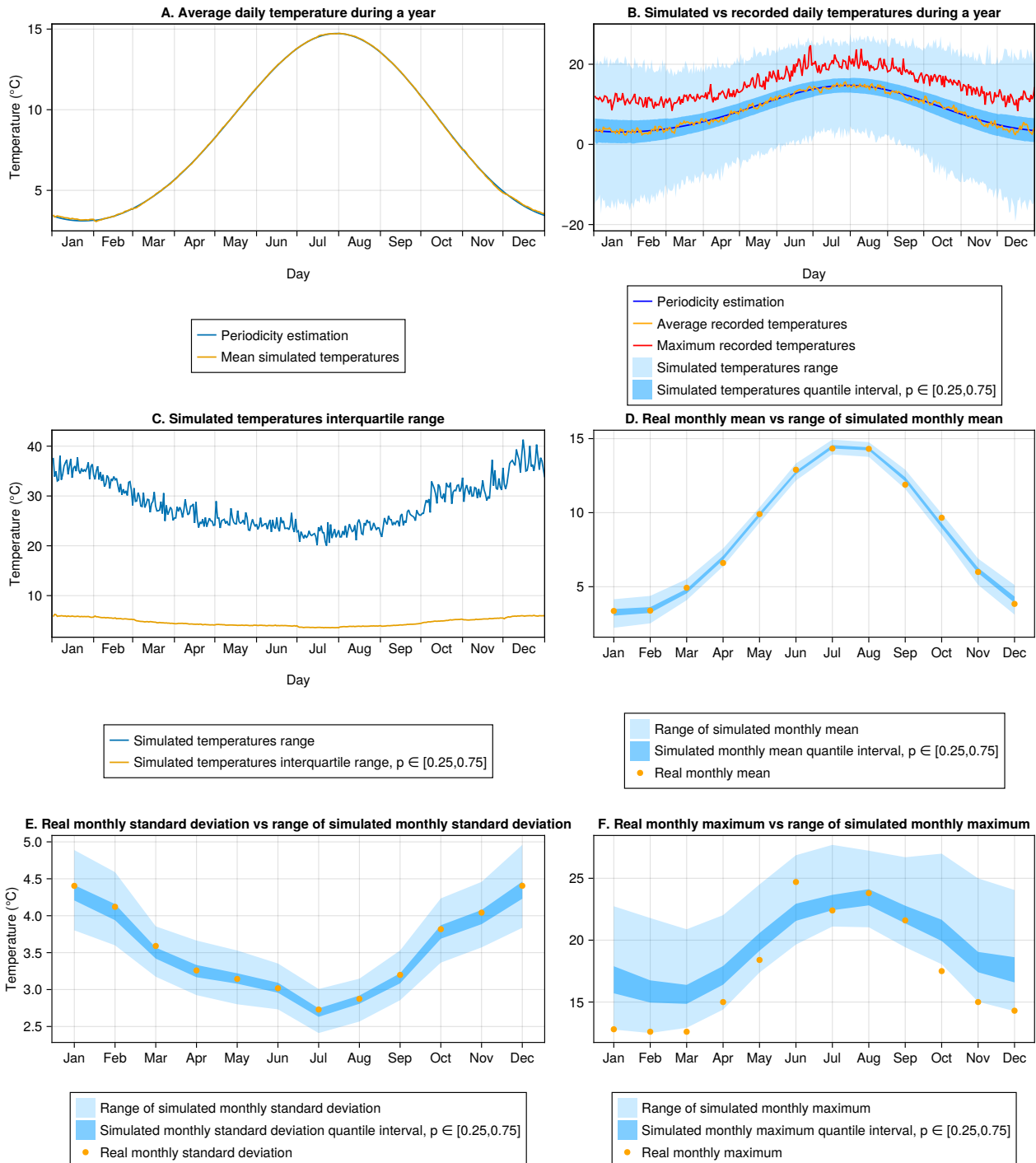


FIGURE 25 – Différentes comparaisons entre la série de températures minimales (TN) journalières enregistrées à la station Nantes-Bougenais et 1000 simulations du générateur de météo entraîné sur les séries appariées de TN et TX avec  $p = 3$  et  $q = 2$ .

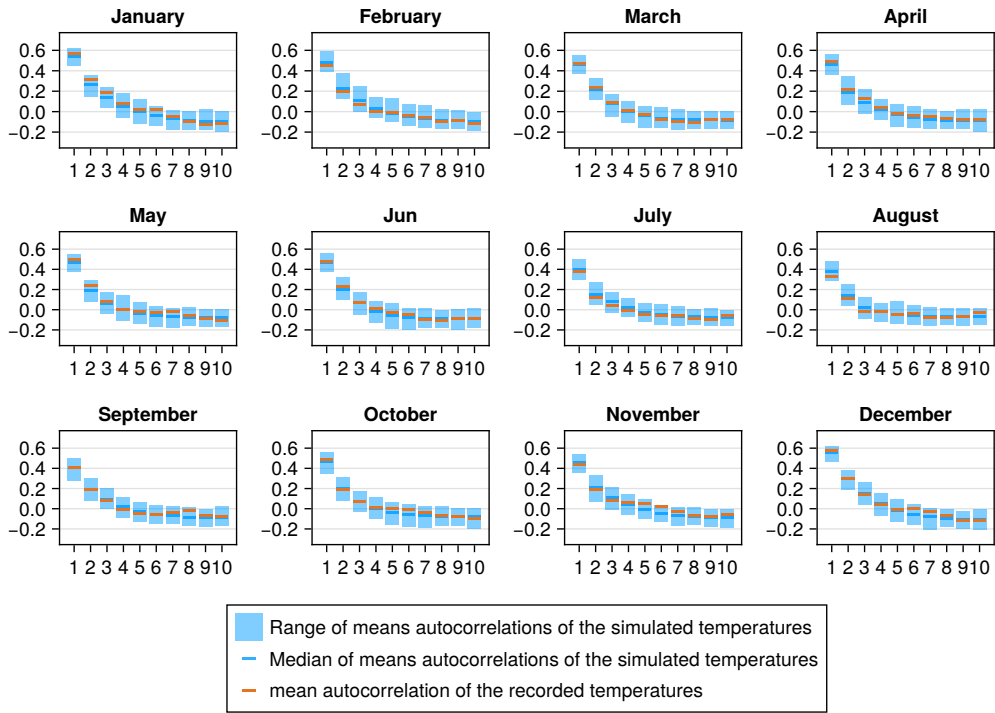


FIGURE 26 – Comparaison entre les ACF mensuelles des résidus de la série de températures minimales (TN) journalières enregistrées à Nantes-Bougenais et celles des résidus des 1000 simulations du générateur de météo entraîné sur les séries appariées de TN et TX avec  $p = 3$  et  $q = 2$ .

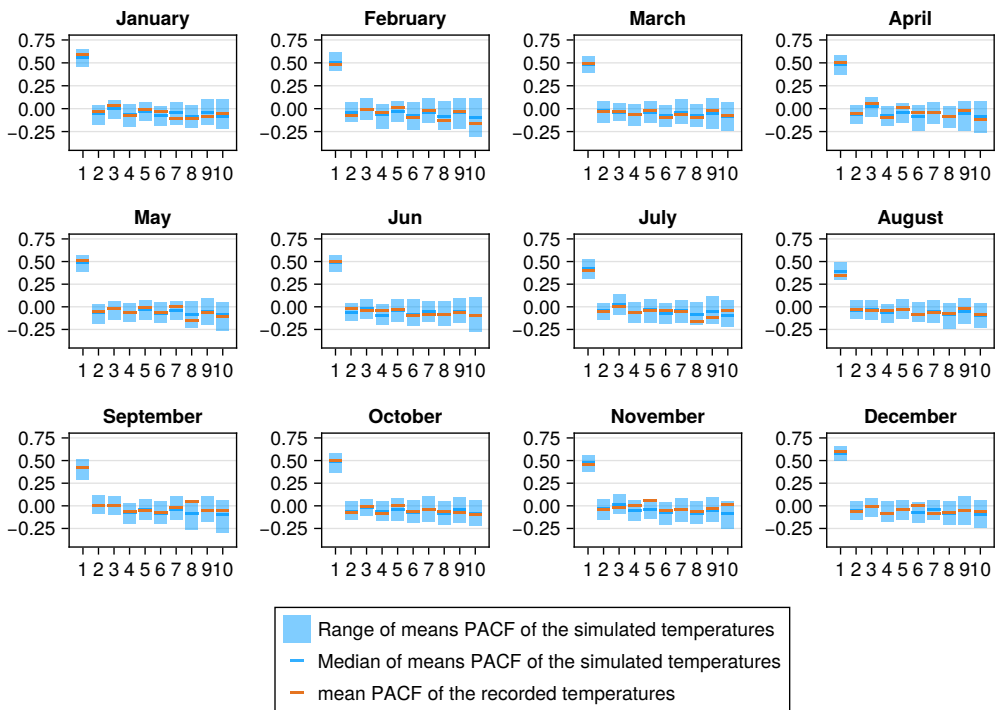


FIGURE 27 – Comparaison entre les PACF mensuelles des résidus de la série de températures minimales (TN) journalières enregistrées à Nantes-Bougenais et celles des résidus des 1000 simulations du générateur de météo entraîné sur les séries appariées de TN et TX avec  $p = 3$  et  $q = 2$ .

## B.2 Résultats pour la séries de températures maximales journalières (TX)

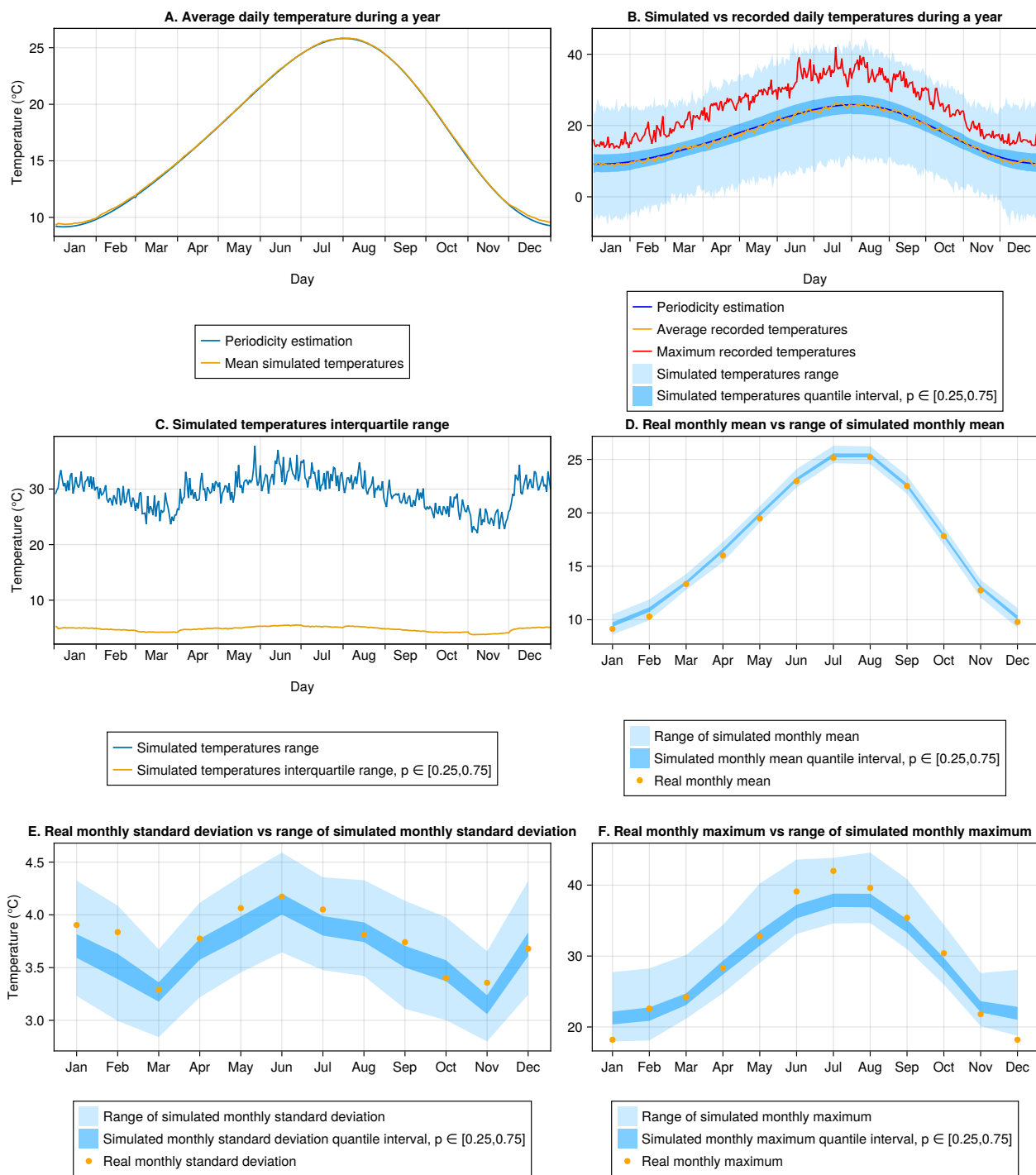


FIGURE 28 – Différentes comparaisons entre la série de températures maximales (TX) journalières enregistrées à la station Nantes-Bougenais et 1000 simulations du générateur de météo entraîné sur les séries appariées de TN et TX avec  $p = 3$  et  $q = 2$ .

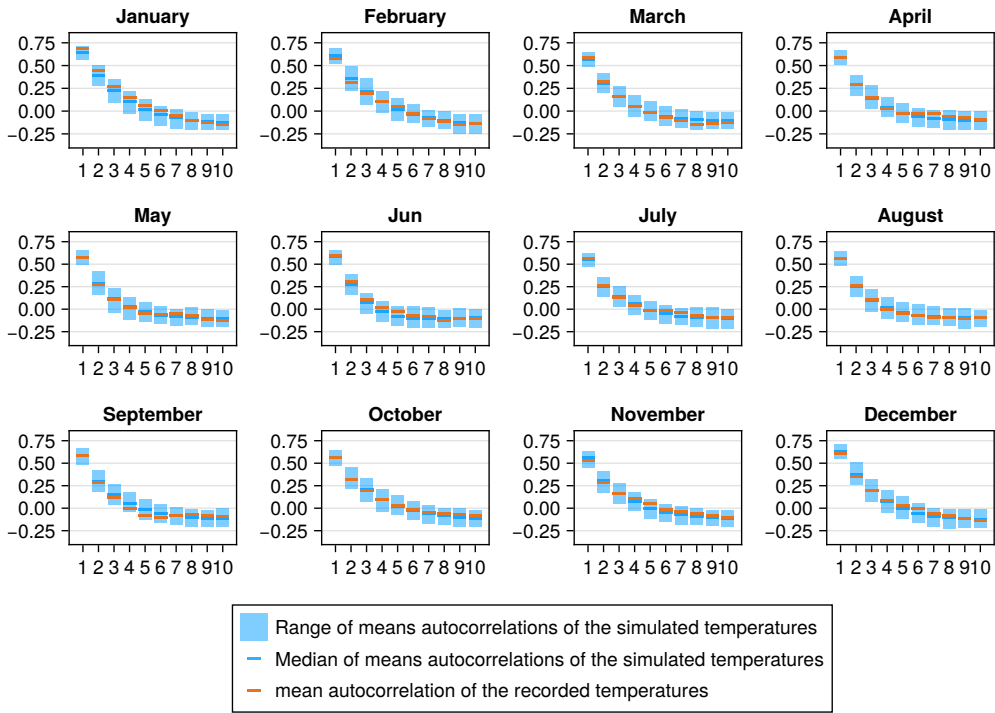


FIGURE 29 – Comparaison entre les ACF mensuelles des résidus de la série de températures maximales (TX) journalières enregistrées à Nantes-Bougenais et celles des résidus des 1000 simulations du générateur de météo entraîné sur les séries appariées de TN et TX avec  $p = 3$  et  $q = 2$ .

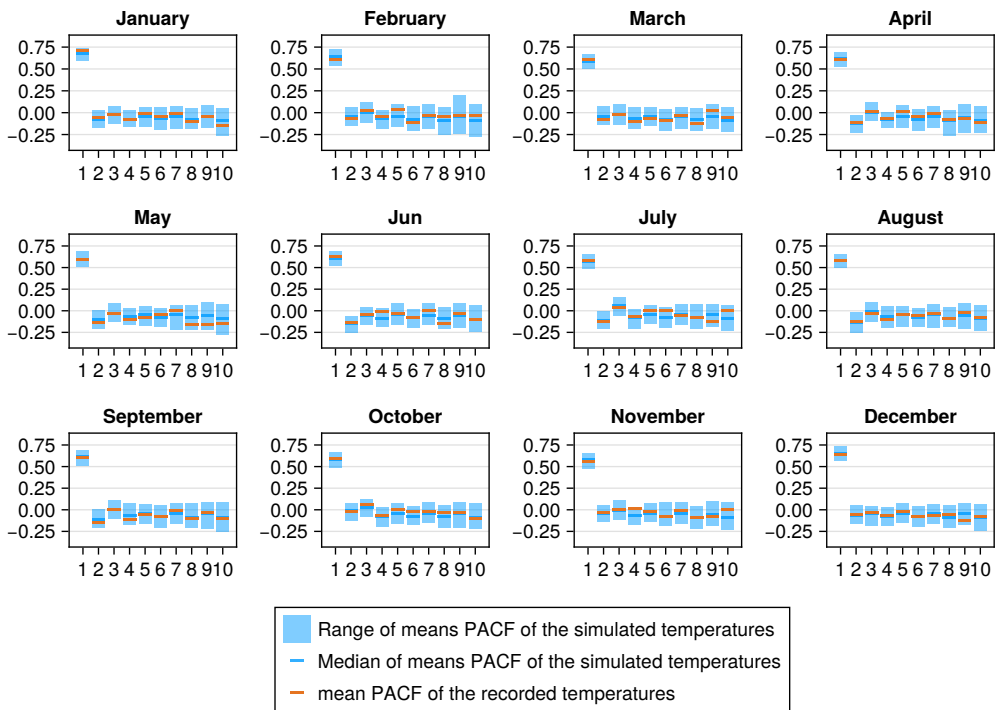


FIGURE 30 – Comparaison entre les PACF mensuelles des résidus de la série de températures maximales (TX) journalières enregistrées à Nantes-Bougenais et celles des résidus des 1000 simulations du générateur de météo entraîné sur les séries appariées de TN et TX avec  $p = 3$  et  $q = 2$ .

# Annexe C Résultats des modèles phénologiques appliquées aux simulations de générateurs météo

## C.1 Générateurs de météo entraîné sur les données de Montpellier-Aéroport

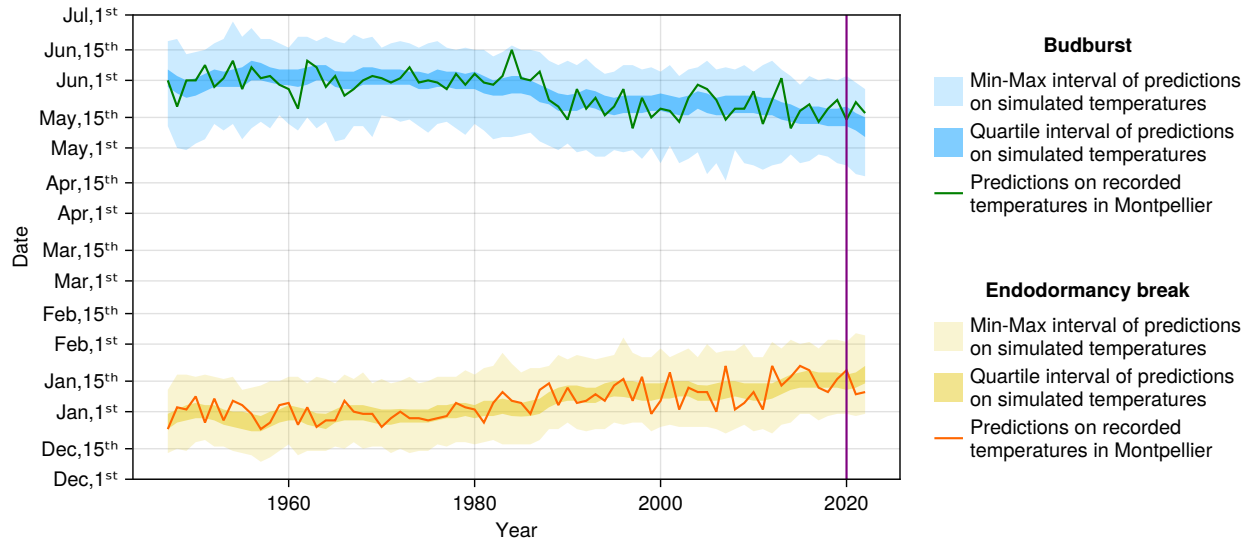


FIGURE 31 – Prédications du modèle phénologique de vigne de la section 3.2 sur le couple (TN, TX) enregistré à la station de Montpellier-Aéroport et sur 1000 simulations du générateur météo entraîné sur ce couple avec  $p = 6$  et  $q = 5$ .

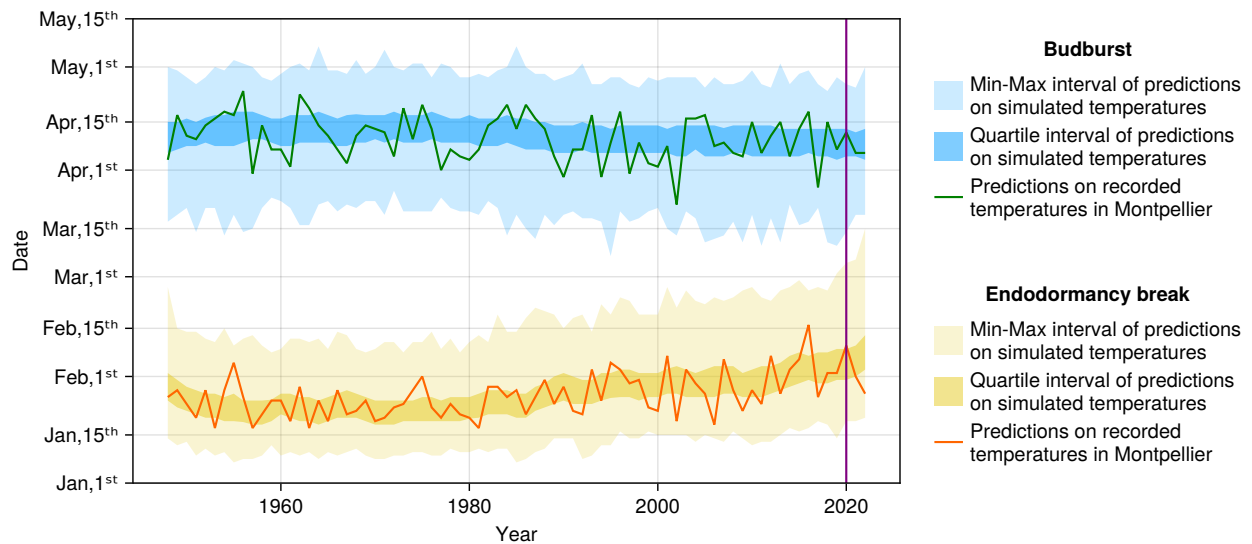


FIGURE 32 – Prédications du modèle phénologique de pomme de la section 3.3 sur le couple (TN, TG) enregistré à la station de Montpellier-Aéroport et sur 1000 simulations du générateur météo entraîné sur ce couple avec  $p = 5$  et  $q = 2$ .

## C.2 Générateurs de météo entraîné sur les données de Nantes-Bougenais

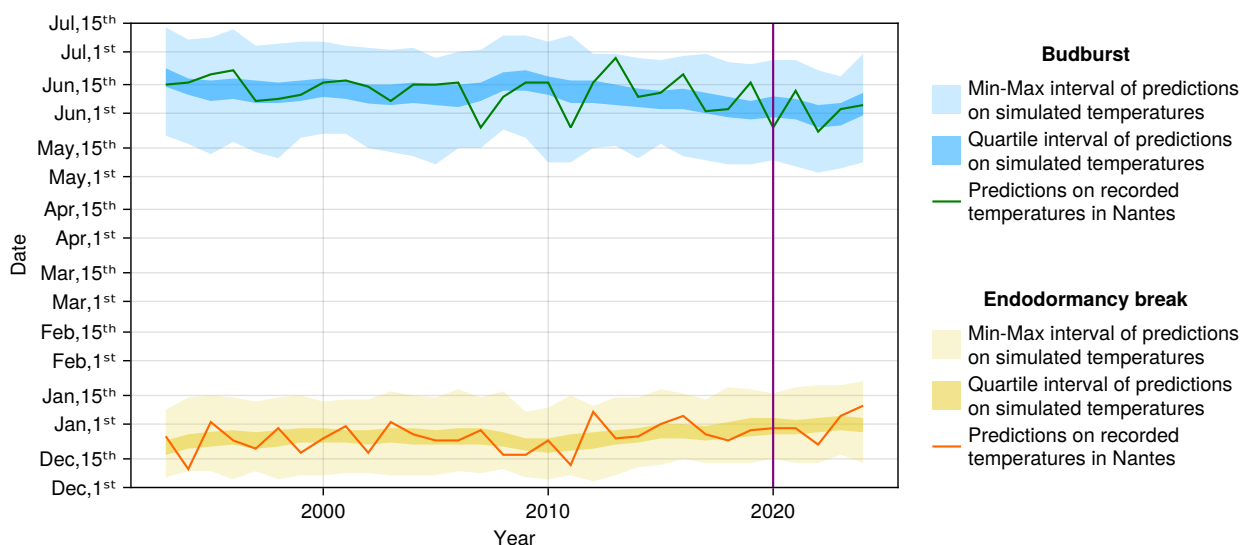


FIGURE 33 – Prédications du modèle phénologique de vigne de la section 3.2 sur le couple (TN, TX) enregistré à la station de Nantes-Bougenais et sur 1000 simulations du générateur météo entraîné sur ce couple avec  $p = 3$  et  $q = 2$ .

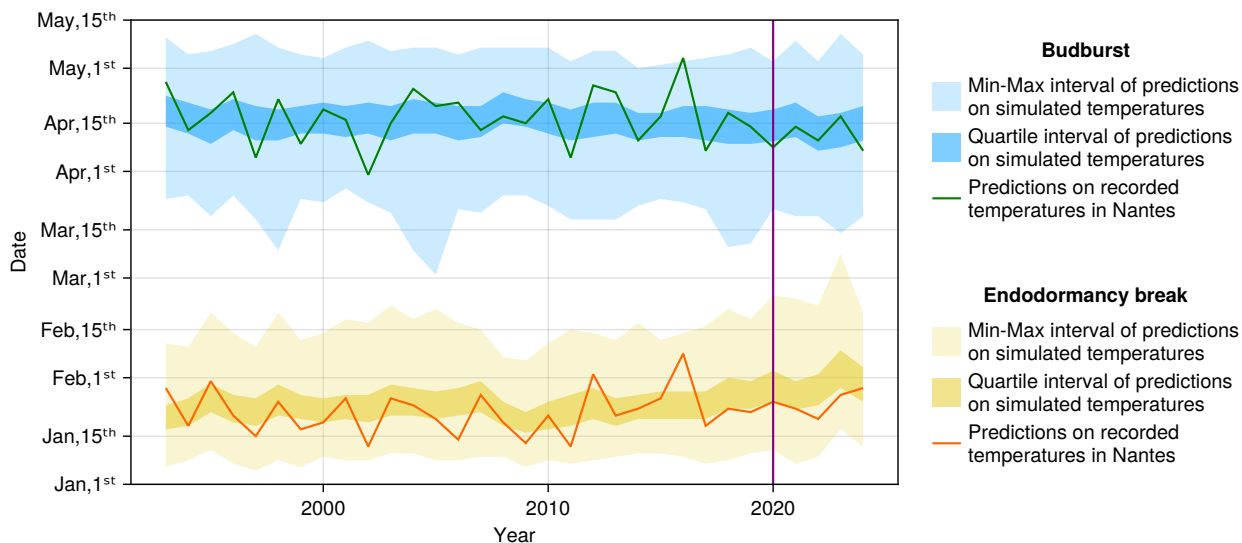


FIGURE 34 – Prédications du modèle phénologique de pomme de la section 3.3 sur le couple (TN, TG) enregistré à la station de Nantes-Bougenais et sur 1000 simulations du générateur météo entraîné sur ce couple avec  $p = 3$  et  $q = 2$ .

## Annexe D Histogrammes des prédictions des modèles phénologiques appliquées aux simulations de générateurs de météo entraînés sur projections climatiques.

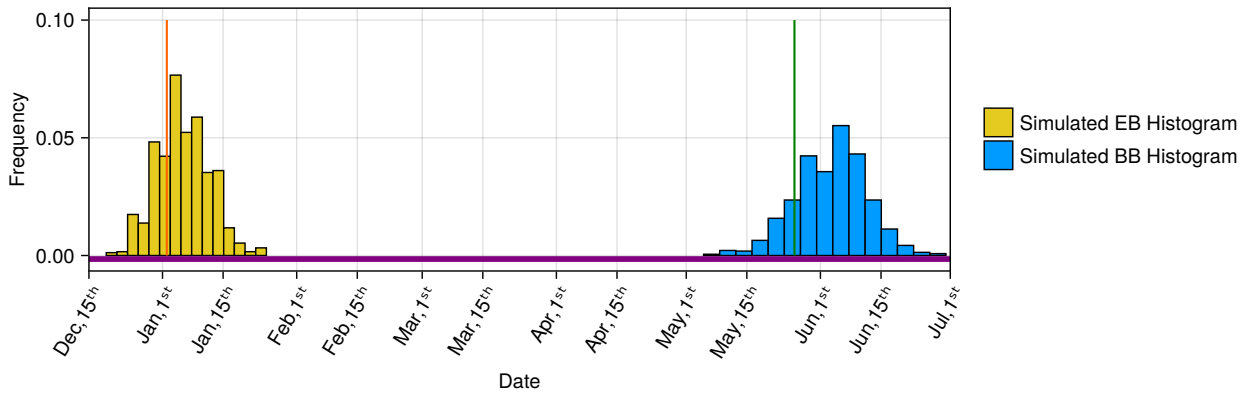


FIGURE 35 – Histogramme des prédictions pour l'année 2080 du modèle phénologique de vigne de la section 3.2 sur 1000 simulations du générateur météo entraîné sur le couple (TN, TX) de la projection climatique aux coordonnées  $7.173^{\circ}, -1.621930^{\circ}$  selon le scénario rc4.5 avec  $p = 3$  et  $q = 2$ .

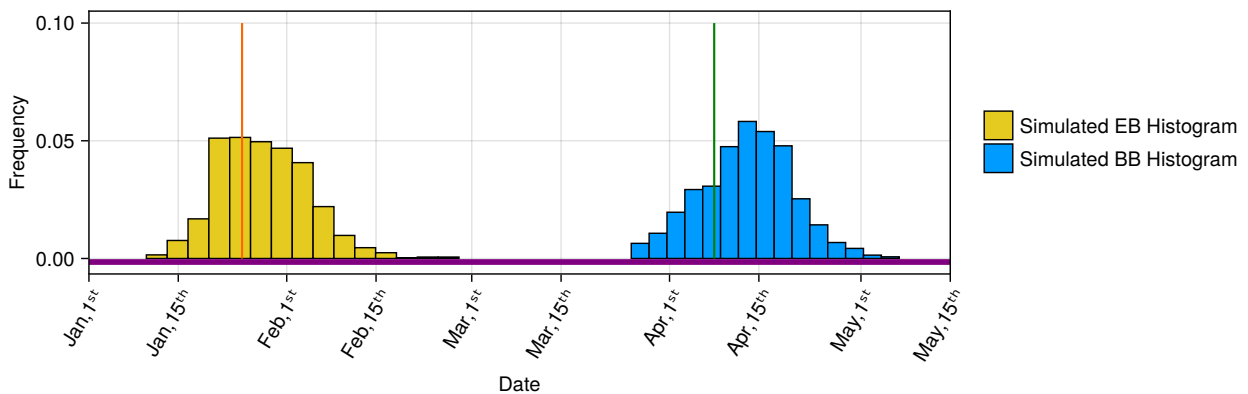


FIGURE 36 – Histogramme des prédictions pour l'année 2080 du modèle phénologique de pomme de la section 3.3 sur 1000 simulations du générateur météo entraîné sur le couple (TN, TG) de la projection climatique aux coordonnées  $7.173^{\circ}, -1.621930^{\circ}$  selon le scénario rc4.5 avec  $p = 3$  et  $q = 2$ .



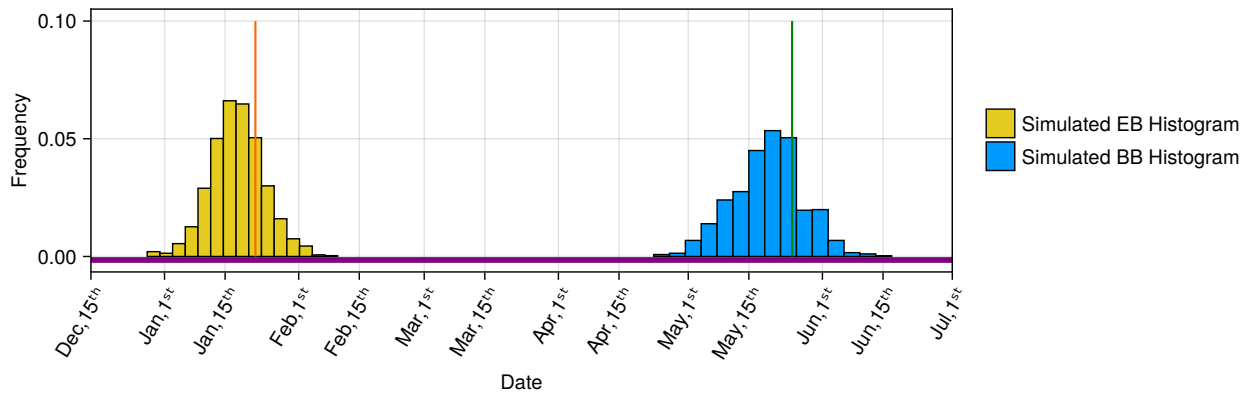


FIGURE 37 – Histogramme des prédictions pour l’année 2080 du modèle phénologique de vigne de la section 3.2 sur 1000 simulations du générateur météo entraîné sur le couple (TN, TX) de la projection climatique aux coordonnées  $7.173^{\circ}, -1.621930^{\circ}$  selon le scénario rcp8.5 avec  $p = 3$  et  $q = 2$ .

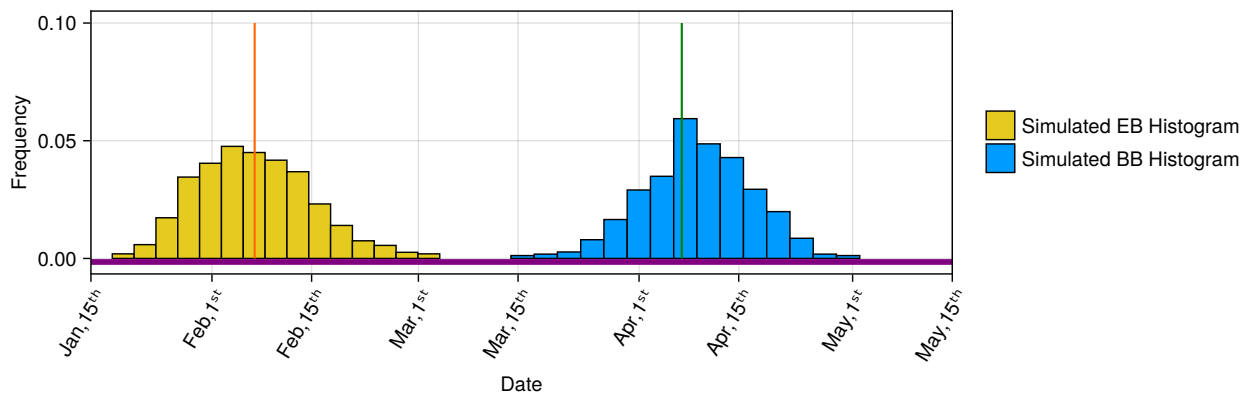


FIGURE 38 – Histogramme des prédictions pour l’année 2080 du modèle phénologique de pomme de la section 3.3 sur 1000 simulations du générateur météo entraîné sur le couple (TN, TG) de la projection climatique aux coordonnées  $7.173^{\circ}, -1.621930^{\circ}$  selon le scénario rcp8.5 avec  $p = 3$  et  $q = 2$ .

## Annexe E Lien du dépôt github du projet

Voici ci dessous le lien du dépôt github dans lequel il y a tous les programmes qui ont servis à réaliser notre étude et des résultats qui ne sont présents ni dans le rapport ni dans l'annexe : [https://github.com/dmetivie/2025\\_Internship\\_SWG\\_Pomme](https://github.com/dmetivie/2025_Internship_SWG_Pomme) Si ce lien ne fonctionne pas n'hésitez pas à me contacter pour que je vous envoie directement les résultats ou les programmes.

## Références

- P. Ailliot, C. Thompson, and 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) :405–426, July 2009. ISSN 0035-9254. doi : 10.1111/j.1467-9876.2008.00654.x. URL <https://doi.org/10.1111/j.1467-9876.2008.00654.x>.
- P. Ailliot, D. Allard, V. Monbet, and P. Naveau. Stochastic weather generators : an overview of weather type models. *Journal de la société française de statistique*, 156(1), 2015.
- H. Akaike. Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory*, pages 267–281, 1973.
- Andrew Claster. User & developer survey. <https://julialang.org/assets/2023-julia-user-developer-survey.pdf>, 2023.
- A. Caffarra, M. Rinaldi, E. Eccel, V. Rossi, and I. Pertot. Modelling the impact of climate change on the interaction between grapevine and its pests and pathogens : European grapevine moth and powdery mildew. *Agriculture, Ecosystems & Environment*, 148 :89–101, 2012.
- R. T. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.
- W. S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association*, 74(368) :829–836, 1979.
- W. S. Cleveland and E. Grosse. Computational methods for local regression. *Statistics and computing*, 1(1) :47–62, 1991.
- C. Cognot, L. Bel, D. Métivier, and S. Parey. A spatio-temporal weather generator for the temperature over France. 2025.
- DRIAS. Les futurs du climat : Les scénarios rep. Site web de DRIAS, section Accompagnement, n.d. URL <https://www.drias-climat.fr/accompagnement/sections/175>. Consulté le 16 Septembre 2025.
- R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica : Journal of the econometric society*, pages 987–1007, 1982.
- G. Evin. Stochastic generators of multi-site daily temperature : Comparison of performances in various applications. *Theoretical and Applied Climatology*, Jan. 2018. URL [https://www.academia.edu/103652321/Stochastic\\_generators\\_of\\_multi\\_site\\_daily\\_temperature\\_comparison\\_of\\_performances\\_in\\_various\\_applications](https://www.academia.edu/103652321/Stochastic_generators_of_multi_site_daily_temperature_comparison_of_performances_in_various_applications).
- G. Evin, A.-C. Favre, and B. Hingray. Stochastic generation of multi-site daily precipitation focusing on extreme events. *Hydrology and Earth System Sciences*, 22(1) :655–672, Jan. 2018. ISSN 1027-5606. doi : 10.5194/hess-22-655-2018. URL <https://hess.copernicus.org/articles/22/655/2018/>.
- I. Funes, X. Aranda, C. Biel, J. Carbó, F. Camps, A. J. Molina, F. de Herralde, B. Grau, and R. Savé. Future climate change impacts on apple flowering date in a mediterranean subbasin. *Agricultural Water Management*, 164 :19–27, 2016.

- I. García de Cortázar-Atauri, N. Brisson, and J. P. Gaudillere. Performance of several models for predicting budburst date of grapevine (*Vitis vinifera* L.). *International Journal of Biometeorology*, 53(4) :317–326, July 2009. ISSN 1432-1254. doi : 10.1007/s00484-009-0217-4. URL <https://doi.org/10.1007/s00484-009-0217-4>.
- T. T. H. Hoang. *MODELING OF NON STATIONARY, NONLINEAR TIME SERIES Application to the definition of trends on mean, variability and extremes of air temperatures in Europe*. phdthesis, Université Paris Sud - Paris XI, Jan. 2010. URL <https://theses.hal.science/tel-00531549>.
- J.-P. Indjehagopian. Les processus stochastiques vectoriels arma : une procédure d'identification. *Revue de Statistique appliquée*, 27(3) :33–45, 1979.
- JuliaLang.org contributors. Julia micro-benchmarks. <https://julialang.org/benchmarks/>, 2024.
- Á. Kőrösi, V. Markó, A. Kovács-Hostyánszki, L. Somay, Á. Varga, Z. Elek, V. Boreux, A.-M. Klein, R. Földesi, and A. Báldi. Climate-induced phenological shift of apple trees has diverse effects on pollinators, herbivores and natural enemies. *PeerJ*, 6 :e5269, 2018.
- J. M. Legave, M. Blanke, D. Christen, D. Giovannini, V. Mathieu, and R. Oger. A comprehensive overview of the spatial and temporal variability of apple bud dormancy release and blooming phenology in Western Europe. *International Journal of Biometeorology*, 57(2) :317–331, Mar. 2013. ISSN 1432-1254. doi : 10.1007/s00484-012-0551-9. URL <https://doi.org/10.1007/s00484-012-0551-9>.
- J.-M. Legave, Y. Guédon, G. Malagi, A. El Yaacoubi, and M. Bonhomme. Differentiated responses of apple tree floral phenology to global warming in contrasting climatic regions. *Frontiers in plant science*, 6 :1054, 2015.
- R. McSweeney. Explainer : What the new ipcc report says about extreme weather and climate change, Aug. 2021. URL <https://www.carbonbrief.org/explainer-what-the-new-ipcc-report-says-about-extreme-weather-and-climate-change/>. Consulté sur carbonbrief.org.
- S. Obakrim, L. Benoit, and D. Allard. A multivariate and space-time stochastic weather generator using a latent Gaussian framework. *Stochastic Environmental Research and Risk Assessment*, July 2025. ISSN 1436-3259. doi : 10.1007/s00477-024-02897-8. URL <https://doi.org/10.1007/s00477-024-02897-8>.
- S. Parey, T. Hoang, and D. Dacunha-Castelle. Validation of a stochastic temperature generator focusing on extremes, and an example of use for climate change. *Climate Research*, 59 (1) :61–75, Feb. 2014. ISSN 0936-577X, 1616-1572. doi : 10.3354/cr01201. URL <http://www.int-res.com/abstracts/cr/v59/n1/p61-75/>.
- A. Parker, I. García de Cortázar-Atauri, M. C. Trought, A. Destrac, R. Agnew, A. Sturman, and C. Van Leeuwen. Adaptation to climate change by determining grapevine cultivar differences using temperature-based phenology models. 2020.
- M. B. Parlange and R. W. Katz. An Extended Version of the Richardson Model for Simulating Daily Weather Variables. *Journal of Applied Meteorology and Climatology*, 39(5) :610–622, May 2000. ISSN 1520-0450, 0894-8763. doi : 10.1175/1520-0450-39.5.610.

- C. W. Richardson. Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, 17(1) :182–190, 1981. ISSN 1944-7973. doi : 10.1029/WR017i001p00182. URL <https://onlinelibrary.wiley.com/doi/abs/10.1029/WR017i001p00182>.
- F. Rousset. La tracc : Le climat d'une france à +2,7°C et à +4°C. Présentation au séminaire CLIMAE, juin 2025. URL <https://metaprogrammes.inrae.fr/media/files/climae/seminaire-2025/sessions-plenieres/tracc>. Séminaire CLIMAE, 19 juin 2025.
- G. Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages 461–464, 1978.
- G. Sgubin, D. Swingedouw, G. Dayon, I. G. de Cortázar-Atauri, N. Ollat, C. Pagé, and C. van Leeuwen. The risk of tardive frost damage in french vineyards in a changing climate. *Agricultural and Forest Meteorology*, 250 :226–242, 2018.
- G. Stupfler and F. Proïa. Cours de séries chronologiques, master 2 data science, université d'angers, 2024.
- T. Sugiura, H. Ogawa, N. Fukuda, and T. Moriguchi. Changes in the taste and textural attributes of apples in response to climate change. *Scientific reports*, 3(1) :2418, 2013.
- C. Van Leeuwen, A. Destrac-Irvine, M. Dubernet, E. Duchêne, M. Gowdy, E. Marguerit, P. Pieri, A. Parker, L. De Resseguier, and N. Ollat. An update on the impact of climate change in viticulture and potential adaptations. *Agronomy*, 9(9) :514, 2019.