

Mémoire

Auteur : Louis, Simon

Promoteur(s) : Doutreloup, Sébastien

Faculté : Faculté des Sciences

Diplôme : Master en sciences géographiques, orientation global change, à finalité approfondie

Année académique : 2022-2023

URI/URL : <http://hdl.handle.net/2268.2/17338>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.



Faculté des sciences
Département de géographie

Développement d'un outil opérationnel de prévision d'apparition de gel printanier dans un vignoble wallon, à l'aide du modèle MAR

Mémoire présenté par : **Simon Louis**

pour l'obtention du titre de

**Master en sciences géographiques,
orientation global change**

Année académique :

2022-2023

Date de défense :

Juin 2023

Président de jury :

Pr. Xavier FETTWEIS

Promoteur :

Dr. Sébastien DOUTRELOUP

Jury de lecture :

Dr. Nicolas GHILAIN

Pr. Guénaël DEVILLET

Remerciements

Je remercie tout d'abord mes collègues d'études, les assistants, les professeurs ainsi que mon entourage m'ayant accompagné durant mes années d'études en géographie.

En effet des études ça ne se réalise/vit pas seul. Être dans un département facultaire de petite taille a quelques désavantages (moins de tuyaux ou de synthèses par exemple) mais permet de mieux connaître les personnes avec qui nous faisons nos études. Cela permet de mieux s'entraider et de se rassurer les uns les autres à l'approche d'examens ou de projets difficiles. Cela permet également, de passer de bons moments de vie car les études ce n'est pas que travailler jour et nuit.

Un autre avantage de la petite taille d'un département facultaire est la proximité avec les professeurs et les assistants, ils peuvent nous suivre et nous épauler dans notre apprentissage de manière efficace. Ceux-ci sont toujours (souvent ?) disposés à répondre à nos questions. Je pense particulièrement à Xavier Fettweis à qui j'ai posé un nombre incalculable de questions ces deux dernières années afin d'améliorer ma programmation. Il a répondu à toutes mes sollicitations et de manière rapide.

En ce qui concerne le mémoire, je remercie Sébastien Doutreloup qui m'a accompagné et guidé dans la création de ce mémoire. Lui aussi a toujours répondu à mes nombreuses questions et rapidement, ce qui m'a permis d'avancer dans mon mémoire tout au long de cette année. De plus, le sujet proposé m'a permis de réaliser de la recherche appliquée ce que j'ai particulièrement apprécié. Enfin je remercie ma famille pour leur soutien durant la réalisation de ce mémoire, particulièrement mon père qui a porté un œil extérieur sur le travail, ce qui m'a permis rendre compréhensible ce que j'avais en tête et mon frère qui a pu me prodiguer des conseils techniques.

Résumé

Avec le changement climatique et la montée des isothermes vers le nord, la Belgique devient un territoire de plus en plus propice à accueillir des vignobles. En effet la Belgique possède les indices bioclimatiques nécessaires à la croissance des vignes. De ce fait, l'activité vigneronne est en croissance en Belgique. Toutefois, la Belgique n'est pas à l'abri des gelées de printemps qui font beaucoup de dégâts aux vignes que ce soit dans le nord de la France ou ailleurs dans le monde. Pour éviter les pertes, les vignerons doivent déployer des outils de défense tels que des bougies chauffantes, des tours à vent, des hélicoptères, de l'aspersion et bien d'autres outils. Ces outils coûtent du temps et de l'argent et ont besoin, pour leur implémentation, de prévisions de gel précises. Ce mémoire a pour but de développer un outil opérationnel de prévision de gel de printemps. Cet outil est créé à partir de deux bases de données, provenant de puces qui ont enregistré la température au niveau du fruit de la vigne sur l'année 2021-2022 ainsi que du MAR (modèle atmosphérique régional, forcé par ERA5). Deux méthodes ont été utilisées pour tenter de prévoir le gel. Ces deux méthodes se sont basées sur les années 2021-2022. Un modèle prédictif a tout d'abord été généré à partir d'une formule modifiée empiriquement : la température de sortie du MAR a ainsi été transformée pour être la plus proche possible de celle des fruits des vignes, estimée grâce aux puces. Ensuite, un second modèle a été créé, via une régression logistique multivariée prenant en compte, en plus de la température minimum du MAR, plusieurs variables pouvant impacter l'apparition de gel. Il en a résulté une amélioration des prévisions des nuits de gel par rapport aux sorties brutes du MAR. En effet sur l'année 2022, à partir de mars, sur 46 nuits, le MAR brut (sans modification) a prédit 7 nuits de gel correctes, 9 nuits de gel manquées, et 30 nuits sans gel correctement prévues alors que le meilleur modèle de ce mémoire (issu d'une régression logistique) a prévu sur ces mêmes 46 nuits, 13 nuit de gel correctes, 3 nuits de gel manquées et 30 nuits sans gel correctement prévues. Le modèle doit encore être testé sur des prévisions futures avec MAR forcé par le modèle global GFS par exemple.

Abstract

With climate change and the northward shift of isotherms, Belgium is becoming an increasingly suitable territory for vineyards. Indeed, Belgium possesses the bioclimatic indices necessary for vine growth. As a result, winegrowing activity is increasing in Belgium. However, Belgium is not immune to spring frosts, which cause a lot of damage to vineyards both in northern France and elsewhere in the world. To avoid losses, winemakers must deploy defense tools such as heating candles, wind towers, helicopters, aspersion, and many other tools. These tools cost time and money and require precise frost forecasts for their implementation. The purpose of this thesis is to develop an operational tool for spring frost prediction. This tool has been created from two databases, obtained from chips that recorded the temperature at the vine fruit level during the 2021-2022 year, as well as from the MAR (regional atmospheric model, forced by ERA5). Two methods were used to attempt frost prediction. These two methods were based on the years 2021-2022. A predictive model was first generated using an empirically modified formula: the MAR output temperature was transformed to be as close as possible to that of the vine fruit, estimated using the chips. Then, a second model was created through multivariate logistic regression, taking into account, in addition to the minimum temperature of the MAR, several variables that could impact the occurrence of frost. This resulted in an improvement in the prediction of frost nights compared to the raw outputs of the MAR. In fact, in 2022, from March onwards, out of 46 nights, the raw MAR (without modification) correctly predicted 7 frost nights, missed 9 frost nights, and correctly predicted 30 nights without frost, whereas the best model in this thesis (derived from logistic regression) predicted on the same 46 nights, 13 correct frost nights, 3 missed frost nights, and correctly predicted 30 nights without frost. The model still needs to be tested on future forecasts with the MAR forced by the global GFS model, for example.

Table des matières

Table des illustrations.....	7
Acronymes	12
I. Introduction.....	13
I.1. Introduction générale.....	13
I.2. La vigne et sa sensibilité au gel.....	14
I.3. Contexte météorologique et local modifiant le risque de gel de la vigne	17
I.4. Systèmes de prévisions.....	19
I.4.1. Modèles utilisant des formules hygrométriques	19
I.4.2. Machine Learning.....	19
I.4.3. Prévision du gel via des observations de terrain et modèle numérique.....	21
I.4.4. Prévision spatiale de zone de gel tardif	25
I.5. Objectifs de ce mémoire.....	25
II. Méthodologie et matériel.....	27
II.1. Enregistrement et traitement des données de température au niveau de la vigne.....	27
II.1.1. Enregistrement des températures au niveau du fruit des vignes	27
II.1.2 Traitement des observations.....	28
II.2. Variables climatiques générées par modélisation (MAR)	32
II.3. Détermination de la température au niveau du fruit de la vigne depuis la température de l'air déterminée par le MAR.....	33
II.4. Analyses complémentaires basée sur l'expérience des acteurs de terrain	35
II.4.1. Approche visuelle de la distribution des données	36
II.4.2. Approche comparative statistique des données.....	36
II.5. Régression logistique	37
II.5.1. Tests de corrélation	37
II.5.2. Préparation et application de la régression logistique.....	37

II.6. Comparaison des deux meilleurs modèles issus de la régression logistique et de la formule de Guyot modifiée (TAS)	40
II.7. Analyse de la qualité des prévisions par nuit sur une année.....	40
II.8. Comparaison du meilleur modèle avec les résultats de la littérature.....	41
III. Résultats et discussions	42
III.1. Détermination de la température actinothermique simulée (TAS).....	42
III.2. Alertes complémentaires	50
III.2.1. Représentation graphique des fonctions de distribution cumulative des alertes... ..	50
III.2.2. Test de Shapiro	51
III.2.3. Test de Mann Whitney	51
III.3. Régression logistique.....	53
III.3.1. Analyse univariée.....	53
III.3.2 Application de la régression logistique.....	56
III.3.2.1. Régression logistique avec la température actinothermique simulée	56
III.3.2.2. Régression logistique avec la température minimum du MAR	58
III.3.2.3. Régression logistique avec les variables de l’alerte vigneron	60
III.3.3. Régression logistique avec sélection de données par régression de LASSO.....	62
III.3.3.1. Application de la régression logistique avec des variables issues de LASSO ..	63
III.3.3.2. Régression logistique avec les variables sélectionnées via LASSO d’après un training set non équilibré	66
III.4. Comparaison des deux meilleurs modèles issus de différentes techniques.....	69
III.5. Analyse de la qualité des prévisions par nuit sur une année	70
III.5.1. Analyse sur la nuit des prévisions réalisées par la régression logistique (LASSO)	70
III.5.2. Analyse sur la nuit des prévisions réalisées par la température actinothermique simulée.....	72
III.5.3. Analyse sur la nuit des prévisions réalisées par la température minimum du MAR (cas de référence)	73

III.6. Limitations des modèles	74
III.6.1. Température actinothermique simulée (TAS).....	74
III.6.2. Régression logistique	75
III.6.3. Limitations communes.....	75
III.7. Comparaison avec la littérature	77
IV. Conclusions et perspectives	78
Bibliographie.....	80
Annexe	85

Table des illustrations

Carte 1 : Localisation des vignobles dans le village de Lincé. Atlas de Belgique, WalOnMap (2023).	27
Carte 2 Emplacement des puces dans les vignobles. Vignoble Henri Simon à gauche et vignoble village à droite, deux vignobles de Sprimont dans le village de Lincé. Selon Leerschool (2022).	28
Figure 1 Repère de l'évolution des stades phénologiques du bourgeon (Baggiolini, 1952)...	15
Figure 2 Les cycles végétatifs et reproducteurs de la vigne. (Reynier, 1991).....	17
Figure 3 Table de contingence pour les 4 postes pour une valeur de l'indice actinothermique inférieur ou égale à -2.5°C sur 50 jours répartis entre 1988 et 1989. (Choisnel et al. 1993)...	23
Figure 4 Différence entre la température minimum des puces et la température actinothermique simulée (courbe bleue) ainsi que la différence entre la température minimum des puces et la sortie température minimum du MAR (courbe orange) sur la période d'intérêt de mars avril mai 2021 (avec des données toutes les heures).	45
Figure 5 Profils horaires de la température minimum des puces (courbe orange), de la température actinothermique simulée (courbe verte) et la température minimale du MAR (courbe bleue) pour mars avril mai 2021.	45
Figure 6 Profils horaires de la température minimale des puces (courbe orange), de la température actinothermique simulée (courbe verte) et la température minimale du MAR (courbe bleue) pour mars avril mai 2022.	46
Figure 7 Comparaison (différence) de la température minimum du MAR avec la température de surface du MAR (TS2) sur mars avril mai 2021 afin de déterminer si l'utilisation de la température de surface est pertinente.	49
Figure 8 Distribution des trois alertes complémentaires via une fonction distributive cumulative. Les alertes sont divisées en deux échantillons : la valeur des alertes sans gel (courbe bleue) et la valeur des alertes lorsqu'il y a gel (courbe orange). Les alertes sont le gradient de température entre 3 et 10 mètres, la vitesse du vent à 2 mètre et la vitesse du vent à 10 mètres.	50
Figure 9 La ROC curve et la Precision-Recall curve du test set avec la température actinothermique simulée comme seule variable indépendante de la régression logistique avec quatre seuils différents. Pour la Precision-Recall curve, la droite horizontale représente le	

seuil de fausses alertes permises par les spécialistes du monde de la vigne (Précision de 70%).	56
Figure 10 La ROC curve et la Precision-Recall curve du test set avec la température minimum du MAR comme seule variable indépendante de la régression logistique avec quatre seuils différents. Pour la Precision-Recall curve, la droite horizontale représente le seuil de fausses alertes permises par les spécialistes du monde de la vigne (Précision de 70%).	58
Figure 11 La ROC curve et la Precision-Recall curve du test set avec les variables de l’alerte vigneron comme variables indépendantes (TAS, gradient de température et vitesse du vent à 10m) de la régression logistique avec quatre seuils différents. Pour la Precision-Recall curve, la droite horizontale représente le seuil de fausses alertes permises par les spécialistes du monde de la vigne (Précision de 70%).	60
Figure 12 La ROC curve et la Precision-Recall curve du test set avec les variables sélectionnées par LASSO (Température minimum du MAR, COD, SLP) comme variables indépendantes de la régression logistique avec quatre seuils différents. Pour la Precision-Recall curve, la droite horizontale représente le seuil de fausses alertes permises par les spécialistes du monde de la vigne (Précision de 70%).	63
Figure 13 La ROC curve et la Precision-Recall du test set avec les variables indépendantes retenues par la régression de LASSO (température minimum du MAR, la densité optique des nuages COD et SLP la pression de surface au niveau de la mer) avec un training set non équilibré. Sur le graphique de droite, la droite horizontale représente le seuil de fausses alertes permises par les spécialistes du monde de la vigne (Précision de 70%).	66
Figure 14 Répartition du type de prévision par nuit (prévision de gel correct (bâtonnet vert), prévision de gel incorrect (bâtonnet rouge), événement gélif manqué (bâtonnet orange)) issue de la régression logistique avec les variables sélectionnées par LASSO (Température minimum du MAR, COD, SLP) lors de l’année 2022. Les prévisions de non gel exactes ne sont pas représentées. Le graphique s’arrête le 15 avril car il n’y plus de gel détecté ensuite pour cette année-là. Le nombre d’évènements gélifs manqués est exprimé en valeur négative afin de faciliter la lecture du graphique.	70
Figure 15 Répartition du type de prévision par nuit (prévision de gel correct (bâtonnet vert), prévision de gel incorrect (bâtonnet rouge), événement gélif manqué (bâtonnet orange)) obtenue par la température actinothermique (formule de Guyot modifiée) lors de l’année 2022. Les prévisions de non gel exactes ne sont pas représentées. Le graphique s’arrête le 15 avril car il n’y plus de gel détecté ensuite pour cette année-là. Le nombre d’évènements gélifs manqués est exprimé en valeur négative afin de faciliter la lecture du graphique.	72

Figure 16 Répartition du type de prévision sur la nuit (prévision de gel correct (bâtonnet vert), prévision de gel incorrect (bâtonnet rouge), événement gélif manqué (bâtonnet orange)) obtenue par la température minimum du MAR lors de l'année 2022. Les prévisions de non gel exactes ne sont pas représentées. Le graphique s'arrête le 15 avril car il n'y plus de gel détecté ensuite pour cette année-là. Le nombre d'évènements gélifs manqués est exprimé en valeur négative afin de faciliter la lecture du graphique. 73

Théorie 1 Définition de l'indice actinothermique. Tiré de l'article de Choïnel (1991)..... 21

Tableau 1 Caractéristiques des puces placées dans le vignobles au niveau des fruits des vignes. 28

Tableau 2 Observations de la qualité du jeu de données. La première colonne correspond à la quantité de données disponible sur le pas de temps du MAR (2021-01-01 1h à 2022-06-30 23h) contre le pas de temps des observations (2021-02-07 à 2022-06-25). La deuxième colonne s'est vue amputée des valeurs extrêmes n'étant pas comprise entre [-20°C ; +45°C]. La troisième colonne est composée des heures d'observation avec au moins 4 observations par heure. La dernière colonne est la différence entre la première et la troisième colonne. 30

Tableau 3 Sélection des meilleurs résultats sur toutes les variations d' ϵ , h et θ pour la température minimum du MAR (avec Monte Carlo) et la température minimum des puces. Le cas de référence correspond à une sortie du MAR non modifiée par la formule de Guyot, les cas 1 et 2, aux meilleurs résultats obtenus avec des valeurs différentes des paramètres (ϵ , h, θ). Pour chacun des cas les trois périodes d'intérêts ont été représentées : l'ensemble des deux ans, le printemps (Mars, avril, mai) 2021 et le printemps 2022. Le meilleur résultat est mis en évidence en vert foncé. L'ensemble des autres résultats est disponible de manière brute en annexe du mémoire. 42

Tableau 4 Sélection des meilleurs résultats sur toutes les variations d' ϵ , h et θ pour la température moyenne du MAR et la température minimum des puces. Le cas de référence correspond à une sortie du MAR non modifiée par la formule de Guyot. Pour chacun des cas les trois périodes d'intérêts ont été représentées, l'ensemble des deux ans, le printemps 2021 (Mars, avril, mai) et le printemps 2022 (Mars, avril, mai)..... 44

Tableau 5 Résultats de la corrélation, RMSE centré, biais, POD, FAR et ACC pour le pixel de 203m et 244m. L'altitude de 244.3 m correspond au pixel adjacent au pixel où se situe le vignoble ayant l'altitude la plus proche de celle du vignoble. De l'autre côté, 203m correspond au pixel du vignoble (c'est-à-dire celui qui en est le plus proche

géographiquement). Pour chacun des cas les trois périodes d'intérêts ont été représentées, l'ensemble des deux ans, le printemps 2021 (Mars, avril, mai) et le printemps 2022 (Mars, avril, mai).	47
Tableau 6 Résultats de la corrélation, RMSE centré, biais, POD, FAR et ACC pour le pixel le plus bas (197 mètres) et le plus haut (341 mètres) avoisinant le pixel où se situe le vignoble pour la température minimum du MAR sans passer par la formule de Guyot modifiée. Pour chacun des cas les trois périodes d'intérêts ont été représentées, l'ensemble des deux ans, le printemps 2021 (Mars, avril, mai) et printemps 2022 (Mars, avril, mai).....	48
Tableau 7 Résultats du test de Shapiro pour les alertes complémentaires.....	51
Tableau 8 Résultats du test de Mann Whitney pour les alertes complémentaires.	51
Tableau 9 Indication de la distribution normale ou pas suivie par l'ensemble des variables indépendantes divisées en deux catégories : leurs valeurs lorsqu'il y a du gel et leurs valeurs lorsqu'il n'y en pas. Avec la valeur du test de Shapiro ainsi que la p-valeur.	53
Tableau 10 Résultat de la corrélation de Spearman qui nous indique si les variables indépendantes sont corrélées ou non avec la variable dépendante : rho donne la corrélation.	54
Tableau 11 Corrélation entre les différentes variables indépendantes à utiliser dans la régression logistique.....	55
Tableau 12 Résultats de la régression logistique sur le test set, avec comme seule variable indépendante la température actinothermique simulée, pour les 4 seuils différents obtenus sur le training set. Pour chaque seuil, une table de contingence est présentée ainsi que la POD et le FAR.	57
Tableau 13 Résultats de la régression logistique sur le test set, avec comme seule variable indépendante la température minimum du MAR, pour les 4 seuils différents obtenus sur le training set. Pour chaque seuil, une table de contingence est présentée ainsi que la POD et le FAR.	59
Tableau 14 Résultats de la régression logistique sur le test set, avec les variables issues de l'alerte vigneron comme variables indépendantes (température actinothermique simulée, gradient de température et vitesse du vent à 10 mètres), pour les 4 seuils différents obtenus sur le training set. Pour chaque seuil, une table de contingence est présentée ainsi que la POD et le FAR.	61
Tableau 15 Valeurs des coefficients pour la régression logistique « alerte vigneron » avec la température actinothermique simulée, le gradient de température et la vitesse du vent à 10 mètres.	61

Tableau 16 Résultats de la sélection de LASSO sur l'ensemble des variables indépendantes sélectionnées avec température minimum du MAR comme variable fortement corrélée conservée.....	62
Tableau 17 Résultats de la régression logistique sur le test set, avec les variables issues la sélection de LASSO comme variables indépendantes (Température minimum du MAR, COD, SLP), pour les 4 seuils différents obtenus sur le training set. Pour chaque seuil, une table de contingence est présentée ainsi que la POD et le FAR.	64
Tableau 18 Valeur des coefficients des variables indépendantes de la régression logistique LASSO : la température minimum du MAR, la densité optique des nuages (COD) et la pression de surface au niveau de la mer (SLP).	64
Tableau 19 Table de contingence ainsi que POD et FAR obtenues via la régression logistique sur le test set avec quatre seuils différents (trouvées sur le training set non équilibré) pour les variables sélectionnées par la régression de LASSO (Température minimum du MAR, COD : la densité optique des nuages et SLP (pression de surface au niveau de la mer)).....	67
Tableau 20 Performances en POD et FAR de la régression logistique sur le training set avec un training set non équilibré et avec les variables indépendantes sélectionnées par LASSO : Température minimum du MAR, COD, SLP.	67
Tableau 21 Deux meilleures modèles issus de deux approches différentes. Les résultats sont testés sur toute l'année 2022 avec la température actinothermique simulée issue de la formule de Guyot modifiée sélectionnée au Tableau 3 ainsi que la régression logistique avec les variables indépendantes données par LASSO (température minimum du MAR, COD, SLP) et en utilisant le seuil lié à la précision.	69
Tableau 22 Toutes les variations pour trouver la meilleure formule de Guyot modifiée avec la température minimum du MAR et en comparant les résultats avec la température minimum des puces. Variation de : θ , ε et h	90
Tableau 23 Résultats de la sélection de LASSO sur l'ensemble des variables indépendantes sélectionnées avec la température actinothermique (TAS) comme variable fortement corrélée conservée.....	91
Théorie 1 Définition de l'indice actinothermique. Tiré de l'article de Choïnel (1991).....	21
Schéma 1 Explication du calcul de la moyenne et de l'écart type en utilisant Monte Carlo pour une heure.....	32

Acronymes

ACC: Accuracy, 34, 39, 42, 43, 44, 48, 49

AUC: Area under the curve, 39, 58, 60, 65, 68

CD: Couverture nuageuse en basse altitude, 37

CM: couverture nuageuse en altitude moyenne, 37, 55, 63

COD: Opacité optique des nuages, 37

CU: couverture nuageuse en haute altitude, 37, 55, 63

FAR: Taux de fausses alertes, 34, 38, 39, 42, 43, 44, 48, 49, 58, 60, 62, 65, 68, 69, 70, 77, 79, 87, 88, 89, 90

FNSEA: Fédération Nationale des syndicats d'exploitants agricoles, 13

FPR: Taux de faux positifs, 39

LWD: Rayonnement infrarouge vers le bas, 37, 55

LWU: Rayonnement infrarouge vers le haut, 37, 55

LASSO: Least Absolute Shrinkage and Selection Operator, 42, 43, 64, 65, 66, 68, 69, 71, 72, 79, 80, 84

MAR: Modèle atmosphérique régional, 25, 26, 29, 30, 31, 32, 33, 34, 35, 37, 38, 39, 40, 41, 42, 43, 44, 45, 47, 48, 49, 50, 52, 55, 59, 60, 63, 65, 67, 68, 69, 70, 74, 76, 77, 83, 86, 87, 91

MARS: Régression multivariée par spline adaptative, 25

NDVI: Normalised Difference Vegetation index, 25

POD: Probabilité de détection, 34, 38, 39, 42, 43, 44, 48, 49, 58, 60, 62, 65, 68, 69, 70, 79, 87, 88, 89, 90

PP: Précipitations, 37, 55, 63

QQ: Humidité spécifique, 37, 55

RMSE: Root mean square error, 35, 42, 43, 44, 48, 49, 87, 88, 89, 90

ROC: Receiver operating characteristic, 39, 57, 58, 59, 60, 61, 62, 64, 65, 67, 68, 84

SLP: Pression de surface au niveau de la mer, 37, 55, 63, 65, 68, 69, 70, 79

SP: Pression de surface, 37, 55, 63, 68

TAS: température actinothermique simulée, 33

TD: Température du point de rosée, 37, 55

UTM: Universal transverse Mercator, 31

I. Introduction

I.1. Introduction générale

La Belgique est un pays où la viticulture occupe une place de plus en plus importante. La superficie consacrée à l'activité a presque quintuplé en 10 ans : l'occupation du territoire par les vignes est passée de 119 ha en 2011 à 527 ha en 2021 (OIV, 2023). En particulier à Liège, l'accroissement des terres exploitées pour la vigne est conséquent : on est passé de 26,6 ha en 2018 à 62,57 ha en 2020 (FGOV Economy, 2021). L'augmentation du nombre de vignes en Belgique est notamment rendue possible par la migration des isothermes vers de plus hautes altitudes et latitudes comme expliqué par l'IPCC (2021). Ce réchauffement climatique permet aux vignes de s'implanter dans des latitudes plus hautes (Fraga et al., 2016). Il a été prouvé que la Belgique rencontre actuellement et de façon de plus en plus nette, les critères climatiques et bioclimatiques nécessaires au développement des vignes. Par exemple, un de ces indices est l'indice héliotherme d'Hulgin. Cet indice très répandu prend en compte la température moyenne et maximum journalière du premier avril jusqu'au 30 septembre ainsi que la latitude et va définir si la région étudiée remplit les critères de chaleur nécessaires au développement des vignes. (Doutreloup et al., 2022).

Bien que la Belgique rencontre ces critères, elle est tout de même victime des gelées de printemps et ce de manière plus significative que les vignobles des régions du nord de la France (Bourgogne, Jura, Alsace, Champagne) (Doutreloup et al., 2022). Le gel de printemps est désastreux pour les rentrées économiques des viticulteurs et des arboriculteurs. En Belgique, Vanel (2021) rapporte que le vignoble des Agaises a perdu la moitié des récoltes à cause du gel en 2017. En France, le Parisien (François, 2021) rapporte que le secteur viticole a perdu 2 milliards d'euros à cause des gelées tardives, ces chiffres sont annoncés par le FNSEA (Fédération Nationale des syndicats d'exploitants agricoles). Au Chili, s'il y avait un système de prévision efficace, 20 millions de dollars par an pourraient être économisés (ces chiffres concernent les répercussions du gel sur toute l'agriculture et pas seulement les vignes) (Santibáñez et al., 1997). Les gelées de printemps sont un phénomène qui touche et endommage de nombreux fruits mais dans le cadre de ce mémoire nous allons nous concentrer sur les dommages occasionnés aux vignes et à leurs fruits. Plus proche de nous, en Suisse, le gel printanier touche durement les vignes qui représentaient 5.6% du revenu agricole en 2016 (M. Meier et al., 2018). D'autres exemples de régions touchées par ce phénomène sont donnés par Molitor (2014): en particulier au Luxembourg, en moyenne, lors des gelées de printemps le rendement des vignes peut être réduit de 39%.

Face à ce problème, deux types de protections existent, ayant un effet complémentaire : la lutte passive et la lutte active contre le gel. La lutte passive comprend la décision de l'emplacement des vignes, le choix du cépage (Poling, 2008), la prévision d'évènements gélifs via différentes méthodes que nous décrirons plus bas. La lutte active, quant à elle, comprend l'utilisation d'outils pour empêcher le gel comme décrit au paragraphe suivant.

Cette lutte active nécessite également des investissements économiques qui viennent s'ajouter aux pertes économiques directes liées à la destruction des récoltes. En l'absence de modèle prédictif fiable, les outils de défense contre le gel sont parfois déployés inutilement. Voici quelques exemples de dispositif de défense trouvés dans la littérature. Aux États-Unis, de l'huile d'amigo est aspergée sur les vignes afin de retarder le débourrement de la vigne et éviter les conséquences du gel printanier (Loseke et al., 2015). Il est aussi possible d'installer des tourelles qui vont brasser l'air afin de casser l'inversion de température (Battany, 2012). Des hélicoptères sont également utilisés aux États-Unis et en France ce qui peut protéger jusqu'à 20ha de vignes (Poling, 2008). Il y a également la technique de l'aspersion des bourgeons qui est une technique très efficace mais qui doit être utilisée en continu et qui consomme énormément d'eau. Dans son article, Poling (2008) donne les quantités pour 3 jours consécutifs de protection contre le gel : approximativement 960 000 L d'eau par ha sont nécessaires. L'aspersion repose sur le principe que l'eau en changeant d'état (lorsqu'elle passe de l'état liquide à l'état solide) va libérer de la chaleur et réchauffer le bourgeon. Enfin, une autre technique est de placer des bougies lorsqu'un risque de gel survient. Cela va permettre de réchauffer l'air autour des vignes et donc d'éviter le gel. Néanmoins, placer des bougies tous les soirs où un risque de gel est possible serait à terme trop coûteux, sauf si la prévision est suffisamment fiable. Cette action est d'autant plus contraignante que les viticulteurs doivent se lever en pleine nuit pour allumer les bougies et ainsi protéger leurs vignes.

Après cette présentation générale de la problématique, nous allons successivement aborder le phénomène du gel de la vigne, le contexte météorologique amenant ce risque de gel et enfin les outils de prédiction existants.

I.2. La vigne et sa sensibilité au gel

Voici une définition de la vigne extraite de la thèse de Madelin (2004) : « La vigne, plante vivace, est une liane comprenant plusieurs éléments : un système racinaire puissant, un tronc frêle et flexible, des feuilles et des bourgeons se développant à partir des nœuds du tronc, et des fruits sous forme de grappe ». Choissnel (2002) définit les trois éléments climatiques indispensables à la croissance de la plante : le rayonnement solaire, la température et l'eau (les

précipitations). Le rayonnement solaire va apporter de la lumière essentielle à la photosynthèse. La température va avoir un rôle important durant toute l'année que ce soit dans la croissance de la plante ou dans sa mise en danger à cause des températures qui peuvent devenir critiques à certains moments de l'année (Galet, 2000). Ces températures critiques peuvent provoquer une destruction d'organe. En hiver, une température inférieure à -15°C va détruire les tissus de la plante. En été, une température supérieure à 42°C peut brûler les grappes. Enfin, au printemps, une température inférieure ou égale à -2.2°C entraîne un gel du bourgeon. Cette température critique peut varier en fonction du cépage (Galet, 2000). Finalement, l'eau va permettre à la plante de réaliser certaines réactions biochimiques et l'absorption d'éléments nutritifs du sol (Pérarnaud & Raynal, 1991).

Les différentes températures critiques, évoquées ci-dessus, sont liées aux différents stades phénologiques du bourgeon. Le bourgeon débute par le stade du bourgeon d'hiver (celui-ci voit sa résistance au froid augmenter de novembre à janvier). Après plusieurs jours avec des températures entre 8 et 12°C , le bourgeon d'hiver se transforme en bourgeon dans le coton (ce stade apparaît à la sortie de l'hiver : au printemps) (Madelin, 2004). Lors de ce stade de croissance, le bourgeon est sensible à des températures de -2.5°C à -3°C contre les -15°C du bourgeon d'hiver. En effet, le bourgeon va se gonfler et contenir plus d'eau et va donc être plus sensible au gel (Dereuddre et al., 1993). Lorsque la première feuille est sortie, on parle alors du stade de pointe verte qui est sensible à une température inférieure à -2.2°C et lorsque que la première feuille est déployée, la sensibilité monte à -1.5° (Willwerth et al., 2014). Les trois derniers stades évoqués sont ceux qui nous intéressent dans ce mémoire. Ces valeurs servent de repère mais varient en fonction du cépage comme mentionné dans le paragraphe précédent. Ces différents stades sont visibles sur la Figure 1. D'autres stades existent mais ne nous intéressent pas dans le cadre de ce mémoire.

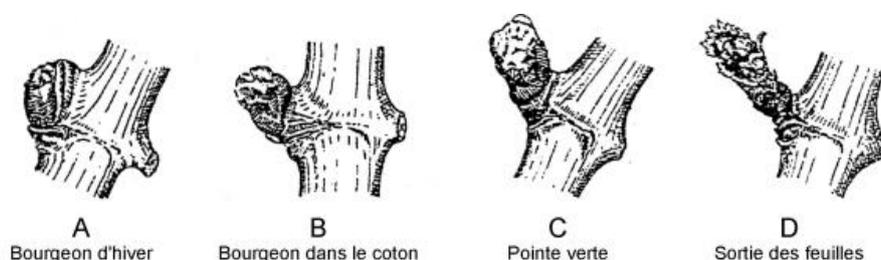


Figure 1 Repère de l'évolution des stades phénologiques du bourgeon (Baggiolini, 1952).

En plus des trois éléments climatiques indispensables à la vigne, des éléments plus locaux vont également impacter la capacité des vignes à se développer sur un territoire particulier. Tout d'abord, la topographie qui va jouer sur les variables climatiques comme l'ensoleillement avec l'altitude, l'exposition, ainsi que les formes topographiques. La végétation qui peut impacter des variables climatiques influençant la vigne. La proximité de surfaces en eau qui peut réguler les températures et augmenter l'humidité (ce qui peut favoriser certaines maladies). Enfin le sol et sa composition vont influencer la conductivité de la chaleur et l'albédo du sol (Madelin, 2004).

Parmi tous les risques climatiques, le gel reste le principal pour les vignes belges (Doutreloup et al., 2022). Le gel peut apparaître dans les espaces intercellulaires, extracellulaires ou encore dans les cellules du végétal. L'eau de ces milieux est chargée en substances dissoutes, ce qui abaisse la température de cristallisation en dessous de 0°C (Pérarnaud & Raynal, 1991). Les espaces intercellulaires seront les premiers touchés par le gel ce qui va entraîner une augmentation de la concentration de cet espace et un déséquilibre osmotique créant un flux d'eau des cellules vers l'espace intercellulaire. Si la perte d'eau est conséquente, la cellule va arrêter de fonctionner, les membranes vont se casser et les cellules vont mourir. La mort de la cellule peut aussi venir de l'action mécanique des cristaux formés (Clément, 1981; Madelin, 2004). Ce qui différencie la température critique en hiver de celle au printemps est la quantité d'eau contenue dans le végétal. En hiver, le végétal est en repos végétatif et les cellules sont moins grandes alors qu'au printemps les cellules sont actives et donc plus grandes. Le niveau de température critique dépend donc du stade phénologique de la plante (Madelin, 2004). On peut voir la phénologie de la vigne sur la Figure 2. La période qui nous concerne se situe en avril : le débourrement (le bourgeon est dans le stade de bourgeon dans le coton) comme déjà expliqué ci-dessus.

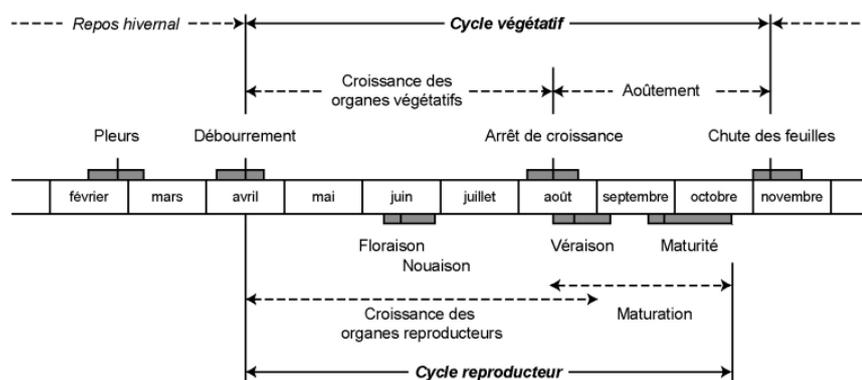


Figure 2 Les cycles végétatifs et reproducteurs de la vigne. (Reynier, 1991)

Il est à noter qu'une question débattue dans la littérature, avec le réchauffement climatique, est la suivante : est-ce que le risque de gel printanier va augmenter ? ou diminuer ? Cette question se pose car bien que le nombre de jours de gel va tendre à diminuer comme le montre l'IPCC (2021), le débourrement pourrait arriver plus tôt (Meier et al., 2018). La réponse ne semble pas être clairement définie et peut varier selon les régions (Meier et al., 2018). Dans le Luxembourg, la fréquence de gel et son impact négatif sembleraient diminuer selon 6 modèles climatiques, car la date du dernier jour de gel avancerait plus que la date de débourrement (Molitor et al., 2014). Néanmoins, une étude réalisée en Bourgogne-Franche-Comté (Gavrilescu et al., 2019) montre cette hétérogénéité spatiale (et justifie ce questionnement) car cette étude montre un résultat différent de celle citée ci-dessus : en effet les modèles phénologiques et climatiques utilisés semblent montrer que la date de débourrement avance (c'est-à-dire : a lieu plus tôt dans l'année) au moins aussi vite que le dernier jour de gel. Le risque de gel printanier est donc maintenu voire amplifié selon cette étude. Dans notre cas, le risque de gel est étudié sur la base d'un modèle climatique ce qui prend l'évolution climatique en compte et donc les modifications des jours de gel ; par contre, nous ne parlerons pas de modèle phénologique.

I.3. Contexte météorologique et local modifiant le risque de gel de la vigne

Les conditions météorologiques qui permettent l'apparition du phénomène de gel se produisent principalement lors de nuits radiatives, de situations advectives ou de refroidissement de l'air par évaporation.

Comme l'explique Beltrando (1998): « Le gel de printemps se produit lors de situations advectives (arrivée d'une masse d'air froid différente de celle qui est en place) ou radiatives

(situation atmosphérique anticycloniques, ciel clair ou peu nuageux, vents faibles) ». La température est en grande partie fonction du bilan radiatif (Choisnel et al., 1993):

$$R_n = (R_D + R_d).(1 - \alpha) - (R_T - R_A)$$

« où R_n est le rayonnement net, R_D le rayonnement solaire direct, R_d le rayonnement solaire diffus, α l'albédo du sol, R_T le rayonnement terrestre (infrarouge) proportionnel à la quatrième puissance de la température du sol d'après la loi de Stephan et Boltzman et R_A le rayonnement atmosphérique parvenant au sol (infrarouge) » (Madelin, 2004).

De ce fait, lors de nuit à ciel clair, le rayonnement terrestre est plus important que le rayonnement atmosphérique atteignant le sol. Le rayonnement net devient négatif ce qui signifie que le sol perd en énergie et donc en chaleur (Madelin, 2004). De ce fait, de l'air froid se forme dans les basses couches de la troposphère et celui-ci, plus dense que l'air chaud, va s'écouler par gravité vers les endroits de plus basses altitudes. Cette écoulement d'air le long de la pente est nommé « vent catabatique ». Même sur une surface plane, un gradient de température va s'installer avec des températures très froides près de la surface. Le refroidissement de l'air par évaporation est un phénomène existant mais minime en ce qui concerne les cas de gelées (Madelin, 2004).

Beltrando (1998) précise bien que l'air froid ne se comporte pas comme l'eau. L'air peut s'accrocher à des aspérités du relief et s'accumuler sur différents niveaux d'épaisseur en fonction de la variation du vent et des obstacles rencontrés. Il en découle qu'un élément important de l'étude des températures d'un vignoble est sa topographie : l'influence des caractéristiques locales est déterminante. Son étude a montré l'interaction entre les vents régionaux et les vents de pente. Lorsque les vents régionaux sont faibles ou modérés (jusqu'à 3 m/s à 10m du sol dans l'étude), le vent de pente peut être bloqué en bas des versants (accumulation d'air froid). Au contraire, quand le vent régional est important, il empêche les vents de pente de se former.

Madelin (2004) souligne des facteurs influençant le bilan radiatif et donc la température minimum. Tout d'abord, l'humidité de l'air d'une part, qui va absorber une partie du rayonnement terrestre et le réémettre vers le sol, et d'autre part, la concentration en humidité qui va élever la température du point de rosée qui, lorsqu'il est atteint, va libérer de la chaleur par son changement de phase. Un air humide va limiter le refroidissement radiatif. La nébulosité va absorber une partie du rayonnement terrestre et va compenser ou annuler un déficit radiatif. Le vent, lorsqu'il est assez fort, va brasser l'air et donc éviter une accumulation d'air froid

proche du sol. La topographie va impacter l'écoulement de l'air comme discuté plus tôt. Enfin le sol dont la composition et le couvert peuvent affecter la température, avec par exemple : des couverts d'herbe qui vont maintenir l'évapotranspiration au début de la nuit ce qui va créer un isolant thermique pour la surface du sol.

I.4. Systèmes de prévisions

Les prévisions de nuit de gel dans les vignobles peuvent se réaliser via différentes techniques, une description de ces différentes techniques va être donnée dans cet état de l'art.

I.4.1. Modèles utilisant des formules hygrométriques

Comme vu précédemment un refroidissement nocturne peut être dû à de l'évaporation. Cette relation va être la base des formules hygrométriques en mettant en relation le refroidissement à une variable liée à l'humidité de l'air comme la température sèche, la température humide, le point de rosée ou l'humidité relative. La relation est basée sur l'analyse de données météo en un point donné. Madelin (2004) montre par exemple, le lien entre l'humidité relative au coucher du soleil et l'écart de température ainsi qu'entre la minimale observée et le point de rosée au coucher de soleil.

Un autre exemple est la formule de Brunt. Comme le décrit Madelin (2004) cette formule repose sur « les équations de conservation de l'énergie et de transfert d'énergie par conduction, convection, rayonnement et changement de phase ». Cette formule n'est applicable que sous les hypothèses énoncées par Brunt (1941) comme d'avoir un rayonnement net constant au cours de la nuit, d'avoir un flux de chaleur latente négligeable, un ciel clair et un faible vent, un sol humide, un apport de chaleur par conduction des plus basses couches atmosphériques vers le sol négligeable et enfin la perte d'une quantité d'énergie constante (par unité de temps). Ces équations ne sont valables que dans certaines conditions atmosphériques et pour un seul endroit.

I.4.2. Machine Learning

Le Machine Learning est un procédé qui va se concentrer sur le lien entre différentes variables (input) pour expliquer un phénomène (output), dans ce cas-ci le gel dans les vignes.

Par exemple, en 2009, Salis et son équipe ont réalisé une étude sur la modélisation du gel dans les vignobles via l'utilisation d'un Neural Network. Ils ont utilisé un "computational multilayer perceptron (MLP) model" qui est une représentation mathématique d'un système

neuronal humain. Un MLP est composé de trois couches, les inputs, les hiddens et les outputs. Globalement, une donnée d'entrée est introduite dans un réseau où celle-ci sera, via différents traitements, modifiée à chaque couche puis envoyée dans la suivante jusqu'à arriver à la sortie du réseau avec une nouvelle valeur. Grâce à cette méthode, des variables indépendantes peuvent être utilisées directement par le modèle (car elles seront expressément traitées pour celui-ci) pour obtenir, par exemple dans ce cas-ci, une combinaison de facteurs expliquant la prédiction de gel.

Pour nourrir ce genre de modèle, il est nécessaire d'avoir une série de variables liées au gel. Cela peut être la température, l'humidité, le point de rosée, la vitesse du vent et la direction de celui-ci ou encore la radiation nette (Ding et al., 2019). Ces variables peuvent être obtenues de différentes manières mais principalement (dans cette étude) par télédétection afin d'avoir une information continue sur tout le terrain étudié. Le modèle va ensuite s'entraîner avec une base de données pour être capable de déterminer une relation entre la température et l'humidité (Sallis et al., 2009). La difficulté avec ce genre de modèle est que le nombre d'inputs de celui-ci devient trop élevé (par exemple 5 variables sur 6 heures vont faire 30 inputs) et que le modèle devient trop complexe. Il faut alors appliquer une simplification en passant par exemple par une classification des variables qui va déterminer à l'aide d'un algorithme l'importance de chaque variable.

Un autre exemple, est l'utilisation de machines à vecteurs de support (Ding et al., 2019). Cette approche se concentre sur la prévision du gel quelques heures avant l'arrivée de celui-ci (3 à 1 heures avant). Le principe est le même, à savoir une analyse des variables explicatives pour expliquer la variable dépendante, mais la méthode change. Tout d'abord, le modèle va être nourri par des variables indépendantes. Ensuite celui-ci va chercher le lien entre la présence de gel et la valeur des variables entrées via des classifications statistiques. En effet, l'utilisation de machines à vecteur de support va permettre de classifier les variables d'entrées selon différentes classes statistiques afin de déterminer l'apparition de gel. Le modèle a montré de bons résultats avec un haut taux de prédiction réussie pour un faible taux de fausses alertes (Ding et al., 2019).

Une des difficultés principales de ce genre de méthode où il y a du machine learning est qu'il est nécessaire d'avoir une très grande quantité de données pour nourrir en suffisance le modèle (Sallis et al., 2009). L'étude de l'équipe de Sallis (2009) n'a pas fourni de résultats très concluants, cependant un mémoire utilisant l'intelligence artificielle et les réseaux neuronaux pour prédire l'aléa de gel a été réalisé à l'Université de Liège par Adrien Leerschool (2022).

Un des deux modèles, développé dans ce mémoire, utilise également un modèle entraîné par des données afin de réaliser des prévisions.

I.4.3. Prévision du gel via des observations de terrain et modèle numérique

Une majorité de la littérature portant sur la détection d'évènements gélifs pour les vignes, étudie le vignoble Champenois. La technique de Choisnel a été utilisée et améliorée au fil des années sur le vignoble. Celle-ci va être décrite avec précision dans ce sous-chapitre. Le squelette de ce système de prévision a été posé par Choisnel et son équipe en 1993.

L'objectif du travail de Choisnel (1993) était de tenir compte des caractéristiques locales (exclusivement la topographie dans ce cas-ci) et de leur influence sur la température. Trois étapes composent ce système : la prévision de la température minimale pour la station météorologique de Reims (à 2 mètres sous abris), l'estimation de la température minimale à 2 mètres quelque part dans le vignoble (un poste) et enfin l'estimation de l'indice actinothermique à 10 cm à l'endroit du poste. Quatre postes ont été sélectionnés dans le vignoble pour l'expérience (Choisnel et al., 1993). Un encadré tiré de l'article de Choisnel, lui-même tiré d'un manuel d'agronomie, explique l'indice actinothermique Théorie 1 :

L'indice actinothermique

En matière de gelées de printemps, la température sous abri météorologique n'est pas une bonne référence. Un thermomètre placé dans les mêmes conditions que les parties les plus exposées des plantes, c'est-à-dire non protégé du rayonnement, loin de tout obstacle, donne un renseignement plus précieux que le thermomètre sous abri. Il a en outre l'avantage d'être beaucoup plus économique et plus facile à installer.

L'indice actinothermique (IA) est la valeur minimale de la température indiquée par un thermomètre exposé à l'air libre à 10 cm au-dessus du sol.

(tiré de V. Perarnaud et N. Raynal, 1991, «Agrométéorologie», Cours et Manuels n° 4, Météo-France, Toulouse).

***Théorie 1** Définition de l'indice actinothermique. Tiré de l'article de Choisnel (1993).*

La première étape qui est la prévision de la température minimale à la station va permettre de prévoir l'indice actinothermique des quatre postes. Cette prévision peut se réaliser avec l'appui d'un logiciel de refroidissement radiatif (logiciel de Cellier) si les conditions atmosphériques le suggèrent. L'étape deux et trois vont servir à trouver des tendances d'écart thermique entre les sites. Car localement, le site a une certaine topographie (interaction entre la

topographie et la température de l'air qui n'est pas représentée dans les modèles de l'époque) qui va influencer positivement ou négativement l'arrivée de gel (Choisnel et al., 1993). Ceci va permettre de réaliser une adaptation locale des températures pour chaque poste. Les tendances sont cherchées pour deux cas :

- La différence entre la température à 2 mètres du « poste » (l'endroit à estimer) et la température à 2 mètres de la station de référence à Reims.
- La différence entre la température à 2 mètres du poste et sa température actinothermique.

Le calcul de la différence du premier cas a été compartimenté selon une classification des vents (7 classes). Ces 7 classes de vents correspondent à 7 orientations du vent possibles localement. Également pour le premier cas, deux cas de nuits ont été définis : les nuits claires (avec comme critère de sélection une fraction d'insolation égale ou inférieur à 80% : ce qui donne une indication sur la couverture nuageuse pour le soir même) et dans l'autre cas, des nuits à nébulosités variables. Ces données ont donc permis une adaptation locale de la prévision. La nuance dans le type de nuit est important car le gel survient souvent lors de nuits radiatives.

Un biais systématique a été déterminé à partir du premier cas et va être intégré dans la prévision locale. Le biais pour le deuxième cas n'est pas dépendant du vent contrairement au premier.

Le logiciel de refroidissement radiatif a besoin de 3 types de données : des données météorologiques issues de prévisions générales, des observations venant de la station principale (température de l'air, point de rosée, vitesse du vent) et enfin des valeurs estimées et des variables météorologiques au coucher du soleil ou une heure après (température de l'air au coucher du soleil, température à 10 cm dans le sol CS¹, température de l'air CS+1 et température du point de rosée de l'air CS+1 (Choisnel et al., 1993). Ce dernier type de donnée sert à initialiser le modèle.

Enfin, la prévision à la station est faite chaque jour pour le lendemain. Celle-ci demande des mesures météorologiques qui doivent être prises la veille (température de l'air, température du point de rosée de l'air et vitesse du vent à 10m du sol). Lors de la mise en place de la prévision, si un cas de situation anticyclonique de ciel clair sans advection est prévu, alors un logiciel de simulation du refroidissement radiatif est utilisé et permettra de simuler le

1 CS : Coucher du soleil

refroidissement nocturne (logiciel de Cellier). Dans le cas contraire, le prévisionniste va réaliser une adaptation locale en tenant compte de la vitesse et de la direction de vent prédites pour l'ensemble de la nuit. Une fourchette de valeurs tirées de cette prévision est fournie pour l'indice actinothermique. Le lendemain une comparaison est effectuée entre la prévision de l'indice actinothermique et la mesure de celui-ci. Ce système a été expérimenté en 1988 et 1989 et donne des résultats pour les quatre stations illustré à la Figure 3.

		Avize		Bouzy	
		OBSERVATION		OBSERVATION	
P R E V I S I O N	O u i	1	1	0	0
	N o n	0	48	0	50
		Chambrecy		Mailly	
		OBSERVATION		OBSERVATION	
P R E V I S I O N	O u i	4	1	0	0
	N o n	0	50	0	50

Figure 3 Table de contingence pour les 4 postes pour une valeur de l'indice actinothermique inférieure ou égale à -2.5°C sur 50 jours répartis entre 1988 et 1989. (Choisnel et al. 1993)

Un défaut avec les prévision de Choisnel et de son équipe (1993) est la difficulté à prévoir le gel lorsque la situation atmosphérique est à la limite entre advective et radiative (Beltrando, 1998.).

Une amélioration de la méthode de Choisnel (1993) a été effectuée lors de la thèse de Madelin (2004). La base de la méthodologie est la même que précédemment mais avec un objectif de spatialiser l'information des gelées printanières et l'utilisation de nouveaux outils. Ces modifications permettent d'avoir un résultat pour l'ensemble du vignoble et non pas seulement aux postes de mesure. Le premier changement par rapport à la technique classique est la détermination de la nébulosité. Ici, les données tri-horaires de nébulosité sont utilisées (pour 21h, 0h, 3h et 6h). Si jamais trois mesures sont inférieures ou égales à deux octats, un ciel clair est déterminé, sinon il s'agit d'un ciel à nébulosité variable. Deuxième changement, l'étude est composée de 19 postes contre 4 dans l'étude précédente. La même démarche a été utilisée pour rendre compte de la spécificité locale des températures (topographie, affectation du territoire) de chaque poste.

La première étape a été d'évaluer la précédente méthodologie. La qualité de la prévision météo à la station (qui conditionne le résultat de l'estimation de l'indice actinothermique) a été évaluée ainsi que la qualité de la prévision du type de masse d'air (nébulosité et vent). La qualité de l'adaptation locale est plutôt bonne lorsque l'on compare les valeurs de mesure et d'estimation des indices actinothermiques, les mauvaises estimations étant liées à une mauvaise prévision météo à la station. Une analyse plus approfondie montre que pour certains postes, la position sur le coteau, allait entraîner une moins bonne estimation de l'indice (principalement les postes en bas de coteau). Enfin la qualité de la prévision de l'aléa gélif a été analysée, en prenant un seuil de -2°C , l'aléa a bien été prévu (du 5/4/2001 au 31/5/2001) dans 8 cas sur 14 et a donné 5 fausses alertes. Le facteur jouant le plus dans les mauvaises estimations semble être lié, encore une fois, à une mauvaise prévision à la station (Madelin, 2004).

La spatialisation des données est faite via des méthodes géostatistiques qui vont donner des surfaces de tendances et des auto-corrélations spatiales du champ de températures minimales. Les surfaces de tendances sont créés grâce aux différentes stations météo qui sont sur le territoire. Une interpolation va être effectuée via un logiciel d'interpolation spatiale SurferR. Malheureusement, aucune structure spatiale (tendance ou autocorrélation) ne ressort sur l'ensemble du printemps. Pour une échelle journalière par contre, des tendances sont observées et sont souvent liées aux facteurs locaux (comme le relief et l'occupation du sol) (Madelin, 2004). L'ensemble des relations entre la température et des éléments topographiques comme l'élévation, la pente ou encore l'exposition est établie une fois de plus grâce aux stations météos présentes sur le terrain. Une spatialisation plus précise de la température de surface en prenant en compte les facteurs locaux a été réalisée. La méthodologie est assez complexe mais elle est décrite dans la thèse (Madelin, 2004). Cette méthodologie n'a pas encore donné de solution satisfaisante.

Une simulation de l'atmosphère via le modèle Meso NH a également été réalisée afin d'appréhender le lien entre des écoulements de l'air et les températures. Le modèle n'a été utilisé que pour deux jours car sa mise en place est lourde. Le modèle sort une température, une vitesse de vent, une humidité et d'autres variables climatiques avec une résolution de 250m dans un carré de 15km^2 pour la nuit du 16 au 17 mars 2003. Les images obtenues mettent en évidence des phénomènes locaux comme les brises de pente ou encore l'accumulation d'air froid en bas des coteaux (Madelin, 2004).

Nous voyons bien à travers cet exemple, le défi que représente la spatialisation de l'information ainsi que la prédiction d'évènements gélifs dus aux caractéristiques locales de

chaque endroit d'un vignoble. Un des deux modèles de ce mémoire appartient à cette catégorie de prévision. Via un modèle atmosphérique régional et des observations de terrain, le but va être de prévoir de manière la plus précise possible les événements gélifs au niveau du fruit de la vigne.

I.4.4. Prévision spatiale de zone de gel tardif

Le but, ici, est de connaître l'étendue de la zone de gel afin de pouvoir informer les vignerons du territoire qu'ils pourront exploiter. En effet, des zones auparavant exploitables peuvent ne plus l'être, mais l'inverse est aussi possible. Il ne s'agit donc pas de prévision du gel au jour le jour mais bien d'une description de l'étendue du territoire cultivable dans les futures conditions climatiques (Gobbett et al., 2020). Les données nécessaires sont des données satellites, de stations météo et de modèles climatiques. Pour lier ces variables au gel, les chercheurs utilisent la régression multivariée par spline adaptative (MARS) qui est une forme de régression proche des régressions linéaires classiques.

MARS va permettre de déterminer les températures les plus froides de la nuit sur la région étudiée. Cette information va être montrée sur une carte d'élévation dont la résolution est de 30 mètres. Des variables comme NDVI (Normalised Difference Vegetation Index) ou la température de surface du sol sont utilisées (obtenues par satellite) ainsi que l'altitude (et toutes les informations que l'on peut en tirer : pente, ensoleillement etc...) (Gobbett et al., 2020). On peut voir une nouvelle fois l'importance de l'altitude dans le risque de gel des vignes.

Une fois la formule ajustée sur des observations passées, la formule va être appliquée à des projections climatiques. Dans une étude récente, en Australie, la découverte des chercheurs est que sur la région étudiée (la vallée de Yara), la ligne de la zone de gel va descendre en altitude : ce qui va permettre aux vignerons d'utiliser les terrains au-dessus de cette zone, en théorie sans danger (Gobbett et al., 2020).

I.5. Objectifs de ce mémoire

Comme nous avons pu le voir, la prévision du gel printanier est essentielle pour les vignerons et est compliquée à effectuer. Par le passé, une première tentative avait été faite pour prédire le gel depuis le modèle atmosphérique régional (MAR) sur un vignoble de Sprimont. Toutefois, des événements gélifs étaient manqués par ces prévisions.

Ce mémoire a pour but de proposer un outil de prévision optimisé via des observations de terrain, les sorties du MAR ainsi que l'expertise d'acteurs du monde viticole de Liège. Si la

prévision du gel peut être effectuée, les vignerons pourront s'organiser pour la disposition de bougies ainsi que la gestion de la main d'œuvre. Ceci leur permettra de perdre moins de temps et d'argent. Ce mémoire est enrichi par l'intervention de ces professionnels afin de développer l'outil le plus efficient pour eux.

Pour se faire un système de prévision basé sur les sorties du Modèle Atmosphérique Régional (MAR) va être construit et testé. Ce système est centré sur l'adaptation empirique d'une formule qui permet de donner la température de l'air à partir de la température d'une puce positionnée à hauteur des 1^{er} bourgeons de la vigne. L'adaptation empirique de la formule sera obtenue en comparant les valeurs données par les puces captant la température sur les vignes durant les années 2021-2022 à Sprimont et celles générées par le MAR forcé par ERA5 avec des sorties horaires sur l'année 2021-2022, en faisant varier différents paramètres non fixés de la formule.

De plus, suite à une interaction avec des personnes travaillant dans la viticulture, et pour tenter d'optimiser la prédiction du gel au niveau du fruit de la vigne, nous allons tester des éléments d'alerte complémentaires proposés par les experts viticoles. Ceux-ci (avec plusieurs variables pouvant impacter la température) seront intégrés dans un modèle de régression logistique multivariée avec les sorties du MAR ajustées ou non. D'autres régressions logistiques avec des sorties du MAR liées à l'apparition du gel seront testées.

Dans un premier temps, ce mémoire aborde les méthodologies employées tant d'un point de vue matériels et protocole de mesure que descriptions des formules et des méthodes statistiques utilisées. Ensuite, les résultats sont présentés accompagnés de discussions. Le mémoire se clôture finalement par des conclusions et perspectives.

II. Méthodologie et matériel

II.1. Enregistrement et traitement des données de température au niveau de la vigne

II.1.1. Enregistrement des températures au niveau du fruit des vignes

Nous voyons les deux vignobles au sud du centre du village de Lincé (Carte 1). Les deux vignobles ont été traités comme étant un seul et même vignoble durant ce mémoire, dès lors, nous ne parlerons plus que d'un seul vignoble. Dix-sept puces ont été placées dans ce vignoble de Sprimont au niveau du fruit des vignes (+/-50-60 cm du sol) de 2021 à 2022 (Carte 2). Ce vignoble s'étend de 235m à 255m d'altitude. Les caractéristiques des températures enregistrées par les puces sont visibles au Tableau 1. Ces puces ont été placées à l'air libre et proches du sol, contrairement aux stations météo. Leur emplacement est visible sur la Carte 2. Ces puces ont enregistré la température toutes les 10 ou 15 minutes (réglable par l'utilisateur). Une fois pleines, celles-ci doivent être récupérées afin de pouvoir en lire les données.

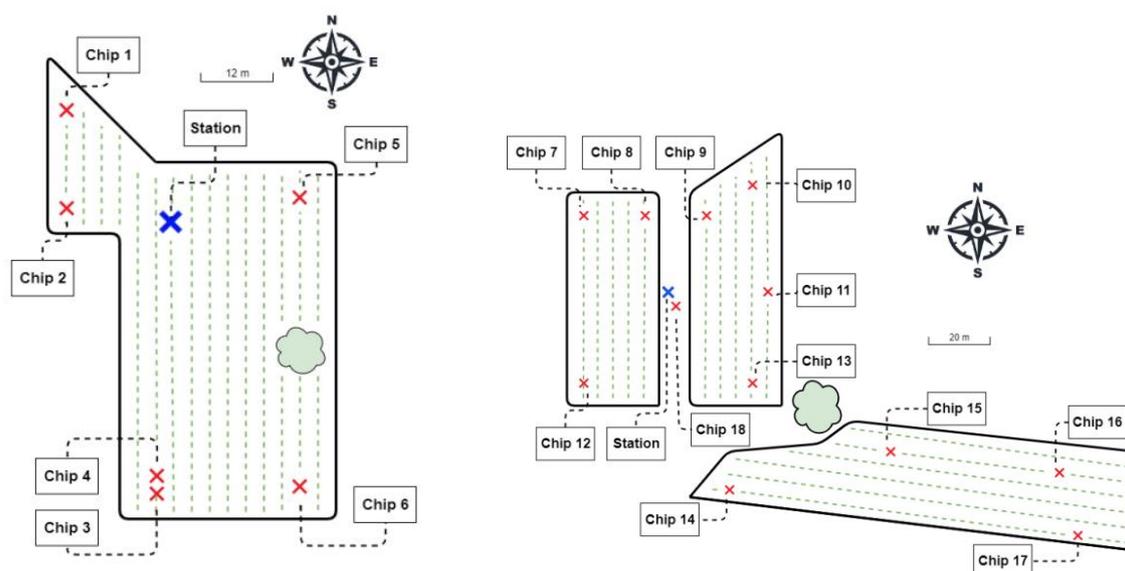


Carte 1 Localisation des vignobles dans le village de Lincé. Le tracé en tiret rouge permet de mieux localiser la zone des vignobles sur la carte. (Atlas de Belgique, 2023; WalOnMap, 2023)

Détails des puces du vignobles

Caractéristique de la puce	Résolution	Étendue des valeurs	Précision
Température	0.06 °C	[-40 °C ; 80 °C]	0.5 °C

Tableau 1 Caractéristiques des puces placées dans les vignobles au niveau des fruits des vignes.



Carte 2 Emplacement des puces dans les vignobles. Vignoble Henri Simon à gauche et vignoble Village à droite, deux vignobles de Sprimont dans le village de Lincé. Selon Leerschool (2022).

II.1.2 Traitement des observations

Étant donné que les observations ont été réalisées sur des pas de temps différents (10 ou 15 minutes), qu'elles ont parfois eu des valeurs aberrantes ou que les puces ont été parfois hors service, un traitement des données obtenues par les puces a été obligatoire afin de pouvoir les utiliser.

Tout d'abord une analyse du jeu de données a été effectuée en regardant pour chaque puce le nombre d'observations par heure et le nombre d'observations sans les données aberrantes. L'analyse a été effectuée uniquement sur les données de nuit car celles-ci représentent notre période d'intérêt. La nuit est définie par les heures où le rayonnement solaire² (SWD : short wave downward) est inférieur à 10 Wm^{-2} . Pour éliminer les données aberrantes,

² Cette valeur a été fixée arbitrairement afin de représenter le lever du soleil.

une limite inférieure de $-20\text{ }^{\circ}\text{C}$ et une limite supérieure de $45\text{ }^{\circ}\text{C}$ ont été placées afin d'éliminer les valeurs ne rentrant pas dans cette fourchette. Étant donné qu'on a travaillé en données horaires, pour se calquer sur les sorties du MAR, le nombre d'observations par heure a été évalué afin de savoir si l'heure allait contenir suffisamment d'informations pour être pertinente ou non. Un premier set de données a ainsi été constitué. Ces données ont été analysées de manière à détecter des données aberrantes et définir un set de données optimal pour créer les modèles prédictifs. Le Tableau 2 montre les résultats principaux de cette analyse et notamment la proportion de valeurs aberrantes. Cette analyse montre la nécessité d'utiliser un set de données nettoyées.

Qualité du jeu de données des puces

	Données disponibles (%)	Données disponibles - valeurs aberrantes (%)	Données avec au minimum 4 observations par heure (%)	Valeurs aberrantes (%)
PUCE 1	90.3	90	90	0.2
PUCE 2	91.5	91.5	91.4	0.1
PUCE 3	58.3	58.3	58.3	0
PUCE 4	91.4	91.4	91.4	0
PUCE 5	87.3	87.3	87.2	0.1
PUCE 6	91.4	91.4	91.4	0
PUCE 7	80.8	80	79.9	0.9
PUCE 8	82.5	82.1	81	1.5
PUCE 9	91.4	91.4	91.3	0.1
PUCE 10	87.3	87.3	87.3	0
PUCE 11	91.4	91.4	91.3	0.1
PUCE 12	82.6	82.6	82.5	0.1
PUCE 13	90.3	90.3	90.2	0.1
PUCE 14	91.4	91.4	91.3	0.1
PUCE 15	70.6	70.6	70.5	0.1
PUCE 16	78.5	78.5	78.4	0.1
PUCE 17	91.4	91.4	91.4	0

Tableau 2 Observations de la qualité du jeu de données. La première colonne correspond à la quantité de données disponible sur le pas de temps du MAR (2021-01-01 1h à 2022-06-30 23h) contre le pas de temps des observations (2021-02-07 à 2022-06-25). La deuxième colonne s'est vue amputée des valeurs extrêmes n'étant pas comprise entre [-20°C ; +45°C]. La troisième colonne est composée des heures d'observation avec au moins 4 observations par heure. La dernière colonne est la différence entre la première et la troisième colonne.

Dans ce contexte, nous avons décidé d'utiliser les données générées et nettoyées par Leerschool (2022). Les observations de ce jeu de données avaient toutes été remises sur un pas de temps de 10 minutes (pour éviter le problème des différents pas de temps de 10 et 15 minutes) en prenant une moyenne pondérée des points qui existaient déjà. Les enregistrements des puces qui étaient en heure d'hiver UTM+2 et d'été UTM+1 ont été synchronisés en heure universelle afin d'être exprimés de la même manière que le MAR.

Pour tenir compte de l'erreur de mesure des puces qui est de 0.5 °C, la méthode de Monte Carlo a été utilisée. Pour chaque observation (toutes les 10 minutes) et pour chacune des 17 puces, 100 valeurs ont été générées en comprenant la température observée en ajoutant une valeur aléatoire entre -0.5 °C et +0.5 °C. Ensuite la moyenne et l'écart type pour chaque heure ont été calculés à partir des 100 valeurs aléatoires des 6 pas de temps de 10 minutes sur l'heure des 17 puces (Schéma 1).

Finalement, étant donné que les températures qui nous intéressent sont les plus froides, le jeu de données final d'observations avec lequel nous avons décidé de travailler pour chaque heure est le suivant : la moyenne calculée pour chaque heure à laquelle il a été retiré deux fois l'écart type. Ceci permet de prendre en compte les températures les plus froides mesurées par les puces sans avoir de valeurs aberrantes (qui seraient trop basses par rapport à la moyenne).

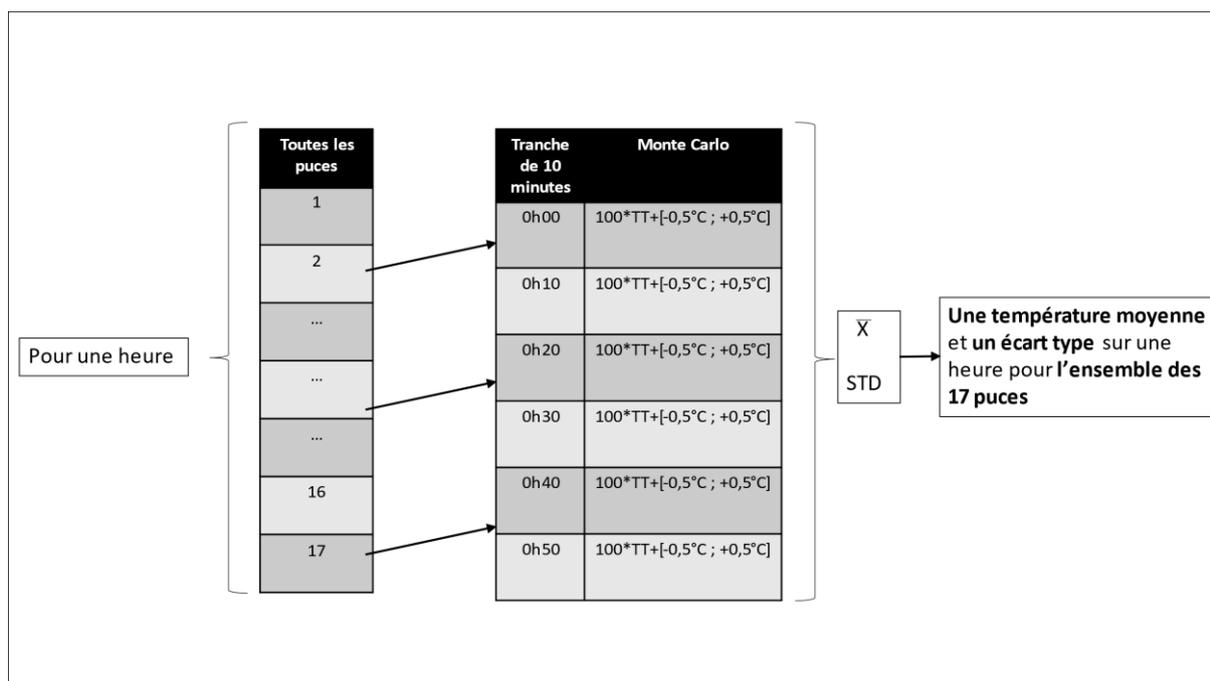


Schéma 1 Explication du calcul de la moyenne et de l'écart type en utilisant Monte Carlo pour une heure.

II.2. Variables climatiques générées par modélisation (MAR)

Les variables climatiques utilisées dans ce travail proviennent du modèle MAR. Elles comprennent notamment les températures, la vitesse du vent et la présence de nuages dans l'atmosphère. Le MAR est un modèle atmosphérique régional à 3 dimensions qui est couplé à un schéma de transfert à une dimension entre la surface, la végétation et l'atmosphère (De Ridder & Gallée, 1998). Bien qu'au départ créé pour les zones polaires (Fettweis et al., 2013), il a été validé pour des régions tempérées telle la Belgique (Doutreloup et al., 2022; Fettweis et al., 2017).

Le modèle a été forcé par les ré-analyses ERA5 (Hersbach et al., 2020) qui ont une résolution horizontale de $0.25^{\circ}C$. Les champs qui ont été forcés sont la température de l'air, la pression, le vent, l'humidité spécifique ainsi que la température de surface de l'eau de mer.

Les sorties du modèle sur la Belgique sont à une résolution de 5km avec une résolution temporelle horaire du premier janvier 2021 au 30 juin (2022). Monte Carlo a été appliqué sur les sorties de température du MAR en considérant une erreur de $[-1^{\circ}C ; 1^{\circ}C]$ avec mille valeurs générées aléatoirement et en prenant la moyenne pour la température minimum ou moyenne du MAR.

II.3. Détermination de la température au niveau du fruit de la vigne depuis la température de l'air déterminée par le MAR

Les températures enregistrées et décrites au point 2.1 correspondent aux températures des puces utilisés et ont été assimilés à la température du fruit de la vigne. Afin d'approcher les températures au niveau des puces (T_T) (ne se trouvant pas dans les conditions d'une station météo classique) à partir des températures générées par le MAR (T_a), nous avons fait varier une formule (Guyot, 2013) comprenant une série de variables et de coefficients prédéfinis (h , α , $E_{S\downarrow}$, ε , $E_{L\downarrow}$, σ , θ). Voici la formule originale :

$$T_T - T_a = \left(\frac{1}{h} [(1 - \alpha) \cdot E_{S\downarrow} + \varepsilon \cdot (E_{L\downarrow} - \sigma \cdot T_T^4)] \right)$$

Et ci-dessous, la formule modifiée :

$$T_T = \left(\frac{1}{h} [(1 - \alpha) \cdot E_{S\downarrow} + \varepsilon \cdot (E_{L\downarrow} - \sigma \cdot (T_a + \theta)^4)] \right) + T_a$$

T_T : Température du puce *Kelvin*

T_a : Température de l'air *Kelvin*

h : Coefficient d'échange convectif du puce thermométrique $W \cdot m^{-2} \cdot K^{-1}$

α : Albédo du puce

$E_{S\downarrow}$: Eclairement du puce par le rayonnement d'origine solaire $W \cdot m^{-2}$

ε : Coefficient mathématique

$E_{L\downarrow}$: Eclairement du puce par le rayonnement de la grande longueur d'onde en provenance de l'environnement $W \cdot m^{-2}$

σ : Constante de Stefan-Boltzmann $W \cdot m^{-2} \cdot K^{-4}$

θ : Nombre entier relatif *Kelvin*

Le but de cette étape a été d'obtenir comme température de sortie (T_T) de la formule modifiée, une température la plus proche possible de celle mesurée par les puces au niveau du fruit de la vigne durant les deux années écoulées (2021-2022) tel que décrit au point 2.1. La température de sortie (T_T) a été appelée « température actinothermique simulée (TAS)» pour

tout le reste du mémoire. Hormis les constantes, l'ensemble des variables ont été sorties du MAR. Les démarches suivantes ont été effectuées :

- deux températures de l'air ont été testées dans la calibration de la formule : la température moyenne du MAR³ (moyenne des températures sur l'heure) et la température minimum du MAR (la température la plus basse modélisée sur l'heure).
- trois paramètres ont varié successivement : ε , h , et θ . L'albédo (α) n'a pas varié car il ne porte que sur l'ensoleillement et nous nous concentrons exclusivement sur les données de nuit ($E_{L\downarrow} < 10$). Tout d'abord, ε a varié de 0.1 à 0.9, ensuite h a varié de 1 à 100 et enfin θ a varié de -5 à +5.

Afin de vérifier l'efficacité des variantes de la formule⁴, des statistiques ont été calculées toutes les heures depuis une table de contingence. Une table de contingence est un tableau qui compare deux jeux de données, en général des observations et des prévisions (dans notre cas : la température observée au niveau des puces et la température actinothermique simulée (sortie de la formule modifiée). Une table de contingence demande un critère afin de séparer les bonnes prévisions des mauvaises. Ici, le critère a été que la température observée soit inférieure à 0°C (donc gel potentiel). Ce tableau nous donne alors les vrais positifs (événement présent dans les observations et prédit par les prévisions), les faux positifs (événement absent dans les observations et prédit par les prévisions), les vrais négatifs (événement absent des observations et non prédit par les prévisions) et les faux négatifs (événement présent dans les observations et non prédit par les prévisions). Une table de contingence a déjà été présentée dans l'introduction (Figure 3).

A partir de cette table de contingence trois statistiques ont été calculées : la probabilité de détection, le taux de fausses alarmes et la précision. La probabilité de détection (POD) permet de savoir combien d'événements (gel) ont été prédits correctement. Le taux de fausses alertes (FAR) donne la proportion d'événements prédits qui ne correspondent pas à des événements observés (Doswell III et al., 1990). L'accuracy (ACC) a également été évaluée.

3 Sauf si l'inverse est précisé, dans ce mémoire, c'est toujours la température « classique » du MAR qui est utilisée c'est-à-dire la température de l'air.

4 Il est à noter que lors de ce mémoire, aucun intervalle de confiance n'est directement calculé : ceci est issu d'une discussion avec les vignerons qui voulaient uniquement savoir si il allait geler ou non sans avoir de degré de confiance. Cependant, une évaluation de la qualité des prévisions est donnée par POD et FAR.

Suite à l’entrevue avec les spécialistes du monde viticole, il a été décidé que le seuil maximal de FAR était de 30%.

$$\begin{aligned} \text{Probabilité de détection:} & \quad \frac{\text{Vrai positifs}}{\text{Faux négatifs} + \text{Vrai positifs}} \\ \text{Taux de fausses alertes :} & \quad \frac{\text{Faux positifs}}{\text{Faux positifs} + \text{Vrai négatifs}} \\ \text{Accuracy :} & \quad \frac{\text{Vrai positifs} + \text{Vrai négatifs}}{\text{Total}} \end{aligned}$$

En plus de la table de contingence, le RMSE (Root Mean Square Error) centré ainsi que le biais des moyennes ont été calculés.

Une série de cas contrôles a en outre été créée pour servir de comparatif aux performances de la TAS et pour sélectionner la sortie optimale du MAR : température de sortie du MAR non modifiée, température de surface du MAR, température du pixel ayant l’altitude la plus proche du vignoble étudié (parmi les 8 pixels entourant le pixel le plus proche du vignoble), la température du pixel le plus haut et le plus bas en altitude. L’altitude moyenne du vignoble se situe à 245m d’altitude.

II.4. Analyses complémentaires basée sur l’expérience des acteurs de terrain

Après discussion avec des acteurs du milieu viticole, deux autres éléments de prévision (alerte) en plus de la température actinothermique simulée ont été proposés. Il a été remarqué que, sur le terrain, le gel survenait souvent quand il y avait une faible vitesse de vent ainsi qu’une inversion de température verticale. Et, en effet, un vent faible et une inversion de température correspondent bien à une nuit radiative ce qui est un des cas possibles pour l’apparition de gel nocturne (Madelin, 2004).

Trois variables (vitesse du vent à 2m, vitesse du vent à 10m, gradient de température entre 3 et 10m⁵) ont été extraites du MAR afin de regarder si lorsque ces événements se produisaient (vent faible et inversion de température), les puces avaient une température inférieures à 0°C ou pas. Le vent a été observé à 2 et 10m car le vent à 2m est beaucoup plus influencé par la surface que celui à 10m.

⁵ Le gradient de température n’est calculé qu’entre 3 et 10 mètres car la modélisation de la température par le MAR commence à 2 mètres et pas plus bas.

II.4.1. Approche visuelle de la distribution des données

Tout d'abord, une approche visuelle a été utilisée en regardant la fonction de distribution cumulative des trois variables lorsque du gel était détecté dans les observations et lorsque qu'il ne l'était pas.

II.4.2. Approche comparative statistique des données

Le test de Shapiro est le test le plus puissant pour montrer la normalité d'une distribution d'observation par rapport à d'autres tests comme celui de Kolmogorov-Smirnov ou encore Anderson Darling (Mason & Graham, 2002). Afin de tester la normalité de la distribution des 3 variables (vitesse du vent à 2m, vitesse du vent à 10 m, gradient de température) lors de gel ou lors de non gel, un test de Shapiro a été effectué. A noter que pour deux des trois variables où le test est effectué, l'échantillon est plus grand que 5000 ce qui pourrait poser problème avec l'exécution de ce test (Mohd Razali & Bee Wah, 2011). La variance de l'échantillon a été comparée à celle d'un échantillon théorique suivant la distribution de la loi normale. Après vérification de la normalité de la distribution et en fonction de celle-ci, un test comparatif de type Mann Whitney ou T de Student a été réalisé, de manière à comparer les valeurs pour les températures inférieures à 0°C (gel) et celles supérieures à 0°C (absence de gel). Globalement, le test de Mann Whitney compare les tendances statistiques des deux échantillons c'est-à-dire la moyenne et l'écart type de ceux-ci afin de voir si ils viennent d'une même population ou non (Mason & Graham, 2002).

II.5. Régression logistique

Une dernière technique a été utilisée pour tenter d'améliorer les résultats de prévision : la régression logistique. Celle-ci a été calculée avec de nombreuses variables indépendantes citées dans la littérature comme ayant un impact dans l'apparition d'événement de gel (Ding et al., 2019; Madelin, 2004b; Sallis et al., 2009) : la TAS, la température minimum brute du MAR, le gradient de température, la vitesse du vent à 2m et à 10m, l'opacité optique des nuages (COD), la présence de nuages à différents niveaux de l'atmosphère (CD : couverture nuageuse en basse altitude, CM : couverture nuageuse en altitude moyenne, CU : couverture nuageuse en haute altitude), la pression de surface au niveau de la mer (SLP), la pression de surface (SP), l'humidité spécifique⁶ (QQ), la température du point de rosée (TD), le rayonnement infrarouge (LWD, LWU) et enfin la présence de précipitations (PP : neige + pluie).

II.5.1. Tests de corrélation

Tout d'abord un test de Shapiro a été réalisé afin de savoir si les variables indépendantes étaient distribuées de manière normale ou non. Comme effectué précédemment, les variables ont été séparées en observation lors de période de gel et de non gel. Ici il n'y a plus de variable de plus de 5 000 observations car l'ensemble des observations ont été calquées sur le vent à 10 mètre qui possède un mois de modélisation (les modélisations s'arrêtent en mai) de moins que les autres variables.

En fonction des résultats du test de Shapiro, une corrélation de Spearman ou de Pearson a été effectuée afin de connaître la corrélation entre les variables indépendantes et la variable dépendante. La corrélation de Spearman qui renvoie une valeur (ρ) variant entre -1 et 1, a été utilisée. Plus la valeur est proche de $|1|$ et plus les variables dépendante et indépendante sont corrélées. Ce test permet de déjà donner une indication sur le pouvoir prédictif de la variable indépendante.

II.5.2. Préparation et application de la régression logistique

La première étape a été de regarder si les variables indépendantes n'étaient pas corrélées entre elles sinon, la condition pré-requise pour effectuer la régression logistique dans les meilleures conditions n'était pas rencontrée (Stoltzfus, 2011). Pour se faire un tableau qui montre la corrélation des variables indépendantes entre elles, a été créé. Le but de cette

⁶ Ici l'humidité spécifique a été utilisée à la place de l'humidité relative pour éviter d'avoir une corrélation de base trop importante avec la température.

régression a été d'expliquer la variable dépendante qui suit la loi de Bernoulli : c'est-à-dire ici, la présence ou l'absence de gel au niveau des vignes.

Lorsqu'on effectue une régression logistique, une partie des données est utilisée pour créer la régression (l'affiner, l'entraîner : training set) le reste sert de test (test set) pour vérifier les résultats de la prévision. Que ce soit pour les test sets ou les training sets, ceux-ci sont composés de les valeurs des sortie du MAR énoncées plus tôt ainsi que la détection de gel observée dans le vignoble pour tous les pas de temps (0 si il n'y en a pas et 1 si il y en a). Étant donné que notre jeu de données de base (5748 observations avec 868 observations lors de gel et 4880 sans gel) possède beaucoup moins d'observations de gel que d'observations de non gel, pour le jeu de donnée qui a servi à entraîner le modèle (l'année 2021), les mois où il n'y a pas de gel ont été enlevés. Ensuite pour obtenir une répartition égale d'événements de gel et de non gel (50/50), 1720 observations aléatoires ont été enlevées dans les mois restants. Ceci permet d'avoir un modèle, a priori, plus robuste pour la prévision de gel et sert de rééchantillonnage. La technique de sous-échantillonnage (expliquée ci-dessus) a été utilisée parmi d'autre techniques de rééchantillonnage car il y avait assez d'observations pour pouvoir l'appliquer et que cela représentait l'approche la plus simple.

Lorsqu'une régression logistique est appliquée, il faut définir un seuil de probabilité qui va définir si le résultat de la régression (donc la prévision) est 1 (gel) ou 0 (non) gel. Par défaut ce seuil se situe à 50%, une probabilité supérieure à 50% donnera du gel et une inférieure n'en donnera pas. Afin d'augmenter les performances de la régression dans la détection de gel, quatre seuils ont été analysés : un seuil où la POD était de 99%, un autre avec une POD de 90%, un autre avec le plus haut Youden Index et enfin un seuil obtenu lorsque la précision est la plus élevée possible (la précision correspond à 1-FAR) tout en gardant une POD la plus élevée possible. En pratique, il a été regardé sur le training set quel seuil allait être obtenu en ne conservant que celui qui allait donner une précision supérieur à 90% et une POD la plus élevé possible. Ces seuils ont été calculés sur le training set (données de 2021) puis ensuite appliqués au test set (données de 2022). L'indice de Youden est défini comme suit (Fluss et al., 2005) :

$$J = \max_c (Se_c + Sp_c - 1)$$

Avec Se étant la sensibilité et Sp étant la spécificité du modèle. L'indice c correspond au seuil choisi. Plus J est proche de 1 et plus la discrimination entre les deux populations (gel et non gel) est bonne. J varie entre 0 et 1. Cet indice permet de définir un seuil à priori optimal (Fluss et al., 2005).

Traditionnellement, pour analyser l'efficacité de la régression logistique une courbe ROC (Receiver operating characteristic) est utilisée ainsi que l'aire sous la courbe dite AUC. La courbe ROC est représentée avec en abscisse le FPR (False Positive Rate). Le FPR est une nouvelle statistique dans ce mémoire et correspond à :

$$FPR = \frac{FalsePositive}{FalsePositive+TrueNegative}.$$

L'ordonnée est définie par le « True Positive Rate » : ce qu'on a appelé POD dans cette étude. Une AUC supérieure à 70% est un modèle applicable qui va correctement discriminer les événements de gel et de non gel, alors qu'une AUC équivalente à 50% revient à avoir autant de chance de tomber sur un vrai positif que sur un faux positif (donc ce n'est pas pertinent) (Broeckx et al., 2019). En plus de l'AUC, la POD, le FAR et la table de contingence ont été représentés.

Dans le cadre de ce mémoire, en plus de la ROC curve, la Precision-Recall curve a été utilisée. Cette courbe permet également de représenter la performance du modèle avec en abscisse « Recall » ce qui correspond à la POD et la « Precision »⁷ qui est égale à 1-FAR⁸. Avec la Precision-Recall curve, une AUC est également calculée et montre également la performance du modèle (Sofaer et al., 2019). Si l'AUC de la Precision-Recall curve a un score de 1, c'est un modèle parfait et plus le score est proche de 0 plus le modèle est mauvais. Ici la Precision-Recall curve est plus intéressante car les statistiques qui nous intéressent sont POD et FAR.

Les deux courbes ont été représentées avec la régression logistique (entraînée au préalable sur le training set (2021)) appliquée sur le test set (2022). Les différents seuils trouvés ont été illustrés sur ces courbes.

Pour l'application de notre régression logistique cinq stratégies ont été utilisées :

- 1) Une régression logistique avec comme seule variable indépendante la TAS afin de servir de cas de référence.
- 2) Une autre régression servant de cas de référence avec la température brute du MAR comme seule variable indépendante.

⁷ Attention, ici, « Precision » est différent de l'accuracy (ACC) de la partie II.3.

⁸ Ici, il y a plusieurs appellations pour une même statistique pour coller aux codes d'un certain contexte dans le monde scientifiques.

- 3) Une régression avec les trois variables suggérées par les spécialistes du monde agricole : la température actinothermique simulée, vitesse du vent à 10m et gradient de température.
- 4) Une régression où les variables indépendantes ont été sélectionnées par LASSO, qui est une méthode de contraction des coefficients (Least Absolute Shrinkage and Selection Operator), parmi toutes celles citées dans l'introduction du chapitre sur la régression logistique⁹. En effet, la régression de LASSO permet de faire une pré-sélection lors d'une régression linéaire avec beaucoup de variables indépendantes en donnant un coefficient allant de -1 à 1. Si la variable obtient une valeur de 0, elle est à enlever (L. Meier et al., 2008). Afin d'avoir le meilleur résultat possible il faut enlever les variables les plus corrélées entre elles. La température minimum du MAR a été conservée car elle avait la plus grande corrélation avec la variable dépendante.
- 5) Enfin, une régression logistique a été réalisé avec les variables donnant le meilleur modèle et avec un dataset non équilibré (sans faire le traitement expliqué en II.5.2.) afin de voir l'utilité de cette démarche et si nous n'étions pas passés à côté d'un modèle plus intéressant.

II.6. Comparaison des deux meilleurs modèles issus de la régression logistique et de la formule de Guyot modifiée (TAS)

Afin de départager les deux types de modèle utilisés dans ce mémoire, à savoir la modification de la formule de Guyot donnant la TAS et la régression logistique, une comparaison finale via des tables de contingence avec les statistiques qui y sont associés a été réalisée sur l'année 2022 pour la détection de gel.

II.7. Analyse de la qualité des prévisions par nuit sur une année

Toutes nos analyses se basent sur des données horaires, or les vignerons n'ont pas un besoin crucial de savoir à quelle heure précisément il va geler, mais bien de savoir s'il y aura

⁹ La régression logistique dont les variables sont sélectionnées par LASSO est appelée régression de LASSO pour tout le reste du mémoire.

du gel ou non durant la nuit¹⁰. Dans cette section, lorsqu'on parle de nuit, on fait référence à la période allant du coucher du soleil le jour « j » jusqu'au lever du soleil le jour « j+1 », période où l'ensoleillement est faible (SWD<10). Afin de vérifier la qualité des prévisions par nuit et ce sur une année entière (2022), de nouveaux calculs ont été effectués. Ces calculs ont été réalisés avec comme modèle pour la prévision de gel, la température actinothermique simulée (issue de la formule de Guyot modifiée), la régression logistique la plus performante (LASSO) et également la température minimum du MAR afin d'avoir un cas de référence. La détection du gel par la prévision ou l'observation, s'est vue attribuer la valeur de 1 et une non détection la valeur de 0. Pour chaque heure, la différence entre la prévision et l'observation a été analysée et ensuite, pour chaque nuit, le nombre de prévisions correctes, de prévisions incorrectes et d'événements gélifs manqués ont été comptabilisés et représentés sous forme graphiques.

Dans cette analyse pour chaque heure : une prévision de gel correcte correspond à un gel observé et prévu par le modèle, une prévision de gel incorrecte correspond à une fausse alerte (le modèle prévoit du gel mais les observations n'en détectent pas), un événement gélif manqué correspond à un gel observé mais pas prévu par le modèle. Ensuite, l'ensemble des prévisions sur une nuit a été analysée, une nuit de gel correctement prévue correspondant à au moins une prévision horaire de gel correcte, une nuit de gel incorrectement prévue correspondant à une nuit avec uniquement des fausses alertes ou des prévisions horaires manquées et enfin une nuit de gel manquée correspondant à une nuit avec uniquement des prévisions horaires d'événement gélif manqué.

II.8. Comparaison du meilleur modèle avec les résultats de la littérature

Enfin, une comparaison des résultats du meilleur modèle (régression logistique LASSO) a été réalisée avec les résultats obtenus dans la littérature en tenant compte des différences entre ces études. Ces comparaisons ont été faites avec Choisnel et Madelin (2002; 2004)

¹⁰ Ici la nuit fait bien référence à la période allant du coucher du soleil le jour « j » jusqu'au lever du soleil le jour « j+1 ».

III. Résultats et discussions

III.1. Détermination de la température actinothermique simulée (TAS)

Nous présentons ici les résultats obtenus avec les valeurs des paramètres générant les modèles les plus performants. L'ensemble des autres résultats, générés par les variations des valeurs des paramètres de la formule de Guyot modifiée, est repris en annexe (Tableau 22).

Température actinothermique simulée, utilisant la température minimum du MAR comparée avec la température minimum des puces

	Corrélation	RMSE	BIAIS	RMSE centré	POD	FAR	ACC	Albedo	ϵ	h	θ	
Réf	0.950	2.881	2.213	1.844	0.385	0.064	0.907					Toutes obs
	0.869	2.530	1.837	1.739	0.404	0.109	0.829					2021-M-A-M
	0.881	3.109	2.549	1.780	0.345	0.024	0.885					2022-M-A-M
Cas 1	0.947	2.009	0.603	1.913	0.624	0.225	0.919	0.2	0.9	11.0	-1.5	Toutes obs
	0.864	1.833	-0.027	1.833	0.734	0.224	0.874	0.2	0.9	11.0	-1.5	2021-M-A-M
	0.870	2.096	0.272	2.078	0.741	0.301	0.900	0.2	0.9	11.0	-1.5	2022-M-A-M
Cas 2	0.947	1.954	0.398	1.910	0.656	0.243	0.919	0.2	0.9	11.0	-1.0	Toutes obs
	0.864	1.845	-0.223	1.833	0.783	0.243	0.876	0.2	0.9	11.0	-1.0	2021-M-A-M
	0.870	2.077	0.071	2.076	0.810	0.299	0.907	0.2	0.9	11.0	-1.0	2022-M-A-M

Tableau 3 Sélection des meilleurs résultats sur toutes les variations d' ϵ , h et θ pour la température minimum du MAR (avec Monte Carlo) et la température minimum des puces. Le cas de référence correspond à une sortie du MAR non modifiée par la formule de Guyot, les cas 1 et 2, aux meilleurs résultats obtenus avec des valeurs différentes des paramètres (ϵ , h, θ). Pour chacun des cas les trois périodes d'intérêts ont été représentées : l'ensemble des deux ans, le printemps (Mars, avril, mai) 2021 et le printemps 2022. Le meilleur résultat est mis en évidence en vert foncé. L'ensemble des autres résultats est disponible de manière brute en annexe du mémoire.

Sur le Tableau 3, nous pouvons voir que la modification de la formule du Guyot améliore considérablement la capacité du MAR à détecter des évènements de gel au niveau du fruit des vignes. Nous remarquons que le modèle est surtout performant en mars, avril, mai plutôt que sur toute l'année. Etant donné que cela constitue notre période d'intérêt, les résultats sont encore plus pertinents. Nous pouvons voir que bien que la corrélation augmente dans les différents cas par rapport au cas de référence, la RMSE centrée semble aussi augmenter pour certaines périodes.

Les statistiques comme POD et FAR vont nous permettre de déterminer si la formule de Guyot modifiée (donc TAS) prévoit mieux le gel ou non : on remarque que malgré une

hausse de 10-20% de FAR (pour le Cas 1 et Cas 2), la POD augmente significativement et permet donc de prévoir plus d'événements gélifs. Le Cas 2 avec la température minimum du MAR comme température à ajuster par la formule de Guyot modifiée, avec un albedo de 0.2, un h de 11, un ε de 0.9 et un θ de -1 représente le meilleur ajustement du modèle dans notre mémoire. On voit en général que l'ACC est très élevée avec les différentes variations du modèle, mais celle-ci est fortement influencée par le nombre important de vrais négatifs (événement sans gel) et ne représente donc pas une statistique pertinente. Le Cas 2 représente donc le modèle de calcul de la température actinothermique simulée sélectionné pour la suite de tout le travail (hormis pour le Tableau 4).

Température actinothermique simulée, utilisant la température moyenne du MAR comparée avec la température minimum des puces

	Corrélation	RMSE	BIAIS	RMSE centré	POD	FAR	ACC	Albedo	ε	h	θ	
Réf	0.948	3.106	2.454	1.903	0.347	0.023	0.904					Toutes obs
	0.862	2.770	2.118	1.785	0.355	0.065	0.823					2021-M-A-M
	0.877	3.362	2.829	1.817	0.310	0.000	0.880					2022-M-A-M
Cas 1	0.966	1.530	0.193	1.518	0.724	0.279	0.951	0.2	0.3	11	0	Toutes obs
	0.908	1.522	-0.382	1.473	0.721	0.279	0.898	0.2	0.3	11	0	2021-M-A-M
	0.907	1.589	0.078	1.587	0.817	0.319	0.949	0.2	0.3	11	0	2022-M-A-M

Tableau 4 Sélection des meilleurs résultats sur toutes les variations d' ε , h et θ pour la température moyenne du MAR et la température minimum des puces. Le cas de référence correspond à une sortie du MAR non modifiée par la formule de Guyot. Pour chacun des cas les trois périodes d'intérêts ont été représentées, l'ensemble des deux ans, le printemps 2021 (Mars, avril, mai) et le printemps 2022 (Mars, avril, mai).

On peut voir sur le Tableau 4 que la POD du Cas 1 est meilleure que le cas de référence une fois de plus. Nous remarquons que les POD sont légèrement supérieures à celle du Tableau 3 ; toutefois, les FAR sont plus importants. Les FAR sont des éléments très importants à prendre en compte car le but initial de ce mémoire est de créer un outil opérationnel, or si il y a trop de fausses alertes, cela ne va pas aider les vignerons qui devront souvent se réveiller pendant la nuit pour rien. Monte Carlo n'a pas été utilisé ici car les résultats étaient de toute façon moins bon que pour le tableau précédent et ce n'est pas la formule modifiée sélectionnée.

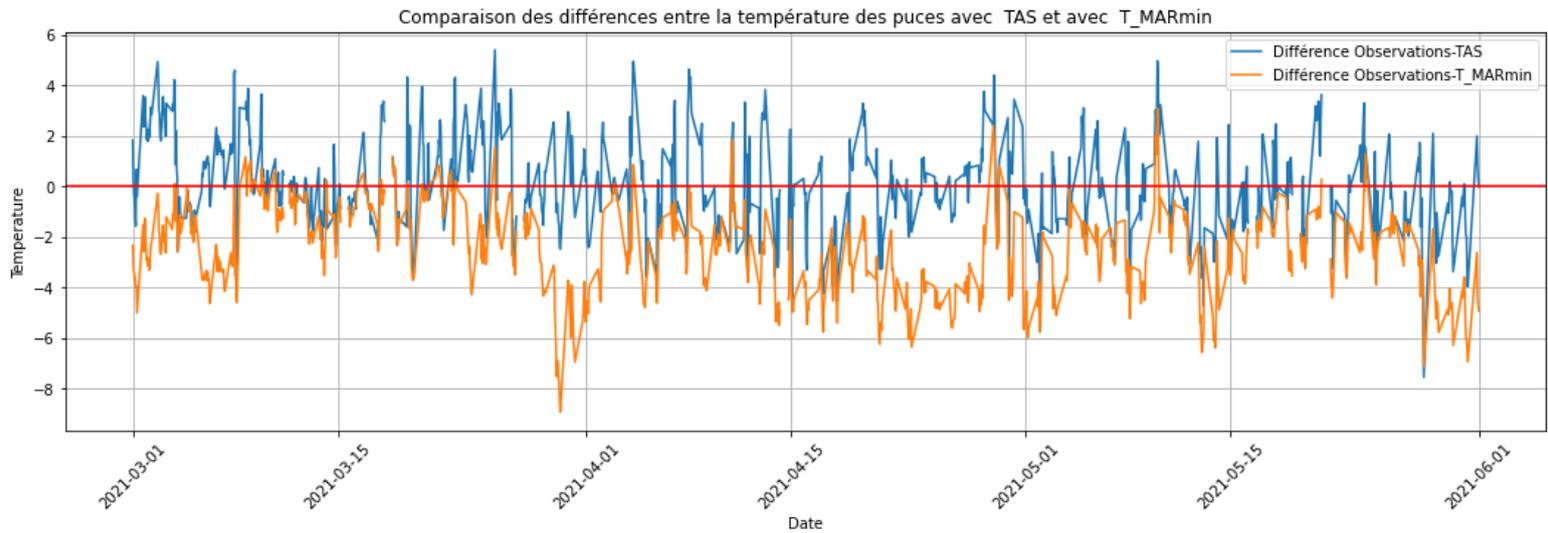


Figure 4 Différence entre la température minimum des puces et la température actinothermique simulée (courbe bleue) ainsi que la différence entre la température minimum des puces et la sortie température minimum du MAR (courbe orange) sur la période d'intérêt de mars avril mai 2021 (avec des données toutes les heures).

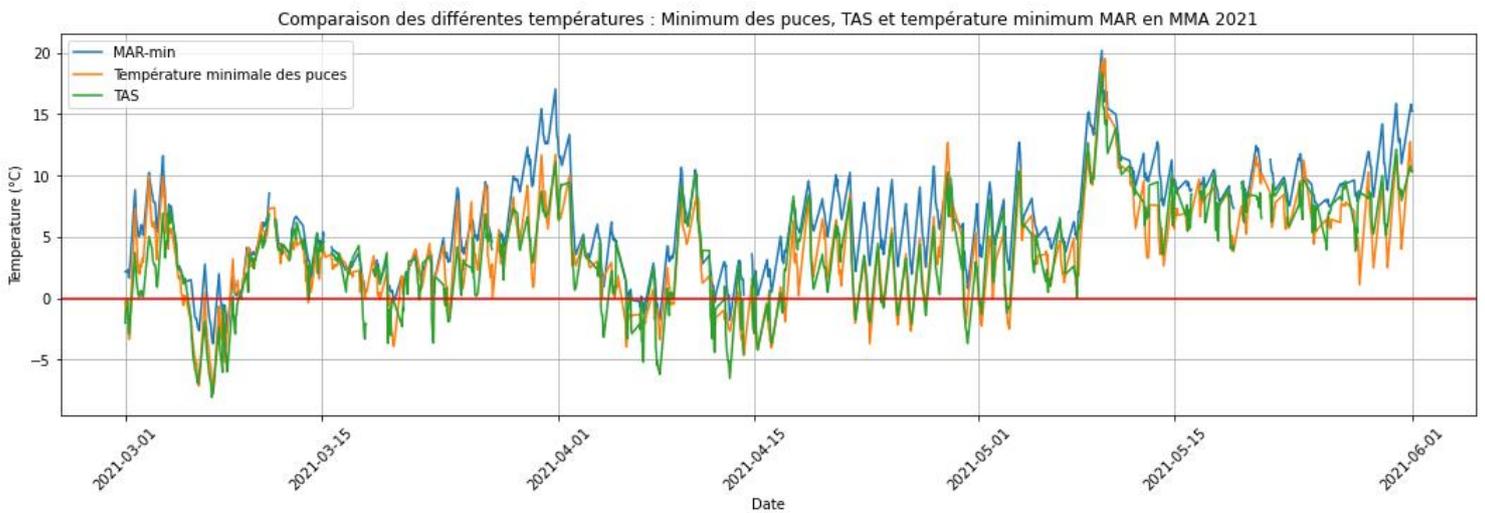


Figure 5 Profils horaires de la température minimum des puces (courbe orange), de la température actinothermique simulée (courbe verte) et la température minimale du MAR (courbe bleue) pour mars avril mai 2021.

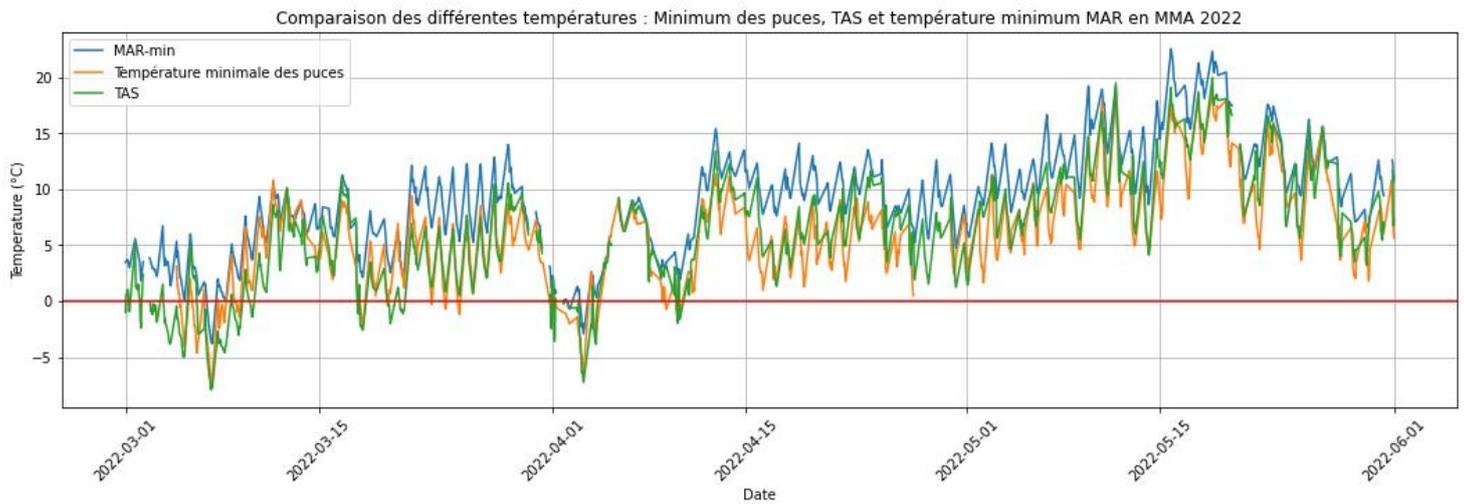


Figure 6 Profils horaires de la température minimale des puces (courbe orange), de la température actinothermique simulée (courbe verte) et la température minimale du MAR (courbe bleue) pour mars avril mai 2022.

Ces deux Figure 5 et Figure 6 nous font remarquer que la température actinothermique simulée n'abaisse pas simplement la température du MAR de manière générale mais suit de façon plus proche le profil de température des puces. Même lorsque la température est positive, la température actinothermique simulée semble mieux suivre la température des observations (des puces) que le MAR.

Comparaison de Tmin MAR et température minimum des puces en fonction de l'altitude et de la proximité des zones sélectionnées pour le MAR.

	Corrélation	RMSE	BIAIS	RMSE centré	POD	FAR	ACC	
244.3 m	0.950	2.881	2.213	1.844	0.385	0.064	0.907	Toutes obs
	0.869	2.530	1.837	1.739	0.404	0.109	0.829	2021-M-A-M
	0.881	3.109	2.549	1.780	0.345	0.024	0.885	2022-M-A-M
203 m	0.949	3.232	2.520	1.869	0.309	0.070	0.905	Toutes obs
	0.901	3.033	2.458	1.782	0.282	0.047	0.856	2021-M-A-M
	0.923	3.813	3.055	1.878	0.216	0.000	0.916	2022-M-A-M

Tableau 5 Résultats de la corrélation, RMSE centré, biais, POD, FAR et ACC pour le pixel de 203m et 244m. L'altitude de 244.3 m correspond au pixel adjacent au pixel où se situe le vignoble ayant l'altitude la plus proche de celle du vignoble. De l'autre côté, 203m correspond au pixel du vignoble (c'est-à-dire celui qui en est le plus proche géographiquement). Pour chacun des cas les trois périodes d'intérêts ont été représentées, l'ensemble des deux ans, le printemps 2021 (Mars, avril, mai) et le printemps 2022 (Mars, avril, mai).

Le Tableau 5 nous montre que le MAR donne une approximation meilleure de la température des puces au pixel d'altitude la plus proche de celle du vignoble qu'au pixel spatialement le plus proche. Cela nous a poussé à utiliser ce pixel (244m d'altitude) comme étant le pixel de référence pour la suite du travail.

Comparaison de Tmin MAR et température minimum des puces en fonction de l'altitude des zones sélectionnées pour le MAR.

	Corrélation	RMSE	BIAIS	RMSE centré	POD	FAR	ACC	
197 m	0.925	3.281	2.328	1.902	0.305	0.026	0.898	Toutes obs
	0.903	2.798	2.041	1.817	0.292	0.016	0.863	2021-M-A-M
	0.826	4.555	3.300	1.865	0.174	0.000	0.882	2022-M-A-M
341 m	0.924	2.741	1.465	1.849	0.487	0.164	0.917	Toutes obs
	0.895	2.277	1.113	1.819	0.474	0.172	0.884	2021-M-A-M
	0.819	3.861	2.176	1.804	0.379	0.035	0.910	2022-M-A-M

Tableau 6 Résultats de la corrélation, RMSE centré, biais, POD, FAR et ACC pour le pixel le plus bas (197 mètres) et le plus haut (341 mètres) avoisinant le pixel où se situe le vignoble pour la température minimum du MAR sans passer par la formule de Guyot modifiée. Pour chacun des cas les trois périodes d'intérêts ont été représentées, l'ensemble des deux ans, le printemps 2021 (Mars, avril, mai) et printemps 2022 (Mars, avril, mai).

Dans la même optique qu'avec le tableau précédent, le Tableau 6 présente les résultats que l'on obtient en prenant le pixel le plus haut et le plus bas. Le pixel à 197m montre des POD inférieures au pixel sélectionné précédemment (244m) bien que les différences soient faibles. En ce qui concerne le pixel le plus haut à 341m, bien que les résultats soient intéressants car un peu meilleurs pour les POD, le FAR de base est trop élevé (hormis pour mars avril mai 2022 où le FAR est excellent) ce qui pose problème si on l'utilise comme valeur de température pour l'ajustement. En effet, le FAR augmente encore lorsque la température est transformée par la formule modifiée afin de trouver la température actinothermique simulée. On remarque également un RMSE centré légèrement plus élevé pour le pixel de 341m que pour le pixel de 244m.

Globalement, nous voyons bien avec le Tableau 5 et Tableau 6 l'influence de l'altitude (et donc du changement de pression) sur la température (Wang et al., 2011). Plus le pixel sélectionné est haut, plus on aura des températures froides et donc plus d'alertes de gel. Le contraire se produit avec le pixel le plus bas.

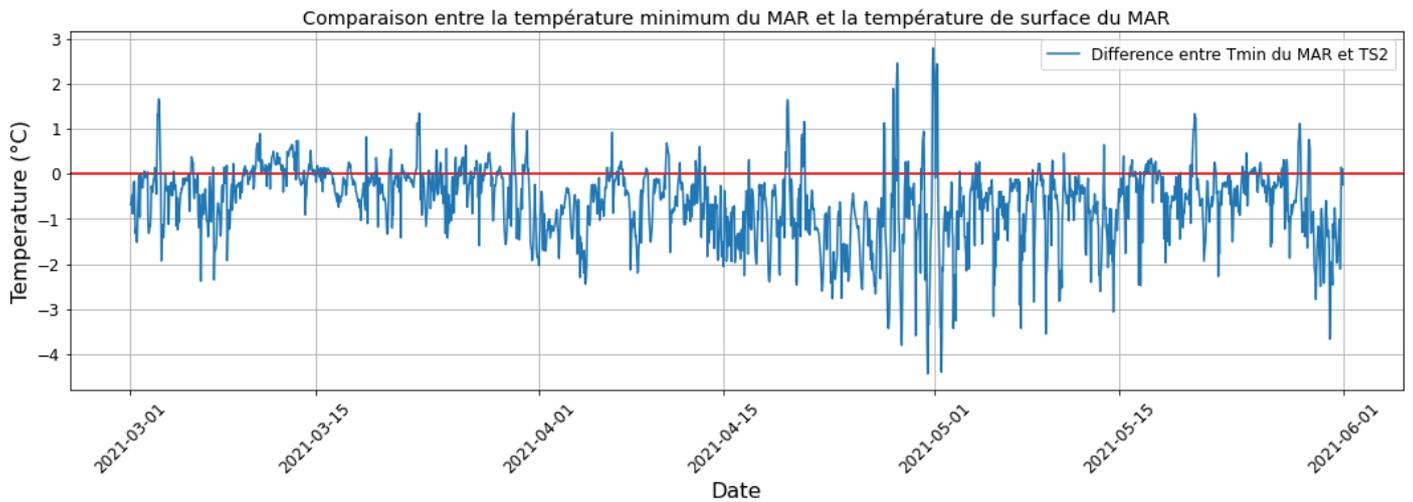


Figure 7 Comparaison (différence) de la température minimum du MAR avec la température de surface du MAR (TS2) sur mars avril mai 2021 afin de déterminer si l'utilisation de la température de surface est pertinente.

Sur la Figure 7, nous voyons que la température de surface du MAR (à ne pas confondre avec la température de l'air) est généralement plus chaude que la température minimum de l'air du MAR. Or, la problématique du départ de ce mémoire est que la température modélisée par le MAR a tendance à rater des événements gélifs. Nous aurions pu travailler avec la température de surface si elle était plus froide que la température minimum MAR.

III.2. Alertes complémentaires

III.2.1. Représentation graphique des fonctions de distribution cumulative des alertes

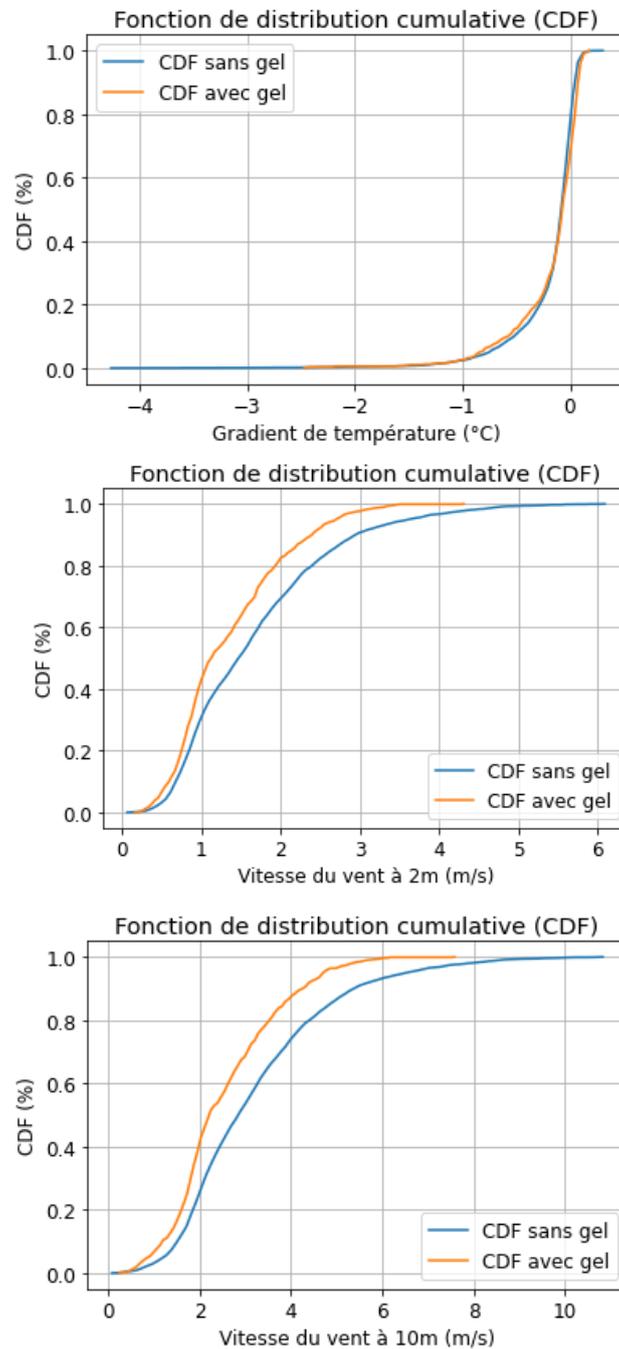


Figure 8 Distribution des trois alertes complémentaires via une fonction distributive cumulative. Les alertes sont divisées en deux échantillons : la valeur des alertes sans gel (courbe bleue) et la valeur des alertes lorsqu'il y a gel (courbe orange). Les alertes sont le gradient de température entre 3 et 10 mètres, la vitesse du vent à 2 mètre et la vitesse du vent à 10 mètres.

III.2.2. Test de Shapiro

Variable	Test de Shapiro	p-valeur
Gradient de température sans gel	0.646	0
Gradient de température avec gel	0.720	1.03E-35
Vitesse du vent sans gel (10 mètres)	0.959	7.54E-15
Vitesse du vent avec gel (10 mètres)	0.921	5.61E-45
Vitesse du vent sans gel (2 mètres)	0.909	0
Vitesse du vent avec gel (2 mètres)	0.936	6.94E-19

Tableau 7 Résultats du test de Shapiro pour les alertes complémentaires.

III.2.3. Test de Mann Whitney

Variable	p-valeur
Gradient de température	0.001
Vitesse du vent (10 mètres)	8.71E-29
Vitesse du vent (2 mètres)	6.63E-19

Tableau 8 Résultats du test de Mann Whitney pour les alertes complémentaires.

Tout d'abord, visuellement, grâce aux CDF de la Figure 8, nous pouvons voir que la variable « gradient de température » n'est pas du tout discriminante : c'est-à-dire qu'elle semble avoir la même valeur qu'il y ait du gel ou non. Nous remarquons même que les gradients les plus négatifs ne sont obtenus que pour les moments où il n'y a pas de gel. Il est à noter qu'un gradient négatif n'est pas forcément lié à des températures négatives près de la surface. Néanmoins, un gradient de température négatif et donc une inversion thermique est une situation favorable à l'apparition de gel (Beltrando, 1998). Il n'existe pas encore d'article ayant analysé le gradient de température du MAR pour nous aider dans nos analyses. En ce qui concerne les deux vitesses de vent, il y a bien un léger décalage entre la courbe avec et sans présence de gel. Seulement, les deux courbes couvrent majoritairement les mêmes valeurs ce qui ne nous aide pas à retenir certaines valeurs de vent pour créer une alerte de gel.

Le Tableau 7 nous a indiqué que nous devons utiliser le test de Mann Whitney (distribution non normale). Notre analyse pour le gradient de température est confirmée : ce

n'est pas une variable discriminante pour le gel. Cependant, le test nous dit que pour les vents (Tableau 8), les deux groupes (avec et sans gel) sont différents. Malheureusement comme expliqué plus haut le fait que les courbes recouvrent les mêmes valeurs ne permet pas d'en retirer des informations utiles. On remarque que la vitesse du vent (à 2m et 10m) lors d'événement non gélif est plus importante : cet élément pourrait être utilisé pour éliminer des fausses alertes. Or, le vent est responsable de deux types de gel : l'inversion thermique (lorsque le vent est faible un gradient se met en place) et le gel advectif ou gelée noire lorsque qu'une masse d'air froide remplace la masse d'air présente actuellement sur le territoire (Beltrando, 1998). Aucun de ces phénomènes n'est suggéré par notre jeu de données.

III.3. Régression logistique

III.3.1. Analyse univariée

Test de Shapiro pour les variables indépendantes

Variables	Test	p-valeur
Gel température minimum	0.892	3.90E-24
Non gel température minimum	0.976	1.06E-27
Gel température actinothermique simulée	0.921	6.19E-21
Non gel température actinothermique simulée	0.977	1.54E-27
Gel gradient de température	0.720	1.03E-35
Non gel gradient de température	0.633	0
Gel vitesse du vent à 10m	0.959	7.54E-15
Non gel vitesse du vent à 10m	0.921	5.61E-45
Gel densité optique des nuages	0.588	4.61E-41
Non gel densité optique des nuages	0.714	0
Gel nuages à altitude moyenne	0.378	0
Non gel nuages à altitude moyenne	0.550	0
Gel nuages à altitude basse	0.733	4.54E-35
Non gel nuages à altitude basse	0.745	0
Gel nuage à haute altitude	0.564	6.68E-42
Non gel nuage à haute altitude	0.687	0
Gel long wave downward	0.931	1.57E-19
Non gel long wave downward	0.984	4.04E-23
Gel long wave upward	0.914	8.21E-22
Non gel long wave upward	0.977	1.87E-27
Gel humidité spécifique	0.949	8.15E-17
Non gel humidité spécifique	0.937	2.50E-41
Gel pression de surface	0.961	1.78E-14
Non gel pression de surface	0.991	3.89E-17
Gel pression au niveau de la mer	0.963	6.31E-14
Non gel pression au niveau de la mer	0.994	5.07E-14
Gel point de rosée	0.897	1.30E-23
Non gel point de rosée	0.977	1.26E-27
Gel précipitations	0.259	0
Non gel précipitations	0.354	0

Tableau 9 Indication de la distribution normale ou pas suivie par l'ensemble des variables indépendantes divisées en deux catégories : leurs valeurs lorsqu'il y a du gel et leurs valeurs lorsqu'il n'y en pas. Avec la valeur du test de Shapiro ainsi que la p-valeur.

Corrélation de Spearman

Variabes	Rho
Température minimum MAR	-0.579
TAS	-0.577
Gradient de température	0.034
Vitesse du vent à 10m	-0.147
COD	-0.173
CM	-0.166
CD	-0.004
CU	-0.176
LWD	-0.413
LWU	-0.567
QQ	-0.566
SP	0.210
SLP	0.239
TD	-0.564
PP	-0.123

Tableau 10 Résultat de la corrélation de Spearman qui nous indique si les variables indépendantes sont corrélées ou non avec la variable dépendante : rho donne la corrélation.

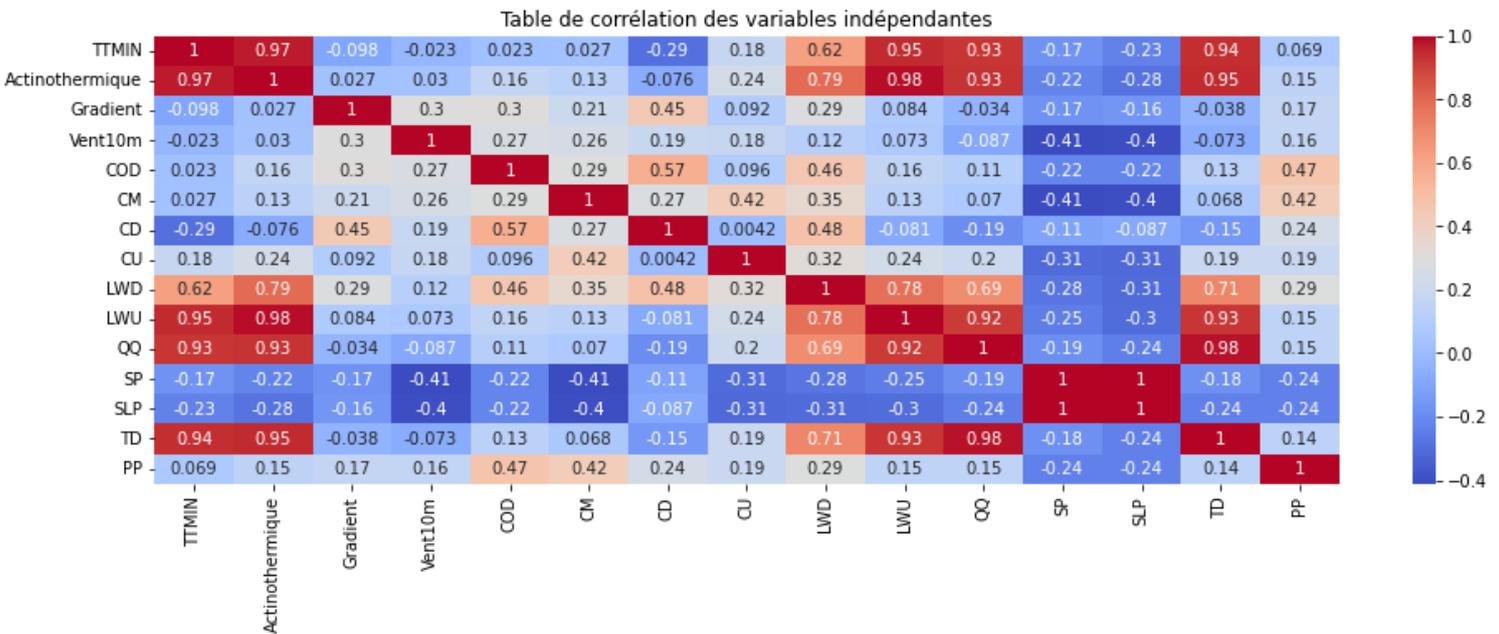


Tableau 11 Corrélation entre les différentes variables indépendantes à utiliser dans la régression logistique.

Le Tableau 9 nous indique que la majorité des variables ne suivent pas une distribution normale de sorte que le test de Spearman a dû être utilisé pour analyser la corrélation des variables indépendantes à la variable dépendante. Nous remarquons directement (Tableau 10) les variables qui ont la plus grosse corrélation avec le gel observé : la température minimum, la température actinothermique simulée, les 2 rayonnements à longue longueur d'onde, la température du point de rosée et l'humidité spécifique. Ces corrélations sont négatives : par exemple lorsque la température augmente, l'apparition du gel diminue. Les émissions à longue longueur d'onde jouent un rôle encore plus important la nuit qu'en journée pour influencer la température.

En regardant le Tableau 11, nous remarquons que les variables les plus corrélées au gel sont également celles qui sont le plus corrélées entre-elles. Généralement lors de régression, nous enlevons les variables qui sont trop corrélées entre elles (>90%). Dès lors deux variables ont été conservées parmi les 5 fortement corrélées (Température minimum, température actinothermique simulée, longue longueur d'onde vers le bas, humidité spécifique et température du point de rosée) : la température minimum et TAS qui vont toutes les deux être intégrée dans une régression logistique.

III.3.2 Application de la régression logistique

III.3.2.1. Régression logistique avec la température actinothermique simulée

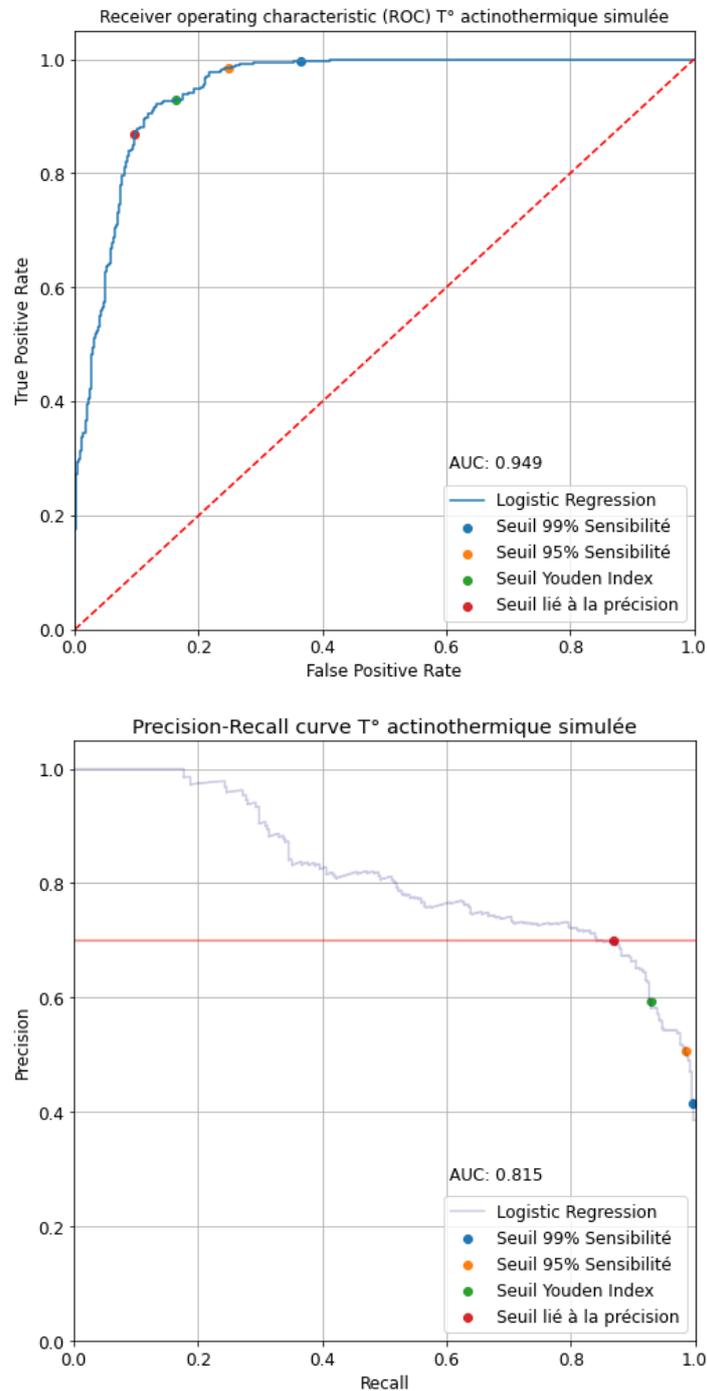


Figure 9 La ROC curve et la Precision-Recall curve du test set avec la température actinothermique simulée comme seule variable indépendante de la régression logistique avec quatre seuils différents. Pour la Precision-Recall curve, la droite horizontale représente le seuil de fausses alertes permises par les spécialistes du monde de la vigne (Précision de 70%).

Température actinothermique simulée

99% Sensibilité				95% Sensibilité			
		Observations				Observations	
		Oui	Non			Oui	Non
Valeurs	Oui	378	541	Valeurs	Oui	374	371
prédites	Non	1	923	prédites	Non	5	1093
POD	1			POD	0.99		
FAR	0.59			FAR	0.50		

Youden Index				Seuil lié à précision			
		Observations				Observations	
		Oui	Non			Oui	Non
Valeurs	Oui	358	283	Valeurs	Oui	318	130
prédites	Non	21	1181	prédites	Non	61	1334
POD	0.94			POD	0.84		
FAR	0.44			FAR	0.29		

Tableau 12 Résultats de la régression logistique sur le test set, avec comme seule variable indépendante la température actinothermique simulée, pour les 4 seuils différents obtenus sur le training set. Pour chaque seuil, une table de contingence est présentée ainsi que la POD et le FAR.

La régression logistique de la température actinothermique simulée seule nous montre de très bon résultats pour la ROC curve (Figure 9) : une AUC de 0.949. Nous pouvons voir sur le graphique de la ROC curve que tous les seuils trouvés sur le training set montre de bonne performance en terme de ROC curve : « True Positive Rate » est souvent bien supérieur au « False Positive Rate ». Cependant la ROC curve ne prend pas en compte les éléments manqués, les faux négatifs hors ceci est un enjeu majeur de notre recherche. L'AUC de la Recall-Precision curve est moins élevée que celle de la ROC curve mais reste performante avec une valeur de 0.815. Ici, tous les seuils ne sont pas performants : il n'y a que le seuil lié à la précision qui montre une précision supérieur à 70% sur le test set (correspond au 30% de FAR). En effet lorsqu'on regarde les tables de contingences (Tableau 12), on voit que le FAR (1-Precision) est souvent trop élevé hormis le seuil lié à la précision qui détecte 84% des évènements de gel. On manque tout de même 61 événements sur 378.

III.3.2.2. Régression logistique avec la température minimum du MAR

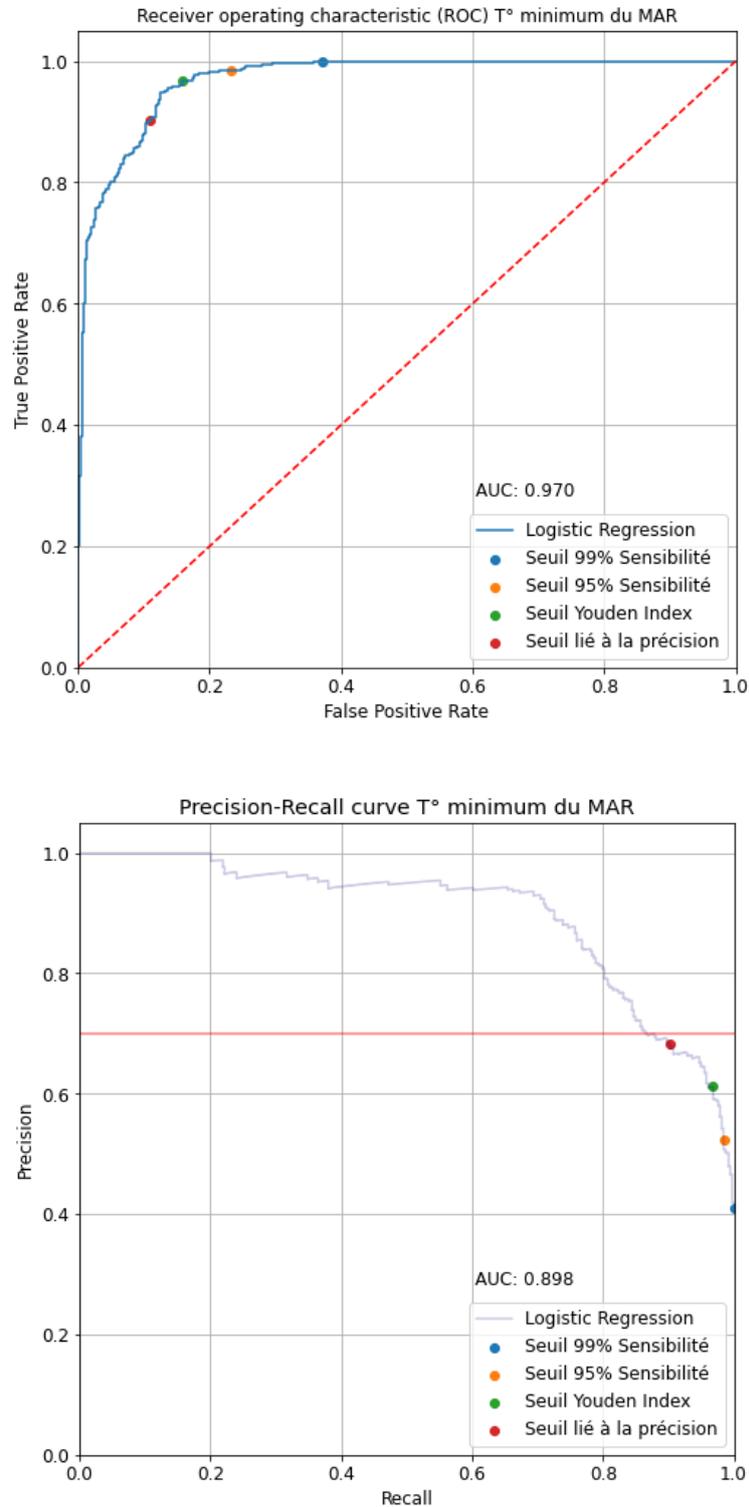


Figure 10 La ROC curve et la Precision-Recall curve du test set avec la température minimum du MAR comme seule variable indépendante de la régression logistique avec quatre seuils différents. Pour la Precision-Recall curve, la droite horizontale représente le seuil de fausses alertes permises par les spécialistes du monde de la vigne (Precision de 70%).

Température minimum du MAR

99% Sensibilité				95% Sensibilité			
		Observations				Observations	
Valeurs prédites		Oui	Non	Valeurs prédites		Oui	Non
	Oui	379	543		Oui	373	333
	Non	0	921		Non	6	1131
	POD	1			POD	0.98	
	FAR	0.59			FAR	0.47	

Youden Index				Seuil lié à précision			
		Observations				Observations	
Valeurs prédites		Oui	Non	Valeurs prédites		Oui	Non
	Oui	367	244		Oui	334	146
	Non	12	1220		Non	45	1318
	POD	0.97			POD	0.88	
	FAR	0.40			FAR	0.32	

Tableau 13 Résultats de la régression logistique sur le test set, avec comme seule variable indépendante la température minimum du MAR, pour les 4 seuils différents obtenus sur le training set. Pour chaque seuil, une table de contingence est présentée ainsi que la POD et le FAR.

En terme de ROC curve pour la température minimum du MAR (Figure 10), l'AUC (0.97) présente un meilleur résultat que pour la température actinothermique simulée seule. Les différents seuils présentent également des résultats concluants pour la ROC curve. En ce qui concerne la Precision-Recall curve, l'AUC (0.89) est également meilleure que pour la température actinothermique simulée. La majorité des seuils présente trop de fausses alertes mais nous avons tout de même le seuil lié à la précision qui montre des résultats performants avec une POD de 0.88 et malheureusement une FAR de 0.32 ce qui est un peu trop haut (Tableau 13). Ici nous manquons que 45 événements de gel sur 379 mais 146 évènements prédits sont de fausses alertes. En ce qui concerne les autres seuils, ce sont les mêmes tendances que le tableau précédent : une très bonne POD mais une trop grosse FAR.

III.3.2.3. Régression logistique avec les variables de l'alerte vigneron

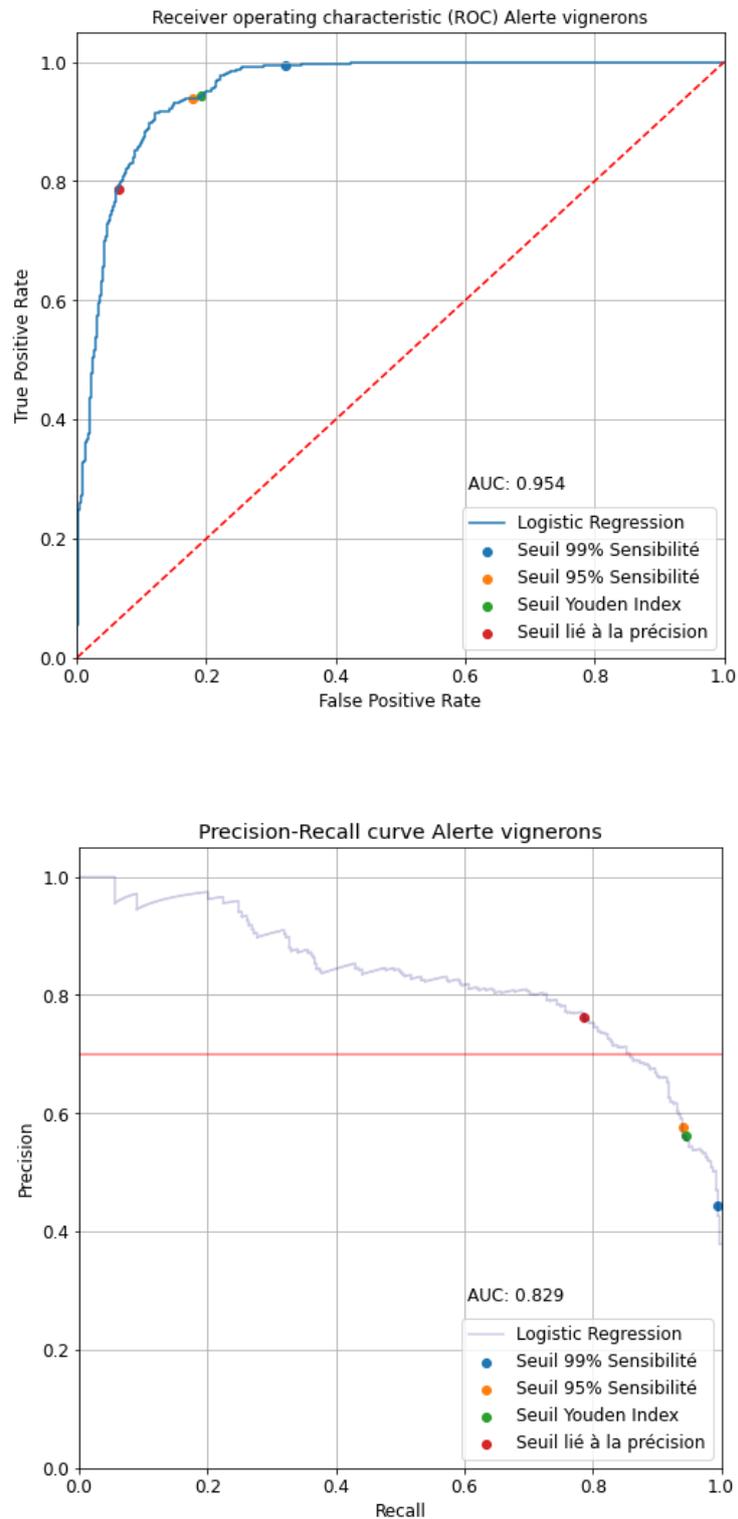


Figure 11 La ROC curve et la Precision-Recall curve du test set avec les variables de l'alerte vigneron comme variables indépendantes (TAS, gradient de température et vitesse du vent à 10m) de la régression logistique avec quatre seuils différents. Pour la Precision-Recall curve, la droite horizontale représente le seuil de fausses alertes permises par les spécialistes du monde de la vigne (Precision de 70%).

Alerte vignerons

99% Sensibilité				95% Sensibilité			
		Observations				Observations	
		Oui	Non			Oui	Non
Valeurs	Oui	377	473	Valeurs	Oui	356	262
prédites	Non	2	991	prédites	Non	23	1202
POD	0.99			POD	0.93		
FAR	0.55			FAR	0.42		

Youden Index				Seuil lié à précision			
		Observations				Observations	
		Oui	Non			Oui	Non
Valeurs	Oui	358	280	Valeurs	Oui	298	93
prédites	Non	21	1184	prédites	Non	81	1371
POD	0.94			POD	0.76		
FAR	0.43			FAR	0.24		

Tableau 14 Résultats de la régression logistique sur le test set, avec les variables issues de l'alerte vignerons comme variables indépendantes (température actinothermique simulée, gradient de température et vitesse du vent à 10 mètres), pour les 4 seuils différents obtenus sur le training set. Pour chaque seuil, une table de contingence est présentée ainsi que la POD et le FAR.

Valeur des coefficients des variables indépendantes pour la régression logistique « Alerte vignerons »

Variable	TAS	Gradient de température	Vitesse du vent à 10 m
Coefficient	-0.83	0.06	-0.11

Tableau 15 Valeurs des coefficients pour la régression logistique « alerte vignerons » avec la température actinothermique simulée, le gradient de température et la vitesse du vent à 10 mètres.

Ici nous pouvons comparer avec la régression logistique de la température actinothermique simulée seule. Nous remarquons que les résultats pour les deux AUC, celle de la ROC curve (0.95) et de la Precision-Recall curve (0.83) (Figure 11) sont légèrement meilleurs que pour la température actinothermique simulée seule. En ce qui concerne les seuils (Tableau 14), la conclusion est la même que pour les deux autres tableaux. Ici le seuil lié à la précision ne donne pas un résultat très performant avec une faible POD avec 81 événements manqués sur 379. Sans surprise nous voyons sur le Tableau 15 que la température est la variable qui impacte le plus la présence ou non de gel avec un coefficient de -0.83, vient ensuite la vitesse du vent (-0.11) et enfin le gradient de température (0.06).

III.3.3. Régression logistique avec sélection de données par LASSO

Sélection de variables via LASSO avec TTmin

Variables	Coefficient
TTMIN	-0.060
Gradient	0.000
Vent10m	0.000
COD	-0.018
CM	0.000
CD	0.000
CU	0.000
SP	0.000
SLP	0.002
PP	0.000

Tableau 16 Résultats de la sélection de LASSO sur l'ensemble des variables indépendantes sélectionnées avec température minimum du MAR comme variable fortement corrélée conservée.

Le Tableau 16 montre que nous pouvons réaliser une régression avec trois variables définies par LASSO : la température minimum du MAR, COD (la densité optique des nuages) et SLP (la pression de surface au niveau de la mer). Il a été testé de laisser la température actinothermique comme variable indépendante fortement corrélée conservée (Tableau 23, en annexe) mais ça donnait un moins bon coefficient pour la température actinothermique simulée (-0.059) comme pour COD (-0.005) et SLP (0) qu'avec la température minimum .

III.3.3.1. Application de la régression logistique avec des variables issues de LASSO

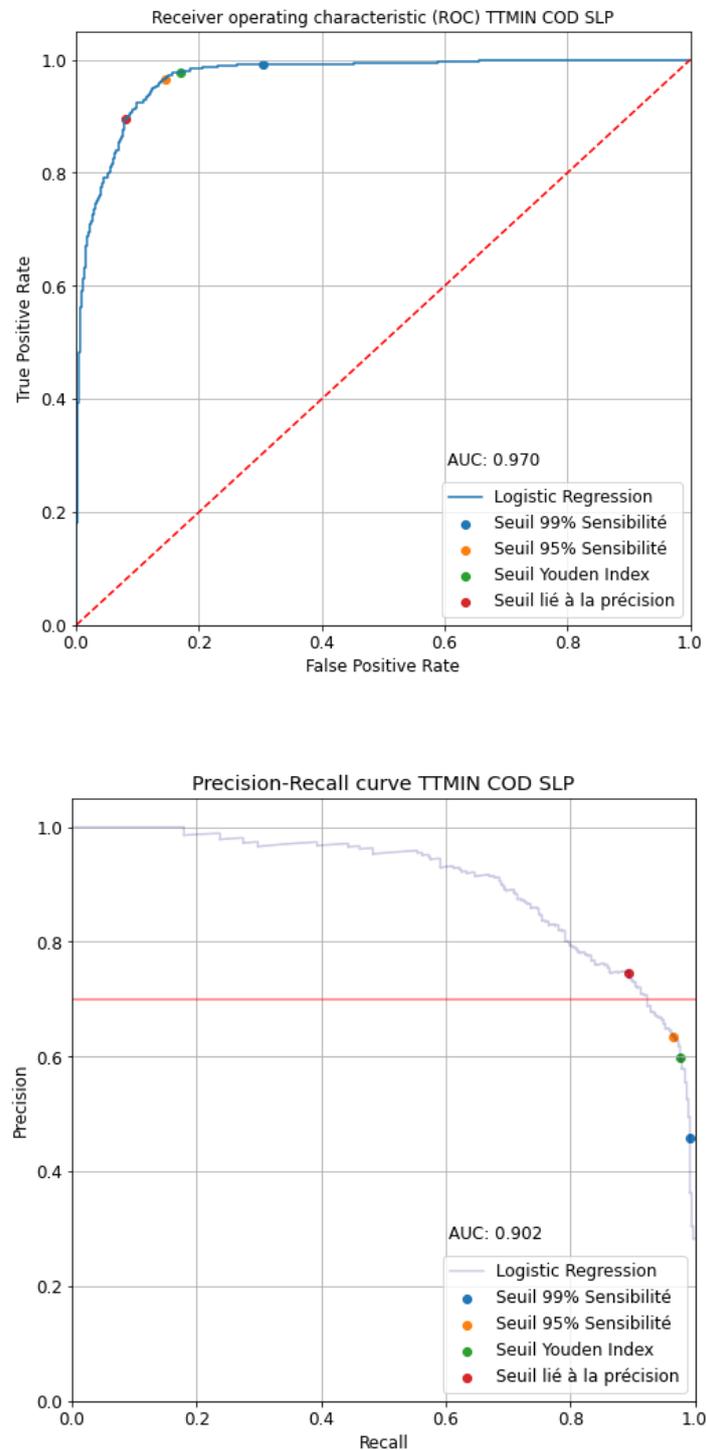


Figure 12 La ROC curve et la Precision-Recall curve du test set avec les variables sélectionnées par LASSO (Température minimum du MAR, COD, SLP) comme variables indépendantes de la régression logistique avec quatre seuils différents. Pour la Precision-Recall curve, la droite horizontale représente le seuil de fausses alertes permises par les spécialistes du monde de la vigne (Précision de 70%).

LASSO : Température minimum MAR, COD et SLP

99% Sensibilité				95% Sensibilité			
		Observations				Observations	
Valeurs		Oui	Non	Valeurs		Oui	Non
prédites	Oui	376	446	prédites	Oui	366	211
	Non	3	1018		Non	13	1253
POD	0.99			POD	0.96		
FAR	0.54			FAR	0.36		

Youden Index				Seuil lié à précision			
		Observations				Observations	
Valeurs		Oui	Non	Valeurs		Oui	Non
prédites	Oui	370	248	prédites	Oui	339	116
	Non	9	1216		Non	40	1348
POD	0.97			POD	0.89		
FAR	0.40			FAR	0.25		

Tableau 17 Résultats de la régression logistique sur le test set, avec les variables issues la sélection de LASSO comme variables indépendantes (Température minimum du MAR, COD, SLP), pour les 4 seuils différents obtenus sur le training set. Pour chaque seuil, une table de contingence est présentée ainsi que la POD et le FAR.

Valeur des coefficients des variables indépendantes pour la régression logistique LASSO

Variables	Température minimum MAR	Densité optique des nuages (COD)	Pression de surface au niveau de la mer (SLP)
Coefficient	-0.80	0.19	-0.01

Tableau 18 Valeur des coefficients des variables indépendantes de la régression logistique LASSO : la température minimum du MAR, la densité optique des nuages (COD) et la pression de surface au niveau de la mer (SLP).

Pour cette combinaison de variables, nous obtenons une AUC (0.97) identique à celle de la température minimum du MAR seule pour la ROC curve et une AUC (0.90) légèrement supérieur à celle de la température minimum pour la Precision-Recall curve (Figure 12). En ce qui concerne les tables de contingence (Tableau 17), le constat est le même que précédemment pour les seuils à l'exception du seuil lié à la précision. En effet, ici nous avons une POD de 0.88 avec une FAR de 0.25 ce qui est le meilleur résultat parmi les 4 régressions logistiques. On arrive à égaler la POD de la température minimum du MAR pour un seuil lié à la précision tout en diminuant de 7% la proportion de FAR. Nous voyons que la densité optique

des nuages (COD) apporte réellement une plus-value. Ceci est logique car plus la densité optique est importante, plus les nuages sont denses et plus ils vont renvoyer le rayonnement infra-rouge ce qui va empêcher la formation d'une nuit radiative (étant donné que le ciel est couvert). À l'inverse, une COD faible indique que les nuages sont peu denses et que donc c'est une nuit à ciel clair/dégagé très favorable à l'apparition de gel (Beltrando, 1998). Quant à la pression de surface au niveau de la mer, lorsqu'il y a une haute pression : cela est signe d'anticyclone et donc d'air descendant, ce qui va assécher l'air ambiant et donc empêcher la formation de nuage : favorable à une nuit radiative et donc à l'apparition de gel (Madelin, 2004). Nous voyons grâce au Tableau 18 que la température minimum du MAR est la variable la plus impactante avec un coefficient de -0.80, ensuite la densité optique avec 0.19 et enfin SLP avec -0.01. L'influence de la pression de surface au niveau de la mer est donc minime.

III.3.3.2. Régression logistique avec les variables sélectionnées via LASSO d'après un training set non équilibré

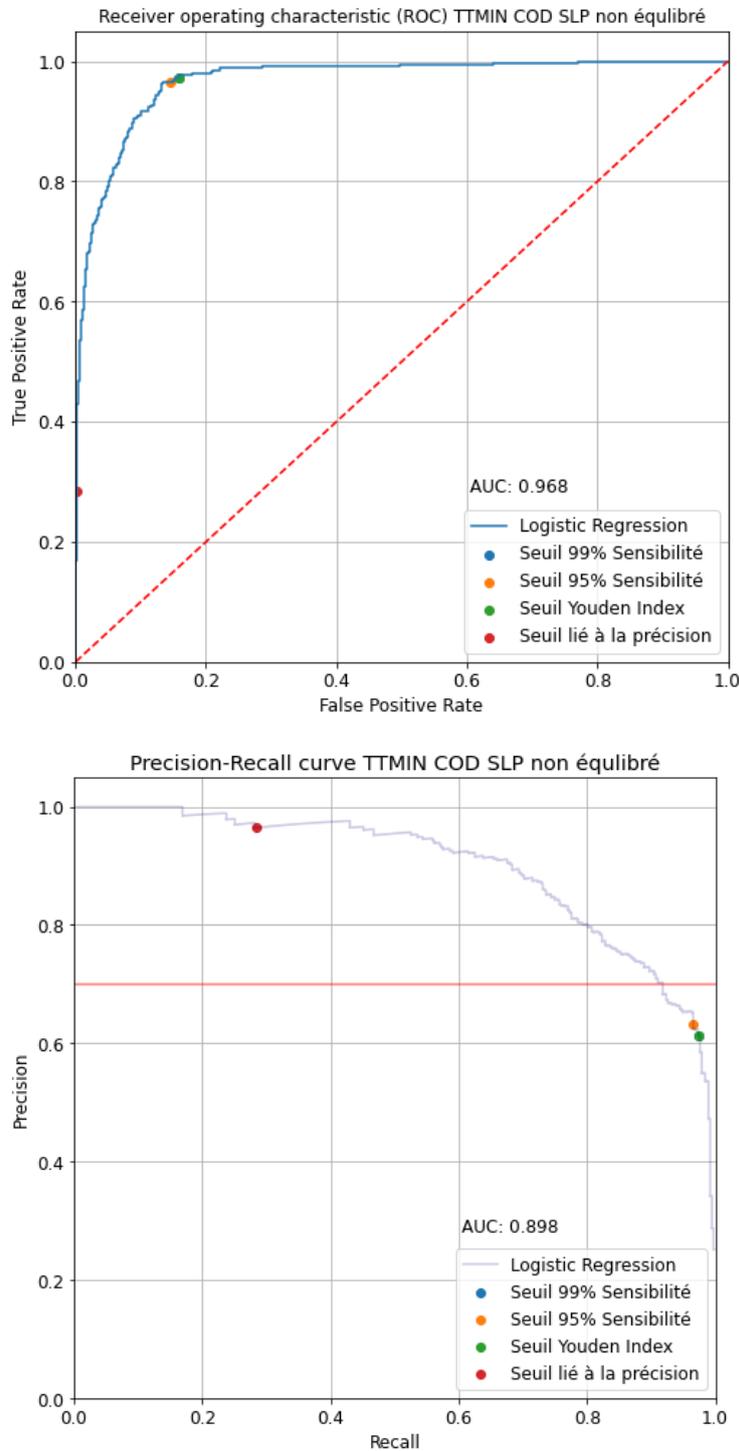


Figure 13 La ROC curve et la Precision-Recall du test set avec les variables indépendantes retenues par la régression de LASSO (température minimum du MAR, la densité optique des nuages COD et SLP la pression de surface au niveau de la mer) avec un training set non équilibré. Sur le graphique de droite, la droite horizontale représente le seuil de fausses alertes permises par les spécialistes du monde de la vigne (Precision de 70%).

Non équilibré : Température minimum MAR, COD et SP

99% Sensibilité				95% Sensibilité			
		Observations				Observations	
		Oui	Non			Oui	Non
Valeurs	Oui	369	232	Valeurs	Oui	366	214
prédites	Non	10	1232	prédites	Non	13	1250
POD	0.97			POD	0.96		
FAR	0.38			FAR	0.36		

Youden Index				Seuil lié à précision			
		Observations				Observations	
		Oui	Non			Oui	Non
Valeurs	Oui	369	232	Valeurs	Oui	108	4
prédites	Non	10	1232	prédites	Non	271	1460
POD	0.97			POD	0.28		
FAR	0.38			FAR	0.03		

Tableau 19 Table de contingence ainsi que POD et FAR obtenues via la régression logistique sur le test set avec quatre seuils différents (trouvées sur le training set non équilibré) pour les variables sélectionnées par la régression de LASSO (Température minimum du MAR, COD : la densité optique des nuages et SLP (pression de surface au niveau de la mer)).

Non équilibré : Température minimum MAR,

COD et SLP [TRAINING SET]

Seuil lié à la précision	
POD	0.82
FAR	0.09

Tableau 20 Performances en POD et FAR de la régression logistique sur le training set avec un training set non équilibré et avec les variables indépendantes sélectionnées par LASSO : Température minimum du MAR, COD, SLP.

Nous voyons avec les AUC de la régression sur un dataset non équilibré (sans l'équilibrage réalisé dans la méthodologie), qu'elles sont légèrement moins bonnes que celles de la régression avec les mêmes variables mais avec un dataset équilibré que ce soit pour la ROC curve ou la Precision-Recall curve (Figure 13). Nous remarquons également que pour les tables de contingences (Tableau 19) les résultats sont soit égaux soit meilleurs (notamment le FAR) pour les 3 premiers seuils (99% sensibilité, 95% sensibilité et Youden Index). Par contre,

le seuil qui jusqu'à présent était le plus performant, le seuil lié à la précision montre des mauvaises performances. Ceci montre qu'équilibrer le dataset était important pour avoir un seuil pertinent en POD et avec un FAR inférieur à 30%. Sur le training set les résultats du seuil lié à la précision sont bons (Tableau 20) mais cela n'est pas confirmé sur le test set.

Nous pouvons en conclure que l'équilibrage du training set ne modifie pas les capacités de prédictions du modèle qui restent très bonnes et nous permet de trouver un seuil particulièrement performant avec un faible FAR et un POD assez élevé (seuil lié à la précision, Tableau 17).

III.4. Comparaison des deux meilleurs modèles issus de différentes techniques

Nous allons maintenant comparer les performances du meilleur modèle issu des régressions logistiques et du modèle de la température actinothermique simulée issue de la modification de la formule de Guyot. Cette comparaison va se réaliser via les tables de contingences sur toute l'année 2022 :

Température actinothermique simulée				Régression logistique (LASSO)			
				Seuil lié à précision			
		Observations				Observations	
		Oui	Non			Oui	Non
Valeurs	Oui	212	88	Valeurs	Oui	339	116
prédites	Non	167	1376	prédites	Non	40	1348
POD	0.56			POD	0.89		
FAR	0.29			FAR	0.25		

Tableau 21 Deux meilleures modèles issus de deux approches différentes. Les résultats sont testés sur toute l'année 2022 avec la température actinothermique simulée issue de la formule de Guyot modifiée sélectionnée au Tableau 3 ainsi que la régression logistique avec les variables indépendantes données par LASSO (température minimum du MAR, COD, SLP) et en utilisant le seuil lié à la précision.

Premièrement, la comparaison s'effectue sur toute l'année 2022 et non pas seulement sur une période de référence (mars avril mai) car la régression logistique doit être testée sur tout l'échantillon. Nous voyons à travers ces résultats (Tableau 21) que la régression logistique (qui a une POD de 0.89 pour un FAR de 0.25) est plus robuste que la température actinothermique simulée seule (POD de 0.56 et FAR de 0.29). Pour confirmer cela, il faudrait des données d'années ultérieures pour pouvoir comparer les deux de manière optimale (sur la période de référence 2023 par exemple).

III.5. Analyse de la qualité des prévisions par nuit sur une année

III.5.1. Analyse sur la nuit des prévisions réalisées par la régression logistique (LASSO)

Répartition du type de prévision pour chaque nuit de l'année 2022 à partir de la régression logistique Lasso

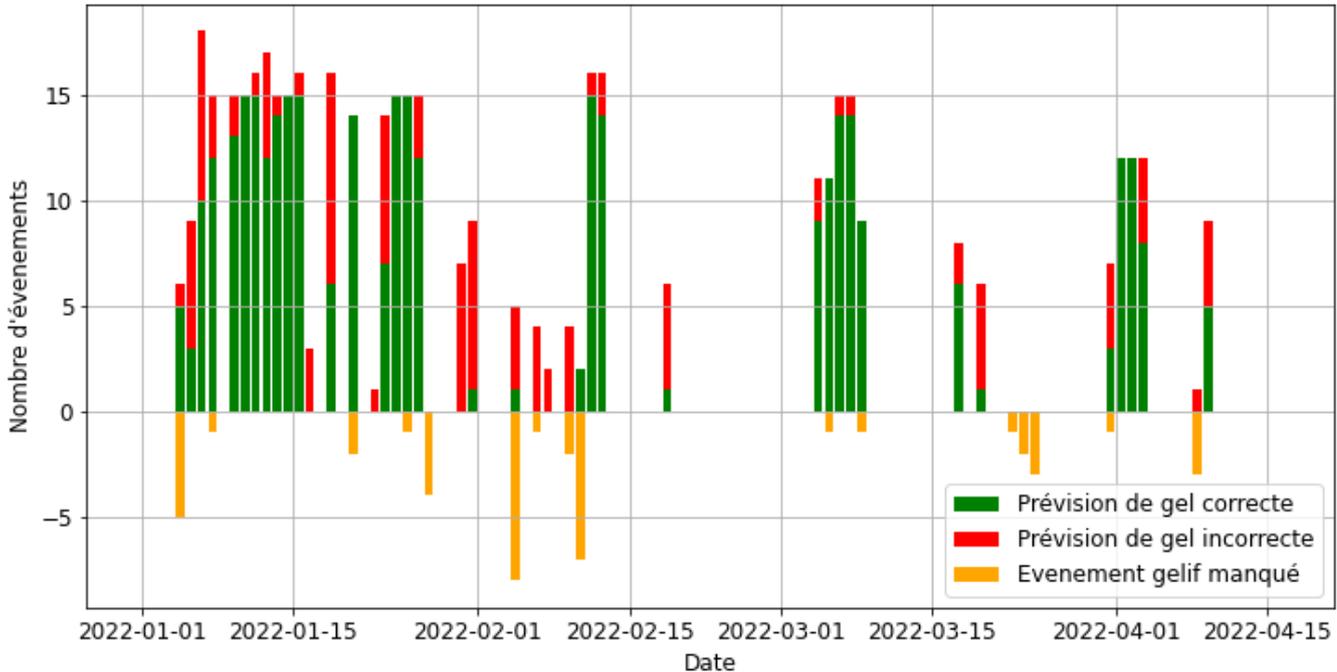


Figure 14 Répartition du type de prévision par nuit (prévision de gel correct (bâtonnet vert), prévision de gel incorrect (bâtonnet rouge), événement gélif manqué (bâtonnet orange)) issue de la régression logistique avec les variables sélectionnées par LASSO (Température minimum du MAR, COD, SLP) lors de l'année 2022. Les prévisions de non gel exactes ne sont pas représentées. Le graphique s'arrête le 15 avril car il n'y plus de gel détecté ensuite pour cette année-là. Le nombre d'évènements gélifs manqués est exprimé en valeur négative afin de faciliter la lecture du graphique.

Nous pouvons voir sur la Figure 14 que 6 combinaisons de prédiction sont possibles :

1. Il n'y a que des prévisions de gel correctes, les vignerons devront se lever pour mettre en place les systèmes de lutte contre le gel car il y aura apparition de gel.
2. Il y a des prévisions de gel correctes ainsi que des prévisions de gel incorrectes : les vignerons se lèveront pour mettre en place le système de lutte et ceci sera tout de même payant car sur la nuit il y a effectivement apparition de gel malgré une partie de prévisions incorrectes.
3. Il y a des prévisions de gel correctes ainsi que des événements gélifs manqués : les vignerons mettront en place les bougies. Malgré le fait que des événements gélifs soient manqués, les vignes seront bien protégées grâce aux prévisions de gel correctes sur la nuit.

4. Il y a uniquement des évènements gélifs manqués : alors les vigneron ne seront pas prévenus et les vignes seront victimes du gel.
5. Il n'y a que des prévisions de gel incorrectes : les vigneron vont se lever et effectuer les préparatifs nécessaires alors qu'il n'y aura pas de gel. Perte de temps et d'argent.
6. Dernier cas, des prévisions de gel incorrectes ainsi que des évènements gélifs manqués : bien qu'aucune prévision exacte n'ait été faite, les vigneron vont tout de même recevoir une alerte de gel causée par les prévisions incorrectes et ne vont donc pas rater le gel (évènements gélifs ratés). Cette combinaison de prévision, bien que mauvaise prise heure par heure, va permettre d'avertir correctement les vigneron si l'on prend l'ensemble de la nuit en compte.

Par exemple, si nous considérons le débourrement début mars, sur 46 nuits à risque de gel, nous voyons 13 nuits où soit les prévisions de gel¹¹ sont bonnes soit les évènements gélifs manqués sont compensés par des prédictions de gel incorrectes (les vigneron se lèveraient pour une bonne raison), 3 nuits où le gel est manqué (risque pour vigne), et 30 nuits sans événement prévus ni observés.

¹¹ Nous rappelons que dans ce mémoire, lorsqu'on parle de gel, nous parlons du gel météorologique : lorsque la température est inférieure à 0°C et non du gel viticole.

III.5.2. Analyse sur la nuit des prévisions réalisées par la température actinothermique simulée.

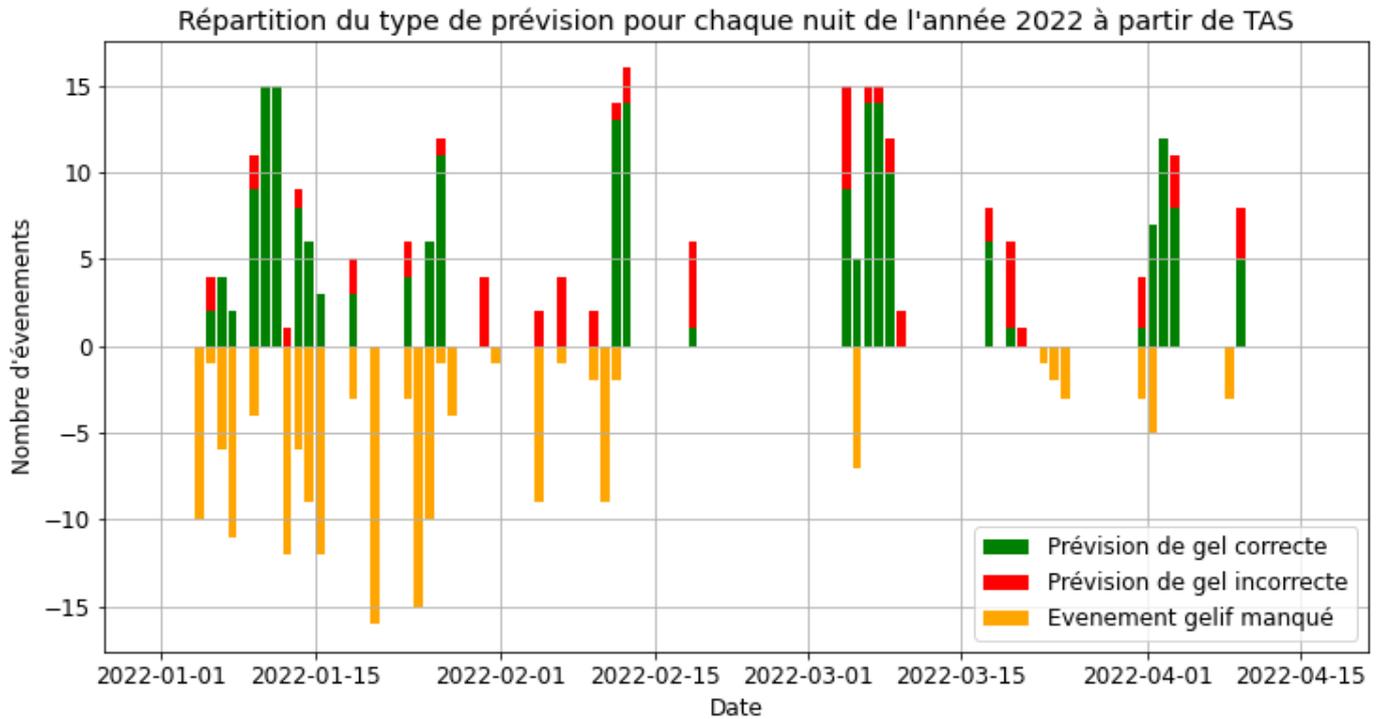


Figure 15 Répartition du type de prévision par nuit (prévision de gel correct (bâtonnet vert), prévision de gel incorrect (bâtonnet rouge), événement gélif manqué (bâtonnet orange)) obtenue par la température actinothermique (formule de Guyot modifiée) lors de l'année 2022. Les prévisions de non gel exactes ne sont pas représentées. Le graphique s'arrête le 15 avril car il n'y plus de gel détecté ensuite pour cette année-là. Le nombre d'évènements gélifs manqués est exprimé en valeur négative afin de faciliter la lecture du graphique.

Une fois de plus si nous considérons le débourrement de la vigne début mars par exemple, sur 46 nuits, il y a 2 nuits de fausses alertes (Figure 15), 4 nuits d'évènements gélif manqués, 12 nuits où, soit le gel est correctement prédits, soit l'alerte est appropriée en raison de la combinaison des prévisions correctes et incorrectes sur la même nuit, et 28 nuits sans gel prévu ou observé.

III.5.3. Analyse sur la nuit des prévisions réalisées par la température minimum du MAR (cas de référence)

Répartition du type de prévision pour chaque nuit de l'année 2022 à partir de la température minimum du MAR

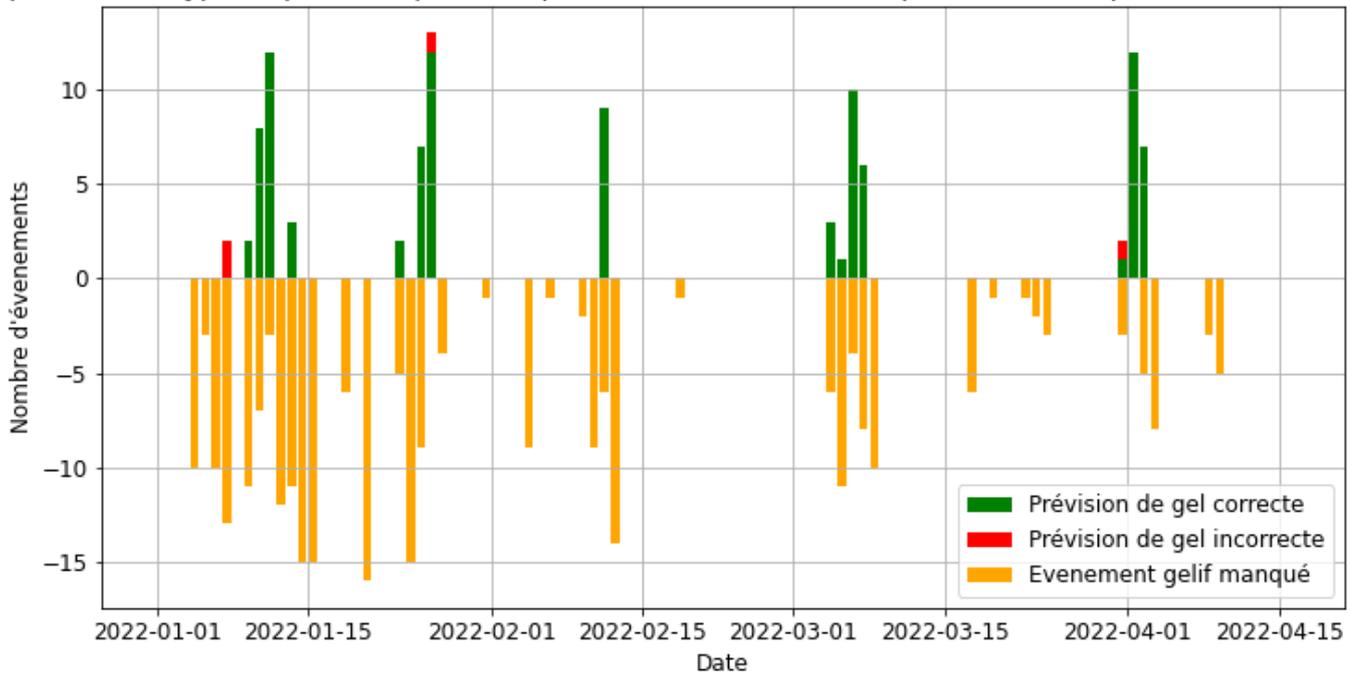


Figure 16 Répartition du type de prévision sur la nuit (prévision de gel correct (bâtonnet vert), prévision de gel incorrect (bâtonnet rouge), événement gélif manqué (bâtonnet orange)) obtenue par la température minimum du MAR lors de l'année 2022. Les prévisions de non gel exactes ne sont pas représentées. Le graphique s'arrête le 15 avril car il n'y plus de gel détecté ensuite pour cette année-là. Le nombre d'évènements gélifs manqués est exprimé en valeur négative afin de faciliter la lecture du graphique.

Enfin, si on regarde pour une sortie du MAR "brute" (la température minimum), nous voyons sur la Figure 16 qu'à partir d'un débourrement début mars, sur 46 nuits, il y a 7 prévisions de gel qui sont correctes, 9 nuits de gel manquées, et 30 nuits sans gel prévu ou observé (les nuits non représentées correspondent à des prévisions de non gel exactes). Ceci nous indique que même lorsqu'on considère le gel sur la nuit et pas pour chaque heure, les simulations brutes de MAR continuent de rater certains évènements gélifs qui ne sont pas ratées par les deux modèles développés dans ce mémoire. Néanmoins, l'avantage de la température minimum du MAR est l'absence de fausses alertes.

Une fois de plus nous voyons que les prévisions d'apparition de gel issues de la régression logistique sont meilleures que celle de la température actinothermique simulée avec une nuit de gel correctement prévue en plus, pas de mauvaises prévisions et un évènement gélif raté en moins. Ces analyses servent à évaluer l'efficacité de l'outil sur une nuit au-delà de son efficacité à prévoir s'il va geler ou non à une heure précise (résultats obtenus à la section

précédente : « II.6. Comparaison des deux meilleurs modèles issus de la régression logistique et de la formule de Guyot modifiée »). Effectivement, nous voyons via ces graphiques que l'outil semble être plus performant sur la nuit que pour une heure précise. Ceci est notamment dû aux compensations qui existent sur la nuit entre les différents types de prévisions (prévision correcte, prévision incorrecte, évènement gélif manqué).

Ces résultats encourageants sont à nuancer : une seule prévision du gel correcte par nuit suffit dans le cas d'une nuit radiative et non pas d'une gelée advective. Car lorsqu'on regarde les prévisions sur la nuit, on ne sait pas précisément à quelle heure la gelée va survenir. Autant l'apparition d'un gel lors d'une nuit radiative est facilement géré par les vignerons autant la gestion d'une gelée advective est beaucoup plus compliquée. De plus, tester la performance des modèles pour des prévisions sur les 4-5 jours futurs est nécessaire afin de voir à quel point la performance de MAR est impactée par le modèle de prévisions qui serait potentiellement le modèle global GFS dans notre cas mais qui est moins précis que les réanalyses utilisées dans ce mémoire. Cette analyse sur les nuits permet de renforcer la confiance dans l'outil opérationnel généré par ces modèles, pour les vignerons. De plus, nous rappelons que cet outil n'est pas là pour remplacer la décision du vigneron mais bien pour l'aiguiller : son expérience personnelle du terrain peut influencer la façon dont une alerte est intégrée dans sa pratique.

Enfin, le modèle TAS et le modèle de la régression logistique sont complémentaires car bien que le modèle logistique prédise mieux l'apparition ou non de gel, la TAS nous donne une indication sur la température prévue. Hiérarchiquement, nous proposons de regarder la prévision du modèle de régression logistique en premier lieu et si une information supplémentaire (la température) est requise, de regarder la TAS. Si le modèle de la régression et la TAS se contredisent, il faut favoriser le modèle de régression logistique.

III.6. Limitations des modèles

III.6.1. Température actinothermique simulée (TAS)

En ce qui concerne la température actinothermique simulée la première limitation vient du fait que l'équation de Guyot n'est pas résolue. Au-delà de la difficulté simplement mathématique de résolution de l'équation (équation du 4^e degré), il faudrait pouvoir connaître les valeurs de h et de ε qui nous sont pour le moment inconnues pour pouvoir résoudre l'équation et enlever l'approximation effectuée grâce à θ . La performance de la température actinothermique simulée est limitée lorsqu'on l'applique sur toute l'année. Ceci est un problème

pour la performance générale du modèle mais n'impacte pas notre modèle prédictif visant à être performant lors du débourrement.

III.6.2. Régression logistique

Contrairement à la température actinothermique simulée, le résultat de la régression logistique est une alerte qui informe si le gel aura lieu ou non et ne donnera aucune indication sur la température. Cela convient aux spécialistes de la vigne mais une information pertinente n'est plus présente, pour connaître l'intensité du gel par exemple. Tout comme pour TAS, la régression a été calibrée sur ce vignoble et n'est donc pas nécessairement transposable sur un autre. Cependant, la détermination de la régression logistique est moins laborieuse que la détermination de la TAS. De plus, la régression logistique montre des meilleurs résultats. Cette approche est préférable si on souhaite connaître les prévisions de gel sur un autre vignoble. Le fait que la régression soit entraînée sur une seule année présente une autre limitation car une année ne représente pas tous les cas d'apparition du gel possible pour les régressions logistiques multivariées¹²: avec plus d'années, le modèle aurait plus d'informations sur les différents cas d'apparition du gel et il serait potentiellement plus performant.

III.6.3. Limitations communes

La première limitation commune est que ces deux modèles ne reposent pas sur des mesures de terrain mais reposent eux-mêmes sur les résultats d'un modèle, celui du MAR. Ceci est une force comme une faiblesse. La force est que l'acquisition de données peut se faire aisément et que nous pouvons prévoir les variables dans le futur. Quand nous comparons aux travaux précédents (Choisnel et al., 1997; Madelin, 2004), l'acquisition de données est beaucoup plus simple et les prévisions peuvent se réaliser sur tous les jours de l'année, puisqu'elles sont générées par le MAR et ne nécessitent pas de mesure laborieuse sur le terrain. La faiblesse est que le modèle de prévision du gel va dépendre de la qualité (l'exactitude) de prévision des variables utiles pour le modèle. Cependant, comme expliqué plus haut dans le travail : le modèle MAR a été validé sur la Belgique (Doutreloup et al., 2022; Fettweis et al., 2017) ce qui minimise cette limitation lorsqu'on travaille sur des domaines belges.

La deuxième limitation est le manque de spatialisation de l'information. Un vignoble est composé d'une multitude de caractéristiques locales déterminantes dans l'apparition du gel

12 Cette limitation ne s'applique pas aux régressions logistiques univariées car la seule relation est entre une variable indépendante et la variable dépendante : or, lorsque la régression est multivariée, il y a plus de possibilité entre les variables indépendantes.

(comme la topographie par exemple) : or, les deux modèles de prévision donnent une seule information) sur 25km² (5km x 5km étant la résolution spatiale du MAR utilisé dans ce travail) qui est censée représenter la température la plus froide du vignoble. Mais à cette résolution spatiale, il est impossible de tenir en compte les caractéristiques locales du site. Cette limitation peut être compensée par l'expérience du vigneron qui sait où est ce point froid et qui doit prendre des mesures de protection là où il pense que cela est pertinent dans son vignoble. En sachant que les bougies que les vignerons utilisent sont chères et qu'il faut donc les utiliser avec parcimonie : une spatialisation de l'information semble capitale (surtout pour des exploitations de taille importante).

La troisième est que ces deux modèles ont été entraînés ou calibrés sur les données de ce vignoble et ne seront donc pas à priori très performants sur de nouveaux vignobles qui ont leurs caractéristiques propres. Il faudrait ré-itérer l'exercice sur ceux-ci.

Néanmoins, les modèles de prévision donnent tout de même une information plus juste que le MAR pour la prévision du gel bien que le taux de FAR soit plus élevé. Les prévisions de Madelin (2004) n'étaient réalisées que pour quelques jours clés et étaient compliquées à mettre en place : là où le modèle de ce mémoire pourrait prévoir une multitude de dates aisément grâce au MAR.

III.7. Comparaison avec la littérature

Si l'on compare tout d'abord avec Choissnel (Figure 3) qui a réalisé des prévisions pour 50 nuits étalés sur deux printemps (1988 et 1989). Premier constat, il n'y avait jamais d'évènement de gel raté chez Choissnel (1993), il a eu à deux endroits différents un jour de fausse alerte et en tout il a réussi à prévoir 5 nuits de gel corrects sur 50 nuits d'observation à 4 endroits différents. Pour nous rapprocher de ce travail, nous allons utiliser le résultat des prévisions de ce mémoire sur les nuits et non pas les heures. Nous rappelons également que Choissnel prenait comme seuil -2.5°C , là où nous utilisons 0°C . Avec le meilleur modèle (régression logistique LASSO) de ce mémoire, sur 46 nuits, nous en observons 0 de fausse alerte contre 1 chez Choissnel, 13 nuits de prédiction correcte et 3 évènements manqués contre 0 chez Choissnel. Étant donné que Choissnel ratait moins d'évènement de gel, son modèle semble meilleur ; cependant, les périodes de comparaison ne sont pas les mêmes ni le seuil d'apparition du gel. Malgré le fait que notre modèle soit plus simple à mettre en place et puisse prévoir plus de dates, le modèle de Choissnel effectue une prévision sur le futur là où nous sommes toujours sur une prévision qui utilise des réanalyses. Lorsque le modèle devra prévoir le futur, le nombre d'erreurs risque d'augmenter.

Ensuite si on compare avec le travail de Madelin qui a pris un seuil de -2°C , l'aléa a bien été prévu (du 5/4/2001 au 31/5/2001) dans 8 cas sur 14 et a donné 5 fausses alertes pour plusieurs stations présentes sur le vignoble champenois. Madelin a donc raté 6 évènements de gel sur 56 nuits. La difficulté de comparaison ici, réside dans le fait que Madelin a effectué le travail sur plusieurs endroits géographiques alors que nous l'avons fait sur un seul, et une fois de plus le seuil d'apparition du gel est différent. Étant donné que nous sommes plus focalisés sur un endroit géographique, il serait normal que nous soyons plus précis mais cela reste à préciser en utilisant un autre vignoble test.

IV. Conclusions et perspectives

En raison de la croissance de vignobles en Belgique (OIV, 2023), le nombre de vignes susceptibles d'être impactées par un gel de printemps au moment du débourrement est en croissance. Dès lors, la création d'un outil opérationnel permettant la prévision du gel pour les vigneron est indispensable afin d'accompagner leur gestion des vignobles. Dans cette optique, ce mémoire a créé deux modèles de prévision du gel horaire sur base des sorties du MAR forcé par les réanalyses ERA5. Le premier modèle, la température actinothermique simulée issue de la formule de Guyot modifiée, a donné les valeurs les plus pertinentes lorsque h valait 11, ϵ valait 0.9 et θ était égale à -1. Cette version de TAS a donné de meilleurs résultats que le MAR pour la prévision de gel avec une POD qui augmente de plus ou moins 30% sur l'ensemble des deux ans et qui voit également le FAR passer à 25%. Ensuite, le deuxième modèle sélectionné parmi ceux générés par régression logistique est celui obtenu par la régression de LASSO (Température minimum du MAR, COD, SLP). Celui-ci a montré de meilleurs résultats encore que le premier modèle. En effet, en prenant la régression logistique avec les variables sélectionnées par LASSO sur l'année 2022 entière, celui-ci a permis de baisser FAR de 4% et d'augmenter POD de 33% par rapport au premier modèle (TAS). Bien qu'on rate moins d'évènements de gel grâce à ces modèles, on en prévoit également plus, mais parfois à tort. Cependant, les spécialistes du monde viticole préfèrent recevoir une fausse alerte plutôt que de rater du gel (bien entendu le nombre de fausses alertes doit être raisonnable, maximum 30% selon les vigneron). De plus, lorsque nous regardons les prévisions non plus de manière horaire mais sur l'ensemble de la nuit, nous nous rendons compte que, pour le meilleur modèle, on n'observe pas de nuit des fausses alertes, 13 nuits où, soit les prévisions de gel sont bonnes soit les prévisions incorrectes et les évènements gélifs manqués se compensent (les vigneron se lèveraient pour une bonne raison) et 3 nuits où le gel est manqué sur 46 nuits, contre 7 prévisions correctes et 9 jours manqués et aucune fausse alerte pour le MAR. Bien que le modèle issu de la régression logistique soit plus performant pour prévoir le gel que TAS, ces deux modèles sont à utiliser ensemble car ils se complètent. Si les deux modèles se contredisent, il semble qu'il faille favoriser les résultats de la régression logistique.

Afin d'approfondir ce travail, la première étape qui permettrait potentiellement d'améliorer le premier modèle (température actinothermique simulée) serait de résoudre l'équation de Guyot afin d'exprimer la température des puces à partir de la température de l'air (donc du MAR). Il faudra trouver les valeurs exactes de la conductivité ainsi que la constante

ε. Ensuite, il faudra analyser les différentes solutions de l'équation et déterminer la/les plus plausibles dans la détermination de la température actinothermique simulée. Une deuxième étape serait de collecter une année de données supplémentaires ce qui permettrait de confirmer ou d'infirmer l'efficacité du deuxième modèle de ce mémoire (la régression logistique). Pour se faire, les données du MAR utilisées devront être des données simulées dans le futur proche forcé par un modèle global de prévision et non plus des réanalyses. De plus, si ce travail était à reproduire sur un autre vignoble, une fois les deux modèles de ce mémoire testés, nous recommanderions la régression logistique au lieu de la température actinothermique simulée qui, dans ce mémoire, est plus performante et est moins contraignante à appliquer. Il serait tout de même intéressant de tester les deux types de modèle sur un autre vignoble car ils sont complémentaires. Enfin et comme déjà mentionné plus haut dans le mémoire, une spatialisation de l'information ou une augmentation de la résolution de l'information augmenterait la robustesse de l'outil et l'application que les vignerons en feraient.

Bibliographie

- Atlas de Belgique. (2023). *Données cartographiques – Atlas de Belgique*. <https://www.atlas-belgique.be/index.php/fr/ressources/donnees-cartographiques/>
- Baggiolini, M. (1952). Les stades repères dans le développement de la vigne et leur utilisation pratique. *Revue Romande d'agriculture, de Viticulture et d'arboriculture*, 8, 4–6.
- Battany, M. C. (2012). Vineyard frost protection with upward-blowing wind machines. *Agricultural and Forest Meteorology*, 157, 39–48.
<https://doi.org/10.1016/j.agrformet.2012.01.009>
- Beltrando, G. (1998). Les gelées printanières en Champagne viticole - Quelques résultats obtenus à partir d'un nouveau réseau de stations automatiques. *La Météorologie*, 1998(21), 30–43. <https://doi.org/10.4267/2042/47040>
- Broeckx, J., Maertens, M., Isabirye, M., Vanmaercke, M., Namazzi, B., Deckers, J., Tamale, J., Jacobs, L., Thiery, W., Kervyn, M., Vranken, L., & Poesen, J. (2019). Landslide susceptibility and mobilization rates in the Mount Elgon region, Uganda. *Landslides*, 16(3), 571–584. <https://doi.org/10.1007/s10346-018-1085-y>
- Brunt, D. (1941). *Physical en dynamical meteorology*. Cambridge University Press.
- Choisnel, E. (2002). Agrométéorologie. In *Encyclopædia Universalis* (Vol. 1, pp. 601–605). Encyclopædia Britannica.
- Choisnel, E., Jourdain, S., Camblan, M., Klis-Lilienthal, I., & Mac Veigh, J. P. (1993). *La Météorologie 8 e série-n° 2-juin 1993*.
- Clément, J.-M. (1981). *Larousse Agricole*. Larousse.
- De Ridder, K., & Gallée, H. (1998). Land Surface–Induced Regional Climate Change in Southern Israel. *Journal of Applied Meteorology*, 37(11), 1470–1485.
[https://doi.org/10.1175/1520-0450\(1998\)037<1470:LSIRCC>2.0.CO;2](https://doi.org/10.1175/1520-0450(1998)037<1470:LSIRCC>2.0.CO;2)
- Dereuddre, J., Audran, J., Leddet, C., Barka, E. A., & Brun, O. (1993). Réponse de la vigne (*Vitis vinifera* L) aux températures inférieures à 0°C. III. Effets d'un refroidissement contrôlé sur des bourgeons au cours du débourrement. *Agronomie*, 13(6), 509–514.
<https://hal.science/hal-00885569>

- Ding, L., Noborio, K., & Shibuya, K. (2019). Frost forecast using machine learning - From association to causality. *Procedia Computer Science*, 159, 1001–1010.
<https://doi.org/10.1016/j.procs.2019.09.267>
- Doutreloup, S., Bois, B., Pohl, B., Zito, S., & Richard, Y. (2022). Climatic comparison between Belgium, Champagne, Alsace, Jura and Bourgogne for wine production using the regional model MAR. *Oeno One*, 56(3). <https://doi.org/10.20870/oenone.2022.56.3.5356>
- Fettweis, X., Franco, B., Tedesco, M., van Angelen, J. H., Lenaerts, J. T. M., van den Broeke, M. R., & Gallée, H. (2013). Estimating the Greenland ice sheet surface mass balance contribution to future sea level rise using the regional atmospheric climate model MAR. *The Cryosphere*, 7(2), 469–489. <https://doi.org/10.5194/tc-7-469-2013>
- Fettweis, X., Wyard, C., Doutreloup, S., & Belleflamme, A. (2017). Noël 2010 en Belgique : neige en Flandre et pluie en Haute-Ardenne. *Bulletin de La Société Géographique*, 68, 97–107.
- FGOV Economy. (2021, June 27). *Un quart de viticulteurs belges en plus en 2020 - FOD Economie (communiqué de presse)*. FOD Economie Salle de Presse.
<https://news.economie.fgov.be/199451-un-quart-de-viticulteurs-belges-en-plus-en-2020>
- Fluss, R., Faraggi, D., & Reiser, B. (2005). Estimation of the Youden Index and its Associated Cutoff Point. *Biometrical Journal*, 47(4), 458–472.
<https://doi.org/10.1002/bimj.200410135>
- Fraga, H., García de Cortázar Aauri, I., Malheiro, A. C., & Santos, J. A. (2016). Modelling climate change impacts on viticultural yield, phenology and stress conditions in Europe. *Global Change Biology*, 22(11), 3774–3788. <https://doi.org/10.1111/gcb.13382>
- François, M. (2021, May 15). Le Parisien. *Gel : Plus de 4 Milliards d'euros de Pertes Estimées Dans La Viticulture et l'arboriculture*.
- Galet, P. (2000). *Précis de viticulture : à l'usage des ingénieurs agronomes, des étudiants en agronomie et en oenologie, des techniciens des lycées agricoles, et des professionnels de la viticulture* (P. Galet, Ed.).

- Gavrilescu, C., Bois, B., Castel, T., Larmure, A., Ouvrié, M., & Richard, Y. (2019). *Analyse spatiale de l'évolution du risque de gel sur la vigne en Bourgogne-Franche-Comté*. 181–186.
- Gobbett, D. L., Nidumolu, U., & Crimp, S. (2020). Modelling frost generates insights for managing risk of minimum temperature extremes. *Weather and Climate Extremes*, 27. <https://doi.org/10.1016/j.wace.2018.06.003>
- Guyot, G. (2013). *Climatologie de l'environnement Cours et exercices corrigés* (Dunod, Ed.; 2e édition).
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., ... Thépaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>
- IPCC, Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Gosh, S., Iksanfar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S. M., Wehner, M., & Zhou, B. (2021). *Weather and Climate Extreme Events in a Changing Climate. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*.
- Leerschool, A. (2022). *A climate forecasting model to assist Belgian wine-growers against bud freezing*. Uliège.
- Loseke, B. A., Read, P. E., & Blankenship, E. E. (2015). Preventing spring freeze injury on grapevines using multiple applications of Amigo Oil and naphthaleneacetic acid. *Scientia Horticulturae*, 193, 294–300. <https://doi.org/10.1016/j.scienta.2015.07.025>
- Madelin, M. (2004). *L'aléa gélif printanier dans le vignoble marnais en Champagne. Modélisation spatiale à une échelle fine des écoulements de l'air et des températures minimales*. <https://theses.hal.science/tel-00008906>
- Mason, S. J., & Graham, N. E. (2002). Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society*, 128(584 PART B), 2145–2166. <https://doi.org/10.1256/003590002320603584>

- Meier, L., Van De Geer, S., & Bühlmann, P. (2008). The group lasso for logistic regression. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 70(1), 53–71. <https://doi.org/10.1111/j.1467-9868.2007.00627.x>
- Meier, M., Fuhrer, J., & Holzkämper, A. (2018). Changing risk of spring frost damage in grapevines due to climate change? A case study in the Swiss Rhone Valley. *International Journal of Biometeorology*, 62(6), 991–1002. <https://doi.org/10.1007/s00484-018-1501-y>
- Mohd Razali, N., & Bee Wah, Y. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. In *Journal of Statistical Modeling and Analytics* (Vol. 2, Issue 1).
- Molitor, D., Caffarra, A., Sinigoj, P., Pertot, I., Hoffmann, L., & Junk, J. (2014). Late frost damage risk for viticulture under future climate conditions: A case study for the Luxembourgish winegrowing region. *Australian Journal of Grape and Wine Research*, 20(1), 160–168. <https://doi.org/10.1111/ajgw.12059>
- OIV. (2023). *Base de données / OIV*. Oiv.Int. <https://www.oiv.int/fr/what-we-do/data-discovery-report?oiv>
- Pérarnaud, V., & Raynal, N. (1991). *Agrométéorologie*. Ecole nationale de la météorologie.
- Poling, E. B. (2008). Spring Cold Injury to Winegrapes and Protection Strategies and Methods. *HortScience*, 43(6), 1652–1662. <https://doi.org/10.21273/HORTSCI.43.6.1652>
- Reynier, A. (1991). *Manuel de viticulture* (Tec & Doc-Lavoisier).
- Sallis, P., Jarur, M., Trujillo, M., & Ghobakhlou, A. (2009). Frost Prediction using a Combinational Model of Supervised and Unsupervised Neural Networks for Crop Management in Vineyards. In *18 th World IMACS / MODSIM Congress*. <http://mssanz.org.au/modsim09>
- Santibáñez F, Morales L, de La Fuente J, Cellier P, & Huete A. (1997). Topoclimatic modeling for minimum temperature prediction at a regional scale in the Central Valley of Chile. *Agronomie*, 17(6–7), 307–314.
- Sofaer, H. R., Hoeting, J. A., & Jarnevich, C. S. (2019). The area under the precision-recall curve as a performance metric for rare binary events. *Methods in Ecology and Evolution*, 10(4), 565–577. <https://doi.org/10.1111/2041-210X.13140>

- Stoltzfus, J. C. (2011). Logistic regression: A brief primer. *Academic Emergency Medicine*, 18(10), 1099–1104. <https://doi.org/10.1111/j.1553-2712.2011.01185.x>
- Vanel, M. (2021, April 12). *Gel dans le vignoble wallon: relativement peu de dégâts, mais rien n'est gagné* - Marc Vanel. <https://www.marcvanel.be/gel-2021-wallonie/>
- WalOnMap. (2023). *Relief de la Wallonie - Combinaison MNT, MNS et Hillshade 2013-2014 / Géoportail de la Wallonie*. <https://geoportail.wallonie.be/catalogue/300a9c30-f7de-474e-98f1-b43f5296e880.html>
- Wang, K., Sun, J., Cheng, G., & Jiang, H. (2011). Effect of altitude and latitude on surface air temperature across the Qinghai-Tibet Plateau. *Journal of Mountain Science* 2011 8:6, 8(6), 808–816. <https://doi.org/10.1007/S11629-011-1090-2>
- Willwerth, J., Ker, K., & Inglis, D. (2014). *Best Management Practices for Reducing Winter Injury in Grapevines*.

Annexe

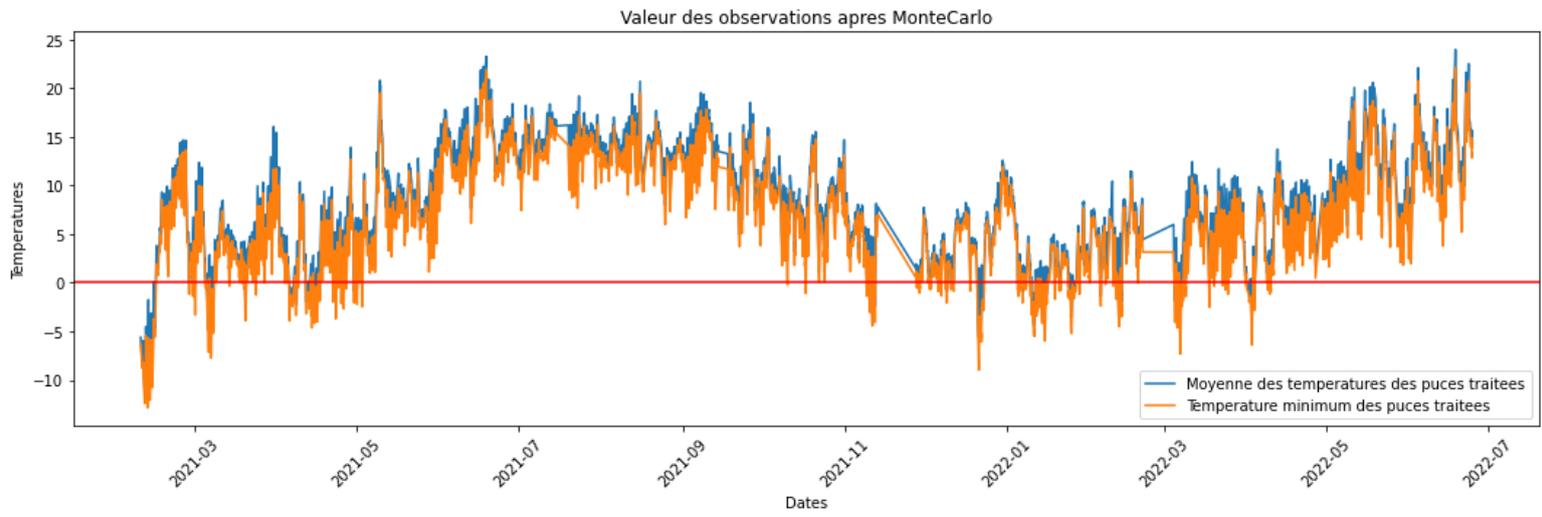


Figure 17 : Aperçu des valeurs de l'ensemble des puces après l'application de Monte Carlo. La température minimum (courbe orange) correspond à la température moyenne (courbe bleue) moins deux écarts type.

Variation de la formule de Guyot modifiée
Température minimum du MAR et minimum des puces

Θ variation										
Corrélation	RMSE	Biais	RMSE centré	POD	FAR	Accuracy	albedo	ε	h	Θ
0.95	2.80	2.01	1.93	0.46	0.12	0.91	0.2	0.9	11	-5
0.87	2.26	1.31	1.84	0.54	0.17	0.85	0.2	0.9	11	-5
0.87	2.67	1.65	2.10	0.59	0.16	0.91	0.2	0.9	11	-5
0.95	2.66	1.81	1.93	0.48	0.14	0.91	0.2	0.9	11	-4.5
0.87	2.16	1.13	1.84	0.57	0.18	0.85	0.2	0.9	11	-4.5
0.87	2.55	1.46	2.09	0.60	0.20	0.91	0.2	0.9	11	-4.5
0.95	2.52	1.61	1.93	0.50	0.15	0.91	0.2	0.9	11	-4
0.87	2.07	0.94	1.84	0.60	0.19	0.86	0.2	0.9	11	-4
0.87	2.44	1.26	2.09	0.61	0.20	0.91	0.2	0.9	11	-4
0.95	2.39	1.41	1.92	0.52	0.17	0.91	0.2	0.9	11	-3.5
0.86	1.98	0.75	1.83	0.64	0.20	0.86	0.2	0.9	11	-3.5
0.87	2.34	1.07	2.09	0.63	0.22	0.91	0.2	0.9	11	-3.5
0.95	2.28	1.21	1.92	0.55	0.19	0.92	0.2	0.9	11	-3
0.86	1.92	0.55	1.83	0.68	0.20	0.87	0.2	0.9	11	-3
0.87	2.26	0.87	2.09	0.64	0.27	0.90	0.2	0.9	11	-3
0.95	2.17	1.01	1.92	0.57	0.19	0.92	0.2	0.9	11	-2.5
0.86	1.87	0.36	1.83	0.70	0.21	0.87	0.2	0.9	11	-2.5
0.87	2.19	0.67	2.08	0.67	0.28	0.90	0.2	0.9	11	-2.5
0.95	2.08	0.81	1.92	0.60	0.21	0.92	0.2	0.9	11	-2
0.86	1.84	0.17	1.83	0.71	0.22	0.87	0.2	0.9	11	-2
0.870	2.133	0.472	2.081	0.707	0.299	0.897	0.2	0.9	11.0	-2.0
0.947	2.009	0.603	1.913	0.624	0.225	0.919	0.2	0.9	11.0	-1.5
0.864	1.833	-0.027	1.833	0.734	0.224	0.874	0.2	0.9	11.0	-1.5
0.870	2.096	0.272	2.078	0.741	0.301	0.900	0.2	0.9	11.0	-1.5
0.947	1.954	0.398	1.910	0.656	0.243	0.919	0.2	0.9	11.0	-1.0
0.864	1.845	-0.223	1.833	0.783	0.243	0.876	0.2	0.9	11.0	-1.0
0.870	2.077	0.071	2.076	0.810	0.299	0.907	0.2	0.9	11.0	-1.0
0.95	1.92	0.19	1.91	0.68	0.26	0.92	0.2	0.9	11	-0.5
0.86	1.88	-0.42	1.83	0.81	0.27	0.87	0.2	0.9	11	-0.5
0.87	2.08	-0.13	2.07	0.84	0.31	0.91	0.2	0.9	11	-0.5
0.95	1.90	-0.01	1.90	0.72	0.28	0.92	0.2	0.9	11	0
0.86	1.93	-0.62	1.83	0.85	0.30	0.86	0.2	0.9	11	0
0.87	2.10	-0.33	2.07	0.84	0.32	0.90	0.2	0.9	11	0
0.95	1.91	-0.22	1.90	0.76	0.30	0.92	0.2	0.9	11	0.5
0.86	2.00	-0.82	1.83	0.87	0.34	0.85	0.2	0.9	11	0.5
0.87	2.14	-0.54	2.07	0.85	0.34	0.90	0.2	0.9	11	0.5
0.95	1.95	-0.43	1.90	0.79	0.31	0.92	0.2	0.9	11	1

0.86	2.09	-1.02	1.83	0.90	0.36	0.84	0.2	0.9	11	1
0.87	2.20	-0.74	2.07	0.86	0.36	0.89	0.2	0.9	11	1
0.95	2.00	-0.64	1.90	0.83	0.32	0.92	0.2	0.9	11	1.5
0.86	2.19	-1.22	1.83	0.91	0.38	0.83	0.2	0.9	11	1.5
0.87	2.27	-0.95	2.06	0.88	0.39	0.88	0.2	0.9	11	1.5
0.95	2.07	-0.85	1.89	0.85	0.35	0.91	0.2	0.9	11	2
0.86	2.31	-1.42	1.83	0.93	0.40	0.82	0.2	0.9	11	2
0.87	2.37	-1.16	2.06	0.90	0.41	0.87	0.2	0.9	11	2
0.95	2.17	-1.07	1.89	0.87	0.36	0.91	0.2	0.9	11	2.5
0.86	2.44	-1.62	1.83	0.94	0.41	0.81	0.2	0.9	11	2.5
0.87	2.47	-1.37	2.06	0.90	0.45	0.85	0.2	0.9	11	2.5
0.95	2.28	-1.28	1.89	0.88	0.39	0.90	0.2	0.9	11	3
0.86	2.58	-1.83	1.83	0.95	0.43	0.80	0.2	0.9	11	3
0.87	2.59	-1.58	2.06	0.90	0.47	0.84	0.2	0.9	11	3
0.95	2.40	-1.50	1.89	0.90	0.40	0.90	0.2	0.9	11	3.5
0.86	2.73	-2.03	1.83	0.96	0.45	0.78	0.2	0.9	11	3.5
0.87	2.72	-1.79	2.06	0.91	0.49	0.84	0.2	0.9	11	3.5
0.95	2.54	-1.71	1.89	0.91	0.42	0.89	0.2	0.9	11	4
0.86	2.88	-2.24	1.83	0.97	0.46	0.77	0.2	0.9	11	4
0.87	2.87	-2.00	2.05	0.91	0.50	0.83	0.2	0.9	11	4
0.95	2.69	-1.93	1.88	0.93	0.45	0.88	0.2	0.9	11	4.5
0.86	3.05	-2.45	1.83	0.98	0.48	0.75	0.2	0.9	11	4.5
0.87	3.02	-2.21	2.05	0.94	0.52	0.81	0.2	0.9	11	4.5

Variation de h

Corrélation	RMSE	Biais	RMSE centré	POD	FAR	Accuracy	albedo	ε	h	Θ
0.33	25.11	-19.26	16.12	0.90	0.81	0.41	0.2	0.8	1	-1
0.24	29.30	-22.88	18.33	0.97	0.67	0.47	0.2	0.8	1	-1
0.38	31.87	-26.68	17.43	0.96	0.80	0.33	0.2	0.8	1	-1
0.95	1.95	0.65	1.83	0.63	0.21	0.92	0.2	0.8	11	-1
0.88	1.73	0.07	1.73	0.74	0.22	0.88	0.2	0.8	11	-1
0.88	2.01	0.41	1.97	0.76	0.27	0.91	0.2	0.8	11	-1
0.96	2.34	1.60	1.70	0.48	0.11	0.92	0.2	0.8	21	-1
0.90	1.94	1.16	1.54	0.53	0.16	0.85	0.2	0.8	21	-1
0.90	2.43	1.70	1.73	0.55	0.09	0.91	0.2	0.8	21	-1
0.96	2.60	1.94	1.72	0.43	0.08	0.91	0.2	0.8	31	-1
0.89	2.22	1.55	1.58	0.46	0.13	0.84	0.2	0.8	31	-1
0.89	2.77	2.16	1.73	0.47	0.05	0.90	0.2	0.8	31	-1
0.96	2.75	2.11	1.75	0.40	0.06	0.91	0.2	0.8	41	-1
0.89	2.39	1.75	1.62	0.42	0.11	0.83	0.2	0.8	41	-1
0.89	2.96	2.40	1.74	0.41	0.04	0.89	0.2	0.8	41	-1
0.95	2.85	2.21	1.77	0.38	0.05	0.91	0.2	0.8	51	-1
0.88	2.50	1.87	1.65	0.39	0.09	0.83	0.2	0.8	51	-1
0.89	3.09	2.54	1.76	0.37	0.02	0.89	0.2	0.8	51	-1
0.95	2.91	2.28	1.79	0.37	0.04	0.91	0.2	0.8	61	-1
0.88	2.58	1.95	1.67	0.38	0.07	0.83	0.2	0.8	61	-1
0.89	3.17	2.63	1.77	0.34	0.02	0.88	0.2	0.8	61	-1
0.95	2.96	2.33	1.80	0.35	0.04	0.90	0.2	0.8	71	-1

0.88	2.63	2.01	1.69	0.36	0.06	0.83	0.2	0.8	71	-1
0.89	3.23	2.70	1.78	0.31	0.00	0.88	0.2	0.8	71	-1
0.95	3.00	2.37	1.81	0.35	0.03	0.90	0.2	0.8	81	-1
0.87	2.67	2.05	1.70	0.36	0.01	0.83	0.2	0.8	81	-1
0.88	3.28	2.76	1.79	0.31	0.00	0.88	0.2	0.8	81	-1
0.95	3.02	2.40	1.82	0.35	0.03	0.90	0.2	0.8	91	-1
0.87	2.71	2.09	1.71	0.35	0.01	0.83	0.2	0.8	91	-1
0.88	3.32	2.80	1.79	0.31	0.00	0.88	0.2	0.8	91	-1

ε variation

Corrélation	RMSE	Biais	RMSE centré	POD	FAR	Accuracy	albedo	ε	h	Θ
0.95	3.28	2.67	1.89	0.30	0.02	0.90	0.2	0.0	11	-1
0.86	3.00	2.38	1.81	0.30	0.00	0.82	0.2	0.0	11	-1
0.88	3.64	3.14	1.85	0.22	0.00	0.86	0.2	0.0	11	-1
0.95	3.03	2.42	1.82	0.34	0.03	0.90	0.2	0.1	11	-1
0.87	2.71	2.09	1.71	0.35	0.01	0.83	0.2	0.1	11	-1
0.88	3.32	2.80	1.78	0.30	0.00	0.88	0.2	0.1	11	-1
0.95	2.80	2.16	1.76	0.39	0.06	0.91	0.2	0.2	11	-1
0.89	2.44	1.80	1.62	0.40	0.10	0.83	0.2	0.2	11	-1
0.89	3.01	2.46	1.74	0.41	0.02	0.90	0.2	0.2	11	-1
0.96	2.58	1.91	1.72	0.44	0.08	0.91	0.2	0.3	11	-1
0.89	2.19	1.51	1.57	0.46	0.13	0.84	0.2	0.3	11	-1
0.89	2.73	2.12	1.72	0.48	0.05	0.91	0.2	0.3	11	-1
0.96	2.39	1.66	1.70	0.47	0.10	0.91	0.2	0.4	11	-1
0.90	1.98	1.23	1.54	0.51	0.15	0.85	0.2	0.4	11	-1
0.90	2.48	1.78	1.73	0.53	0.09	0.91	0.2	0.4	11	-1
0.96	2.22	1.41	1.70	0.51	0.13	0.92	0.2	0.5	11	-1
0.90	1.81	0.94	1.54	0.58	0.18	0.86	0.2	0.5	11	-1
0.89	2.27	1.43	1.76	0.62	0.11	0.92	0.2	0.5	11	-1
0.96	2.08	1.15	1.73	0.56	0.14	0.92	0.2	0.6	11	-1
0.89	1.71	0.65	1.58	0.65	0.19	0.87	0.2	0.6	11	-1
0.89	2.11	1.09	1.81	0.71	0.14	0.93	0.2	0.6	11	-1
0.95	1.99	0.90	1.77	0.60	0.18	0.92	0.2	0.7	11	-1
0.89	1.68	0.36	1.64	0.70	0.20	0.87	0.2	0.7	11	-1
0.89	2.02	0.75	1.88	0.76	0.21	0.92	0.2	0.7	11	-1
0.95	1.95	0.65	1.83	0.63	0.21	0.92	0.2	0.8	11	-1
0.88	1.73	0.07	1.72	0.74	0.22	0.88	0.2	0.8	11	-1
0.88	2.01	0.41	1.97	0.76	0.27	0.91	0.2	0.8	11	-1
0.95	1.95	0.40	1.91	0.66	0.24	0.92	0.2	0.9	11	-1
0.86	1.84	-0.22	1.83	0.78	0.24	0.88	0.2	0.9	11	-1
0.87	2.08	0.07	2.08	0.81	0.30	0.91	0.2	0.9	11	-1
0.94	2.01	0.15	2.00	0.68	0.29	0.91	0.2	1	11	-1
0.85	2.02	-0.51	1.96	0.83	0.32	0.85	0.2	1	11	-1
0.86	2.21	-0.27	2.20	0.85	0.32	0.91	0.2	1	11	-1
0.94	2.11	-0.11	2.11	0.70	0.34	0.90	0.2	1.1	11	-1
0.83	2.24	-0.80	2.10	0.85	0.37	0.83	0.2	1.1	11	-1
0.85	2.41	-0.61	2.33	0.89	0.39	0.88	0.2	1.1	11	-1
0.93	2.25	-0.36	2.23	0.71	0.37	0.90	0.2	1.2	11	-1

0.81	2.50	-1.09	2.25	0.88	0.40	0.81	0.2	1.2	11	-1
0.84	2.65	-0.95	2.47	0.90	0.45	0.85	0.2	1.2	11	-1
0.92	2.43	-0.61	2.36	0.72	0.40	0.89	0.2	1.3	11	-1
0.80	2.78	-1.38	2.41	0.89	0.43	0.79	0.2	1.3	11	-1
0.83	2.92	-1.29	2.62	0.90	0.50	0.83	0.2	1.3	11	-1
0.92	2.63	-0.86	2.49	0.72	0.44	0.88	0.2	1.4	11	-1
0.78	3.07	-1.67	2.59	0.90	0.47	0.76	0.2	1.4	11	-1
0.81	3.22	-1.63	2.78	0.90	0.54	0.80	0.2	1.4	11	-1
0.91	2.85	-1.11	2.63	0.73	0.48	0.86	0.2	1.5	11	-1
0.76	3.38	-1.96	2.76	0.90	0.49	0.75	0.2	1.5	11	-1
0.80	3.54	-1.97	2.94	0.90	0.58	0.76	0.2	1.5	11	-1
0.90	3.09	-1.37	2.78	0.73	0.50	0.85	0.2	1.6	11	-1
0.74	3.70	-2.25	2.94	0.91	0.51	0.72	0.2	1.6	11	-1
0.79	3.88	-2.31	3.11	0.90	0.61	0.74	0.2	1.6	11	-1
0.89	3.34	-1.62	2.93	0.74	0.52	0.84	0.2	1.7	11	-1
0.72	4.02	-2.54	3.13	0.91	0.53	0.71	0.2	1.7	11	-1
0.78	4.22	-2.65	3.28	0.90	0.64	0.70	0.2	1.7	11	-1
0.88	3.61	-1.87	3.09	0.74	0.54	0.83	0.2	1.8	11	-1
0.70	4.35	-2.83	3.32	0.91	0.54	0.69	0.2	1.8	11	-1
0.76	4.57	-3.00	3.45	0.90	0.65	0.70	0.2	1.8	11	-1
0.87	3.87	-2.12	3.25	0.75	0.56	0.82	0.2	1.9	11	-1
0.68	4.69	-3.12	3.51	0.92	0.55	0.69	0.2	1.9	11	-1
0.75	4.93	-3.34	3.63	0.90	0.67	0.67	0.2	1.9	11	-1

Monte Carlo: θ variation

Corrélation	RMSE	Biais	RMSE centré		Accuracy	albed				
			POD	FAR		o	ε	h	Θ	
0.95	2.80	2.01	1.93	0.47	0.12	0.91	0.2	0.9	11	-5
0.87	2.27	1.32	1.84	0.54	0.17	0.85	0.2	0.9	11	-5
0.87	2.67	1.65	2.09	0.59	0.16	0.91	0.2	0.9	11	-5
0.95	2.66	1.81	1.93	0.48	0.13	0.91	0.2	0.9	11	-4.5
0.87	2.16	1.13	1.84	0.57	0.18	0.85	0.2	0.9	11	-4.5
0.87	2.55	1.46	2.09	0.60	0.20	0.91	0.2	0.9	11	-4.5
0.95	2.52	1.61	1.93	0.50	0.15	0.91	0.2	0.9	11	-4
0.87	2.07	0.94	1.84	0.59	0.19	0.85	0.2	0.9	11	-4
0.87	2.44	1.26	2.09	0.61	0.20	0.91	0.2	0.9	11	-4
0.95	2.39	1.41	1.92	0.52	0.17	0.91	0.2	0.9	11	-3.5
0.86	1.99	0.75	1.83	0.65	0.20	0.86	0.2	0.9	11	-3.5
0.87	2.34	1.07	2.09	0.63	0.21	0.91	0.2	0.9	11	-3.5
0.95	2.28	1.21	1.92	0.55	0.19	0.92	0.2	0.9	11	-3
0.86	1.92	0.56	1.83	0.68	0.20	0.87	0.2	0.9	11	-3
0.87	2.26	0.87	2.09	0.64	0.27	0.90	0.2	0.9	11	-3
0.95	2.17	1.01	1.92	0.57	0.19	0.92	0.2	0.9	11	-2.5
0.86	1.87	0.36	1.83	0.70	0.21	0.87	0.2	0.9	11	-2.5
0.87	2.19	0.67	2.08	0.67	0.28	0.90	0.2	0.9	11	-2.5
0.95	2.08	0.81	1.92	0.60	0.21	0.92	0.2	0.9	11	-2
0.86	1.84	0.17	1.83	0.71	0.22	0.87	0.2	0.9	11	-2
0.87	2.13	0.47	2.08	0.72	0.30	0.90	0.2	0.9	11	-2
0.95	2.01	0.60	1.91	0.63	0.23	0.92	0.2	0.9	11	-1.5

0.86	1.83	-0.03	1.83	0.73	0.22	0.87	0.2	0.9	11	-1.5
0.87	2.10	0.27	2.08	0.74	0.31	0.90	0.2	0.9	11	-1.5
0.95	1.95	0.40	1.91	0.66	0.24	0.92	0.2	0.9	11	-1
0.86	1.85	-0.22	1.83	0.79	0.24	0.88	0.2	0.9	11	-1
0.87	2.08	0.07	2.08	0.81	0.30	0.91	0.2	0.9	11	-1
0.95	1.92	0.19	1.91	0.68	0.26	0.92	0.2	0.9	11	-0.5
0.86	1.88	-0.42	1.83	0.81	0.27	0.87	0.2	0.9	11	-0.5
0.87	2.08	-0.13	2.07	0.84	0.31	0.91	0.2	0.9	11	-0.5
0.95	1.90	-0.01	1.90	0.72	0.28	0.92	0.2	0.9	11	0
0.86	1.93	-0.62	1.83	0.85	0.30	0.86	0.2	0.9	11	0
0.87	2.10	-0.33	2.07	0.84	0.32	0.90	0.2	0.9	11	0
0.95	1.91	-0.22	1.90	0.76	0.30	0.92	0.2	0.9	11	0.5
0.86	2.00	-0.82	1.83	0.87	0.35	0.84	0.2	0.9	11	0.5
0.87	2.14	-0.54	2.07	0.85	0.34	0.90	0.2	0.9	11	0.5
0.95	1.95	-0.43	1.90	0.79	0.31	0.92	0.2	0.9	11	1
0.86	2.09	-1.02	1.83	0.90	0.36	0.84	0.2	0.9	11	1
0.87	2.20	-0.74	2.07	0.86	0.36	0.89	0.2	0.9	11	1
0.95	2.00	-0.64	1.90	0.83	0.33	0.92	0.2	0.9	11	1.5
0.86	2.19	-1.22	1.83	0.92	0.38	0.83	0.2	0.9	11	1.5
0.87	2.27	-0.95	2.06	0.87	0.39	0.88	0.2	0.9	11	1.5
0.95	2.07	-0.85	1.89	0.85	0.35	0.91	0.2	0.9	11	2
0.86	2.31	-1.42	1.83	0.93	0.40	0.82	0.2	0.9	11	2
0.87	2.36	-1.16	2.06	0.89	0.41	0.87	0.2	0.9	11	2
0.95	2.17	-1.07	1.89	0.87	0.36	0.91	0.2	0.9	11	2.5
0.86	2.44	-1.62	1.83	0.94	0.41	0.81	0.2	0.9	11	2.5
0.87	2.47	-1.37	2.06	0.90	0.45	0.86	0.2	0.9	11	2.5
0.95	2.28	-1.28	1.89	0.88	0.39	0.90	0.2	0.9	11	3
0.86	2.58	-1.83	1.83	0.95	0.43	0.79	0.2	0.9	11	3
0.87	2.59	-1.58	2.06	0.90	0.46	0.85	0.2	0.9	11	3
0.95	2.40	-1.50	1.89	0.90	0.40	0.90	0.2	0.9	11	3.5
0.86	2.73	-2.03	1.83	0.96	0.45	0.78	0.2	0.9	11	3.5
0.87	2.72	-1.79	2.06	0.91	0.49	0.84	0.2	0.9	11	3.5
0.95	2.54	-1.71	1.89	0.91	0.42	0.89	0.2	0.9	11	4
0.86	2.88	-2.24	1.83	0.97	0.46	0.77	0.2	0.9	11	4
0.87	2.87	-2.00	2.05	0.91	0.50	0.83	0.2	0.9	11	4
0.95	2.69	-1.93	1.88	0.93	0.45	0.88	0.2	0.9	11	4.5
0.86	3.05	-2.45	1.83	0.98	0.48	0.75	0.2	0.9	11	4.5
0.87	3.02	-2.21	2.05	0.94	0.52	0.81	0.2	0.9	11	4.5

Tableau 22 Toutes les variations pour trouver la meilleure formule de Guyot modifiée avec la température minimum du MAR et en comparant les résultats avec la température minimum des puces. Variation de : θ , ε et h . Les valeurs en vert sont les plus pertinentes et celles en rouge/orange celles que nous ne voulons pas.

**Sélection de variables via LASSO avec
TAS**

Variables	Coefficient
TAS	-0.059
Gradient	0.000
Vent10m	0.000
COD	-0.005
CM	0.000
CD	0.000
CU	0.000
SP	0.000
SLP	0.000
PP	0.000

Tableau 23 Résultats de la sélection de LASSO sur l'ensemble des variables indépendantes sélectionnées avec la température actinothermique (TAS) comme variable fortement corrélée conservée.