

Rapport technique :

QUALITOMIR :

***Qualité des tomates destinées à
la transformation par
l'infrarouge moyen***

Avancement du projet et réalisations
lors de la saison 2024



I. QUALITOMIR	1
I.1 Rappel du projet.....	1
I.2 Objectifs	1
I.3 Réalisations écrites.....	2
II. Méthodologie pour l'analyse des échantillons	3
II.1 Récolte des échantillons	3
II.2 Préparation des échantillons	3
II.3 Transformation des échantillons.....	3
II.4 Analyses de références.....	4
III. Méthodologie pour l'acquisition spectrale.....	4
III.1 Matériel utilisé	4
III.2 Analyses chimiométriques.....	5
IV. Modélisation de la qualité (PLSR)	5
IV.1 Stratégies de modélisation	5
IV.2 Echantillons frais	7
IV.3 Echantillons HB et CB.....	9
IV.4 Produits finis.....	10
V. Discrimination des échantillons.....	11
V.1 En fonction de l'année.....	11
V.2 En Fonction du type.....	12
CONCLUSION	14
ANNEXES.....	15
Table des Figures & Tableaux :	17
Bibliographie :.....	18

I. QUALITOMIR

I.1 Rappel du projet

Le projet QUALITOMIR (Qualité des tomates destinées à la transformation par l'infrarouge) dirigé par la SONITO au sein du PNDAR du CTIFL, s'inscrit dans l'Action Elémentaire 1 qui consiste à « Renforcer la souveraineté, la valeur et la qualité des fruits et légumes par la captation des innovations variétales, l'étude des préférences consommateurs et la maîtrise de la variabilité dans l'itinéraire. » Plus précisément, la tâche 4 de cet AE 1 consiste en la « Mise en œuvre d'outils haut débit de caractérisation de la qualité et de sa variabilité sur légumes transformés ». Ce projet mené par la SONITO s'étend jusqu'en 2027 et doit permettre en partie de répondre à l'objectif opérationnel 2 (Caractériser et maîtriser la variabilité des produits par le développement d'outils à moyen ou haut débit de phénotypage. Maintenir la qualité des produits lors de la phase post-récolte).

A terme, la mise en place d'une base de données spectrales IR doit permettre d'aider les agriculteurs ainsi que les techniciens des OP dans leurs choix de variétés vis-à-vis des contraintes agronomiques de leurs zones de production mais également de mieux répondre aux demandes des industriels en termes de caractéristiques des tomates livrées. Durant la phase de transformation des tomates, l'objectif est d'apporter un outil fiable et robuste aux industriels pour permettre une surveillance en continu de la qualité et de son évolution au cours du temps.

I.2 Objectifs

Tout d'abord, l'utilisation de la spectroscopie infrarouge doit permettre de former une base de données spectrale. Cette base de données a pour but de faciliter la caractérisation des différentes variétés de tomates destinées à la transformation. De manière générale, avec l'arrivée des machines possédant un fort niveau de calcul, l'utilisation de spectre IR à transformer de Fourier s'est fortement développée. Elle apparait aujourd'hui comme une méthode fiable et robuste permettant d'obtenir des données rapidement sur la qualité des tomates par quantification de certains de ses composés. Ainsi, il est aujourd'hui possible de remplacer les différentes techniques d'analyses traditionnelles par la spectroscopie infra rouge (réflexion totale atténuée). Bien entendu, il est primordial de posséder en premier lieu des données acquises avec les techniques classiques pour permettre de faire des comparaisons avec les données spectrales.

En entrée d'usine, la matière brute est caractérisée pour différents paramètres de qualité. La matière sèche soluble est l'un des paramètres les plus contrôlés, mais il existe également d'autres critères dont la place devient de plus en plus importante. Parmi eux, on peut citer le pourcentage de matière sèche totale, qui prend aussi en compte des composés insolubles (Bureau et al., 2020), l'acidité titrable et la viscosité. L'utilisation de la spectroscopie dans l'infrarouge moyen doit donc permettre de caractériser les tomates brutes qui arrivent en usines pour être transformées. En comparaison avec les méthodes classiques d'analyses, l'utilisation du MIR présente des avantages au niveau du gain de temps et de la précision des résultats. Il n'a plus de biais liés à l'expérimentateur ou au protocole réalisé.

Durant la transformation des tomates, cet outil doit également permettre de surveiller l'évolution des produits au cours du temps. Un paramètre clé de réussite pour une bonne transformation est la viscosité. Ce paramètre est fortement dépendant de la température à laquelle les fruits sont broyés et de la vitesse à laquelle l'eau est évaporée (Bureau et al., 2020 ; Page et al., 2012 ; Barrett et al., 1998). Comme la totalité des différents mécanismes qui influent sur la viscosité ne sont pas connus, les industriels adaptent leur process de cuisson en fonction de leurs observations et de leurs mesures. Il en est de même pour les caroténoïdes et notamment le lycopène qui est l'un des composés les plus abondants dans la tomate, mais dont la quantité disponible dans les produits transformés varie en fonction du type de cuisson (Page et al., 2019).

Ainsi, l'utilisation du MIR assurerait une surveillance en continu de l'évolution du produit et faciliterait la mise en place d'actions correctives durant la cuisson. En sortie d'usine, avant commercialisation des produits, la spectroscopie IR doit aussi assurer un rôle de vérification de la qualité. Cet outil peut être utilisé à 3 étapes durant la chaîne de transformation (Bureau et al., 2020).

L'objectif est d'améliorer les modèles de prédiction en ajoutant de nouveaux échantillons provenant des années suivantes. En suivant la chronologie de l'arrivée des échantillons aux usines, les modèles seront adaptés à la prédiction de ces nouveaux échantillons.

I.3 Réalisations écrites

Un article scientifique relatif aux travaux menés en 2022 et 2023 a été soumis et validé pour publication dans *Acta Horticulturae*. Cette production a aussi été présentée lors du WPTC (Congrès mondial de la tomate d'industrie), à Budapest en Hongrie en juin 2024.

II. Méthodologie pour l'analyse des échantillons

II.1 Récolte des échantillons

Pour rappel, durant la saison 2022, 93 échantillons provenant de 36 variétés avaient été récoltés dans les bassins de production Sud-est et Sud-Ouest. Au cours de l'année 2023, 152 nouveaux échantillons ont été collectés. Cette diversité d'échantillons provenait de 38 variétés différentes. La matière brute était issue majoritairement d'essais de variétés, mais également d'essais sur systèmes de culture en paillage biodégradable, en réduction de l'irrigation et en réduction des intrants de manière plus générale.

Durant la saison 2024, 205 nouveaux échantillons frais ont été analysés. Il provenait de 37 variétés différentes. En plus de ces échantillons frais, 29 produits finis ont aussi été analysés. Contrairement aux années précédentes, certains échantillons frais ne provenaient pas directement des parcelles. Leur prélèvement s'est effectué dans les bennes des camions lors de leur arrivée aux usines de transformation. Enfin, certains échantillons HB et CB (Hot Break et Cold Break) ont été directement prélevés durant la transformation en usine, après la phase de cuisson. Pour les échantillons prélevés sur parcelle : chaque échantillon provenait d'un mélange de 20 fruits qui avaient été récoltés sur 3 micro-parcelles de 4 plants. Ainsi, chaque échantillon est issu de 12 plants identiques en termes de modalités. Les récoltes ont été effectuées durant la période de juillet à fin septembre.

II.2 Préparation des échantillons

Environ 1kg de chaque échantillon (15/20 tomates) a été séparé en 3 parties équitables pour pouvoir ensuite réaliser les analyses sur échantillons frais et transformés (HB et CB.) Pour la préparation préalable des échantillons pour les caractériser en frais et en CB, les tomates coupées grossièrement ont été mixées avec un blender à puissance maximale pendant 30 secondes. Une partie de la purée pour la caractérisation en frais a été congelée dans des pots à échantillons de 15ml pour la réalisation d'analyses complémentaires à l'INRAE d'Avignon.

II.3 Transformation des échantillons

Pour la réalisation des échantillons HB et CB, le protocole de transformation était basé sur la méthode fournie par Page et al. (2019). Les échantillons HB sous forme de cubes grossiers étaient chauffés dans un micro-ondes pendant 4 minutes à 900 watts puis broyés et mis en bocaux de verre de 37cl. La méthode pour faire des échantillons CB était la même, mis à part que

les échantillons étaient déjà broyés pour permettre une macération pendant 30s (activation des enzymes responsables de la dégradation des pectines) avant d'être chauffés au micro-ondes. Pour les deux types d'échantillons, les analyses ont été réalisées 24 heures après la cuisson pour laisser la viscosité se stabiliser (Page et al., 2012). Un schéma récapitulatif du protocole réalisé est présent en **Erreur ! Source du renvoi introuvable..**

II.4 Analyses de références

Différentes mesures ont été réalisées sur les purées de tomates fraîches. La matière sèche soluble a été caractérisée durant la saison à l'aide d'un réfractomètre PAL-1 (ATAGO). Chaque mesure a été répétée 5 fois et la moyenne a été calculée (°BRIX). Certaines mesures présentaient un fort écart type qui était provoqué par l'émulsion à la suite du passage au blender. D'autres analyses ont été faites au mois de novembre à l'INRAE d'Avignon. La matière sèche soluble a été caractérisée une seconde fois pour corriger la valeur de la première mesure sur certains échantillons. La matière sèche (%) a été mesurée par pesées et passage dans une étuve pendant 72h. Enfin, l'acidité titrable a été mesurée par titrage acido-basique avec une solution de NaOH (0.1 N). Le résultat a été exprimé en mmol H⁺/100g de matière fraîche. Pour les échantillons HB et CB, seule la viscosité a été déterminée. La viscosité a été mesurée à l'aide d'un consistomètre de Bostwick avec une cuvette et un réglet mesurant 24cm.

Durant la saison 2024, les mesures de la matière sèche soluble, de l'acidité titrable et de la matière sèche totale ont été réalisées sur tous les échantillons frais ainsi que les produits finis. Comme pour les années précédentes, les échantillons transformés en laboratoire (HB et CB) n'ont été caractérisés que pour la viscosité.

III. Méthodologie pour l'acquisition spectrale

III.1 Matériel utilisé

L'acquisition spectrale a été réalisée au CTCPA (Centre Technique de Conservation des Produits Agricoles) à l'aide d'un spectromètre ALPHA II (Bruker) sur le logiciel OPUS 8.7.31. L'outil était équipé d'un accessoire à réflectance totale atténuée composé d'un cristal en sélénium de zinc (ZnSe) à 6 réflexions. Un détecteur DLATGS (Deuterated Lanthanum Alpha Alanine doped TriGlycine Sulphate) a enregistré les spectres à température ambiante. L'acquisition spectrale a été réalisée sur une plage allant de 4000 cm⁻¹ à 600 cm⁻¹ (2,5 μm à 25μm) avec une résolution de 4cm⁻¹, une vitesse de balayage de 10kHz et un arrière-plan de 24 scans. Un blanc a été réalisé entre chaque mesure d'échantillons en lançant une analyse avec le cristal vierge lavé et essuyé.

III.2 Analyses chimiométriques

La totalité de l'analyse des données spectrales a été réalisée sur le logiciel CHEMFLOW 20.05. Dans un premier temps, la plage spectrale a été réduite de 2000 cm^{-1} à 800 cm^{-1} car elle correspond à l'empreinte digitale des composés à analyser avec notamment les vibrations fondamentales C-C et C-O (Ajib, 2013). Pour corriger les effets additifs et multiplicatifs, le prétraitement par SNV (Standard Normal Variate) a été appliqué. Dans certains cas, il a été possible d'utiliser la dérivé seconde en suivant l'algorithme de Savitzky-Golay. Différents outils mathématiques ont été utilisés comme l'ACP et le clustering hiérarchique pour fournir une première approche exploratoire des données. Dans un second temps, la régression par les moindres carrés partiels (PLSR) avec validation croisée (4 blocs, 10 répétitions en aléatoire) a permis la mise en place des modèles prédictifs pour les paramètres de qualité. Les échantillons « outliers » ont été enlevés des modèles par observation de leur Q et T^2 . Le choix du nombre de variables latentes des modèles a été fait avec différents éléments tel que le critère de Durbin Watson et la variance des coefficients de régression. La qualité des modèles de régression a été vérifiée en observant les valeurs du $R^2\text{CV}$, $R^2\text{P}$, RMSECV et du RMSEP, puis par le calcul du RPD. Enfin, l'analyse factorielle discriminante (AFD) a permis de discriminer les échantillons en fonction de l'année de récolte, de leur provenance et de leur état (frais, HB et CB). Dans ce nouveau rapport technique, le cheminement et la méthodologie utilisée pour traiter les échantillons n'ont pas fondamentalement changé. Les résultats obtenus avec les analyses exploratoires ne seront pas présentés car l'objectif premier de ces travaux est de prédire les paramètres de qualité et de pouvoir séparer les échantillons.

IV. Modélisation de la qualité (PLSR)

IV.1 Stratégies de modélisation

Tableau 1 : Présentation des différentes stratégies de modélisation.

Paramètres :	Type	Calibration et validation croisée (échantillons pour créer le modèle)	Validation externe (échantillons à prédire)
MSS	Frais	2022 à 2024	Aucune
MSS	Frais	2022 + 2023	2024
MSS	Frais	2022+ 2023 + Précoce 2024	Tardive 2024
AT	Frais	2023	2024
AT	Frais	2024	2023
AT	Frais	2023 + 2024	Aucune
AT	Frais	2023 + Précoce 2024	Tardive 2024
MST	Frais	2023	2024
MST	Frais	2024	2023
MST	Frais	2023 + 2024	Aucune
MST	Frais	2023 + Précoce 2024	Tardive 2024
Viscosité	HB & CB	2023	2024
Viscosité	HB & CB	2024	2023
Viscosité	HB & CB	2023 + 2024	Aucune
Viscosité	HB & CB	2023 + Précoce 2024	Tardive 2024
SSC	Produits finis	2024	Aucune
AT	Produits finis	2024	Aucune
MS	Produits finis	2024	Aucune

Les différentes stratégies présentées dans le Tableau 1 ne sont pas exhaustives, il est toujours possible de séparer les échantillons de différentes manières. Par exemple, il a été décidé de ne pas faire de modèle antichronologique pour la matière sèche soluble alors que c'est le cas pour les autres paramètres. Plus le nombre d'années augmente, plus les champs de modélisation se multiplient. Cette augmentation des possibilités permet de maximiser les chances d'obtenir des segmentations d'échantillons robustes, mais augmente aussi les risques de ne pas bien traiter l'information. Si chaque stratégie de modélisation est couplée avec des prétraitements différents et l'utilisation de bandes spectrales alors les possibilités deviennent encore plus grandes. Il

devient alors très difficile de faire les choix les plus judicieux. Pour cela, peut importe les paramètres modélisés, un seul prétraitement et une bande spectrale ont été retenus.

De plus, l'utilisation de la spectroscopie IR, dans ce contexte, a pour but de permettre une utilisation industrielle répondant aux besoins de certains professionnels. Ainsi, certaines stratégies pourraient paraître plus performantes, tout en n'étant pas adaptées aux attentes relatives aux conditions d'utilisation de l'outils IR dans les usines de transformation de tomates.

IV.2 Echantillons frais

Tableau 2 : Performance des modèles prédictifs sur les échantillons frais des années 2022 à 2024

Paramètres	Calibration & Validation croisée	Validation externe	Variables latentes	R ² CV	RMSEC	RMSECV	R ² P	RMSEP
MSS	2022 à 2024	Aucune	7	0.92	0.162	0.169	Na	Na
MSS	2022 et 2023	2024	7	0.93	0.164	0.175	0.88	0.171
MSS	2022, 2023 et précoce 2024	Tardive 2024	7	0.92	0.161	0.171	0.86	0.174
AT	2023 et 2024	Aucune	11	0.94	0.214	0.242	Na	Na
AT	2023	2024	8	0.92	0.223	0.260	0.73	0.697
AT	2024	2023	5	0.92	0.262	0.302	0.83	0.495
AT	2023 et précoce 2024	Tardive 2024	8	0.90	0.262	0.290	0.81	0.442
MST	2023 et 2024	Aucune	6	0.79	0.310	0.339	Na	Na
MST	2023	2024	6	0.90	0.295	0.301	0.75	0.452
MST	2024	2023	5	0.72	0.309	0.324	0.71	0.432
MST	2023 et précoce 2024	Tardive 2024	5	0.78	0.298	0.328	0.71	0.411

Les différents modèles prédictifs réalisés en 2024 ont démontré l'importance de l'ajout de nouveaux échantillons dans l'amélioration de la robustesse. De manière générale, la prédiction des échantillons de nouvelles années peut être plus difficile. Lorsque la part d'échantillons dans

le jeu de validation externe est très différente de ceux présents dans le jeu de calibration et de validation croisée, les performances des caractéristiques prédictives sont diminuées.

Pour la matière sèche soluble, les 3 modèles testés étaient performants avec $R^2CV > 0.92$ La racine carré des erreurs moyennes de validation croisées était toujours inférieure à 0.18° BRIX. Lors de la validation externe, il y avait $R^2P > 0.86$ et des RMSEP similaires à ceux de la validation croisée. Ces différents modèles montrent que l'infra rouge moyen est propice à prédire des échantillons variés, qui diffèrent par leur année de récolte. Cette succession d'années permettra d'améliorer encore les modèles dans le temps. L'utilisation de 3 années différentes permet de démontrer qu'avec un grand nombre d'échantillons, les modèles deviennent de plus en plus robustes.

La prédiction de l'acidité titrable a elle aussi été encourageante. La modélisation de ce paramètre n'a pour le moment été réalisée que sur des échantillons provenant de deux années consécutives. Les valeurs des R^2CV et $RMSECV$ montre une bonne performance des modèles en validation interne. En revanche, la prédiction des échantillons de 2024 en utilisant ceux de 2023 s'est montrée moins performante ($R^2P = 0.73$ & $RMSEP = 0.697$) en comparaison avec la prédiction des échantillons de 2023 avec les données de 2024 ($R^2P = 0.83$ & $RMSEP = 0.495$). Actuellement, seul les outliers les plus extrêmes (mauvaise mesure avec l'appareil de référence) ont été supprimés des modèles, et des ajustements des modèles permettront d'améliorer ces 2 modèles.

Ensuite, la prédiction de la matière sèche totale a été plus complexe, avec des modèles moins performants. Pour les échantillons de 2023, le modèle est performant en validation interne, mais ne permet pas de déterminer avec précision les valeurs en matière sèche totale des échantillons de 2024. Lorsque les échantillons de 2024 sont utilisés pour former et entraîner les modèles de régression, les prédictions ne sont pas assez performantes. C'est le seul paramètre pour lequel une grosse différence est observée entre les 2 années d'acquisition des données. Cependant il ne s'agit pas seulement d'une trop grande variabilité des échantillons de 2024 par rapport à ceux de 2023 car dans ce cas il aurait été possible d'obtenir un modèle performant en interne avec uniquement les données de 2024. La méthodologie pour réaliser les mesures de références n'a pourtant pas changer, et les données ont été revérifiées plusieurs fois. Il en est de même pour la méthodologie d'acquisition des spectres. Une explication possible serait un dérèglement des balances de précision au cours de la semaine de mesure, car celles-ci avaient été testées avant

le début des analyses. Il reste donc des améliorations à réaliser sur ces modèles pour les rendre plus performants, si cela est possible.

IV. 3 Echantillons HB et CB

Tableau 3 : performance des modèles prédictifs sur les échantillons HB et CB préparés en laboratoire et provenant d'une usine de transformation.

Paramètres	Calibration & Validation croisée	Validation externe	Variables latentes	R ² CV	RMSEC	RMSECV	R ² P	RMSEP
Viscosité	2023 et 2024	Aucune	10	0.89	2.25	2.39	Na	Na
Viscosité	2023	2024	9	0.90	1.98	2.17	0.80	3.65
Viscosité	2024	2023	6	0.75	3.72	3.95	0.74	3.83
Viscosité	2023 et précoce 2024	Tardive 2024	5	0.89	2.29	2.40	0.82	3.40

Lors de la saison 2023, les analyses ont démontré qu'il n'était pas possible de séparer les lots d'échantillons HB et CB pour les traiter séparément comme 2 échantillons distincts en prédisant HB avec CB par exemple. Ils ont donc été regroupés et les différentes stratégies de modélisation de sont concentrées sur la segmentation chronologique de l'acquisition des échantillons.

L'utilisation des échantillons de 2023 et 2024 permet d'obtenir de bonnes prédictions en validation internes. L'utilisation des échantillons de 2023 pour prédire ceux de 2024 est aussi une stratégie performante avec $R^2P = 0.8$ et $RMSEP = 3.65$. Il a été plus compliqué de réaliser l'opération en sens inverse, avec $R^2P = 0.74$ et $RMSEP = 3.83$. En ne prenant en compte que les modèles qui permettent d'avoir une validation externe, l'ajout des données précoces de 2024 s'est montré bénéfique pour accroître la performance de la prédiction ($R^2P = 0.82$ et $RMSEP = 3.40$). Il est important de rappeler qu'une variabilité supplémentaire a été introduite avec la méthode de cuisson des échantillons qui était différente pour ceux provenant des usines de transformation, légèrement différentes du protocole réalisé en laboratoire avec des micro-ondes.

IV.4 Produits finis

Tableau 4 : Performance des modèles prédictifs sur les échantillons de produits finis

Paramètres	Calibration & Validation croisée	Validation externe	Variables latentes	R ² CV	RMSEC	RMSECV	R ² P	RMSEP
MSS	2024	Aucune	6	0.99	0.122	0.134	Na	Na
TA	2024	Aucune	6	0.99	0.114	0.129	Na	Na
MST	2024	Aucune	6	0.95	0.148	0.176	Na	Na

L'analyse des produits finis n'avait pas encore été possible pour les années précédentes. Ces échantillons proviennent des stocks de traçabilité restant dans les usines de transformation de tomates. Contrairement aux échantillons frais ainsi que cuits en laboratoire, ils ne sont pas monovariétaux. Les paramètres analysés ici, sont ceux qui intéressent les industriels pour caractériser la qualité de leur produit avant commercialisation. Les 3 modèles réalisés sont très performants. Cependant, il n'est pas encore possible de valider leur robustesse pour 2 principales raisons. La première est qu'ils ne sont basés que sur une seule année, ce qui diminue l'hétérogénéité et la variabilité. Deuxièmement, il n'a pas été simple de s'en procurer un nombre conséquent. Les modèles réalisés sont donc basés sur des petites quantités d'échantillons, qui ne présentent pas de forte hétérogénéité par leur nombre, bien que ces différents produits soient très variés les uns des autres (jus, triple concentré...)

V. Discrimination des échantillons

V.1 En fonction de l'année

Tableau 5 : Matrice de classification en fonction des années de récoltes

	Référence 2022	Référence 2023	Référence 2024
Prédiction 2022	66	1	3
Prédiction 2023	7	361	34
Prédiction 2024	20	69	602

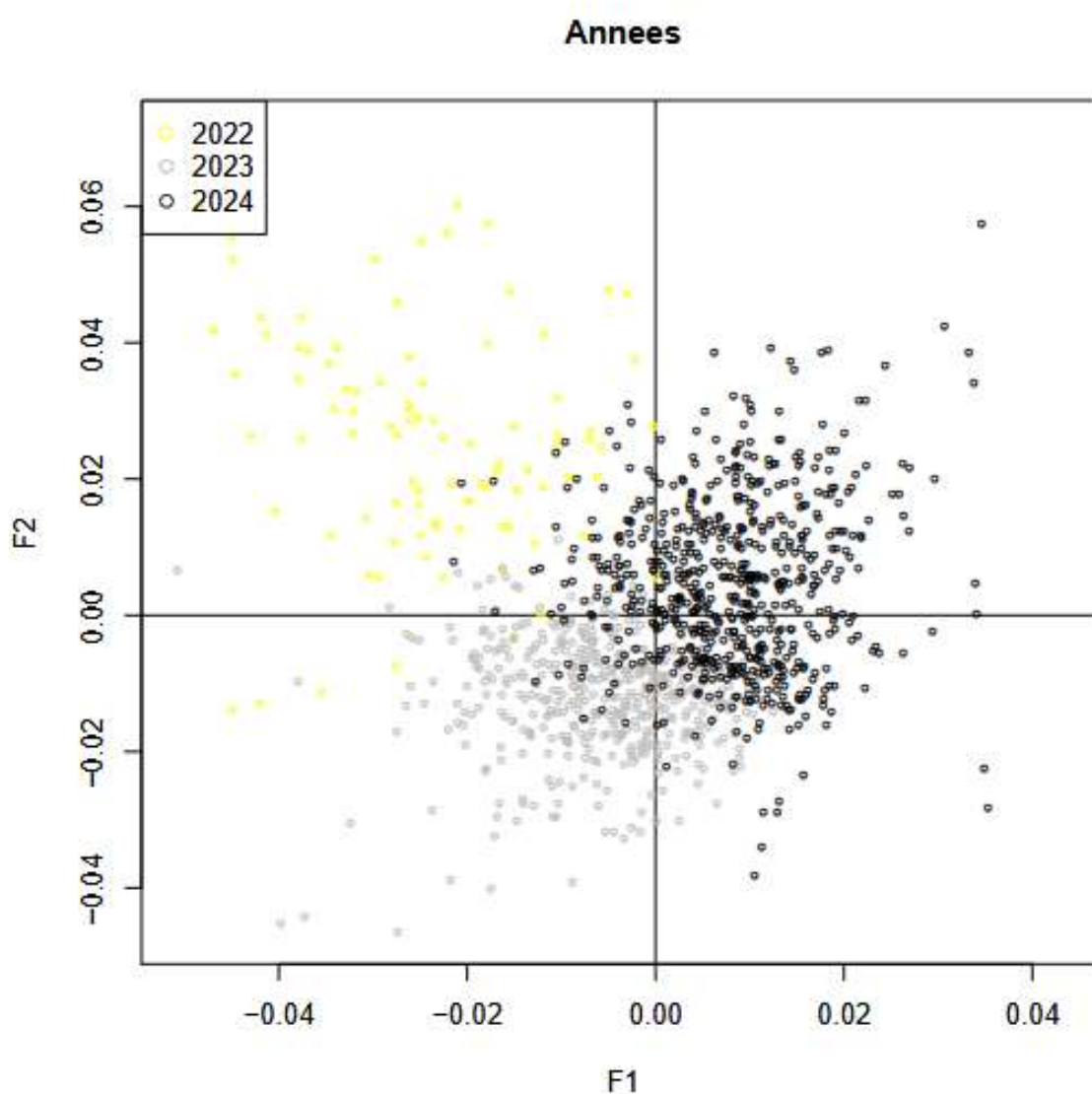


Figure 1 : Représentation des échantillons sur le repère F1 et F2, coloration en fonction de l'année de récolte

La spectroscopie infrarouge permet également de séparer des échantillons en fonction de plusieurs critères. L'objectif est d'améliorer la traçabilité des échantillons en obtenant des

informations sur leur origine, la présence d'adultérations, leur année de récolte ou encore déterminer de manière précise le type de transformations réalisés. Plusieurs approches mathématiques ont été étudiées pour réaliser ce travail de discrimination. La PLS-DA s'est montrée efficace pour séparer les échantillons. Cette méthode à l'avantage de pouvoir être réalisée directement sur la totalité des variables spectrales, car elle est adaptée à des variables fortement corrélées entre elles. La FDA ou AFD (Analyse factorielle discriminante) a aussi permis d'avoir de faibles erreurs de classification. Pour pouvoir utiliser l'AFD, il est nécessaire de préparer le jeu de données. Il est obligatoire que les différentes variables retenues ne soient pas corrélées entre elles. Il est possible de sélectionner les variables les plus importantes pour la discrimination avec la méthode Cosvel, mais ce n'est pas la technique utilisée dans notre cas. Pour la réalisation de la FDA, une première approche exploratoire avec une ACP a permis de former 20 nouvelles variables (les composantes principales de l'ACP). La FDA a donc été appliquée directement sur les scores de l'ACP en fonction des 20 composantes principales. L'avantage de l'utilisation des composantes de l'ACP est qu'elles sont, par définition, non corrélées les unes par rapport aux autres. Pour rappel, la première est formée de sorte à représenter le maximum de variabilité du jeu de donnée alors que la seconde se forme de manière orthogonale à la première et donc représente les résidus non pris en compte par la première.

La FDA s'est montrée très efficace pour séparer les échantillons suivant les années de récoltes. L'erreur de classification en validation croisée est de 12,7%, ce résultat est correct mais il doit être possible d'obtenir une meilleure classification. Le tableau 5 présente le nombre d'échantillons qui ont été correctement discriminés. La principale source d'erreur de classification provient d'une inversion entre certains échantillons de 2023 et de 2024. On peut émettre l'hypothèse qu'ils étaient plus proches en comparaison avec ceux de 2022, cependant une forte part des échantillons de 2022 a été classifiée comme provenant de 2024.

V.2 En Fonction du type

Tableau 6 : Matrice de classification en fonction du type de transformation

	Référence F	Référence HB	Référence CB
Prédiction F	446	1	2
Prédiction HB	4	335	2
PrédictionCB	0	6	335

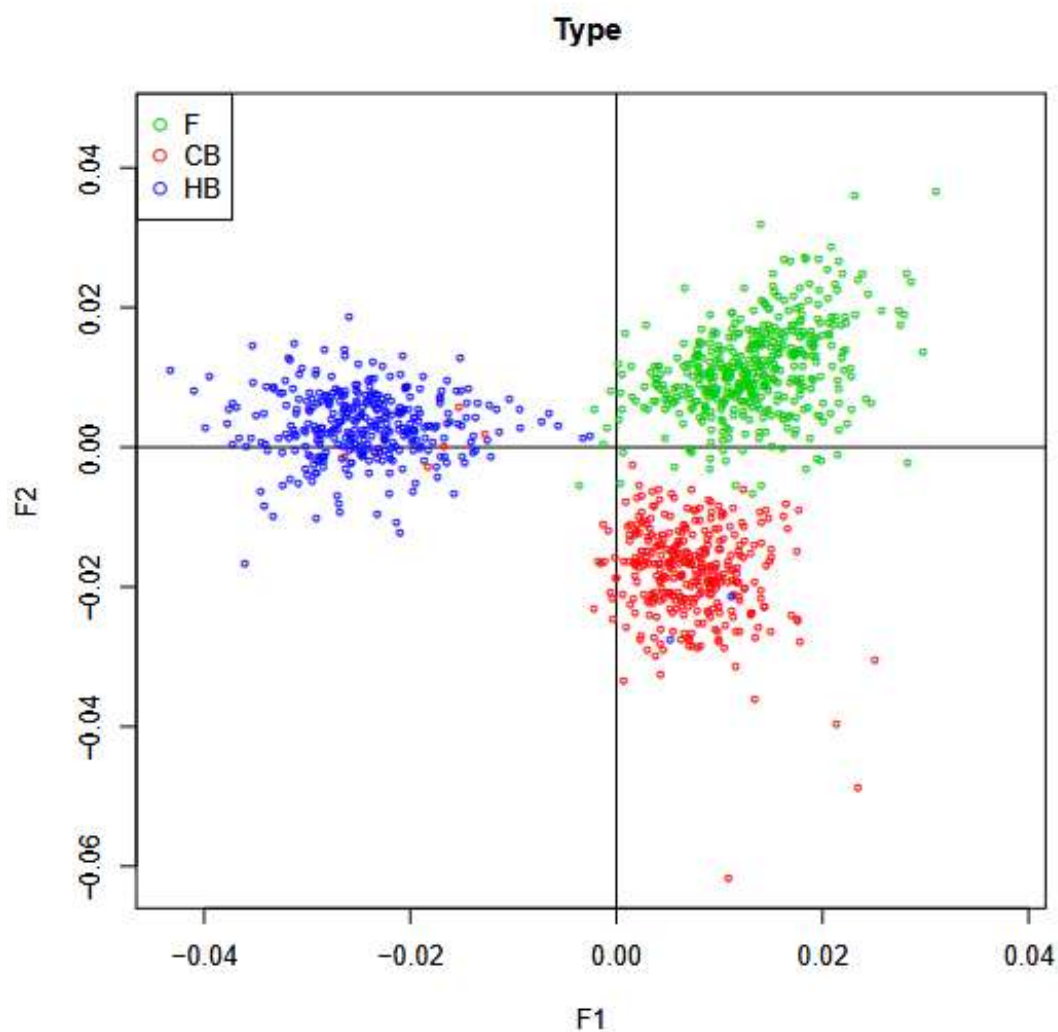


Figure 2 : Représentation des échantillons sur le repère F1 et F2, coloration en fonction de l'année et du lieu de plantation.

La FDA a permis de séparer efficacement les échantillons en fonction de leur type. Les produits n'ont pas été introduits dans ce modèle discriminant car leur nombre n'était pas suffisant. L'erreur moyenne de classification en validation croisée est de 1,6%, ce qui est très faible. Le Tableau 6 présente le nombre d'échantillons bien classés. Sur 448 échantillons frais, uniquement 4 n'ont pas été attribués au bon groupe. Pour les échantillons HB, 335 ont été bien classés sur 342, et pour les échantillons CB, 4 seulement n'ont pas été bien classés.

Ces résultats montrent la capacité de l'analyse infrarouge couplé avec des outils de chimiométrie à séparer et à regrouper des échantillons n'ayant pas eu les mêmes conditions de cuissons.

CONCLUSION

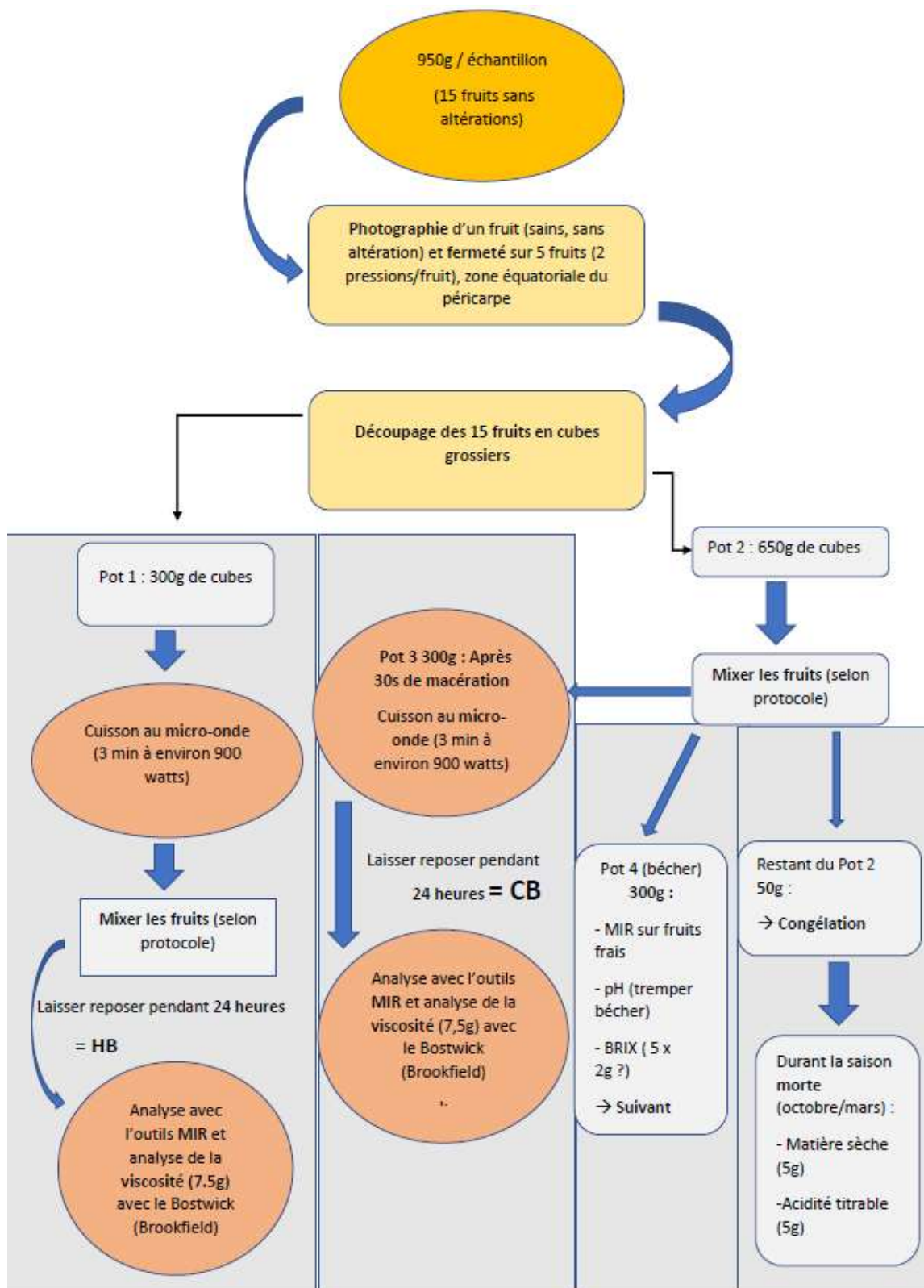
Pour conclure, l'utilisation de la spectroscopie dans l'infrarouge moyen est une technique fiable pour la prédiction des paramètres de la qualité des tomates destinées à la transformation.

L'augmentation de la diversité des échantillons permet d'améliorer les modèles et de les rendre plus robustes. Certains paramètres de qualité tels que la matière sèche totale sont plus difficiles à prédire et des ajustements doivent encore être réalisés pour pouvoir réellement utiliser la spectroscopie infra rouge en conditions industrielles.

La saison 2022 a permis de découvrir l'outil et de déterminer comment régler les modèles pour les rendre plus performants. La saison 2023 a montré qu'il était possible d'utiliser cet outil sur des échantillons nouveaux, ce qui a également été fait durant la saison 2024 sur un plus grand nombre de paramètres de qualité. Lors des saisons suivantes, il va maintenant être nécessaire de déterminer comment rendre les modèles utilisables sur des échantillons dont les analyses spectrales ont été réalisées sur des appareils différents sans perdre en performance.

ANNEXES

ANNEXE 1 : Protocole réalisé pour les analyses en frais, et sur produits transformés.



ANNEXE 2 : Schéma de la division des échantillons en un jeu d'apprentissage (étalonnage et validation) pour la validation croisée et un jeu de test pour valider le modèle

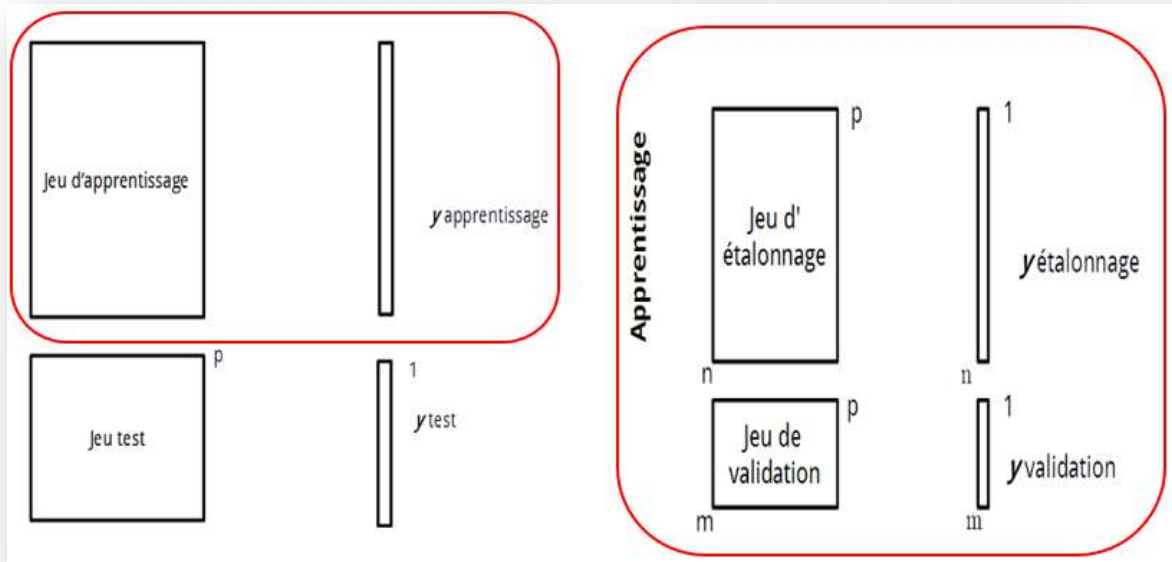


Table des Figures & Tableaux :

Figure 1 : Représentation des échantillons sur le repère F1 et F2, coloration en fonction de l'année de récolte.....	11
Figure 2 : Représentation des échantillons sur le repère F1 et F2, coloration en fonction de l'année et du lieu de plantation.	13
Tableau 1 : Présentation des différentes stratégies de modélisation.....	6
Tableau 2 : Performance des modèles prédictifs sur les échantillons frais des années 2022 à 2024	7
Tableau 3 : performance des modèles prédictifs sur les échantillons HB et CB préparés en laboratoire et provenant d'une usine de transformation.....	9
Tableau 4 : Performance des modèles prédictifs sur les échantillons de produits finis	10
Tableau 5 : Matrice de classification en fonction des années de récoltes.....	11
Tableau 6 : Matrice de classification en fonction du type de transformation.....	12
ANNEXE 1 : Protocole réalisé pour les analyses en frais, et sur produits transformés.	15
ANNEXE 2 : Schéma de la division des échantillons en un jeu d'apprentissage (étalonnage et validation) pour la validation croisée et un jeu de test pour valider le modèle.....	16

Bibliographie :

- Ajib, B. (2013). Contribution à la modélisation de la qualité de l'orge et du malt pour la maîtrise du procédé de maltage. <https://hal.univ-lorraine.fr/tel-01750656>
- Barrett, D. M., Garcia, E. F. Y., & Wayne, J. E. B. (1998). Textural Modification of Processing Tomatoes. *Critical Reviews In Food Science And Nutrition*, 38(3), 173258. <https://doi.org/10.1080/10408699891274192>
- Bureau, S., De Castro Vilas Boas, A. A., Giovinazzo, R., Jaillais, B., & Page, D. L. (2020). Toward the implementation of mid-infrared spectroscopy along the processing chain to improve quality of the tomato based products. *LWT*, 130, 109518. <https://doi.org/10.1016/j.lwt.2020.109518>
- Page, D. L., Labadie, C., Reling, P., Bott, R., Garcia, C., Gaillard, C., Fourmaux, B., Bernoud-Hubac, N., Goupy, P., Georgé, S., & Caris-Veyrat, C. (2019). Increased diffusivity of lycopene in hot break vs. cold break purees may be due to bioconversion of associated phospholipids rather than differential destruction of fruit tissues or cell structures. *Food Chemistry*, 274, 500509. <https://doi.org/10.1016/j.foodchem.2018.08.062>
- Page, D. L., Van Stratum, E., Degrou, A., & Renard, C. M. (2012). Kinetics of temperature increase during tomato processing modulate the bioaccessibility of lycopene. *Food Chemistry*, 135(4), 24622469. <https://doi.org/10.1016/j.foodchem.2012.06.028>
- Cheemook, Grain 3 ACP, Ondalys, Sébastien Preys, Eric Latrille, INRA de Narbonne, France
Serge Rudaz et Julien Boccard, Université de Genève, Suisse
- Cheemook, Grain 9: Régression PLS Douglas Rutledge, Delphine Jouan-Rimbaud, Bouveresse AgroPa-risTech, Paris (France)