

Une situation pédagogique sur microcontrôleur pour faire émerger les besoins en analyse de données : le tri des matières plastiques

Résumé Pour introduire l'importance du traitement de données multiples, cet article propose aux étudiants un scénario pédagogique basé sur le recyclage de matières plastiques. Ayant au préalable acquis les compétences nécessaires à l'utilisation des microcontrôleurs de type Arduino et ayant déjà élaboré et validé un colorimètre mono-longueur d'onde, un nouveau challenge est proposé aux étudiants. Ce défi consiste à pouvoir trier des feuilles plastiques de différentes couleurs, ainsi que des assemblages de feuilles, à partir de mesures d'absorbance réalisées avec une diode trois couleurs (c'est-à-dire avec des mesures à trois longueurs d'onde différentes). Une fois les données acquises, toute la difficulté réside dans la façon de faire ressortir l'information utile. D'une démarche par tâtonnement, les étudiants sont invités à découvrir l'analyse en composantes principales (ACP). Passer à un spectrophotomètre à huit longueurs d'onde, lui aussi mis en place à très faible coût sur un microcontrôleur Arduino, se révèle ensuite trivial. Nous voilà prêts pour aborder d'autres techniques analytiques et aller plus avant dans la compréhension fine des outils de traitement des données.

Mots-clés **Arduino, recyclage, tri, analyse de données, analyse en composantes principales (ACP).**

Si la « chimie analytique » au sens même de transformation de la matière en vue de quantification d'une espèce chimique a été au cœur des pratiques professionnelles pendant les XIX^e et XX^e siècles, une mutation considérable de ces pratiques est advenue dans les cinquante dernières années avec un développement important des instruments de mesure basés sur des concepts physiques. Après une phase de développement au sein de laboratoires de recherche, les techniques électrochimiques puis chromatographiques, les spectrométries optiques, de masse, RMN, que ce soit pour l'analyse élémentaire ou moléculaire, sont maintenant largement déployées dans les laboratoires académiques ou industriels.

En parallèle, les compétences attendues dans le domaine de l'analyse physico-chimique ont fortement évolué et peuvent aujourd'hui s'exprimer autour de quatre axes principaux :

- définir la stratégie analytique face au besoin,
- optimiser la chaîne et les processus analytiques,
- traiter les données,
- assurer la qualité du résultat.

Cette situation invite donc à repenser l'enseignement de l'analyse en chimie qui, s'il doit toujours s'appuyer sur de la réactivité entre espèces ou sur du partage entre phases pour la préparation et le traitement d'échantillons, doit de plus en plus être orienté vers l'usage des techniques instrumentales, associé à l'exploitation des données produites. Avec l'arrivée de microcontrôleurs à très bas coûts, un changement de pratique pédagogique a été proposé et largement décrit dans les colonnes de *L'Actualité Chimique* [1-5]. Les étudiants deviennent alors acteurs dans l'élaboration des instruments de mesure et dans la mise en place de l'ensemble des étapes conduisant à l'information pertinente pour l'utilisateur. Ainsi, la problématique de l'étalonnage peut être abordée au travers de la pH-métrie [1] et de la colorimétrie [2] suite à la construction d'instruments adaptés à ces usages. Puis après adjonction de pompes, les instruments peuvent être utilisés pour réaliser des titrages potentiométriques (acide-base, redox, précipitation ou complexation) et des dosages de composés colorés en solution. Les mesures obtenues avec les instruments développés peuvent aussi être utilisées pour traiter des sujets tels que

la thermodynamique [5] ou la cinétique chimique [3] avec un travail préalable des étudiants quant à la définition de la stratégie de mesure à déployer face à la problématique abordée, tout en garantissant la qualité finale (exactitude) du résultat de mesure.

Toutefois, si un résultat brut de mesure peut être directement utile pour une problématique spécifique (par exemple le suivi de l'évolution d'un système en fonction du temps via une mesure d'absorbance), il existe tout un panel de situations pour lesquelles le résultat brut d'une mesure n'est pas en lui-même porteur de sens. Dans ce cas, il sera nécessaire de réaliser la mesure de plusieurs grandeurs physiques et c'est l'ensemble constitué des résultats de ces différentes mesures qui sera, on l'espère, porteur d'informations. Il s'agira alors de faire ressortir de l'information de l'ensemble des mesures réalisées, et on entre ainsi de plain-pied dans le domaine du traitement des données appliqué à l'analyse chimique, qui pourra au besoin se décliner jusqu'aux outils d'intelligence artificielle.

Aborder le traitement de données avec des étudiants est un véritable défi didactique car très rapidement (trop rapidement) des jeux de données complexes sont présentés, et leur traitement est conduit avec des logiciels dédiés. Pour introduire l'importance de l'exploitation de données multiples, nous proposons un scénario pédagogique basé sur le retraitement de matières plastiques. Ce scénario est présenté sous forme de projet à des étudiants de licence physique-chimie (L3) dans le cadre d'une unité d'enseignement intitulée « De l'instrument à la mesure » et aux étudiants du master « Analyse et contrôle » dans le cadre de l'unité d'enseignement « Du signal à l'instrument d'analyse ». Les étudiants ont au préalable acquis les compétences nécessaires à l'utilisation des microcontrôleurs et ont déjà élaboré et validé un colorimètre mono-longueur d'onde constitué d'une diode électroluminescente de couleur spécifique et d'une photodiode [2]. La mise en situation pour ce nouveau projet est alors réalisée à partir du document présenté en *encadré*. Cette situation est une transposition des systèmes de séparation et tri des plastiques qui existent à l'échelle industrielle, réalisés grâce à des mesures dans le proche infrarouge (NIR) : une caméra

Encadré

Document pour la mise en situation des étudiants

Dans le cadre d'un processus de retraitement de résidus de polymères, des morceaux de feuilles plastiques colorées et transparentes doivent être automatiquement reconnus afin d'être orientés vers un récipient spécifique en fonction de leur couleur.

Il peut arriver que deux ou trois feuilles passent simultanément sur le système : si les feuilles sont de la même couleur, elles seront orientées vers le récipient correspondant à leur couleur ; si les feuilles sont de couleurs différentes, elles seront orientées vers un récipient dédié aux feuilles « mélangées » pour être remises sur le système de tri après avoir été physiquement séparées les unes des autres.

La partie mécanique du système de tri ne sera pas abordée dans le projet ; vous devrez par contre mettre en place le système de mesure et de traitement de données qui permettra de décider de l'orientation des feuilles vers le bac dédié. Le système de reconnaissance sera basé sur des mesures d'absorbance en utilisant une LED trois couleurs pouvant émettre à 640, 550 et 450 nm. Outre le microcontrôleur et les éléments électroniques nécessaires à l'élaboration du colorimètre à trois longueurs d'onde, vous avez à votre disposition des morceaux de feuilles colorées (codées R, J, B, V) pour mettre au point votre instrument (figure 1).

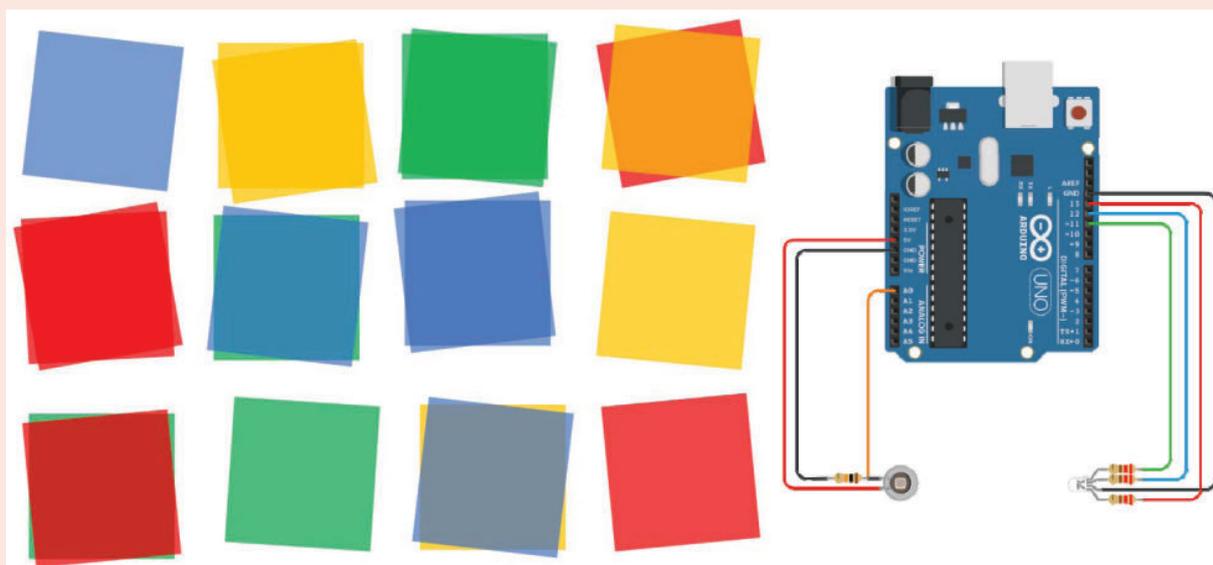


Figure 1 - Combinaisons de feuilles colorées et schéma de montage d'un colorimètre utilisant en émission une diode électroluminescente à trois couleurs.

effectue une mesure sur les particules circulant sur une bande transporteuse, l'unité de contrôle analyse les données, identifie la composition de chaque particule du mélange et détermine leur position sur la bande transporteuse, puis décide de retirer les particules indésirables à l'aide d'éjecteurs projetant de l'air sur les particules à éliminer [6].

Acquisition et premières représentations des données

La démarche globalement adoptée par les étudiants conduit à mesurer l'absorbance aux trois longueurs d'onde 640, 550 et 450 nm pour différents assemblages, dépendant de la couleur et du nombre de feuilles (par exemple avec 1R, 2R, 3R, 1J, 2J, 3J, 1B, 2B, 3B, 1V, 2V, 3V puis 1R+1B, 1R+1J, 1R+1V, 1B+1J...) et incluant des répétitions de mesure. Le jeu de données utilisé dans cet article est disponible en téléchargement [7], le blanc ayant été fait en l'absence de feuilles entre la LED et la photodiode. Ce jeu de données est séparé en deux groupes : un groupe de résultats obtenus avec des assemblages connus de feuilles de même couleur (figure 2), et un groupe qui servira de groupe test pour la reconnaissance d'assemblages à trier (figure 3).

Avec des feuilles transparentes, il peut apparaître étonnant que des feuilles rouges absorbent la lumière émise par une diode colorée en rouge, c'est-à-dire émettant autour de 640 nm. De la même façon, des feuilles bleues absorbent la lumière émise par une diode colorée en bleu vers 450 nm. Les feuilles colorées qui ont été utilisées n'étant pas totalement

transparentes, et pas forcément placées de façon perpendiculaire au faisceau lumineux, une partie de la lumière est donc perdue par diffusion ou par réflexion, laissant croire à une « absorbance » de la lumière par la feuille (être capable d'identifier de telles sources de biais de mesure est un élément important qui se révélera aussi pertinent lors de mesures sur des solutions). Dans le contexte des objectifs fixés, les pertes de lumière par diffusion et le nombre d'interfaces feuille-air augmentant proportionnellement avec le nombre de feuilles, les risques de biais dans l'analyse ultérieure des données seront toutefois minimes.

Les mesures d'absorbance aux trois longueurs d'onde pour des feuilles inconnues (figure 3) invitent ensuite les étudiants à définir une stratégie permettant de répondre à la problématique posée. La classification des feuilles inconnues doit être possible de façon automatisée, mais comment procéder ?

Il apparaît évident qu'une seule mesure d'absorbance à une longueur d'onde choisie ne permet pas de reconnaître le type de feuille, et qu'il sera nécessaire d'utiliser simultanément les trois valeurs d'absorbance acquises. La première piste qui apparaît au cours des discussions consiste à utiliser une comparaison des absorbances aux différentes longueurs d'onde : quelle est la longueur d'onde (λ_{\max}) qui conduit à la plus grande valeur d'absorbance ? À la plus petite valeur d'absorbance (λ_{\min}) ?

Mais force est de constater que si la longueur d'onde du maximum d'absorbance couplée à la longueur d'onde du minimum d'absorbance pourrait permettre de reconnaître des assemblages constitués de feuilles de mêmes couleurs,

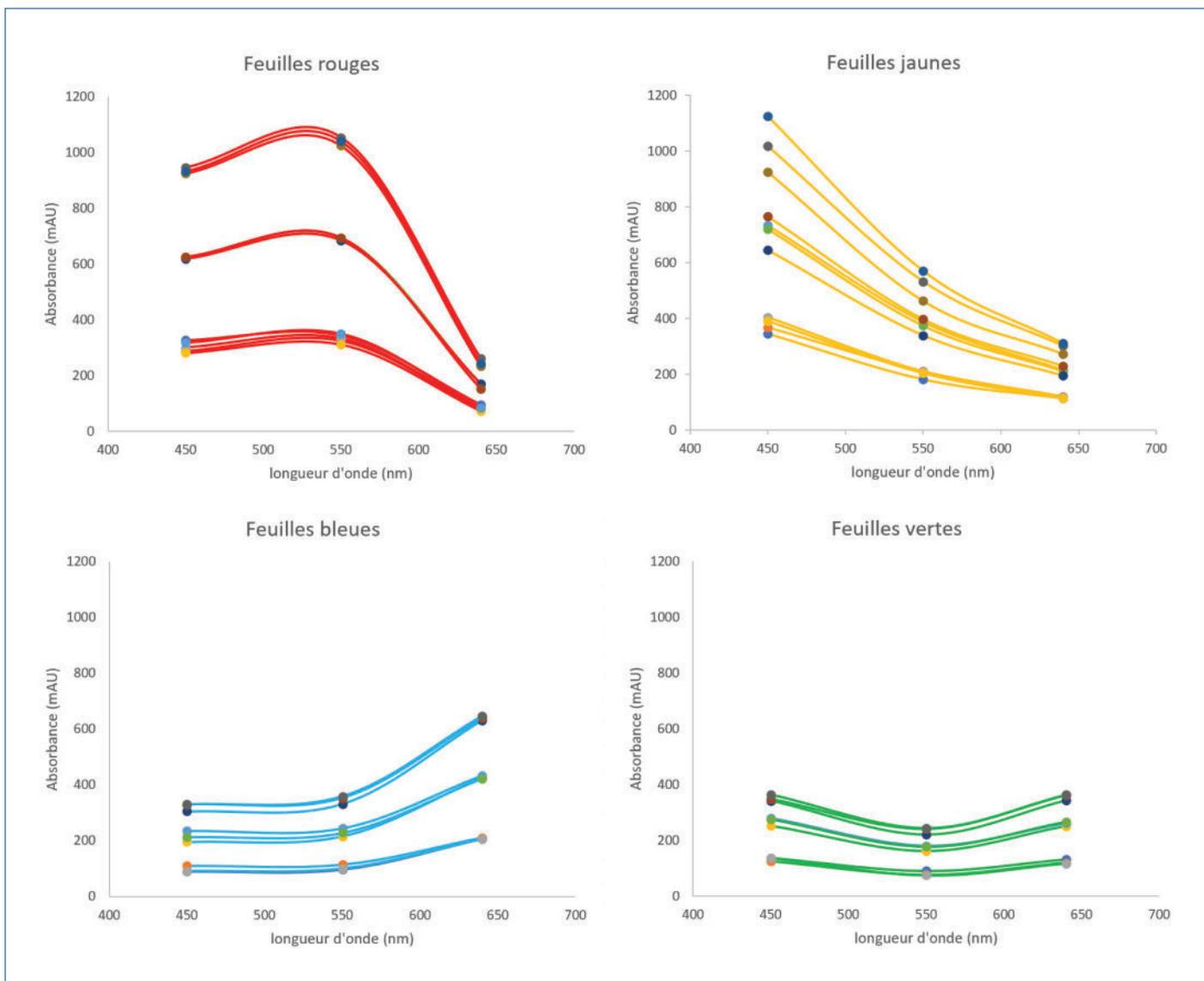


Figure 2 - Absorbance des feuilles colorées (une, deux ou trois feuilles) en fonction de la longueur d'onde émise par la diode électroluminescente.

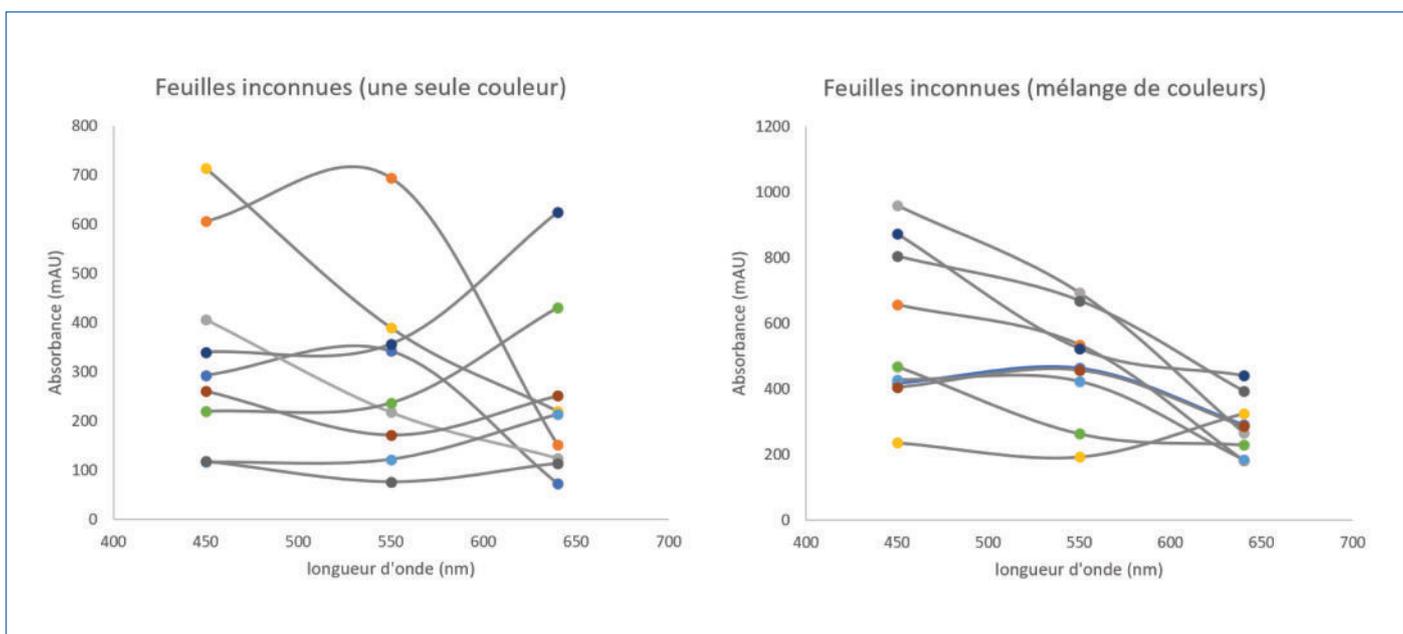


Figure 3 - Absorbance des feuilles devant être triées. Une séparation en deux groupes a été réalisée pour faciliter la lecture ; les points colorés ne correspondent pas aux couleurs de feuilles plastiques utilisées.

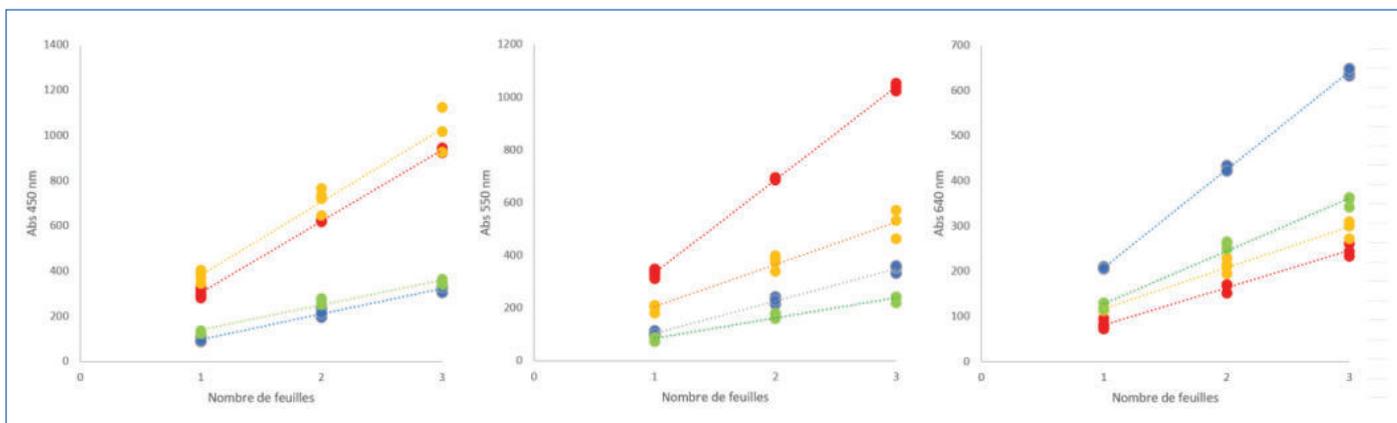


Figure 4 - Absorbance en fonction du nombre de feuilles pour les trois longueurs d'onde de mesure 450, 550 et 640 nm.

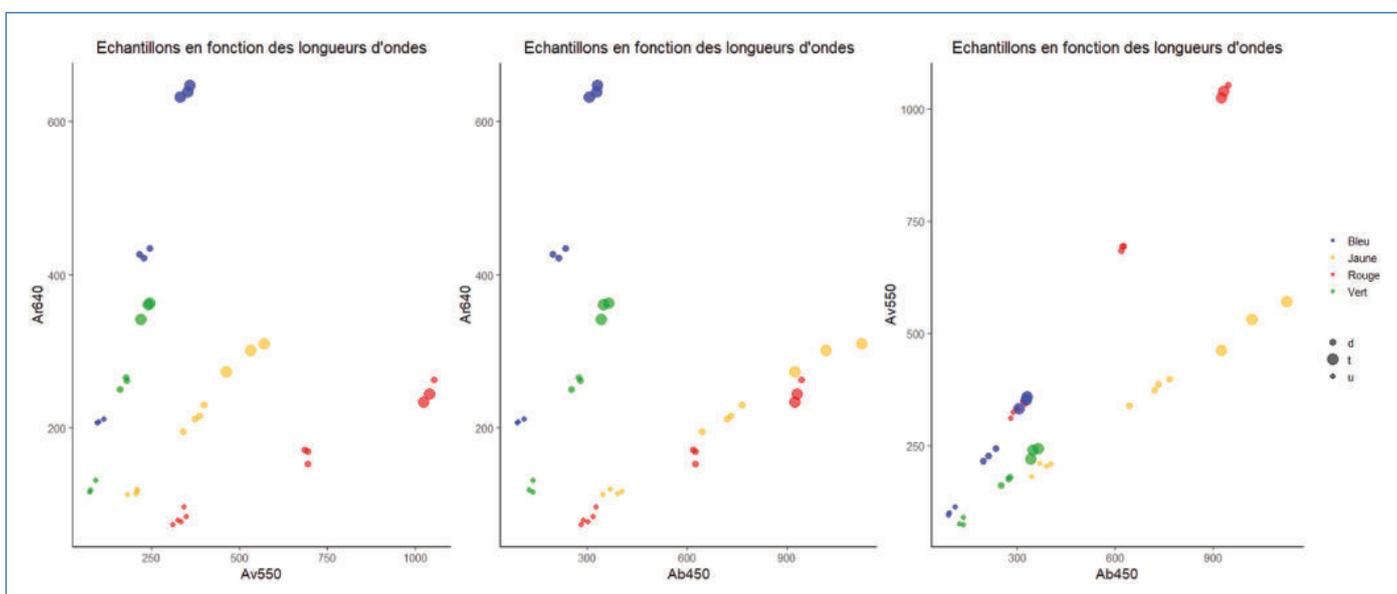


Figure 5 - Représentation des données d'absorbance à une longueur d'onde en fonction de l'absorbance à une autre longueur d'onde.

ce jeu d'informations (λ_{\max} , λ_{\min}) est insuffisant dans le cas d'assemblage de feuilles de couleurs différentes et conduit à une mauvaise identification de ce type d'assemblage. Il faut donc aller plus loin.

Traitement des données basées sur la connaissance de la loi de Beer-Lambert

La connaissance de la loi de Beer-Lambert oriente naturellement les étudiants vers une représentation du type $Abs_{\lambda} = f(\text{nombre de feuilles})$ telle que celles reportées sur la figure 4. Les morceaux de feuilles colorées étant issus de pochettes plastiques transparentes découpées sous la forme de carrés de 3 cm de côté, l'épaisseur de chaque morceau est constante et une relation linéaire est obtenue entre absorbance et nombre de feuilles.

Généralement, les étudiants connaissent et ont utilisé une méthode permettant de déterminer la concentration de deux colorants en solution (par exemple les colorants jaune et bleu présents dans un sirop de menthe) à partir de deux mesures d'absorbance à deux longueurs d'onde différentes. Si cette méthode se révèle fort efficace lorsque l'on connaît les deux colorants utilisés pour préparer la solution à analyser (le sirop), elle ne peut être rendue opérationnelle dans le cas de colorants non préalablement identifiés. Cette piste ne peut

donc être suivie et une autre stratégie doit être imaginée... De façon assez rapide, une représentation du type $A_{\lambda_1} = f(A_{\lambda_2})$ apparaît dans les groupes d'étudiants (figure 5). Mais là encore, s'il semble possible d'identifier le nombre de feuilles pour des feuilles dont on connaît la couleur (ce qui était déjà le cas via la figure 4), il est impossible d'identifier la couleur de la feuille ou des feuilles identiques à la source du phénomène d'absorption.

La frustration apparaît alors dans les différentes équipes d'étudiants : il semble que l'information soit présente (sinon le professeur ne nous aurait pas proposé ce défi) mais accéder à l'information n'apparaît pas de façon spontanée.

Si le point de blocage est trop important pour être levé seul par les étudiants, une piste de recherche peut être proposée en revenant sur les données de la figure 2 : lorsque le nombre de feuilles de même couleur augmente, les valeurs d'absorbance observées aux trois longueurs d'onde varient-elles de façon régulière? Cela permet de revenir sur la loi de Beer-Lambert exprimée pour chacune des longueurs d'onde ($A_{\lambda_1} = \epsilon_{\lambda_1} \cdot l \cdot c$; $A_{\lambda_2} = \epsilon_{\lambda_2} \cdot l \cdot c$; $A_{\lambda_3} = \epsilon_{\lambda_3} \cdot l \cdot c$) pour faire émerger le fait que pour chacun des diagrammes de la figure 2, les trois valeurs d'absorbance varient de façon proportionnelle au nombre de feuilles (la longueur l du trajet optique étant proportionnelle au nombre de feuilles), mais aussi que si l'absorbance à une longueur d'onde multipliée par deux suite

Tableau - Absorbance (mAU) à trois longueurs d'onde pour un nombre de feuilles rouges variable et rapport des valeurs d'absorbance mesurées.

Nb Rouge	Abs 640	Abs 550	Abs 450	Abs 640/Abs 550	Abs 640/Abs 450	Abs 550/Abs 450
1	96	343	326	0,280	0,294	1,052
2	169	694	625	0,244	0,270	1,110
2	171	685	619	0,250	0,276	1,107
2	152	694	624	0,219	0,244	1,112
3	262	1053	945	0,249	0,277	1,114
1	79	325	288	0,243	0,274	1,128
1	77	333	301	0,231	0,256	1,106
1	73	311	281	0,235	0,260	1,107
1	84	349	317	0,241	0,265	1,101
3	233	1 024	924	0,228	0,252	1,108
3	244	1 040	932	0,235	0,262	1,116
moyenne				0,241	0,266	1,106
écart-type				0,016	0,014	0,019

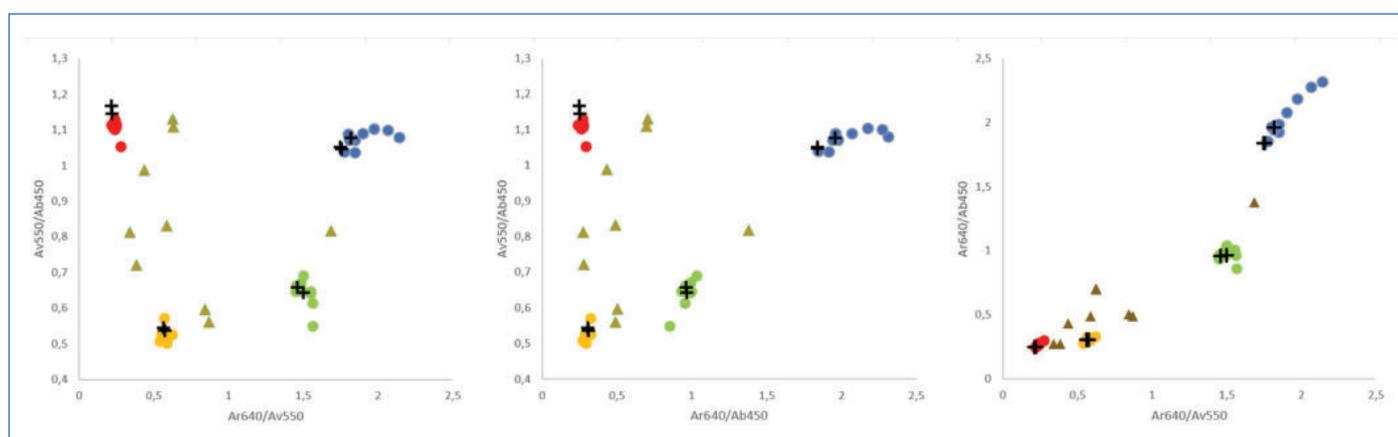


Figure 6 - Représentation de chaque feuille, ou combinaison de feuilles dans un système de coordonnées utilisant des rapports d'absorbance (rond : feuilles colorées – pas de mélange ; croix : feuilles colorées inconnues – pas de mélange ; triangles : feuilles colorées mélangées).

à un accroissement du nombre de feuilles, les valeurs d'absorbance aux autres longueurs d'onde sont aussi multipliées par deux. Ainsi l'information présente dans le spectre n'est pas uniquement liée à la position du maximum d'absorbance, mais aussi à la proportion fixe entre les valeurs d'absorbance observées à plusieurs longueurs d'onde pour un produit particulier (feuille ou solution). Ce dernier principe est d'ailleurs largement utilisé en chromatographie : avec un détecteur UV fonctionnant à deux longueurs d'onde, on vérifie que le signal apparaissant lors de l'élution d'un soluté varie tout au long du pic chromatographique de la même façon sur les deux longueurs d'onde suivies. Une absence de proportionnalité entre les deux signaux est alors un signe de co-élution de plusieurs composés sous le pic chromatographique.

Si on considère le cas des feuilles rouges présenté en figure 1, il est ainsi possible de caractériser le spectre des feuilles rouges grâce aux valeurs des rapports d'absorbance (voir tableau).

Ce type de calcul conduit à tracer plusieurs diagrammes dans lesquels chaque feuille est représentée par un point dans un système de coordonnées utilisant les rapports d'absorbance (figure 5). On retrouve par exemple sur le diagramme de gauche de la figure 5 les feuilles rouges qui sont toutes

proches du point de coordonnées (Abs 640/Abs 550 = 0,241 ; Abs 550/Abs 450 = 1,106).

Toutes les feuilles rouges, jaunes, vertes et bleues et les assemblages de feuilles de même couleur sont regroupés sur chacun des diagrammes. Toutefois, l'éclatement des données dans l'espace de représentation n'est pas identique dans les trois diagrammes et les grandeurs représentées sur les abscisses et ordonnées doivent donc être judicieusement sélectionnées. On peut par exemple constater que sur le diagramme de droite, les points représentatifs des feuilles semblent être beaucoup plus alignés que dans les autres figures (il y a certainement une corrélation entre les grandeurs portées sur les axes de ce diagramme, nous y reviendrons ultérieurement). En ajoutant les données associées au groupe test (croix et triangles sur la figure 6), on constate que les croix, correspondant à des feuilles qui n'étaient pas identifiées au préalable, sont bien positionnées proches de l'ensemble des points identifiant une couleur particulière. Les triangles, qui eux correspondent à des assemblages de feuilles de couleurs différentes, sont décalés par rapport aux autres points. Trois mesures d'absorbance peuvent donc permettre, via le calcul de rapport d'absorbances et la proximité avec les points caractéristiques de chacune des couleurs, de reconnaître la couleur

des feuilles et d'identifier si un assemblage de feuilles de couleurs différentes passe sur le système de mesure.

Le challenge a été relevé ! Mais ceci s'est fait au prix d'un traitement de données qui s'appuyait sur une connaissance préalable de tout le processus de mesure et de ses caractéristiques pour choisir un traitement adapté et finalement une représentation la plus informative possible. Mais si nous avons utilisé un autre domaine spectral ? Si nous avons un nombre plus important de données d'absorbance (8, 20 ou 100 valeurs pour chaque type de feuilles, ceci pour des longueurs d'onde différentes) ? Comment aurions-nous pu procéder ? Devons-nous avoir une approche par tâtonnement comme nous l'avons fait ici, ou pouvons-nous essayer d'obtenir de l'information dans un jeu de données de façon rationnelle ? Et c'est là que la méthodologie de traitement de données peut être introduite, sans forcément entrer dans tout le détail des calculs, mais en montrant l'usage de ce type de méthodologie.

Traitement des données par analyse en composantes principales

L'analyse en composantes principales (ACP) s'applique à des tableaux à deux dimensions croisant des individus (I lignes) et des variables quantitatives (K colonnes) : à l'intersection de la ligne i et de la colonne k se trouve la valeur x_{ik} de la variable k pour l'individu i. Le tableau qui nous intéresse dans notre exemple contient (K = 3) colonnes, ou variables, et (I = 40) lignes, ou individus correspondant à toutes les configurations observées connues (une, deux ou trois feuilles, les feuilles d'un assemblage étant toutes de même couleur). Les valeurs d'absorbance en mAU pour chacune des trois variables et des quarante individus sont les données x_{ik} . Ce sont en fait les données qui ont été préalablement présentées sur la figure 2. Un premier traitement de données consiste à rechercher l'éventuelle corrélation entre variables. Ainsi, considérant deux variables, on peut évaluer leur liaison par l'intermédiaire d'un coefficient de corrélation linéaire (c'est le r de la droite de régression par la méthode des moindres carrés) :

$$r(k', k'') = \frac{\text{covariance}(k', k'')}{\sqrt{\text{variance}(k') \cdot \text{variance}(k'')}}}$$

Pour une meilleure représentation des données (facilité d'usage, de lecture) et obtenir de nouvelles variables décorréliées, l'analyse en composantes principales permet de résumer l'ensemble des variables de départ (ici les absorbances aux différentes longueurs d'onde) par des variables synthétiques appelées composantes principales qui sont des combinaisons linéaires des variables de départ. L'ACP va ainsi transformer les K variables quantitatives initiales, qui sont plus ou moins corrélées, en K nouvelles variables quantitatives non corrélées, appelées composantes principales :

$$PC_1 = w_{11}X_1 + w_{12}X_2 + \dots + w_{1K}X_K$$

L'objectif sera ainsi de créer un nouveau repère orthonormé où chaque axe sera une combinaison linéaire des variables de départ et contiendra une part de l'information initiale. Dans notre cas, on obtient trois composantes principales :

$$PC_1 = w_{11} \cdot Abs450 + w_{12} \cdot Abs550 + w_{13} \cdot Abs640$$

$$PC_2 = w_{21} \cdot Abs450 + w_{22} \cdot Abs550 + w_{23} \cdot Abs640$$

$$PC_3 = w_{31} \cdot Abs450 + w_{32} \cdot Abs550 + w_{33} \cdot Abs640$$

Les coefficients w_{ij} sont calculés de telle sorte que la première composante principale (PC1) possède la plus grande variance,

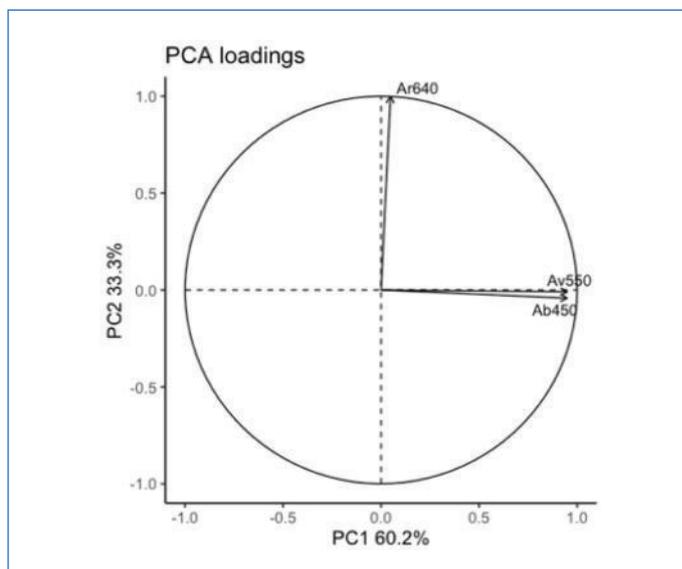


Figure 7 - Projections des variables (« loadings ») de l'ACP sur les plans PC1-PC2.

c'est-à-dire que les points représentant les individus sur cet axe sont les plus éloignés possible les uns des autres. La seconde composante (PC2) possède la deuxième plus grande variance, et ainsi de suite. Si on considère les deux premières composantes, la variance globale de ces deux composantes est la somme des variances de chacune des deux composantes. Par exemple, si la première composante représente 60 % et la deuxième 33 % de la variance totale, la projection des individus sur ce plan (PC1-PC2) représente 93 % de la variance totale. Le détail des calculs conduisant à la construction des composantes principales et la détermination des coefficients w_{ij} ne sera pas développé ici, ni dans un premier temps avec les étudiants.

Le premier type de projection concerne les variables initiales dans l'espace des composantes principales. Ce graphique dit des « loadings » permet de visualiser le degré de corrélation entre variables (variables corrélées, anti-corrélées ou indépendantes). Ainsi, dans le plan des deux composantes principales considérées, si deux variables sont proches géométriquement, cela signifie que ces deux variables sont corrélées. Si elles sont diamétralement opposées, les variables sont anti-corrélées, et si les variables sont orthogonales (positionnées en angle droit), elles sont indépendantes. En appliquant une ACP sur les données issues de notre exemple, on constate sur la figure 7 que Abs550 et Abs450 sont fortement corrélées sur le plan PC1-PC2 et que Abs640 est indépendante des deux autres.

Le deuxième type de projection, dit des « scores », concerne les individus projetés dans l'espace des composantes. Ce type de graphique est en grande partie à l'origine de la très grande popularité de l'ACP, et permet de visualiser des proximités entre les individus. Les « scores » représentent ainsi des informations qui sont intéressantes pour comprendre la structuration des données (proximité, groupes, points aberrants...). Sur la projection issue de nos données (figure 8), les individus sont séparés du point de vue de leur couleur mais aussi en fonction du nombre de feuilles utilisées.

Dans cette représentation, on peut alors ensuite positionner de nouveaux individus non présents dans les données initiales, dits points tests, qui peuvent soit résulter d'une ou plusieurs feuilles de couleur identique (1 à 9), soit être le résultat d'un assemblage de feuilles de couleurs différentes (10 à 18).

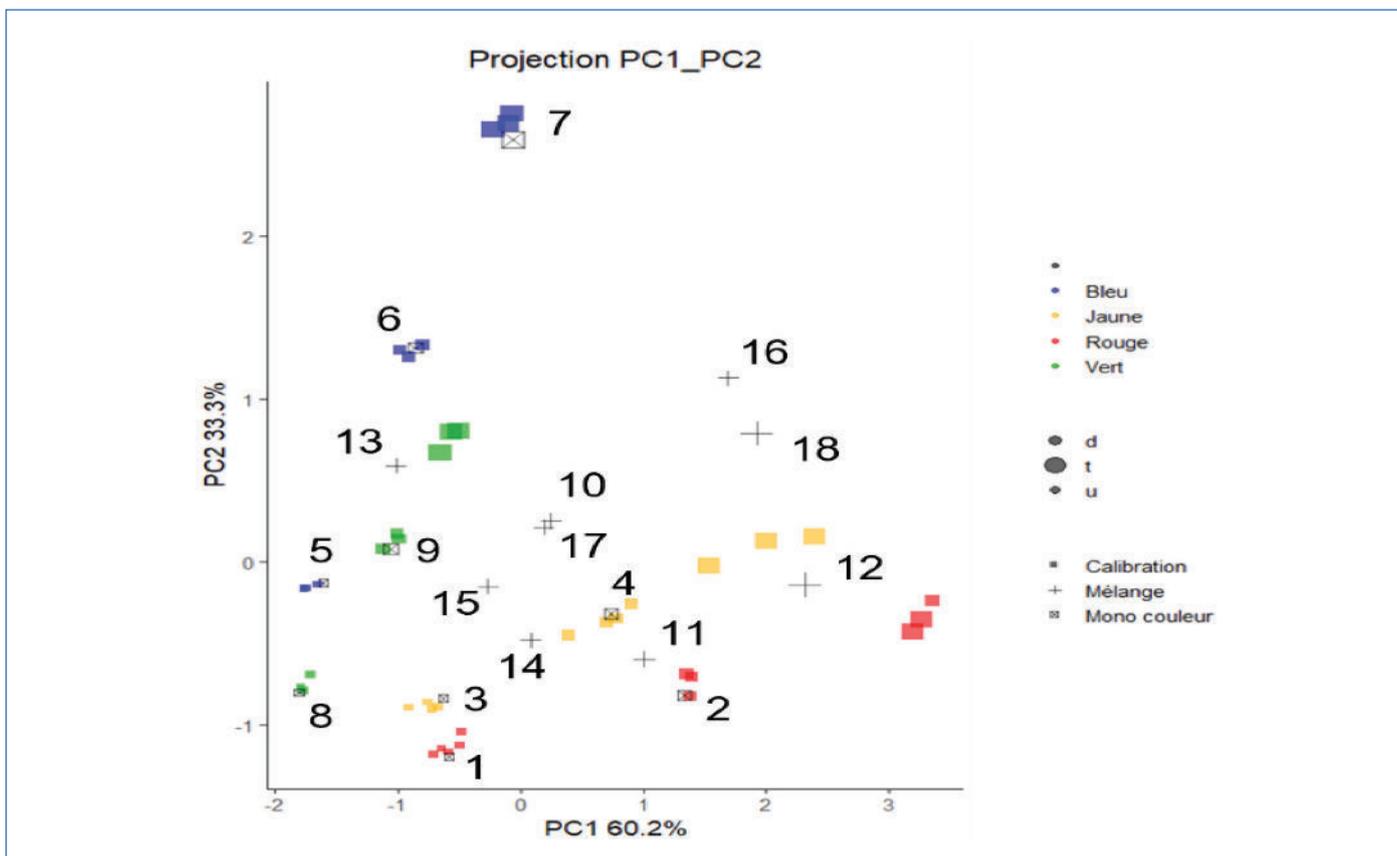


Figure 8 - Projections des individus (« scores ») de l'ACP sur les plans PC1-PC2 avec les individus supplémentaires (de 1 à 18).

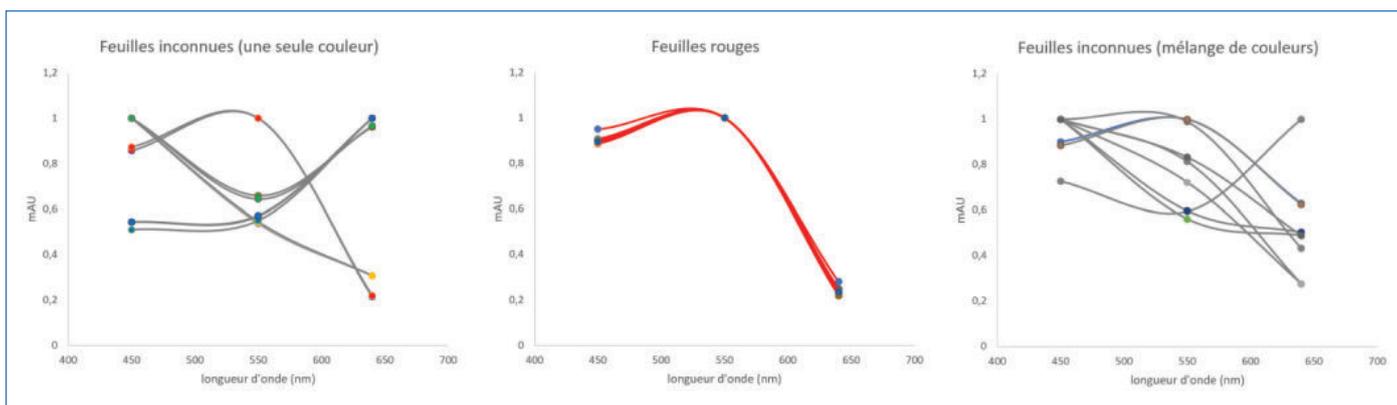


Figure 9 - Spectres normalisés.

Si les individus tests semblent globalement bien projetés sur PC1-PC2 dans leur zone respective, la proximité avec certaines zones pourrait apparaître problématique (par exemple, on pourrait attribuer par erreur le point 14 à deux feuilles jaunes, ou le point 13 à trois feuilles vertes).

Il est peut-être temps de revenir à notre question de départ : nous voulons isoler les feuilles ou assemblages de feuilles de même couleur pour leur retraitement, et repérer les assemblages de feuilles de couleurs différentes. Mais finalement, peu nous importe qu'il y ait une, deux ou trois feuilles dans l'assemblage analysé ; on souhaite trier feuille ou assemblage de feuilles par couleur, et c'est tout ! Si on réinterroge avec les étudiants la relation entre spectre et colorant présent dans la feuille, ceci conduit à identifier que tous les spectres représentés sur la figure 2 sont identiques à un facteur près pour chaque type de feuille. D'un point de vue du traitement numérique de l'information, cela revient à dire que tous les

spectres de chaque diagramme de la figure 2 sont identiques si l'axe vertical est normalisé à la plus grande absorbance du spectre considéré.

On peut constater sur le diagramme du milieu de la figure 9 que tous les spectres normalisés issus d'assemblages de feuilles rouges (1, 2 ou 3) sont en effet très proches. Il semblerait donc aisé de repérer les spectres normalisés identiques pour les assemblages tests constitués de feuilles identiques (diagramme de gauche figure 9), mais encore faut-il arriver à distinguer les écarts sur les spectres issus d'assemblages de feuilles de couleurs différentes (diagramme de droite figure 9). Bon, plutôt que de faire cela à la main, faisons la normalisation des données, réalisons une ACP, et rajoutons les ellipses de confiance associées à chaque groupe identifié par l'ACP (figure 10). Les résultats présentés sont en adéquation totale avec notre objectif initial : chacun des groupes d'individu, feuille unique ou assemblage de feuilles de couleur identique,

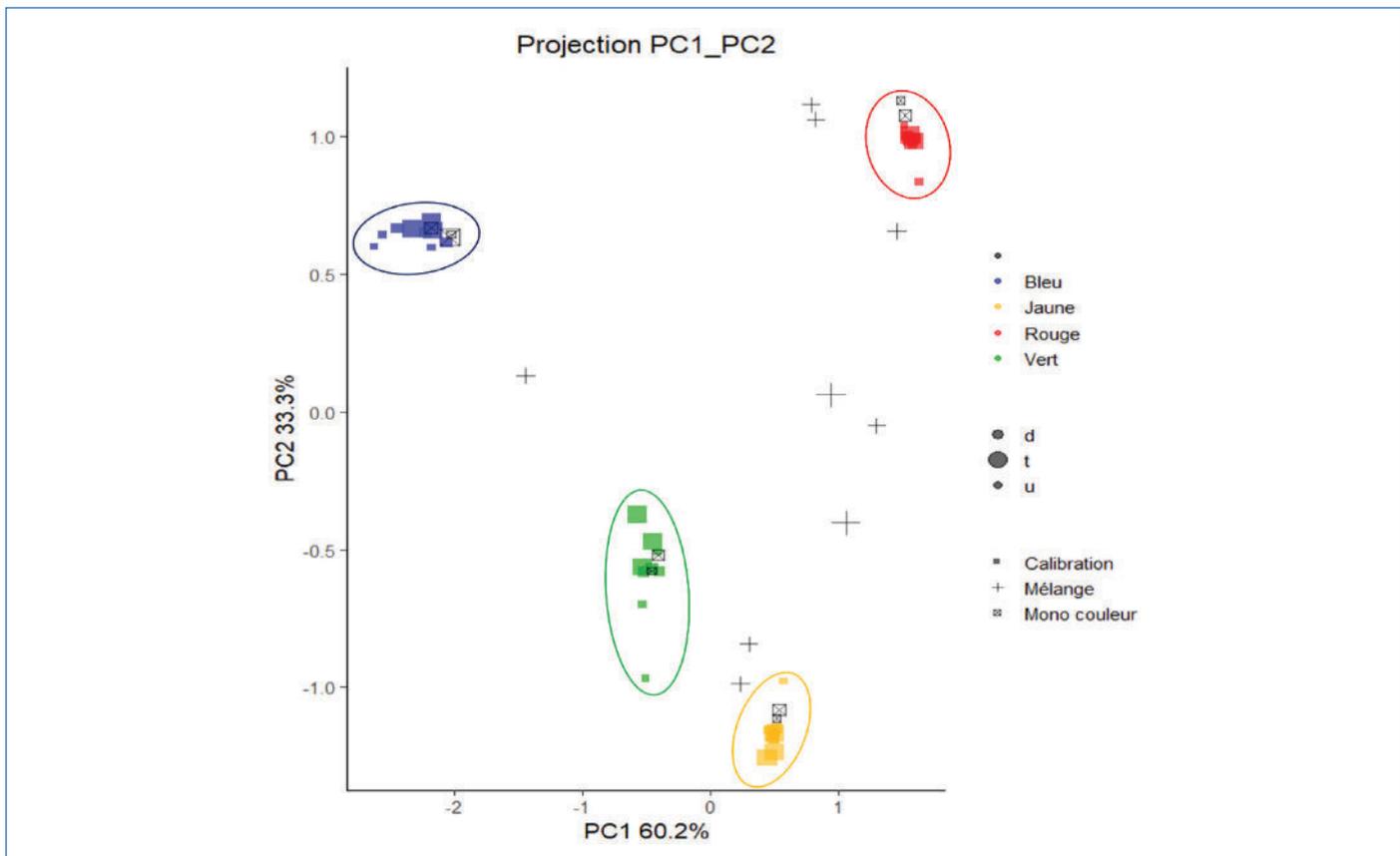


Figure 10 - Projections des individus (« scores ») de l'ACP sur les plans PC1-PC2 avec les individus supplémentaires (de 1 à 18), après normalisation des spectres. Les ellipses représentent un intervalle de confiance à 95 % pour appartenir à un groupe.

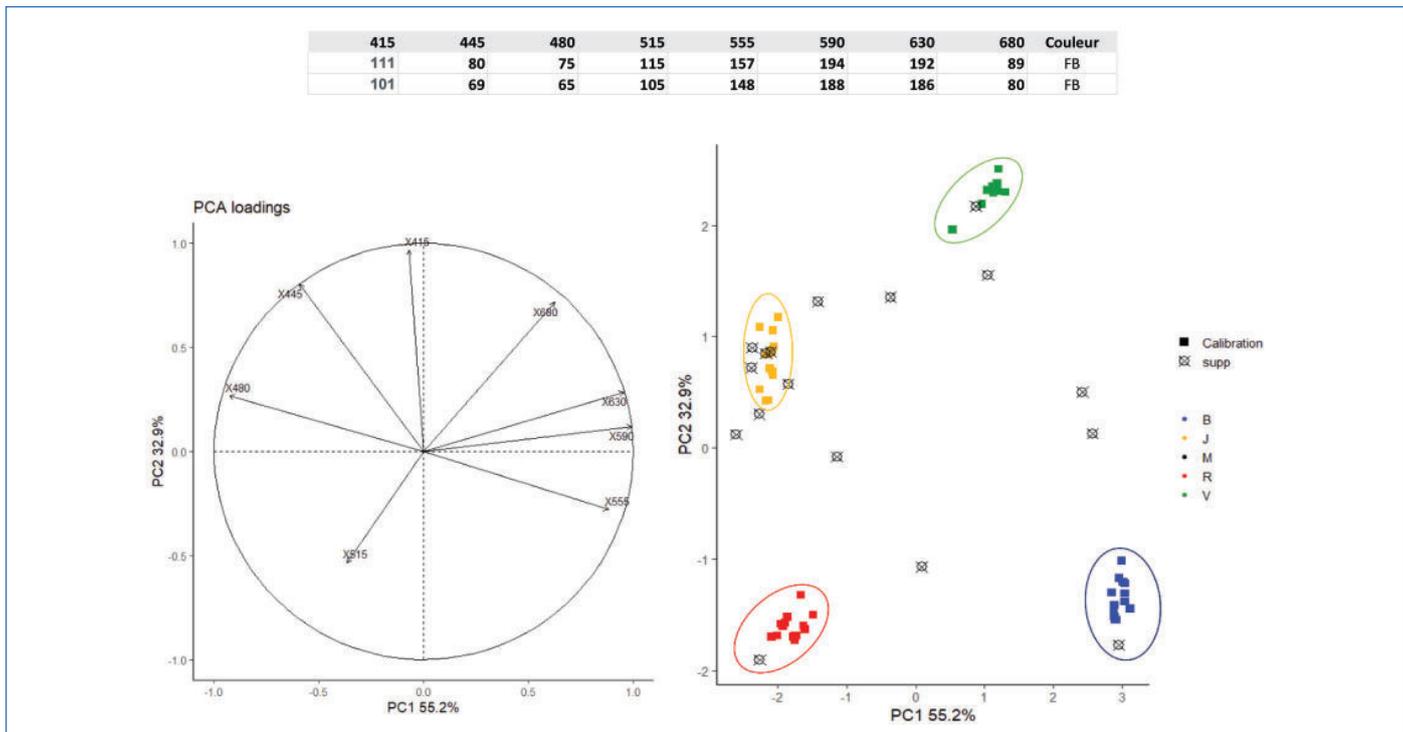


Figure 11 - Forme du tableau de données, « loadings » et « scores » sur les plans PC1-PC2, avec les individus tests en gris, après normalisation des spectres. Les ellipses représentent l'intervalle de confiance à 95 % pour appartenir à un groupe.

sont projetés dans une zone spécifique de l'espace issu des deux composantes principales. Les individus tests sont projetés, soit dans un groupe, ce qui permet de déterminer leur couleur, soit éloignés de chacun des quatre groupes car constitués d'un assemblage de feuilles de couleurs différentes. C'est gagné !

Changement d'échelle dans la quantité de données à traiter

Avant de passer à des spectrophotomètres de laboratoire, les étudiants peuvent aussi être invités à construire un colorimètre à huit longueurs d'onde sur la base d'un capteur

AS7341 et à procéder à des mesures d'absorbance avec cet outil (on retrouve une nouvelle fois l'avantage de l'élaboration d'instruments à partir de matériel à très bas coût permettant de multiplier les instruments, et à tous les étudiants de réaliser leurs propres mesures).

Sans aucun a priori, si ce n'est la normalisation des spectres, l'ACP conduit à une très bonne discrimination pour les individus supplémentaires (figure 11). Tout ceci est réalisé avec quelques lignes de codes en Python (voir ressources complémentaires) ou sous R [7], sans avoir à sélectionner à la main des longueurs d'onde particulières et opérer des calculs par une succession d'essai-erreur jusqu'à peut-être obtenir une méthode adaptée au problème ciblé. Il est indispensable de souligner que dans toutes les figures qui sont présentées dans cet article, les couleurs des points n'ont été placées que pour une meilleure lisibilité des figures. On aurait pu ne pas spécifier les couleurs de feuilles, et les points sur les figures auraient été positionnées par l'ACP exactement au même endroit !

L'information était dans les données obtenues, et un outil adapté de traitement des données a été introduit pour la faire ressortir. Avec cette mise en situation, nous avons fait entrer les étudiants dans le domaine du traitement de données qui est actuellement au cœur des outils de décision dans notre société. Il nous faudra ensuite, en tant qu'enseignant, proposer des situations variées permettant à chacun d'appréhender ce domaine, par exemple en appliquant ces outils à des systèmes plus complexes tels que des données de proche infrarouge ou de GCxGC-MS/MS par exemple.

Des ressources complémentaires (fichiers Excel avec les données et code Python pour faire une ACP) sont téléchargeables librement sur <https://new.societechimiquedefrance.fr/divisions/enseignement-formation/article-ac-une-situation-pedagogique-sur-microcontroleur-pour-faire-emerger-les-besoins-en-analyse-de-donnees-le-tri-des-matieres-plastiques/>

- [1] J. Randon, Repenser l'enseignement des sciences analytiques au travers de la construction et l'évaluation d'instruments : un titrateur automatique pour les dosages acido-basiques, *L'Act. Chim.*, **2020**, 451, p. 41-48.
- [2] J. Randon, Construire un colorimètre et évaluer l'incertitude des méthodes de dosage par étalonnage, *L'Act. Chim.*, **2020**, 452, p. 30-34.
- [3] J. Randon, Des activités à la maison pour développer les compétences scientifiques ?, *L'Act. Chim.*, **2020**, 453, p. 7-8.
- [4] J. Randon, J. Piard, S. Guy, Comment l'information jaillit de la lumière : retour sur les notions cachées liées à une « mesure d'absorbance », *L'Act. Chim.*, **2020**, 455, p. 52-58.
- [5] J. Randon, X. Bataille, Enquête à la maison : calorimétrie et contrôle qualité, *L'Act. Chim.*, **2021**, 463, p. 45-48.
- [6] www.usinenouvelle.com/expo/separation-et-tri-des-plastiques-par-pr-p416761789.html, www.youtube.com/watch?v=QjuXVlrp5uY (consultés le 28/06/2023).
- [7] [https://fr.wikipedia.org/wiki/R_\(langage\)](https://fr.wikipedia.org/wiki/R_(langage)) (consulté le 07/07/2023).

Jérôme RANDON*, professeur d'université, **Yohan CLEMENT**, ingénieur d'études, et **Pierre LANTERI**, professeur émérite des universités, Université Claude Bernard Lyon 1, Institut des Sciences analytiques, Villeurbanne.

* randon@univ-lyon1.fr



L'Union des professeurs de physique et de chimie

Vous présente

son nouveau site : <http://www.udppc.asso.fr>



L'association	Espaces collège et lycée
L'Uodppc ?, Tarifs, Enquêtes,	Actualités, Les journées..., Ressources
Charte graphique, Olympiades,	Espace labo
Partenaires, Positions, Nous soutenons	Actualités, Ressources
Publications	Divers
Le Bup, Nous avons lu,	Agenda, congrès, réseaux sociaux,
L'arpenteur du web, Appel aux auteurs,	Sites académiques
Parus au BO, Ressources	

mais l'Uodppc, c'est aussi...

...la publication numérique mensuelle avec impression papier trimestrielle



...la consultation du Bup en ligne par articles et par numéro avec BupDoc

Du 1^{er} janvier au 31 décembre 2023 :

- ◆ Pour tous : 1907 → 2018
- ◆ Pour les abonnés : 2019 → 2023



...un congrès organisé chaque année par une académie différente



Siège social et courrier : 42 rue Saint-Jacques - 75005 PARIS
Tél. : 01 40 46 83 80 - secretariat.national@udppc.asso.fr



Compléments à l'article « Une situation pédagogique sur microcontrôleur pour faire émerger les besoins en analyse de données : le tri des matières plastiques », par J. Randon (*L'Act. Chim.*, 2023, 487, p. 37)

Des ressources complémentaires (fichiers excel avec les données et code Python pour faire une ACP) sont téléchargeables librement sur <https://new.societechimiquedefrance.fr/divisions/enseignement-formation/article-ac-une-situation-pedagogique-sur-microcontroleur-pour-faire-emerger-les-besoins-en-analyse-de-donnees-le-tri-des-matieres-plastiques/>

Données acquises pour la construction des modèles

Structure des données

Colonnes 1 à 4 : nombre de feuilles de chacune des quatre couleurs.

Colonnes 5 à 7 : absorbance en mAU pour chacune des longueurs d'onde.

R	J	B	V	Ar640	Av550	Ab450
1	0	0	0	96	343	326
2	0	0	0	169	694	625
2	0	0	0	171	685	619
2	0	0	0	152	694	624
3	0	0	0	262	1053	945
1	0	0	0	79	325	288
1	0	0	0	77	333	301
1	0	0	0	73	311	281
1	0	0	0	84	349	317
3	0	0	0	233	1024	924
3	0	0	0	244	1040	932
0	1	0	0	113	181	345
0	2	0	0	215	387	733
0	3	0	0	301	532	1017
0	1	0	0	120	210	368
0	1	0	0	117	209	404
0	1	0	0	114	204	390
0	2	0	0	211	374	721
0	2	0	0	195	339	644
0	2	0	0	230	398	766
0	3	0	0	273	463	925
0	3	0	0	310	570	1125
0	0	1	0	206	96	89
0	0	1	0	211	114	110
0	0	2	0	427	216	196
0	0	2	0	434	244	235
0	0	2	0	422	228	213
0	0	1	0	207	100	91

0	0	3	0	632	332	305
0	0	3	0	639	352	329
0	0	3	0	648	359	330
0	0	0	1	131	90	136
0	0	0	1	119	76	124
0	0	0	1	116	74	135
0	0	0	2	250	161	251
0	0	0	2	261	180	279
0	0	0	2	265	177	274
0	0	0	3	342	220	341
0	0	0	3	361	240	348
0	0	0	3	363	244	364

Autres données pour test d'usage

R1	J1	B1	V1	Ar640	Av550	Ab450
1	0	0	0	73	342	293
2	0	0	0	152	694	606
0	1	0	0	125	218	406
0	2	0	0	219	389	714
0	0	1	0	213	122	116
0	0	2	0	431	237	220
0	0	3	0	625	356	340
0	0	0	2	251	172	261
0	0	0	1	114	76	118

1	0	1	0	292	464	418
1	1	0	0	181	535	658
1	2	0	0	265	693	960
0	0	1	1	325	193	236
1	0	0	1	185	423	428
0	1	0	1	229	263	469
0	2	1	0	440	522	874
1	0	1	0	286	458	405
1	1	1	0	393	669	804