



université
de **BORDEAUX**

Maquette logicielle : création automatisée d'un rapport PDF de visualisation interactive de courbes expérimentales

À consulter avec une version récente d'Adobe Reader

Stage de 2 mois, Licence 3 Mathématiques ingénierie, Université de Bordeaux

Raphaël Cousin
raphael.cousin90@gmail.com

Sous la direction de
Laurent Duval, IFP Energies Nouvelles

Tuteur de stage
Yves Coudière

21 août 2016

Résumé

L'analyse de données expérimentales fait intervenir la visualisation de valeurs numériques et des paramètres expérimentaux associés. Pour l'analyse exploratoire de ces données, il est important de représenter les relations entre les différentes courbes numériques, pour fournir des pistes quant aux outils statistiques à employer pour des analyses plus approfondies. Ce rapport explore des méthodes et outils permettant de générer, à partir de valeurs numériques, un document PDF incluant des fichiers graphiques. Ce document contient un ensemble de graphes 2D et 3D interactifs, offrant des possibilités de visualisation et de calcul employant JavaScript. Partant de données mises en forme, un logiciel de scripting (par exemple Matlab, R, Python) génère des graphiques enrichis au format U3D et PRC via le logiciel libre Asymptote et produit un code \LaTeX , construisant le document PDF souhaité. Ce document présente dans une première partie les enjeux et besoins pour la visualisation des données considérées, et dans la seconde les outils mis en œuvre.

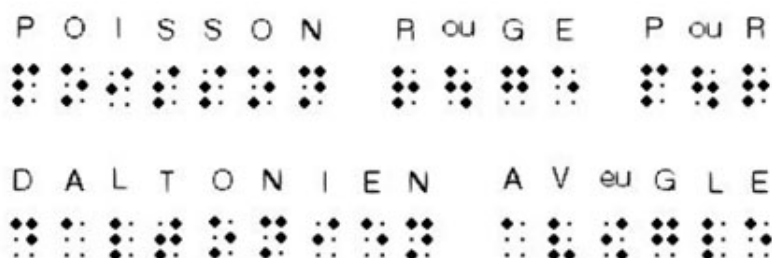


Table des matières

1	Cas concret : rapport interactif pour données expérimentales	3
1.1	Enjeux et besoins	3
1.2	Résultats obtenus	3
1.3	Critiques et perspectives	5
2	Implémentations : logiciels et utilisations	6
2.1	Présentation des logiciels employés	6
2.2	Exemples et utilisations	6
2.3	Méthode d'automatisation	14

1 Cas concret : rapport interactif pour données expérimentales

1.1 Enjeux et besoins

Le recours à des volumes croissants de données expérimentales [6], notamment en expérimentation haut-débit, génère des besoins de traitement de plus en plus importants. Pour l'analyse expérimentale de ces données, la visualisation [6, 10] reste une étape importante, bien que parfois négligée [3, 4].

Les données choisies pour ce stage (extrait en figure 1) proviennent de différentes expérimentations physico-chimiques menées à IFP Energies nouvelles (IFPEN). Elles visent à qualifier et quantifier la performance de réactions chimiques, en fonction de combinaisons de paramètres expérimentaux (typiquement de nature, proportions de réactifs, pré-traitements). L'objectif final est la détermination des conditions expérimentales fournissant la meilleure performance réactionnelle. Ces données comportent des variables de types mixtes (numériques et catégorielles) en nombre conséquent (10 à 30). Une représentation visuelle parlante de leurs combinaisons 2 à 2 et 3 à 3 est donc un enjeu de premier ordre, du diagnostic expérimental à l'analyse de relations latentes entre variables. Pour ce rapport, nous avons sélectionné un sous-ensemble représentatif de ces variables.

2,389291	0,487719	0,128198	0,622682	0,093289	0,493855	2,278257	1,484719	76,538955	75,588884	0,888951	56,427547	72,168957	46,354935	0,888879	57,388482	73,277821	47,482796	0,622675	0,642298	0,36797	0,74888
3,358461	0,448553	0,167832	0,185184	0,188715	0,042518	1,929717	1,046154	76,238482	75,238862	1,088416	79,828832	64,354222	38,748479	79,838518	68,348386	57,723939	0,848735	1,553216	1,354387	0,74217	
2,448252	0,321934	0,148523	0,393735	0,127321	0,157718	1,887935	76,272842	75,19393	1,088972	36,219184	71,488183	45,388913	1,513884	58,144451	79,321488	45,354484	0,888739	0,888828	0,498828	0,72717	
0,651517	0,220883	0,155514	0,138729	0,188932	0,164472	1,555932	1,188467	76,244386	75,254887	0,818889	72,533318	48,538911	42,368889	0,832155	74,238454	48,454527	0,671488	0,491284	0,471521	0,77946	
0,645281	0,724518	0,888154	0,758387	0,328786	1,747558	0,388718	75,851824	75,333333	0,158265	56,888567	68,43824	45,5488	0,837985	54,538212	68,8387	42,372832	0,835587	0,835587	0,35858	0,35858	
1,197785	1,082839	0,151824	0,518935	0,122775	0,148884	2,838963	0,354462	75,582814	75,488923	2,045889	34,888823	62,47893	39,888893	1,432874	54,472825	66,387428	49,39382	0,647241	0,393826	0,58816	
2,324775	0,853819	0,157457	0,588741	0,171125	0,164539	3,088828	2,884464	75,471813	74,152883	1,618238	76,448812	79,735147	45,39153	1,432558	58,888727	79,482734	48,428814	1,518725	1,454422	0,588246	0,5882
2,388854	0,642784	0,148523	0,393735	0,127321	0,157718	1,887935	76,272842	75,19393	1,088972	36,219184	71,488183	45,388913	1,513884	58,144451	79,321488	45,354484	0,888739	0,888828	0,498828	0,72717	
0,482579	0,145895	0,148885	0,558155	0,182582	0,378274	1,736244	0,838578	75,44385	74,61882	0,162783	76,384815	62,48236	48,555344	1,432874	54,472825	66,387428	49,39382	0,647241	0,393826	0,58816	
1,258374	1,584328	0,124781	0,788841	0,124781	0,124781	1,247813	1,247813	75,48893	75,48893	0,851516	37,29382	71,56179	0,838887	71,56179	0,838887	71,56179	0,838887	0,838887	0,838887	0,838887	
2,545244	0,127181	0,171475	0,168826	0,171475	0,171475	1,247813	1,247813	75,48893	75,48893	0,851516	37,29382	71,56179	0,838887	71,56179	0,838887	71,56179	0,838887	0,838887	0,838887	0,838887	
2,884827	1,378884	0,128464	0,788841	0,124781	0,124781	1,247813	1,247813	75,48893	75,48893	0,851516	37,29382	71,56179	0,838887	71,56179	0,838887	71,56179	0,838887	0,838887	0,838887	0,838887	
2,574787	0,178848	0,191895	0,168826	0,171475	0,171475	1,247813	1,247813	75,48893	75,48893	0,851516	37,29382	71,56179	0,838887	71,56179	0,838887	71,56179	0,838887	0,838887	0,838887	0,838887	
2,428371	0,164887	0,148523	0,393735	0,127321	0,157718	1,887935	76,272842	75,19393	1,088972	36,219184	71,488183	45,388913	1,513884	58,144451	79,321488	45,354484	0,888739	0,888828	0,498828	0,72717	
2,251548	0,161484	0,151824	0,518935	0,122775	0,148884	2,838963	0,354462	75,582814	75,488923	2,045889	34,888823	62,47893	39,888893	1,432874	54,472825	66,387428	49,39382	0,647241	0,393826	0,58816	
0,497853	1,351955	0,128198	0,622682	0,093289	0,493855	2,278257	1,484719	76,538955	75,588884	0,888951	56,427547	72,168957	46,354935	0,888879	57,388482	73,277821	47,482796	0,622675	0,642298	0,36797	0,74888
2,124583	1,354733	0,119388	0,144487	0,138887	0,164548	2,418848	1,338818	76,48727	76,88888	1,438816	36,162793	76,478841	47,281973	1,438816	36,162793	76,478841	47,281973	1,438816	36,162793	76,478841	
0,179191	1,555385	0,175966	0,172464	0,188835	0,164548	2,418848	1,338818	76,48727	76,88888	1,438816	36,162793	76,478841	47,281973	1,438816	36,162793	76,478841	47,281973	1,438816	36,162793	76,478841	
2,354815	0,185738	0,187335	0,174225	0,144835	0,184884	1,827427	1,194848	76,48432	76,35388	0,151235	38,88888	76,387184	47,388788	0,151235	38,88888	76,387184	47,388788	0,151235	38,88888	76,387184	
2,381884	1,531945	0,157243	0,538823	0,189758	0,152867	2,587979	79,864281	76,84887	1,346224	37,15242	79,38718	1,3887457	1,3887457	1,3887457	1,3887457	1,3887457	1,3887457	1,3887457	1,3887457	1,3887457	
0,324735	1,538178	0,124848	0,788841	0,124781	0,124781	1,247813	1,247813	75,48893	75,48893	0,851516	37,29382	71,56179	0,838887	71,56179	0,838887	71,56179	0,838887	0,838887	0,838887	0,838887	
1,418237	0,183195	0,151824	0,518935	0,122775	0,148884	2,838963	0,354462	75,582814	75,488923	2,045889	34,888823	62,47893	39,888893	1,432874	54,472825	66,387428	49,39382	0,647241	0,393826	0,58816	
0,421948	0,142852	0,185581	0,548535	0,193883	0,124818	2,397452	1,723813	79,73289	76,38847	2,885897	1,515888	48,588225	2,275885	38,89235	62,58878	48,588424	0,485471	0,485552	0,85882	0,5888	
2,825842	1,388195	0,184248	0,788841	0,124781	0,124781	1,247813	1,247813	75,48893	75,48893	0,851516	37,29382	71,56179	0,838887	71,56179	0,838887	71,56179	0,838887	0,838887	0,838887	0,838887	
2,248418	1,516238	0,187854	0,888888	0,144737	0,144737	1,444737	1,444737	76,38833	76,38833	1,293818	37,15145	75,16142	48,42554	0,847744	39,12883	58,387458	1,438718	1,438718	1,438718	1,438718	
1,571822	1,438718	0,187854	0,888888	0,144737	0,144737	1,444737	1,444737	76,38833	76,38833	1,293818	37,15145	75,16142	48,42554	0,847744	39,12883	58,387458	1,438718	1,438718	1,438718	1,438718	
2,288422	1,582488	0,188177	0,888888	0,144737	0,144737	1,444737	1,444737	76,38833	76,38833	1,293818	37,15145	75,16142	48,42554	0,847744	39,12883	58,387458	1,438718	1,438718	1,438718	1,438718	
1,325833	1,582135	0,148391	0,651834	0,187842	0,152239	2,354814	1,45181	79,48812	79,39382	0,148883	48,48883	62,39452	48,48883	2,888876	1,388816	67,38888	45,38881	1,438816	1,438816	1,438816	
1,557374	1,882181	0,151824	0,518935	0,122775	0,148884	2,838963	0,354462	75,582814	75,488923	2,045889	34,888823	62,47893	39,888893	1,432874	54,472825	66,387428	49,39382	0,647241	0,393826	0,58816	
1,551575	1,551884	0,188888	0,187483	0,188888	0,188888	1,888888	1,888888	76,38833	76,38833	1,293818	37,15145	75,16142	48,42554	0,847744	39,12883	58,387458	1,438718	1,438718	1,438718	1,438718	
1,48145	0,181457	0,228747	0,152185	0,142818	0,151887	1,191849	2,484878	79,81474	76,38833	2,824876	1,588832	48,58833	1,588833	1,588833	1,588833	1,588833	1,588833	1,588833	1,588833	1,588833	
2,512235	1,423937	0,155141	0,151834	0,151834	0,151834	1,518341	1,518341	79,73289	76,73289	1,418341	38,73289	76,73289	1,418341	38,73289	76,73289	1,418341	38,73289	76,73289	1,418341	38,73289	
0,338845	0,167488	0,152239	0,152239	0,152239	0,152239	1,522391	1,522391	76,73289	76,73289	1,418341	38,73289	76,73289	1,418341	38,73289	76,73289	1,418341	38,73289	76,73289	1,418341	38,73289	
0,364794	0,168136	0,145521	0,145521	0,145521	0,145521	1,455211	1,455211	76,73289	76,73289	1,418341	38,73289	76,73289	1,418341	38,73289	76,73289	1,418341	38,73289	76,73289	1,418341	38,73289	
1,588391	1,588391	0,188391	0,188391	0,188391	0,188391	1,883911	1,883911	76,73289	76,73289	1,418341	38,73289	76,73289	1,418341	38,73289	76,73289	1,418341	38,73289	76,73289	1,418341	38,73289	
1,125772	0,178848	0,188888	0,188888	0,188888	0,188888	1,888888	1,888888	76,73289	76,73289	1,418341	38,73289	76,73289	1,418341	38,73289	76,73289	1,418341	38,73289	76,73289	1,418341	38,73289	
1,155717	0,188155	0,188155	0,188155	0,188155	0,188155	1,881551	1,881551	76,73289	76,73289	1,418341	38,73289	76,73289	1,418341	38,73289	76,73289	1,418341	38,73289	76,73289	1,418341	38,73289	

FIGURE 1 – Extrait du fichier de données à visualiser.

1.2 Résultats obtenus

La figure suivante présente un extrait représentatif du rapport généré à partir de ces données.

Points									
Support :	E12	F636	F646	D2770	D2811	N7734	N7754	N10181	N288624
Traitement :	T1	T2	T3	T4					
Précurseur :	P0	P1	P2	P0 et P1					
Teneur métal 1 :	0.125	0.25	0.5	0.75	1	2	4		
Teneur métal 2 :	0	0.5	1						

Le rapport complet contient $C_{12}^3 = 220$ pages (voir fichier joint¹) sur le même modèle. Chacune des pages représente les résultats numériques de 3 produits générés lors de 127 expériences. Un total 12 produits des expériences est considéré, ainsi que 5 paramètres expérimentaux.

Ces produits sont représentés 2 à 2 en 2D (haut, et en bas à gauche) avec la possibilité de sélectionner les paramètres expérimentaux associés, ainsi qu'en 3D (en bas à droite, cliquer pour activer).

1.3 Critiques et perspectives

Cette première représentation des données numériques nous permet d'observer certains comportements globaux (notamment une dépendance linéaire dans la production d'éther et d'éthylène); on peut aussi observer et comparer les résultats en fonction des paramètres expérimentaux associés. Mais l'information devient rapidement illisible lorsque l'on superpose plusieurs paramètres expérimentaux. Le choix des représentations est assez sommaire (cercles de couleur et de taille différentes imbriquées) et mérite amélioration.

Une représentation 3D permet de confirmer ou d'infirmer certaines des intuitions en 2 dimensions. Cependant, étant moins habituelle, elle demande également une certaine pratique de la part des utilisateurs. Par ailleurs, un tel rapport pèse quelques 400 MB et requiert une dizaine d'heures de compilation à l'heure actuelle, cf. section 2.3.

Le programme générant ce rapport (voir dossier joint)² fournit une base pour créer des rapports PDF pour d'autres types de données; en s'appuyant sur des techniques de visualisation (citées en références), on peut aussi construire (section 2.2) une palette de représentations et l'appliquer à l'ensemble du rapport. Mais ce premier essai reste incomplet, et est encore loin de dévoiler clairement les liens subtils cachés dans un ensemble de données. Il serait intéressant d'ajouter des fits (figure 9) et informations statistiques (figure 10), d'afficher les points aberrants (section 2.3) et éventuellement d'ajouter des légendes textuelles caractérisant des singularités. Afin d'offrir d'avantage de lisibilité, les pages du rapport pourraient être ordonnées en fonction de certains critères, par exemple des relations entre produits possédant les mêmes comportements (tendance linéaire, hyperbolique...), et ensuite classés hiérarchiquement en fonction de la pertinence des corrélations. Pour apporter ces informations supplémentaires, il sera nécessaire d'intégrer des outils d'analyse numérique au rapport. Dans le même esprit, il serait envisageable de réaliser des fonctions d'analyse des données (régressions linéaires, t-SNE[9]...) afin de concentrer l'information pour en extraire l'essentiel. Au-delà de la visualisation, les techniques permettant de réaliser des PDF enrichis (son, 3D, interactions) sont rarement utilisées car encore trop peu connues, le chapitre suivant présente plusieurs logiciels permettant la mise en œuvre de PDF enrichis. La réalisation d'un PDF complet par commande demande du temps, une autre perspective serait de développer une interface simple pour réaliser ces PDF riches plus simplement.

1. En raison de la nature des données, le rapport de visualisation interactive produit à l'occasion de ce stage est à diffusion interne IFPEN uniquement.

2. En raison de la nature des données, le rapport de visualisation interactive produit à l'occasion de ce stage est à diffusion interne IFPEN uniquement.

2 Implémentations : logiciels et utilisations

2.1 Présentation des logiciels employés

Les principaux outils logiciels déployés durant ce stage sont les suivants :

- pdf est un format portable, il a la réputation de fournir produit fini, visualisable sur la plupart des machines Il peut contenir textes, sons, images, modèles 3D, interactions et calculs.
- L^AT_EX est un langage libre pourvu d’une communauté active, il permet de générer des PDF riches notamment avec l’utilisation de packages :
 - animate (voir tutoriel),
 - media9/movie15 (voir tutoriel),
 - asy (voir exemples),
 - eforms (voir tutoriel).
- Asymptote est un logiciel libre permettant de générer des figures en 2 et 3 dimensions.
- Ghostscript est un logiciel libre permettant le traitement des formats de fichiers PostScript.
- Texmaker est un logiciel libre permettant de compiler le langage L^AT_EX.
- Adobe Reader est un logiciel libre permettant de lire les PDF à contenu riche (3D, animation, JavaScript).
- Matlab est un logiciel payant permettant de réaliser des calculs, dans notre cas, il servira surtout à automatiser l’écriture des scripts (Asymptote, L^AT_EX, JavaScript) correspondant à nos données (cela fonctionnerait tout aussi bien en C, fortran, python).

2.2 Exemples et utilisations

Asymptote [7, 2] permet de construire des axes (mots-clés : *xaxis*, *yaxis*), en figure 2 :

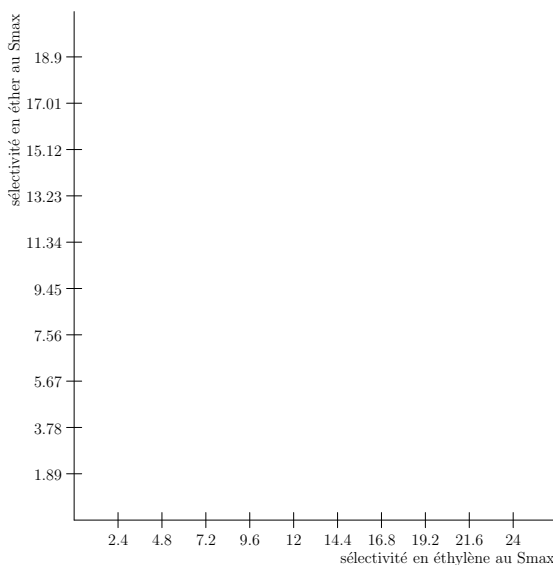


FIGURE 2 – Exemple d’axes de figures.

d'y représenter des points par leurs coordonnées (mots-clés : *dot*), en figure 3 :

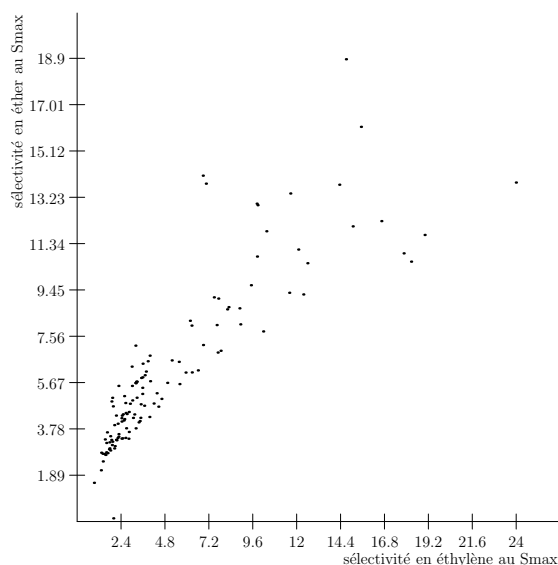


FIGURE 3 – Exemple de nuage de points.

de construire des formes (mots-clés : *draw*, *unitcircle*, *path*, *polygon*, *cross*), en figure 4 :

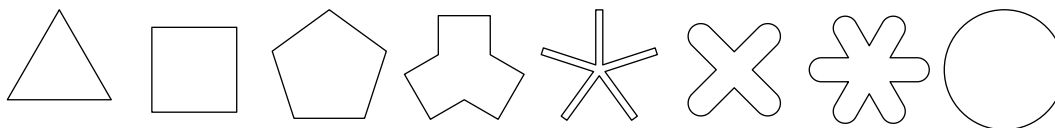


FIGURE 4 – Exemple de formes.

d'en modifier la taille (mots-clés : *scale*), en figure 5 :

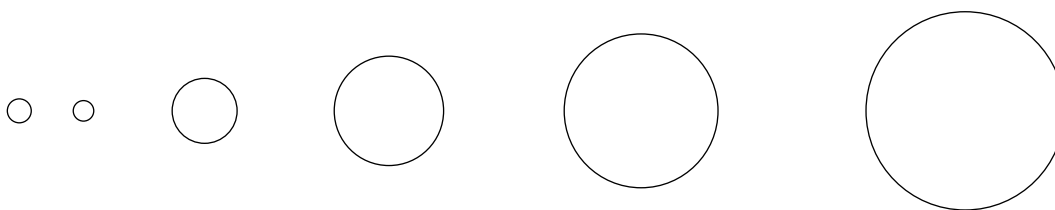


FIGURE 5 – Exemple de choix de tailles.

de choisir la couleur et le remplissage (mots-clés : *filldraw*, *pen*, *pattern*, *opacity*), en figure 6 :

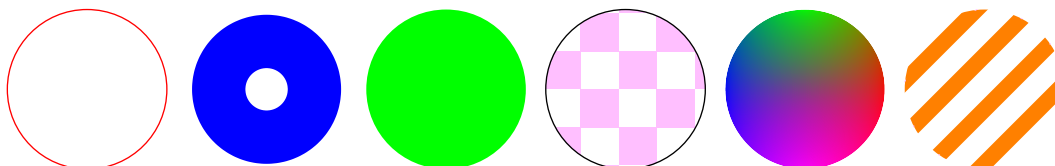


FIGURE 6 – Exemple de couleurs.

d'ajouter du texte (mots-clés : *string*, *label*, *pen*), en figure 7 :

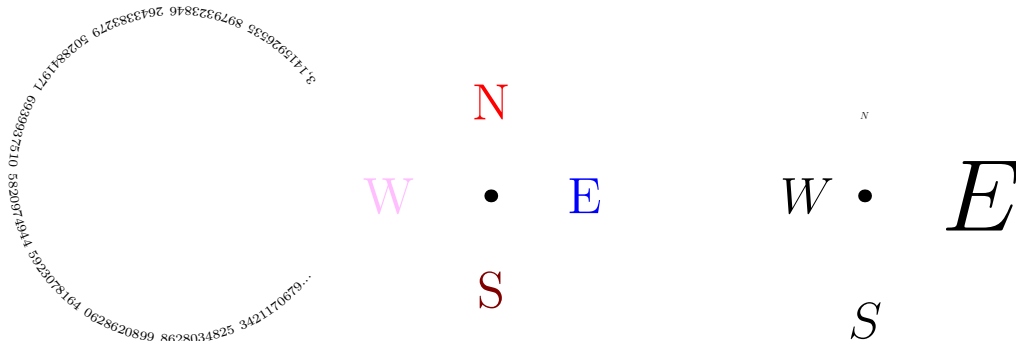


FIGURE 7 – Exemple de texte.

On peut ensuite placer ou déplacer ces figures dans le plan (mots-clés : *shift*), figure 8 :

FIGURE 8 – Exemple représentation forme et couleur dans le plan

On peut aussi tracer des courbes (mots-clés : *graph*), en figure 9 :

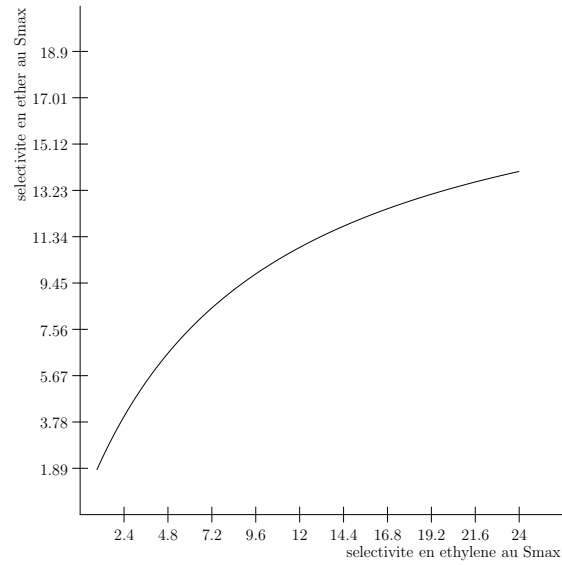


FIGURE 9 – Exemple de graphe.

et réaliser des diagrammes (mots-clés : *path*, *transform*, *for*) en figure 10 :

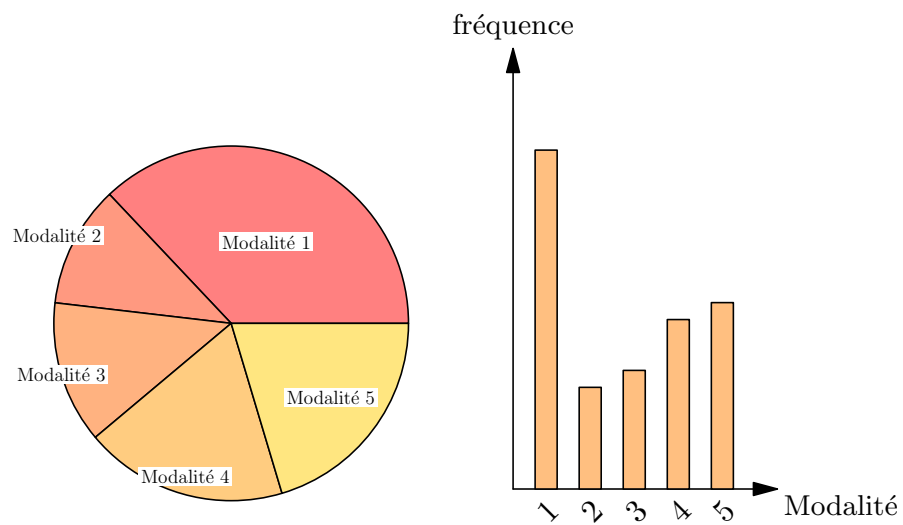


FIGURE 10 – Exemple de diagramme.

eforms [8] permet d'utiliser JavaScript.

On peut réaliser des outils de calcul.

écrivez votre calcul

(exemples : $3+2/6$ ou `Math.pow(35,2)` pour 35^2)

résultat

On peut faire appel à des éléments extérieurs (données web, gps, sorties programmes).
cliquer pour afficher l'heure

animate [1] permet de réaliser des animations (linéaires) à partir d'image 2D (mots-clés : *animateinline*), en figure 11 :

FIGURE 11 – Exemple d'animation.

On peut aussi transformer les images en calques et les superposer (mots-clés : *timeline*), en figure 12 :

FIGURE 12 – Exemple d’animation avec calques.

En utilisation simultanée, l’idée est de réaliser des images avec Asymptote, de les superposer avec animate et sélectionner les calques avec eforms, en figure 13 :

FIGURE 13 – Exemple de sélection calques



Un autre exemple interactif est donné ci-après en figure 14 :

FIGURE 14 – Exemple interactif 1.

FIGURE 15 – Exemple interactif 2.












media9 [5] Asymptote permet de créer des figures dans l'espace au format PRC (mots-clés : *three*, *surface*, *grid3*, *axis*), et media9 permet d'intégrer ce format au fichier PDF (mots-clés : *includemedia*), en figure 16 :

FIGURE 16 – Exemple de format prc

On peut ensuite construire une "toolbar maison" pour sélectionner les vues ; par exemple projeter le graphe sur le plan d'abscisse ou d'ordonnée nulle $x=0$ $y=0$

media9 permet aussi d'utiliser des scripts JavaScript, de réaliser des animations 3D, d'ajouter des sons et des interactions .

2.3 Méthode d'automatisation

Métal 1	Quantité 1		Quantité 2		Quantité 3		
Métal 2	Quantité 1		Quantité 2		Quantité 3		Quantité 4 
Supports	Numéro 1		Numéro 2		Numéro 3		Numéro 4 
Points	Approximation		Outlier potentiel				

La réalisation d'une image interactive comme ci-dessus demande du temps ; avant tout, il faut décider du :

- choix des représentations (*exemple : les variables 1 et 2 seront représentées par leur position, la variable 3 par la forme du point, la 4 par sa couleur, la 5 par un label textuel, un fit par moindres carrés*).
- choix des interactions possibles (*exemple : on choisira de pouvoir voiler/dévoiler certaines colonnes, calculer la médiane des points sélectionnés*).

Une fois ces décisions prises, on génère les images nécessaires, on les intègre à un calque avec animate et on écrit le code JavaScript correspondant aux interactions. Il faut ensuite intégrer ces informations à L^AT_EX pour compiler notre PDF.

Pour quelques images et interactions, l'écriture peut être réalisée manuellement, pour un travail plus important, il est nécessaire d'utiliser un programme qui automatisera la tâche.

Dans le cas du rapport , plusieurs fichiers textes sont préremplis :

- un fichier texte comprenant un jeu de légendes Asymptote,
- un fichier texte à trous comprenant des scripts Asymptote pour gérer les repères et la position des points,
- un fichier texte à trous comprenant le code JavaScript,
- un fichier texte à trous comprenant les calques animate,
- un fichier texte à trous comprenant script \LaTeX de mise en page et intégration des informations précédentes,

Trois programmes Matlab (voir fichiers joints)³ remplissent les espaces à partir des données, compilent puis assemblent le document \LaTeX et enfin créent le PDF. L'intérêt de disposer de plusieurs programmes est par exemple de pouvoir modifier la mise en page sans devoir recompiler toutes les images.

Dans le document PDF final, le parti pris de ce programme a été de solliciter la mémoire plutôt que le calcul. Ce qui donne un document fluide mais lourd. En effet, il y est généré le nombre d'images correspondant aux interactions. *Par exemple : on a 10 points, on veut pouvoir les voiler/dévoiler individuellement, ce qui nous donne $2^{10} = 1024$ images à créer.* Le logigramme d'exécution de ces tâches est présenté en figure 17.

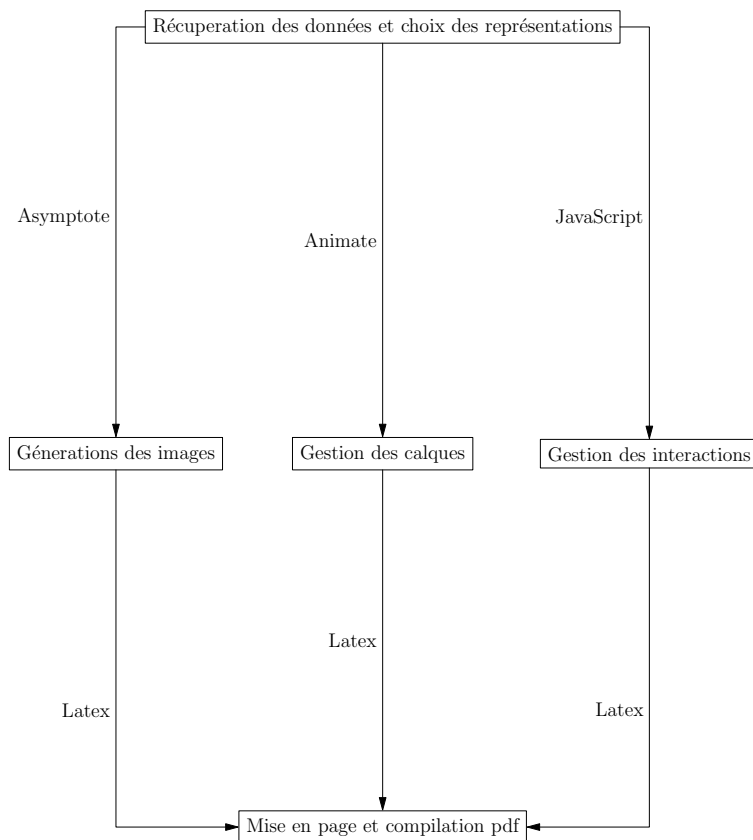


FIGURE 17 – Implémentation des tâches de génération de fichier PDF.

3. En raison de la nature des données, le rapport de visualisation interactive produit à l'occasion de ce stage est à diffusion interne IFPEN uniquement.

References

- [1] animate, 2010. <http://wiki.mathematex.net/doku.php?id=wiki:latex:modules:animate>. 10
- [2] asy.marris, 2014. <http://asy.marris.fr/asymptote/index.html>. 6
- [3] F. J. Anscombe. Graphs in statistical analysis. *Am. Stat.*, 27:17–21, 1973. 3
- [4] R. Dennis Cook and Sanford Weisberg. Graphs in statistical analysis: Is the medium the message? *Am. Stat.*, 53:29–37, 1999. 3
- [5] Alexander Grah. *The media9 Package*, 2016. 13
- [6] C. D. Hansen and C. R. Johnson, editors. *The visualization handbook*. Academic Press, 2005. 3
- [7] Andy Hammerlindl John Bowman Tom Prince. *Asymptote: the Vector Graphics Language*. <http://asymptote.sourceforge.net>, 2016. 6
- [8] D. P. Story. *eforms and insdljs Documentation*. AcroTEX.Net, 2010. 10
- [9] L. J. P. van der Maaten and G. E. Hinton. Visualizing high-dimensional data using t-SNE. *J. Mach. Learn. Res.*, 9:2579–2605, 2008. 5
- [10] Matthew O. Ward, Georges Grinstein, and Daniel Keim. *Interactive Data Visualization: Foundations, Techniques, and Applications*. A. K. Peters/CRC Press, 2nd edition, 2015. 3