HAL
open science

Semaine d'Etude Mathématiques et Entreprises 6 : Analyse statistique des défauts en électronique analogique

Julien Fageot, Morgane Henry, Alexis Huet, Gildas Mazo, Stéphane Veys

To cite this version:

Julien Fageot, Morgane Henry, Alexis Huet, Gildas Mazo, Stéphane Veys. Semaine d'Etude Mathématiques et Entreprises 6 : Analyse statistique des défauts en électronique analogique. 2013. hal-00933235

HAL Id: hal-00933235

https://hal.science/hal-00933235v1

Preprint submitted on 14 Feb 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SEMAINE D'ETUDE MATHS-ENTREPRISES

A Grenoble du 10 au 14 juin 2013

Analyse statistique des défauts en électronique analogique

Julien Fageot[a], Morgane Henry[b], Alexis Huet[c],
Gildas Mazo[b,d] & Stéphane Veys[b]

[a] *Laboratoire d'imagerie biomédicale, Ecole Polytechnique Fédérale de Lausanne*
[b] *Laboratoire Jean Kuntzmann, CNRS UMR 5224, Université de Grenoble*
[c] *Institut Camille Jordan, CNRS UMR 5208, Université de Lyon*
[d] *Inria Rhône-Alpes*

Sujet proposé par :

HALIAS Technologies

Correspondant : Laurent Testard

1 Introduction

Nous nous intéressons à des données issues de mesures de tensions sur des circuits électroniques analogiques. Plus précisément, il s'agit de proposer une analyse de courbes représentant l'évolution en fonction du temps des tensions en différents nœuds d'un circuit électronique. De tels circuits, constitués d'éléments miniaturisés, se caractérisent par leur très grand nombre de nœuds (potentiellement plusieurs dizaines de millions), et donnent ainsi potentiellement lieu à un très grand nombre de courbes de tension, qu'il s'agit ici d'étudier.

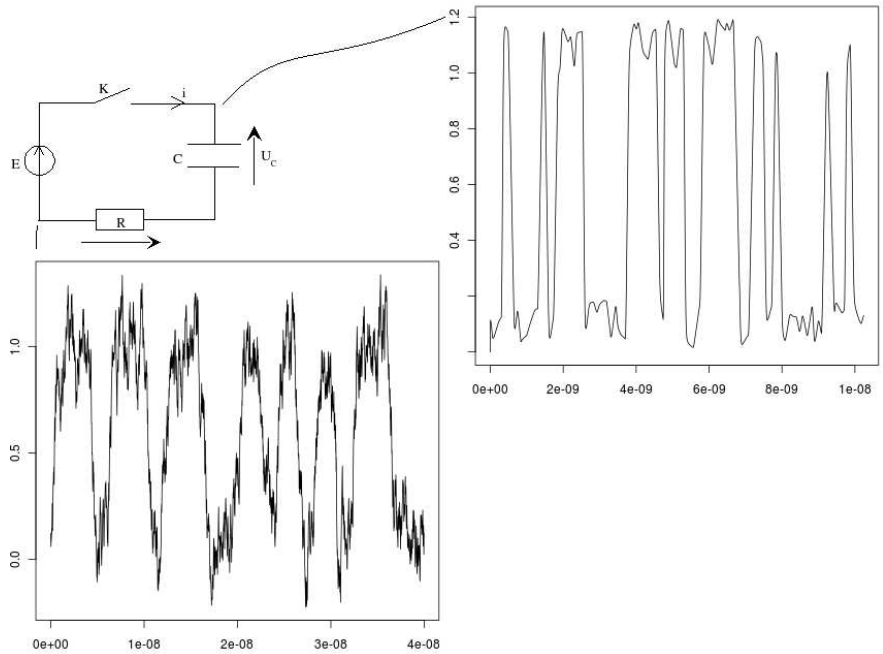


FIGURE 1 – Données : courbe de tension aux nœuds d'un circuit électronique

Comme on le voit dans la figure 1, les données qui nous intéressent sont des mesures de tension correspondant à des commutations dans le circuit électronique (on alterne successivement entre les valeurs 0 et $V_{\max}$). C'est un signal de type créneau. Cependant, les courbes de tension présentent en pratique de nombreuses "imperfections" : les temps de commutation ne sont pas instantanés, il y a aussi des décalages entre les temps de commutation alors qu'ils devraient être quantifiés, etc. Ces défauts proviennent à la fois des propriétés physiques du circuit et des instruments de mesure de tension.

Concrètement, nous disposons de deux types de données sur ce projet : (i) des données simulées à l'aide d'un logiciel spécialisé destinées à identifier des indicateurs généraux pertinents avec une palette de comportements différents susceptibles d'être observés en pratique et (ii) des données industrielles utilisées pour l'étude préalable d'une puce électronique avant sa phase de fabrication présentant un problème spécifique qu'il nous était demandé d'identifier automatiquement sur les courbes.

Notre objectif est de proposer une analyse automatisée de la qualité des courbes. Plus précisément, nous proposons ici des méthodes statistiques d'analyse de données capable de :

— Identifier d'éventuels patterns dans les courbes (classification),
— Isoler les courbes présentant des "anomalies" (détection de courbes suspectes).

Méthodologie — utilisation d'indicateurs

Nous proposons ici une méthode d'analyse s'adaptant au type de données proposé. Les étapes de cette méthode sont les suivantes :

— Identifier les caractéristiques jugées pertinentes des signaux,
— Définir les indicateurs permettant de capturer ces caractéristiques,
— Définir une métrique à partir de ces indicateurs,
— Classifier les données à partir de cette métrique.

Plus formellement, supposons qu'on dispose de N signaux $(f_1, ..., f_N)$ (avec N grand en pratique, éventuellement plusieurs dizaines de millions) qui représentent des discrétisations de fonctions analogiques, *i.e.* f_n est un vecteur de taille fixe M (quelques centaines ou quelques milliers de points). On va définir K indicateurs (avec $K \ll M$) résumant les principales caractéristiques des signaux. On perd certes une partie de l'information de nos signaux mais nous espérons définir des indicateurs suffisamment pertinents qui contiennent l'information importante (et donc idéalement on négligerait l'information superflue). Imaginons donc qu'on dispose d'une fonction indicateurs I qui associe a un signal f ses indicateurs $(c_1(f), ..., c_K(f))$. Ces indicateurs seront soit des caractéristiques intrinsèques du signal (une estimation de sa période, de ses temps de commutation, etc.), soit des estimations de la qualité du signal (niveau du bruit, variance des indicateurs intrinsèques). On donne sur la figure 2 quelques exemples d'indicateurs : τ^+ représente le temps de montée et τ^- le temps de descente lors des commutations, τ_c représente une période, ou temps d'horloge du signal et σ_{bruit} la variance du bruit sur une zone où le signal est supposé être constant.

Ainsi, notre jeu de données initial est résumé par les $K \times N$ quantités $c_k(f_n)$ pour $(k, n) \in \{1, ..., K\} \times \{1, ..., N\}$.

On peut alors utiliser les indicateurs pour répondre à différents buts :

— Qualitatif : visualisation des données (au besoin par une analyse en composante principale (A.C.P.) si $K > 3$ notamment).
— Quantitatif : par le choix d'une méthode de classification, on peut chercher à discriminer les données (K-moyennes, modèle de mélange de gaussienne, etc.).
— Dans le cas d'une question précise sur l'analyse des courbes, on peut définir un indicateur spécialisé pour cette tâche qui permet par exemple de repérer un certain type d'anomalie (ce principe sera utilisé sur des données industrielles dans la suite de ce travail).

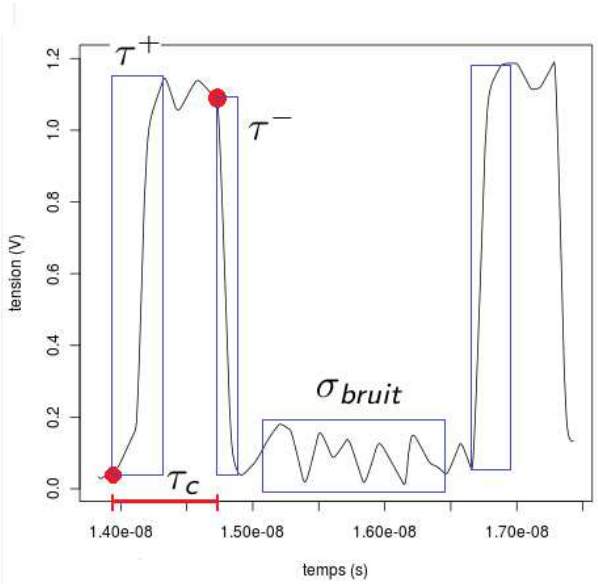


FIGURE 2 – Quelques indicateurs

Plan

Ce rapport s'organise maintenant comme suit. Dans les Sections 2 et 3, nous proposons des indicateurs généraux suivant des approches géométriques (Section 2) ou statistiques (Section 3). Dans ces deux sections, nous respectons la méthodologie proposée, en l'adaptant aux données électroniques que nous cherchons à analyser. Dans la Section 4, nous étudions un problème de nature légèrement différente. L'objectif est d'identifier des anomalies dans des courbes de tension fournies par l'entreprise Freescale. Ces anomalies sont précisément connues de l'utilisateur qui cherche ici un moyen automatique d'identifier ces anomalies afin d'obtenir une mesure précise de la période des signaux dont il dispose. Nous proposons concrètement des méthodes de détection des anomalies par la définition d'un indicateur adapté.

2 Recherche d'indicateurs et classification

L'objectif de cette partie est d'exhiber des indicateurs pertinents permettant de classer les différents signaux. Pour cela, on cherche à estimer les composantes caractéristiques du signal comme le taux de bruit ou la durée d'une phase de montée.

Nous avons choisi d'utiliser la méthode suivante, qui se décompose en deux étapes :

1. identification des instants de montée I^+, de descente I^-, de plateaux bas I^0 et de plateaux hauts I^1 (voir figure 3) ;
2. déduction de données caractéristiques pour chacun des ensembles I^+, I^-, I^0 et I^1.

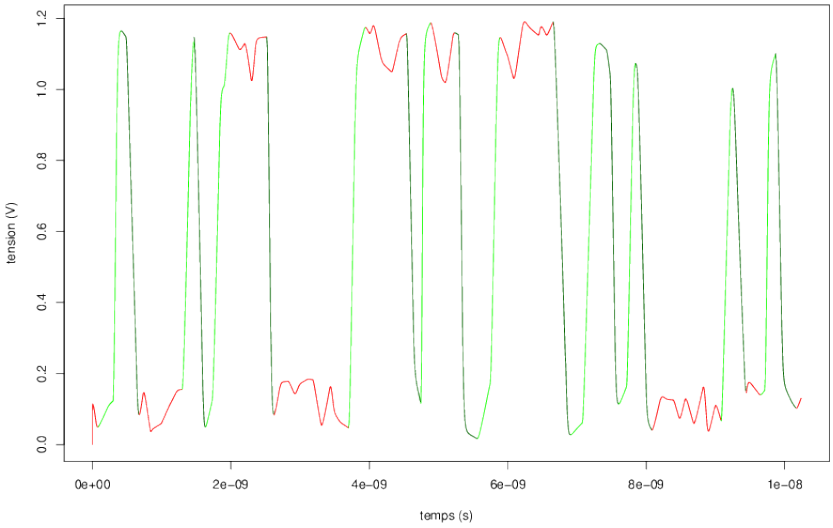


FIGURE 3 – Portion de signal où sont identifiés les ensembles I^+ (en vert clair), I^- (en vert foncé), I^0 et I^1 (en rouge).

2.1 Identification des instants de montée, de descente et des plateaux

La première étape est la plus délicate et peut être abordée selon des points de vue différents. Les méthodes que l'on utilise se basent sur la structure particulière de l'ensemble des signaux considérés :

— amplitudes des variations importantes sur la durée des phases de montée et de descente (de l'ordre de l'amplitude maximale du signal global).

— variations rapides pendant les phases de montée et de descente.

Un problème est que ces deux propriétés structurelles peuvent ne pas être caractéristiques des phases de montée et de descente (voir aussi figure 4) :

— dans le cas de présence de glitch, l'amplitude peut aussi être de l'ordre de l'amplitude maximale du signal global (problème traité en détail section 4) ;

— dans le cas d'un signal fortement bruité, des variations rapides ont lieu pendant les instants de plateaux.

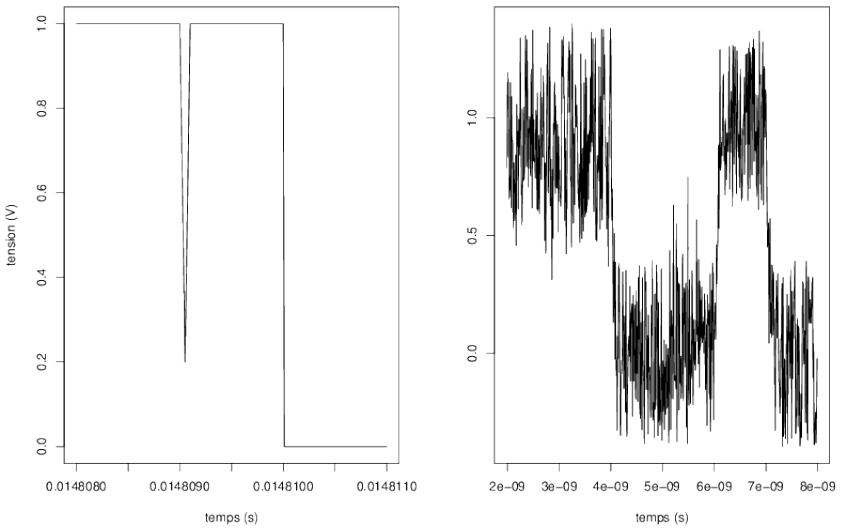


FIGURE 4 – Exemple de présence de glitch et d'un signal fortement bruité.

Nous décrivons ci-dessous deux méthodes, avant de vérifier leur pertinence sur différents jeux de données :
— une méthode par accroissements puis seuillage,
— une méthode mixte amplitude/accroissements.

2.1.1 Méthode par accroissements puis seuillage

Un signal est la donnée d'un ensemble discret $(t_i, x_i)_{i\in\{1,\dots,n\}} \in (\mathbb{R}\times\mathbb{R})^n$ tel que la suite $(t_i)_{i\in\{1,\dots,n\}}$ est strictement croissante. Pour $i \in \{1,\dots,n\}$, on dit alors que la tension x_i est associé à l'instant t_i.

Les accroissements du signal (resp. les accroissements normalisés) sont alors définis pour tout $i \in \{1,\dots,n-1\}$ par (respectivement) :

$$\Delta_i = \frac{x_{i+1}-x_i}{t_{i+1}-t_i} \qquad \text{et} \qquad \tilde{\Delta}_i = \frac{\Delta_i}{\max_{j\in\{1,\dots,n-1\}} \Delta_j}.$$

La figure 5 représente une portion d'un signal et son accroissement associé.

Description de la méthode. La méthode par seuillage consiste à rechercher un pourcentage $\alpha > 0$ tel que tous les indices i de montée et de descente vérifient $|\tilde{\Delta}_i| > \alpha$ et tous les instants de plateaux $|\tilde{\Delta}_i| \leq \alpha$. Un tel pourcentage existe lorsqu'il n'y a pas de variations rapides pendant les instants de plateaux.

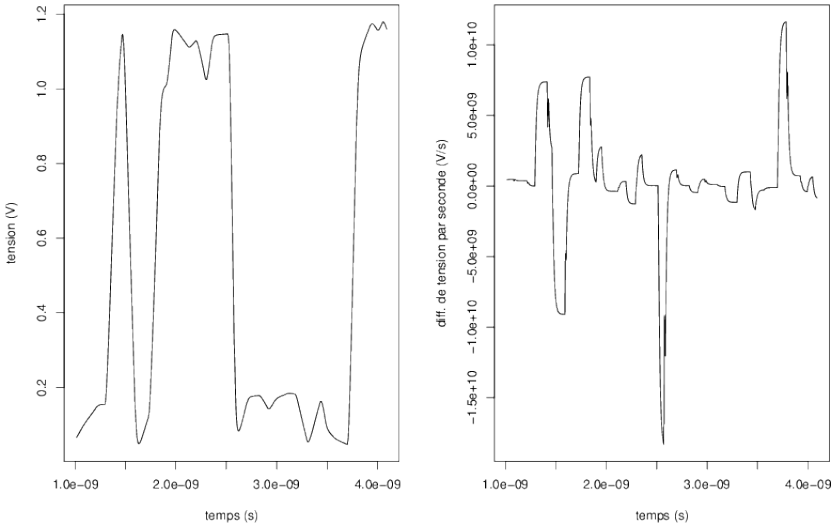


FIGURE 5 – Portion d'un signal et de son accroissement associé.

Choix du seuil.

Seuil minimal. On fait l'hypothèse supplémentaire que le signal est sur un plateau pour au moins 10% des indices (cette hypothèse est vérifiée dans tous les signaux étudiés). Sous cette hypothèse, on définit alors le seuil minimal α_0 vérifiant :

$$\frac{\text{card}\{i; |\tilde{\Delta}_i| < \alpha_0\}}{n-1} > 10\%.$$

Minimisation d'une fonction. On sait que le signal alterne pentes fortes ascendantes et pentes fortes descendantes. Cela se traduit sur les accroissements par des indices où $\tilde{\Delta}_i$ est grand et positif, suivi par des indices où $\tilde{\Delta}_i$ est proche de zéro et par des indices où $\tilde{\Delta}_i$ est grand et négatif.

Ainsi, pour chaque seuil $\alpha \in [\alpha_0, 1]$, on associe la fonction s_α qui à chaque indice i associe l'entier :

$$s_\alpha(i) = \begin{cases} -1 & \text{si } |\tilde{\Delta}_i| \geq \alpha \text{ et } \tilde{\Delta}_i < 0, \\ 0 & \text{si } |\tilde{\Delta}_i| < \alpha, \\ 1 & \text{si } |\tilde{\Delta}_i| \geq \alpha \text{ et } \tilde{\Delta}_i > 0. \end{cases}$$

On obtient une suite de longueur $n-1$ du type :

i	1	2	3	4	5	6	7	8	9	10	11	12	13	...
$s_\alpha(i)$	1	1	0	0	-1	-1	0	0	1	1	0	1	1	...

On sélectionne alors uniquement les indices $(i_1, \ldots, i_l) \subset \{1, \ldots, n-1\}$ où la fonction s varie $(s_\alpha(i) \neq s_\alpha(i-1))$ et est non nulle, avec la convention $i_1 = 1$, pour obtenir une suite du type :

i	1	5	9	12	...
$s_\alpha(i)$	1	-1	1	1	...

On note enfin $n_\alpha \in \{0, \ldots, l\}$ le nombre d'indices qui ne vérifie pas l'alternance entre -1 et 1, et le seuil choisi est alors celui qui minimise cette quantité (en privilégiant la valeur la plus grande en cas d'égalités) :

$$\alpha_* = \text{argmin}_{\alpha \geq \alpha_0} n_\alpha.$$

La figure 6 montre les étapes successives de la méthode pour le seuil fixé $\alpha = 0.1$. La figure 7 compare les seuils 0, 0.08 et 0.1 sur une portion de signal.

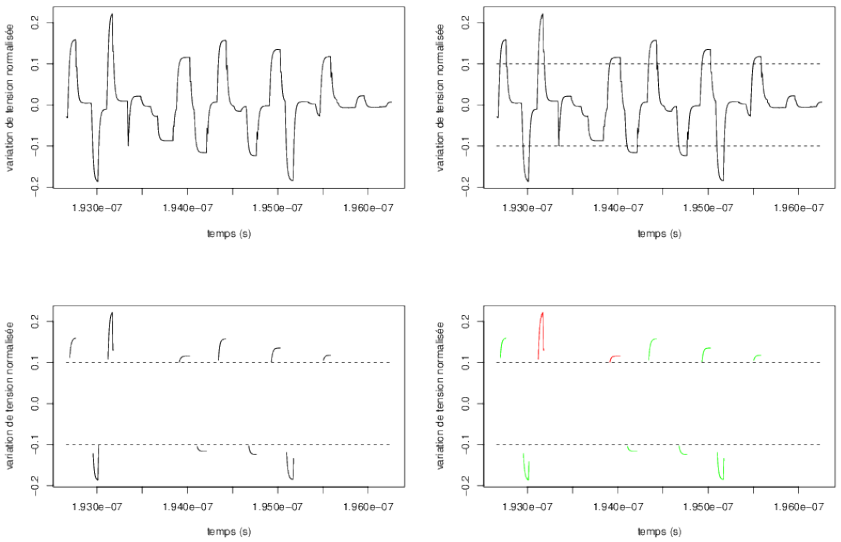


FIGURE 6 – Etapes successives de la méthode pour le seuil $\alpha = 0.1$. Le premier graphique est le signal des accroissements normalisé. On fixe ensuite le seuil à $\alpha = 0.1$ et on ne conserve que les indices vérifiant $|x_i| > \alpha$, ce qui permet de déduire la fonction s_α. On compte ensuite le nombre de morceaux qui ne vérifie pas l'alternance pour obtenir $n_\alpha = 1$: sur le graphique, le deuxième morceau rouge ne vérifie pas l'alternance puisqu'il est de même signe que le morceau précédent.

Résultats sur différents signaux. On calcule le seuil et n_α sur trois signaux, un non bruité, un légèrement bruité et un fortement bruité. On observe dans le tableau suivant le seuil choisi automatiquement et une mesure du bruit donnée par n_α/n (compris entre 0 et 1).

fichier	non bruité	peu bruité	fortement bruité
α_*	99%	5%	α_0
n_{α_*}/n	0	0.1	0.03

n_{α_*}/n est une mesure du bruit puisqu'elle permet de quantifier la proportion de pics dont l'accroissement est au moins de l'ordre de grandeur de l'accroissement des phases de montée ou de descente.

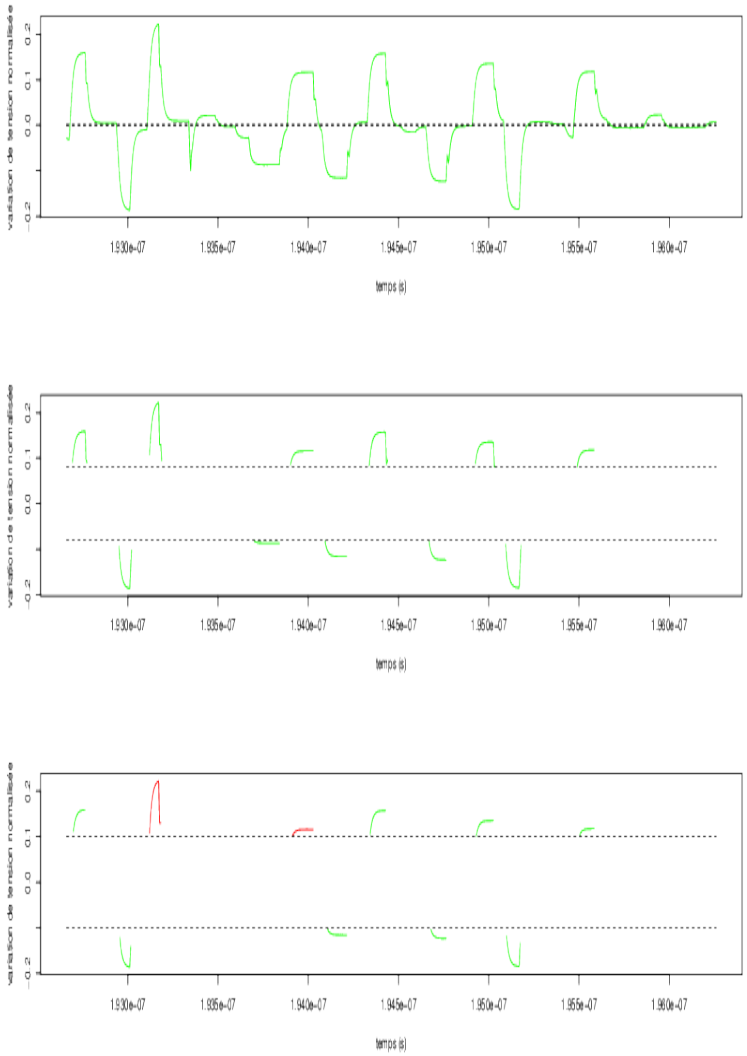


FIGURE 7 – Visualisation de n_α pour trois seuils différents (0, 0.08 et 0.1) sur une portion de signal. Dans le premier cas l'alternance est vérifiée mais $\alpha < \alpha_0$ le seuil minimal, ce seuil n'est donc pas convenable. Dans le deuxième cas, $n_\alpha = 0$ et le seuil est convenable. Dans le troisième cas, $n_\alpha = 2$.

On observe que dans le cas des signaux non bruité et peu bruité, le seuil est choisi au-delà de α_0 et la mesure du bruit augmente en fonction du taux de bruit du signal. Dans le cas d'un signal fortement bruité, le seuil choisi est le seuil minimal autorisé et la mesure du bruit n'est pas pertinente. On visualise sur la figure 8 une portion des trois signaux ainsi que l'identification des instants de montée/descente et de plateaux obtenus. On observe que l'identification est non exploitable dans le cas fortement bruité et est convenable dans les cas peu ou non bruités.

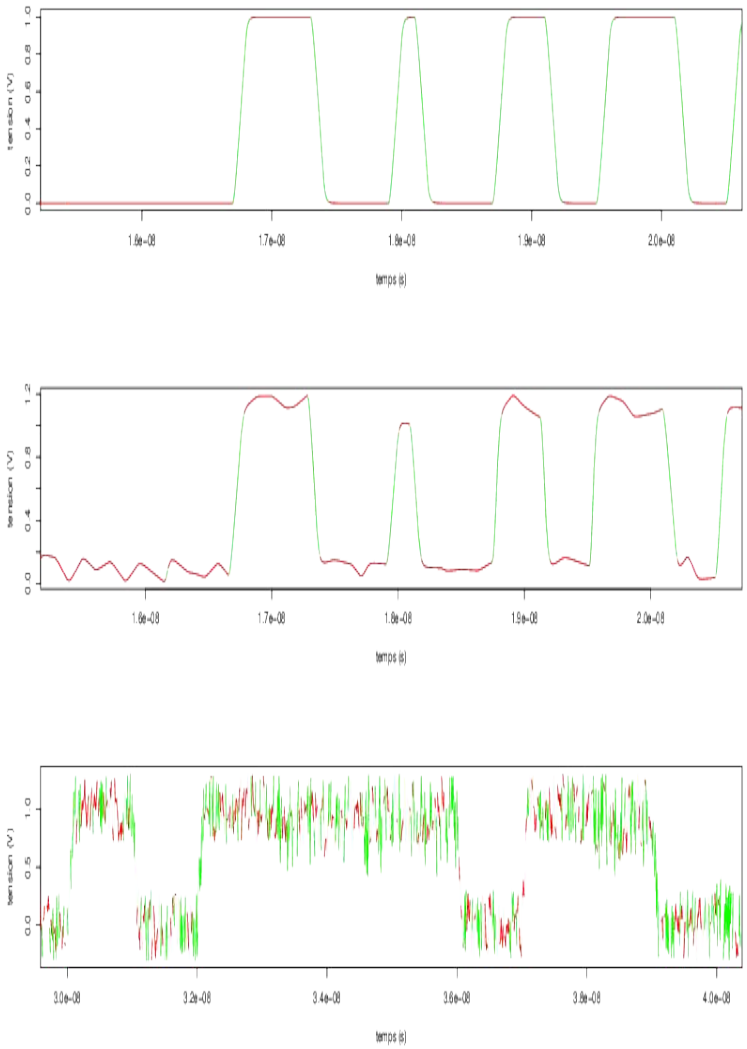


FIGURE 8 – Portion des signaux non bruité, légèrement bruité et fortement bruité, où sont identifiés les ensembles I^+ et I^- (en vert), ainsi que I^0 et I^1 (en rouge).

Conclusion sur cette méthode. La méthode par accroissements puis seuillage permet d'identifier automatiquement les ensembles I^+, I^-, I^0 et I^1 dans le cas d'un signal pour lequel les accroissements dus au bruit sont inférieurs aux accroissements des instants de montée/descente. Cela est vérifié dans le cas de signaux non ou peu bruités, mais ne convient pas pour un signal trop fortement bruité.

On a ici utilisé uniquement le fait que le signal considéré a des variations rapides pendant les phases de montées et de descentes. Pour obtenir une méthode plus robuste au bruit, il faut aussi considérer que le signal a des amplitudes

importantes sur la durée des phases de montées et de descentes. Cette méthode mixte amplitude/accroissements est traitée dans la section 2.1.2.

2.1.2 Méthode mixte amplitude/accroissements

La méthode mixte amplitude/accroissements utilise le fait que le signal a des accroissements d'amplitude importante sur la durée des phases de montée et de descente. La mesure de ces amplitudes est faite sur chaque morceau où l'accroissement est de signe constant. Cela permet de rendre plus robuste l'identification des ensembles I^-, I^+, I^0 et I^1 même en présence de bruit relativement important (précisément tant que l'amplitude des bruits n'est pas de l'ordre de l'amplitude des montées/descentes).

Description de la méthode. On reprend les notations de la section 2.1.1 et on choisit le seuil $\alpha = 0$. La fonction s_α est alors à valeurs dans $\{-1, 1\}$ et alternante. On obtient $(i_1, \ldots, i_l) \subset \{1, \ldots, n-1\}$ les indices où fonction s_α varie (c'est-à-dire passe de -1 à 1 ou de 1 à -1).

Pendant une montée (resp. une descente) du signal, le signe de l'accroissement est strictement positif (resp. négatif) et l'amplitude est élevée. Cela nous à conduit à mesurer, pour chaque $k \in \{1, \ldots, l\}$, l'amplitude a_k du signal sur l'intervalle $\{i_k, \ldots, i_{k+1} - 1\}$ (avec la convention $i_{l+1} = n$ dernier indice du signal). Cette amplitude est donnée par $|x(i_{k+1} - 1) - x(i_k)|$ et doit être proche de l'amplitude globale du signal dans le cas d'une montée ou d'une descente.

Un algorithme de classification, par exemple par k-moyennes avec 2 classes [1] permet de séparer les amplitudes importantes (montées/descentes) des amplitudes faibles (bruit du signal), la mesure de la qualité se faisant par un ratio variance expliquée sur variance totale. On en déduit alors les ensembles I^-, I^+, I^0 et I^1.

Raffinement de la méthode dans les cas très bruités. Deux types de bruits sont possibles :

1. Si les variations du bruit sont très fortes, il arrive que pendant une montée ou une descente le signe de l'accroissement ne soit pas constant. On peut alors filtrer le signal pour éliminer ces variations parasites (par exemple avec un filtrage par convolution contre la loi uniforme d'intervalle centré en zéro). Remarquons que le filtrage est ici utilisé uniquement pour l'identification des ensembles I^-, I^+, I^0 et I^1 (pas pour calculer les caractéristiques du signal).
2. Si les amplitudes du bruit sont fortes mais inférieures aux amplitudes de montées/descentes, l'ensemble $(a_k)_{k=\{1,\ldots,l\}}$ ne se divise pas clairement en deux classes et l'algorithme des k-moyennes n'est pas utilisable. On préfère alors utiliser un algorithme de type *expectation-maximization* [2].

Remarque. Dans le cas de présence de glitchs, cette méthode ne les détecte pas directement : ils sont soit éliminés dans le cas d'un filtrage, soit considérés comme des instants de montée/descente.

Résultats sur différents signaux. On applique la méthode sur trois signaux, un moyennement bruité, un bruité et un très bruité. On applique l'algorithme standard, et dans le cas où la mesure de la qualité n'est pas bonne, on utilise un filtrage et l'algorithme *expectation-maximization* gaussien pour identifier les ensembles I^+, I^-, I^0 et I^1 (visualisé sur la figure 9).

On visualise sur la figure 10 une portion des trois signaux ainsi que l'identification des instants de montée/descente et de plateaux obtenus. On observe que l'identification est concluante dans les deux premiers cas. Dans le troisième cas, l'identification est exploitable mais des instants de plateau sont identifiés comme une montée ou une descente (à cause du filtre mis en place). Dans tous les cas, toutes les périodes de montées et de descentes sont identifiées.

2.2 Déduction de données caractéristiques du signal

2.2.1 Caractéristiques utilisées

Après avoir identifié les instants de montées, de descentes et de plateaux, on en déduit des données caractéristiques pour chacun des ensembles identifiés. Les données caractéristiques qu'on l'on utilise pour classer les signaux sont les suivantes :

- nombre de montées (utilise I^+) et de descentes (utilise I^-).
- coefficient moyen d'une pente de montée (utilise I^+), d'une pente de descente (utilise I^-),
- écart type pendant les phases de plateaux hauts (utilise I^1), pendant les phases de plateaux bas (utilise I^0),
- éventuellement les durées de chaque plateau haut (utilise I^1), de chaque plateau bas (utilise I^0).

2.2.2 Tests des indicateurs

Les indicateurs ont été testés sur trois jeux de données : le premier pour des signaux où seul le temps d'horloge varie, le deuxième où seul le niveau de bruit varie et le troisième où seule la vitesse de montée varie. Les trois méthodes sont testées et conduisent aux mêmes types de graphiques, regroupés sur la figure 11.

2.2.3 Classification des signaux

On dispose d'un jeu de données contenant 300 signaux. Les paramètres variant sont le temps d'horloge (parmi 2 valeurs), la durée d'une montée (parmi 4 valeurs), la durée d'une descente (parmi 4 valeurs), le taux de bruit (parmi 2 valeurs). On cherche à classer les signaux selon ces paramètres. On calcule alors les différents indicateurs avec la méthode exposée dans la section 2.1.2.

On trace sur la figure 12 l'écart type estimé du plateau inférieur en fonction du nombre de montées estimés par unité de temps, et sur la figure 13 l'écart type estimé du plateau inférieur en fonction de la pente montante estimée.

On observe sur la figure 12 un partitionnement clair en quatre classes, permettant d'une part de différencier les temps d'horloge (à gauche *vs* à droite) et d'autre part de différencier les taux de bruit (en bas *vs* en haut).

Sur la figure 13, le partitionnement suivant les pentes montantes n'est pas net. Une raison probable est le fait que la méthode considère certains instants de plateau comme des instants de montée/descente (déjà observé sur la figure 10).

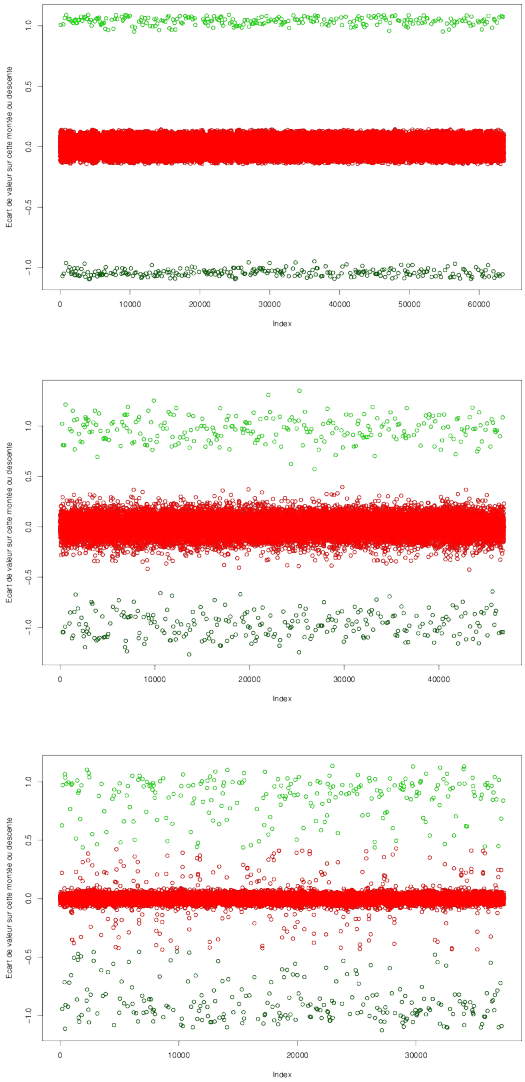


FIGURE 9 – Chaque graphique correspond à un signal bruité, de moyennement (en haut) à très bruité (en bas). Pour chaque graphique, l'abscisse représente l'indice d'une phase ascendante ou une phase descendante du signal (parmi $\{1, \ldots, l\}$) et l'amplitude (signée) associée est représentée en ordonnées. On utilise ensuite un algorithme de classification pour identifier les indices avec une amplitude faible (en rouge) et ceux avec une amplitude élevée (en vert). L'algorithme par k-moyennes a été utilisé dans les deux premiers signaux et l'algorithme d'*expectation-maximization* gaussien dans le troisième.

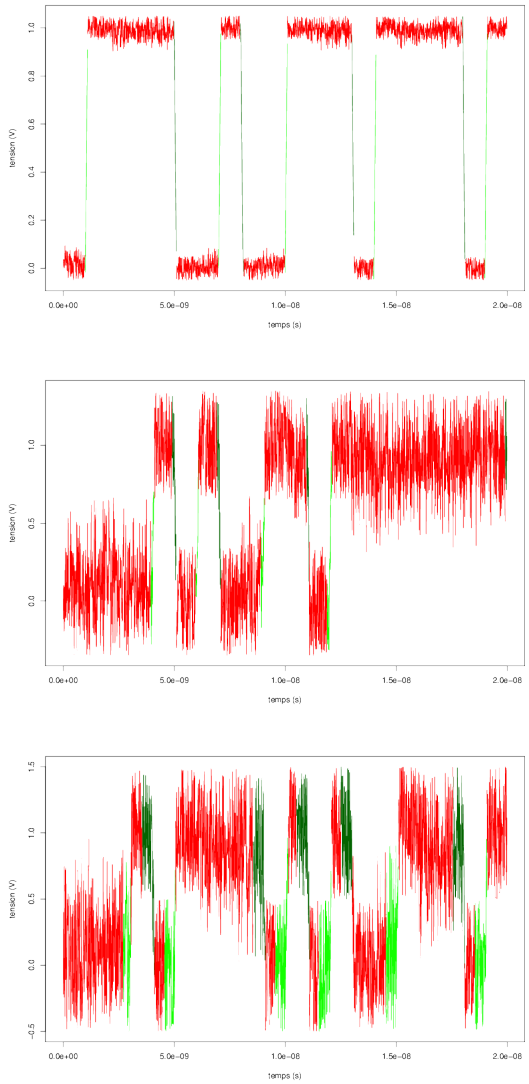


FIGURE 10 – Chaque graphique correspond à un signal bruité, de moyennement (en haut) à très bruité (en bas). Pour chaque graphique, on a identifié les ensembles I^+ (en vert clair), I^- (en vert foncé), I^0 et I^1 (en rouge), avec la méthode de la section 2.1.2. Ces graphiques se déduisent de la classification effectuée sur la figure 9.

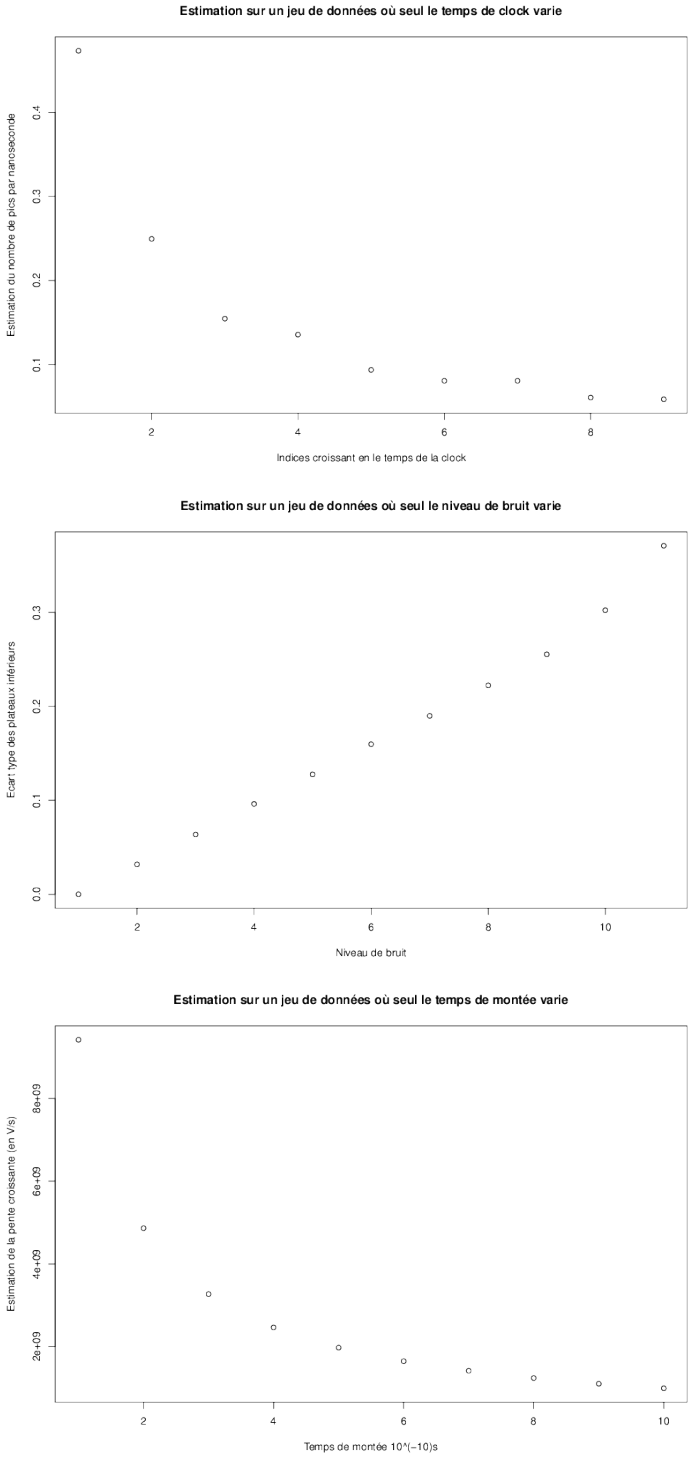


FIGURE 11 – Test de trois indicateurs sur trois jeux de données différents. A noter que seul le deuxième jeu de données contient du bruit.

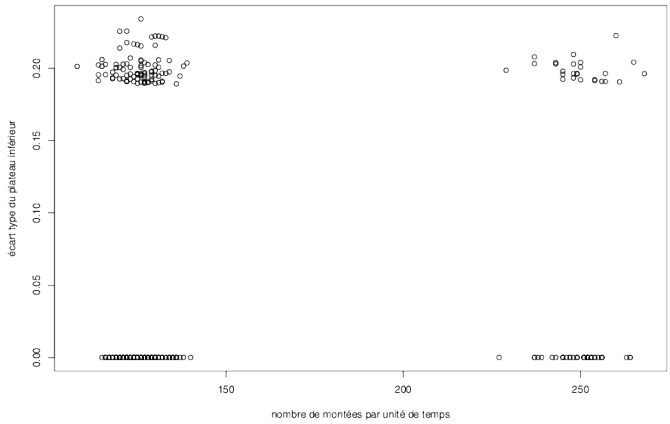


FIGURE 12 – Ecart type estimé du plateau inférieur en fonction du nombre de montées estimés par unité de temps sur le jeu de donnée décrit dans la section 2.2.3.

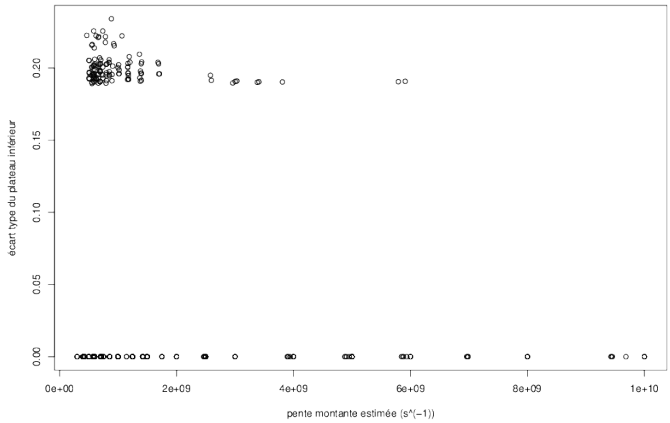


FIGURE 13 – Ecart type estimé du plateau inférieur en fonction de la pente montante estimée sur le jeu de donnée décrit dans la section 2.2.3.

3 Recherche d'indicateurs par modélisation stochastique

Dans cette section nous souhaitons également calculer les indicateurs de la section 2 en proposant une approche mettant en jeu la modélisation stochastique des courbes.

3.1 Notations et modélisation

Notons T_i, $i = 1, \ldots, n$ ($T_i < T_{i+1}$) les pas de temps auxquels sont mesurés des tensions $X(t_i) = X_i$, $i = 1, \ldots, n$. Les lettres majuscules sont là pour rappeler le caractère aléatoire d'un signal. Nous notons les accroissements

$$\Delta_i = (X_{i+1} - X_i)/(T_{i+1} - T_i),$$

$i = 1, \ldots, n-1$ et les accroissements normalisés $\tilde{\Delta}_i = \Delta_i / \max_{j=1,\ldots,n-1}(\Delta_j)$, $i = 1, \ldots, n-1$. A un pas de temps donné T_i, le signal X_i est dans une des trois phases suivantes

— $\tilde{\Delta}_i < 0$: phase descendante, notée $(-)$,
— $\tilde{\Delta}_i = 0$: phase de plateau, notée (0),
— $\tilde{\Delta}_i > 0$: phase ascendante, notée $(+)$.

Notons $Y_i^{(-)} = \mathbf{1}(\tilde{\Delta}_i < 0)$, $Y_i^{(0)} = \mathbf{1}(\tilde{\Delta}_i = 0)$, $Y_i^{(+)} = \mathbf{1}(\tilde{\Delta}_i > 0)$ les variables aléatoires de Bernoulli de paramètres $\pi^{(-)}$, $\pi^{(0)}$, $\pi^{(+)}$ respectivement. Supposons que, étant donné une réalisation $Y_i^{(-)} = y_i^{(-)}$, $Y_i^{(0)} = y_i^{(0)}$, $Y_i^{(+)} = y_i^{(+)}$, l'accroissement $\tilde{\Delta}_i$ est la somme d'un signal (déterministe) et d'un bruit de la forme suivante :

$$\tilde{\Delta}_i = \mu^{(-)} y_i^{(-)} + \mu^{(0)} y_i^{(0)} + \mu^{(+)} y_i^{(+)} + \varepsilon_i, \tag{1}$$

où les ε_i sont des variables aléatoires indépendantes et identiquement distribuées de loi normale $N\left(0, \sigma^{(-)2} y_i^{(-)} + \sigma^{(0)2} y_i^{(0)} + \sigma^{(+)2} y_i^{(+)}\right)$. Par exemple, si $y_i^{(-)} = 0$, $y_i^{(0)} = 0$, $y_i^{(+)} = 1$, on a $\tilde{\Delta}_i = \mu^{(+)} + \varepsilon_i$ avec $\varepsilon_i \sim N(0, \sigma^{(+)2})$. Remarquons qu'une hypothèse sous-jacente à la modélisation (1) est que les accroissements moyens $\mathbf{E}(\tilde{\Delta}_i | Y_i^{(\star)} = 1) = \mu^{(\star)}$, $\star = -, 0, +$ ne dépendent pas des indices i des pas de temps. Les paramètres $\sigma^{(\star)}$, $\star = -, 0, +$ peuvent s'interpréter comme une mesure du bruit présent dans les données.

L'estimation des paramètres $\mu^{(\star)}$, $\sigma^{(\star)}$ et $\pi^{(\star)}$, $\star = -, 0, +$ s'effectue par maximum de vraisemblance via un algorithme de type EM [2]. Elle est implémentée par exemple dans le paquet `mclust` du logiciel R [3].

Un avantage de la modélisation (1) est que des indicateurs d'accroissements et de mesure de bruits sont fournis immédiatement et automatiquement. De plus, pour un pas de temps donné T_i, il est possible de connaître automatiquement l'état de la phase du signal qui est donné par

$$\underset{\star=-,0,+}{\operatorname{argmax}} \quad P(Y_i^{(\star)} = 1 | \tilde{\Delta}_i = \tilde{\delta}_i)$$

où $\tilde{\delta}_i$ est la valeur de l'accroissement observée au pas de temps i. L'estimation des états des phases permet de détecter automatiquement les instants de montée et de descente.

3.2 Résultats

Nous avons modélisé cinq courbes à l'aide de (1) et avons estimé les paramètres des modèles qui sont reportés dans le tableau 1. Un des courbes a donné lieu à un avertissement lors de la lecture de nos scripts c'est pourquoi nous reportons les résultats pour quatres courbes seulement.

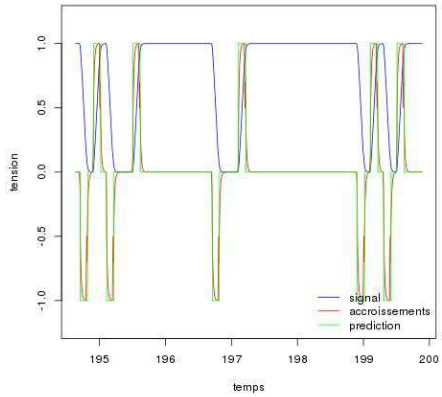
courbe	descendante	nulle	ascendante	bruit
1	-0.87	0.00	0.87	0.11
2	-0.57	0.00	0.58	0.08
3	-0.15	0.00	0.14	0.29
4	-0.48	0.00	0.48	0.07

Table 1 – Estimation des paramètres. La colonne "descendante" correspond à $\mu^{(-)}$, "nulle" correspond à $\mu^{(0)}$, "ascendante" correspond à $\mu^{(+)}$ et "bruit" correspond à la somme des écart-types $\sigma^{(-)} + \sigma^{(0)} + \sigma^{(+)}$.

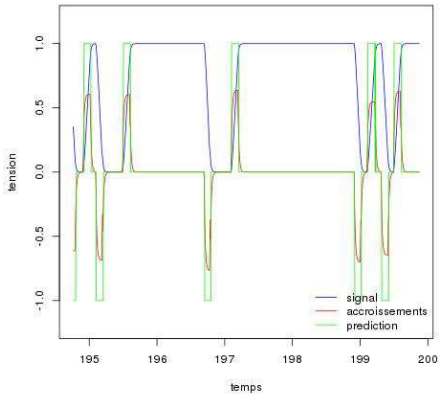
La figure 14 représente les signaux bruts, les accroissements normalisés et la détection des états des phases des milles derniers pas de temps pour les courbes analysées. La figure 15 montre les histogrammes des accroissements normalisés avec leurs densités de probabilité estimées. Le tableau 1 montre que les accroissements moyens estimés pour les phases de plateau sont bien égaux à zéro. On peut voir également que les moyennes sont égales en valeur absolue pour les phases ascendantes et descendantes. La colonne "bruit" nous apprend que la courbe 3 est plus bruitée que les autres, ce qui est confirmé visuellement (voir par exemple la figure 14c). D'après les figures 14a, 14b et 14d, nous pouvons voir que la détection des états des signaux a parfaitement réussi pour les courbes 1, 2 et 4. Les figures 15a, 15b et 15d confirment cette bonne impression où les trois populations du modèle de mélange sont bien différenciées. La courbe 3 est, nous l'avons vu, plus bruitée que les autres. L'histogramme correspondant 15c montre que les trois populations ne se différencient pas aussi bien que pour les courbes 1, 2, et 4. Cependant la figure 14c montre que la détection des états des phases du signal est globalement satisfaisante, bien que certains bruits ont été détectés comme du signal.

3.3 Discussion

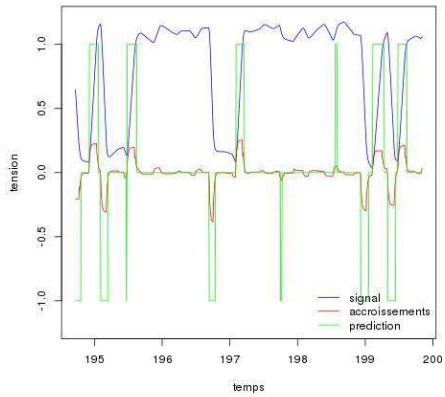
L'avantage de l'approche proposée dans cette section est qu'elle permet un traitement automatique, simple et rapide des signaux. Une fois le modèle établi, les indicateurs sont calculés conjointement avec l'estimation des paramètres. Pour bénéficier de ces avantages, il faut accepter l'hypothèse d'homogénéité qui est faite (les accroissements moyens $\mu^{(\star)}$ ne dépendent pas des indices des pas temps) sur les courbes. Les résultats montrés ci-dessus sont toutefois encourageants.



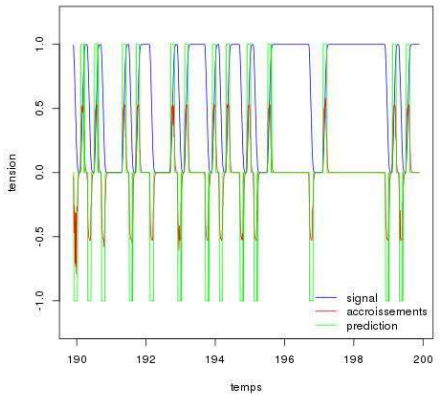
(a) Courbe 1



(b) Courbe 2

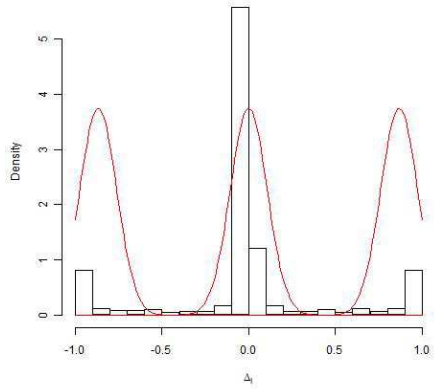


(c) Courbe 3

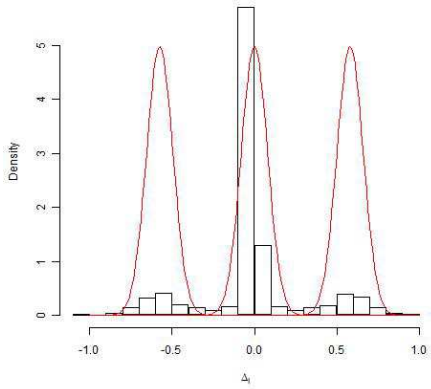


(d) Courbe 4

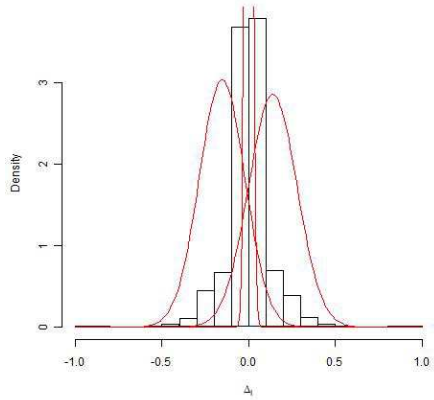
FIGURE 14 – Courbes et leur modélisation. Le signal est représenté par la courbe bleue et les accroissements normalisés par la courbe rouge. La courbe verte représente l'estimation des états des phases du signal. Les unités de temps sont exprimées en nanosecondes.



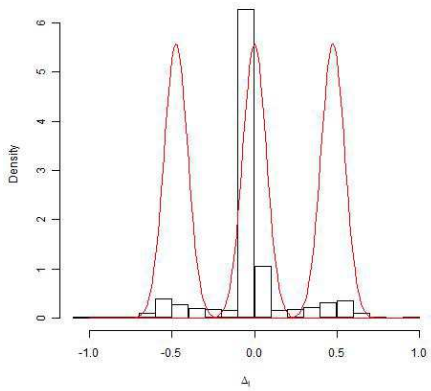
(a) Courbe 1



(b) Courbe 2



(c) Courbe 3



(d) Courbe 4

FIGURE 15 – Histogramme des accroissements normalisés $\tilde{\Delta}_i$ ainsi que leur densité estimée.

4 Détection de glitchs

On se place à présent dans le cadre de l'analyse de l'ensemble des signaux résultats de simulation fournis par des partenaires industriels (Freescale). Ces signaux sont tous de type créneaux et très peu bruités, comme le signal présenté sur la figure 16.

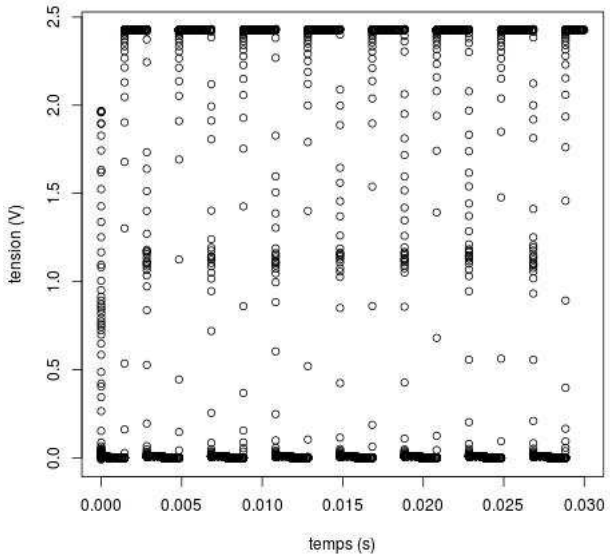


FIGURE 16 – Type de signal fourni par Freescale.

Le circuit se comporte fonctionnellement comme attendu, mais les résultats produits peuvent être accompagnés d'erreurs qui empêcheront leur bonne utilisation : apparition de fréquences parasites qui empêchent la bonne évaluation de la fréquence. Notre approche est l'étude des résultats de la simulation de ces circuits pour déterminer la fréquence d'horloge. Il s'agit ici de mettre en place des méthodes automatiques pour la détection de pics de fréquences qui perturbent le calcul de la fréquence du signal. Nous appellerons dans la suite ces pics des glitchs. Les résultats seront validés sur des jeux de données fournis par les partenaires industriels. Il faudra ensuite réfléchir à une extension de notre approche pour des données plus générales, ayant plusieurs fréquences ou des fronts ascendants ou descendants non verticaux.

Les signaux auxquels on s'intéresse sont de type créneaux, similaires à celui présenté sur la figure 16. Les glitchs perturbant la fréquence que nous devons trouver sont quant à eux de la forme représentée sur la figure 17. Comme nous pouvons le voir sur cette figure, les glitchs sont très proches d'un front et ne sont généralement pas visibles par simple observation du signal.

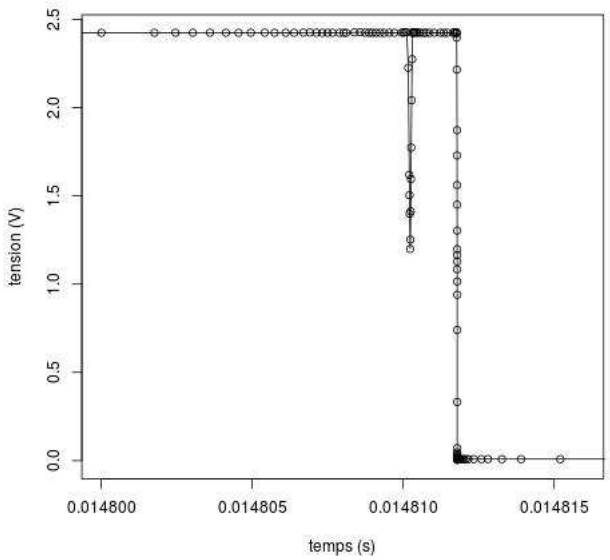


FIGURE 17 – Exemple de glitch.

Notons T_{initial} et T_{final} respectivement le temps initial et final de la simulation. Soit $\mathcal{T}$ l'ensemble des temps pour lesquels le signal est défini. Afin de ne pas traiter les données pendant le temps de chauffe du circuit, nous ne prenons en compte que les données après 8 ms de simulation, et définissons le sous-ensemble $\mathcal{T}' = \{t \in \mathcal{T}\ ;\ t > 8 \text{ ms}\} = \{t_i\}_{i\in\{1,\ldots,N\}}$. Nous proposons de créer un ensemble $\mathcal{S}_{\text{front}}$ contenant les temps pour lesquels le signal présente un front. L'ensemble $\mathcal{S}_{\text{front}}$ est défini par $\mathcal{S}_{\text{front}} = \{t \in \mathcal{T}'\ ;\ \alpha_- < signal(t) < \alpha_+\}$, où les seuils α_+ et α_- sont donnés, pour une tolérance tol_α fixée, par $\alpha_+ = \max_{x\in\mathcal{T}} signal(x) - tol_\alpha$ et $\alpha_- = \min_{x\in\mathcal{T}} signal(x) + tol_\alpha$. L'ensemble $\mathcal{S}_{\text{front}}$ est représenté par les points rouges sur la figure 18. Nous ne considérons donc que les points sur les fronts ascendant ou descendant.

Remarque 1 *L'échantillonnage étant non uniforme, un problème que l'on peut rencontrer est une absence de points sur un front. Pour y remédier, il suffit de sur-échantillonner le signal avant de créer $\mathcal{S}_{\text{front}}$. On construirait par exemple l'ensemble $\mathcal{T}' = \{t_i\}_{i\in\{1,\ldots,2N\}}$ par ajout de points milieux.*

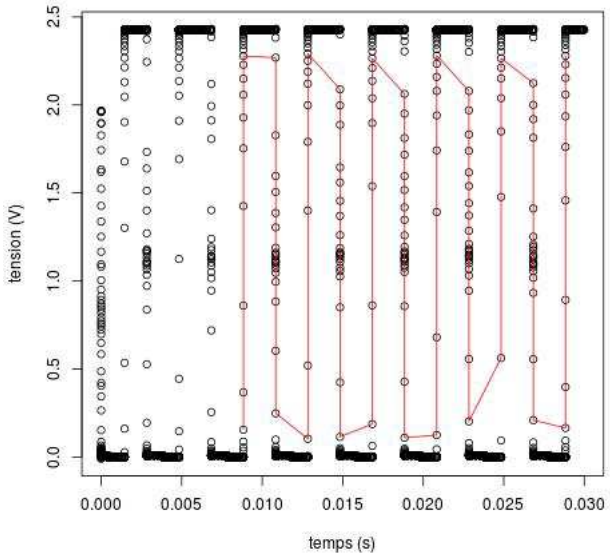


FIGURE 18 – Représentation de l'ensemble $\mathcal{S}_{\text{front}}$ en rouge.

Remarque 2 *Une autre méthode possible serait d'enlever les paliers hauts ou bas seulement quand on sait que les glitchs se situent sur les paliers hauts ou bas. Par exemple en ne levant que les paliers hauts, l'ensemble égal à* $\mathcal{S}_{\text{front}}$ *est* $\mathcal{S}_{\text{front}} = \{t \in \mathcal{T}' \ ; \ signal(t) < \alpha_+\}$. *Ainsi quand on ne considère que le premier point de chaque front, on n'a beaucoup moins de points à considérer.*

Nous allons maintenant introduire l'ensemble $\hat{\mathcal{S}}_{\text{front}}$, qui est l'ensemble $\mathcal{S}_{\text{front}}$ auquel nous avons enlevé les doublons afin de n'avoir qu'une seule valeur du temps par front. Ce nouvel ensemble est défini par

$$\hat{\mathcal{S}}_{\text{front}} = \{t_i \in S_{fronts} \ ; \ t_i \neq t_j \quad \forall i, j = 1, ..., N_{fronts}\},$$

Nous pouvons dès à présent calculer plusieurs périodes du signal de la manière suivante : $\mathcal{P} = \{t_i - t_{i-2} \ ; \ \ j = 2, \ldots, N_{\text{front}} \ \ \forall t \in \hat{\mathcal{S}}_{\text{front}}\}$. A partir de l'ensemble des périodes du signal calculées $\mathcal{P}$, une période moyenne Mp_s peut être calculée. Enfin, un test sur cette moyenne nous permet de dire si le signal possède ou non un glitch. En effet, si une des périodes de l'ensemble $\mathcal{P}$ est trop éloignée de Mp_s, pour une tolérance $tol_{\mathcal{P}}$ fixée, alors il y a au moins un glitch dans le signal.

Algorithme 4.1 ($presence_glitch = DetectionGlitch(tol_\alpha, tol_{\mathcal{P}})$)

$\alpha_+ \leftarrow \max_{x \in \mathcal{T}} signal(x) - tol_\alpha$

$\alpha_- \leftarrow \min_{x \in \mathcal{T}} signal(x) + tol_\alpha$

$\mathcal{T}' \leftarrow \{t \in \mathcal{T} \ ; \ t > 8\text{ms}\}$

$\mathcal{S}_{\text{front}} \leftarrow \{t \in \mathcal{T}' \ ; \ \alpha_- < signal(t) < \alpha_+\}$

$\hat{\mathcal{S}}_{\text{front}} \leftarrow \{t \in \mathcal{S}_{\text{front}} \ ; \ t_i \neq t_j \quad \forall i, j = 1, ..., N_{\text{front}}\}$

$\mathcal{P} = \{t_i - t_{i-2}\ ;\ \ j = 2, \ldots, N_{\text{front}}\ \ \forall t \in \hat{\mathcal{S}}_{\text{front}}\}$
$Mp_s \leftarrow mean(\mathcal{P})$
$presence_glitch \leftarrow false$
Pour tout $p \in \mathcal{P}$
  Si $|Mp_s - p| > tol_{\mathcal{P}}$
    $presence_glitch \leftarrow true$
  Fin si
Fin pour

Ensuite, il faut repérer les glitchs. Pour cela, on a gardé les indices des temps et on peut donc pour une période p trop éloignée de la moyenne retrouver les indices des temps qui ont servi à la calculer. Le premier de ces temps est donc celui où se trouve le glitch.

4.1 Résultats sur un jeu de données industriels

Regardons les cas des jeux de données fournies par le partenaire industriel Freescale pour identifier les glitchs.

Dans le jeu de données, nous avons 474 signaux à analyser. Nous obtenons après utilisation de notre algorithme :

— Nombre de périodes calculées : 2050.

— Période moyenne du signal sans enlever les glitchs : 3.699855 ms.

L'histogramme les périodes calculées est donné par la figure 19.

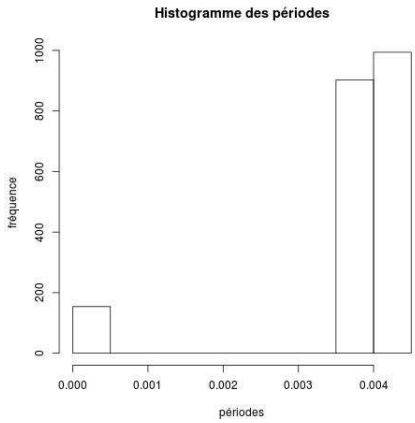


FIGURE 19 – Histogramme des périodes calculées pour le jeu de données 1.

On peut observer que certaines périodes sont très en dessous de la moyenne. On en déduit donc qu'il y a des glitchs dans le signal où se trouve ces périodes et on obtient :

— Nombres glitchs repérés : 154.

— Période moyenne du signal en enlevant les glitchs : 4.000285 ms.

— Distance moyenne entre les glitchs et la pente suivante du signal : 1.048035 μs.

Nous avons donc un algorithme capable de repérer les glitchs et de les contourner pour calculer la période d'un signal. Nous voulons à présent regarder dans quelle mesure nous pouvons utiliser notre méthode dans des cas plus complexes.

4.2 Limites de la méthode proposée

Regardons à présent ce que donne l'algorithme 4.1 dans le cas d'un signal ayant plusieurs périodes, comme illustré par la figure 20.

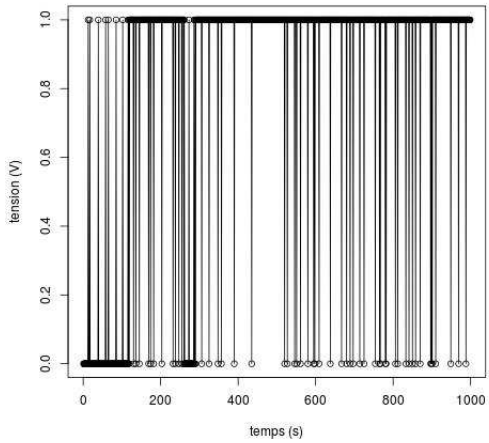


FIGURE 20 – Signal multipériodes.

En appliquant l'algorithme 4.1 sans modification, nous obtenons l'histogramme des périodes détectées (figure 21). Cet histogramme n'est clairement pas satisfaisant.

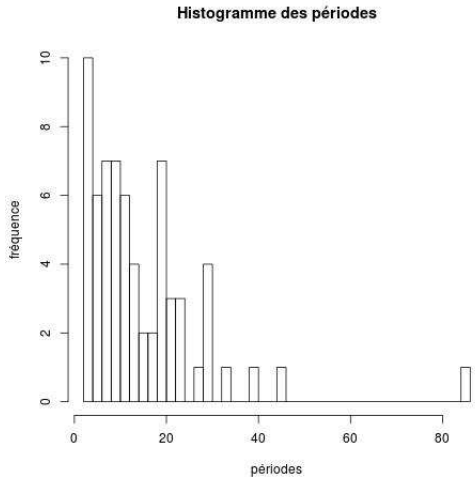


FIGURE 21 – Histogramme des périodes détectées.

Nous proposons donc ici une piste pour étendre l'algorithme 4.1 la détection de glitchs dans le cas de signaux ayant plusieurs périodes. Dans un premier temps, il s'agit de déterminer le nombre de périodes du signal. Cette étape est cruciale pour la suite. En effet, l'idée principale est de générer non pas un seul ensemble $\mathcal{P}$, mais n_p ensembles ($\mathcal{P}_i$, $1 \leq i \leq n_p$) où n_p est le nombre de périodes détectées. Il s'agit ensuite de calculer, pour chaque ensemble $\mathcal{P}_i$ la période moyenne et d'appliquer le test proposé par l'algorithme. Nous pouvons réécrire l'algorithme 4.1 afin de l'adapter au cas de signaux ayant plusieurs périodes :

Algorithme 4.2 ($presence_glitch = DetectionGlitch(tol_\alpha, tol_\mathcal{P})$)

$n_p \leftarrow$ *nombre de périodes du signal*

$\alpha_+ \leftarrow \max_{x \in \mathcal{T}} signal(x) - tol_\alpha$

$\alpha_- \leftarrow \min_{x \in \mathcal{T}} signal(x) + tol_\alpha$

$\mathcal{T}' \leftarrow \{t \in \mathcal{T} \ ; \ t > 8 \text{ ms}\}$

$\mathcal{S}_{\text{front}} \leftarrow \{t \in \mathcal{T}' \ ; \ \alpha_- < signal(t) < \alpha_+\}$

$\hat{\mathcal{S}}_{\text{front}} \leftarrow \{t \in S_{fronts} \ ; \ t_i \neq t_j \quad \forall i, j = 1, ..., N_{fronts}\}$

$\mathcal{P} = \{t_i - t_{i-2} \ ; \quad j = 2, \ldots, N_{\text{front}} \quad \forall t \in \hat{\mathcal{S}}_{\text{front}}\}$

$i \leftarrow 1$

Pour $i \leq n_p$

  $Mp_s^i \leftarrow mean(\mathcal{P}_i)$

Fin pour

$presence_glitch \leftarrow false$

$i \leftarrow 1$

Pour $i \leq n_p$

  Pour tout $p \in \mathcal{P}_i$

Si $|Mp_s^i - p| > tol_{\mathcal{P}_i}$

$presence_glitch \leftarrow true$

Fin si

Fin pour

Fin pour

4.3 Discussion

Cette méthode est donc efficace pour les signaux qui nous ont été fournis. Cependant, dans le cas de signaux très bruités, beaucoup de glitchs seront détectés et il sera difficile de trouver la fréquence car il y aurait beaucoup de fréquences différentes. De plus, cette méthode ne détecte que les pics sur les paliers du signal. Si on doit trouver des glitchs sur les fronts montants ou descendants, on ne pourrait pas utiliser la méthode développée. Enfin, le seuillage est très important pour déterminer ce qui est un glitch de ce qui est du bruit et ne doit pas être pris en compte. La détermination du paramètre de seuillage dépend de manière cruciale de l'application industrielle visée et doit être fixée en fonction du type de défaut que les signaux peuvent présenter sans affecter le fonctionnement du circuit.

5 Conclusion

Nous avons proposé des méthodes d'analyse adaptées à l'étude de l'évolution temporelle des tensions aux noeuds d'un circuit électronique analogique. Résumons ici les principaux points :

1. Proposition d'une méthode générale de traitement de données analogiques.
2. Définition d'indicateurs repérant des caractéristiques ciblées et jugées pertinentes.
3. Application à des données simulées et des données industrielles (recherche de glitchs).

Nous avons cherché à mettre en avant les avantages et les limites des outils développés, et proposé quelques pistes de développements à partir de ce premier travail.

Références

[1] James MacQueen, *Some methods for classification and analysis of multivariate observations*, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, 281–297, Univ. of California Press, 1967.

[2] Arthur P. Dempster, Nan M. Laird, and Donald B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, Journal of the Royal Statistical Society. Series B (Methodological), **39**, 1–38, 1977.

[3] `http://www.r-project.org`.