
Réseaux Bayésiens Dynamiques : Méthodologie pour l'inférence sur les Schémas de Mode d'Action en Toxicologie

Frédéric BOIS¹, Wang GAO^{1,2}, Ghislaine GAYRAUD²

1. INERIS

Unité METO, Parc ALATE BP2, 60550 Verneuil-en-Halatte, France
frederic.bois@ineris.fr, wang.gao-EXTERNE@ineris.fr

2. LMAC, Université de Technologie de Compiègne

57 Avenue de Landshut CS 60203 Compiègne, France
wang.gao@utc.fr, ghislaine.gayraud@utc.fr

RÉSUMÉ. En toxicologie, un schéma de mode d'action (AOP : Adverse Outcome Pathway) est un cadre conceptuel qui décrit qualitativement les connaissances existantes concernant les liens entre deux points d'ancrage : un événement initiateur moléculaire (MIE : Molecular Initiating Event) et un résultat défavorable (AO : Adverse Outcome) à un niveau d'organisation biologique pertinent pour l'évaluation du risque. La version quantitative d'un AOP, le qAOP, promet d'être un outil puissant pour l'évaluation des risques, grâce notamment à sa capacité de prédiction. Cet article présente une méthode de modélisation originale de qAOPs par les réseaux bayésiens dynamiques.

ABSTRACT. In toxicology, an Adverse Outcome Pathway (AOP) is a conceptual framework that qualitatively describes the existing knowledge on the links between the two anchor points: Molecular Initiating Event (MIE) and Adverse Outcome (AO) at a level of biological organisation relevant for risk assessment. The transformation of an AOP to its quantitative version, qAOP allows to build a powerful risk assessment tool, thanks to its ability to quantitatively predict the AO. This paper presents a new method for modelling qAOP using Dynamic Bayesian Networks (DBN).

MOTS-CLÉS: Réseaux Bayésiens Dynamiques, Systèmes Dynamiques Linéaires, Schémas de Mode d'Action en Toxicologie (AOP), Toxicologie, Méthode MCMC, Inférence bayésienne.

KEYWORDS: Dynamic Bayesian Networks, linear dynamical systems, Adverse Outcome Pathway (AOP), Toxicology, MCMC Method, Bayesian inference.

DOI:10.3166/RIA.28.1-11 © 2018 Lavoisier

1. Introduction

En toxicologie, un schéma de mode d'action défavorable (AOP : Adverse Outcome Pathway) est un cadre conceptuel de représentation graphique qualitative qui décrit les connaissances existantes concernant les liens entre deux points d'ancrage: i) un événement initiateur moléculaire (MIE : Molecular Initiating Event) et ii) un résultat défavorable (AO : Adverse Outcome) à un niveau d'organisation biologique, pertinent pour l'évaluation du risque. L'organisation systématique de l'information dans les cadres AOP a le potentiel d'améliorer la prise de décision réglementaire grâce à une plus grande intégration et à une utilisation plus pertinente des données mécanistes. (Leist *et al.*, 2017)

Les travaux récents s'orientent vers la quantification d'AOP (qAOP) rendue possible grâce aux méthodes de modélisation mathématique. Nous proposons dans cet article une méthode originale de modélisation de qAOPs par des réseaux bayésiens dynamiques (DBN) capables d'appréhender les dépendance en temps et entre nœuds des AOPs. Notre modèle DBN (cf système d'équation (2)) généralise les DBN linéaires gaussiens (cf équation (1)) usuellement considérés dans la littérature sous le nom de modèle d'espace-état linéaire-gaussien (linear-gaussian state-space model) (Durbin, Koopman, 2012). Après un bref rappel des concepts de base en section 1, nous comparons et discutons les modèles DBN en section 2. En section 3, nous appliquons notre modèle au cadre AOP. Nous terminons par une conclusion et des perspectives en section 4.

1.1. AOP et ses variantes

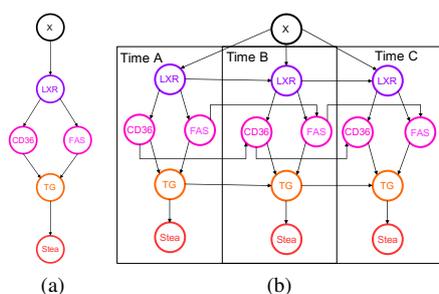


Figure 1. Réseau d'AOPs pour la stéatose hépatique avec un nœud externe X représentant un produit chimique qui induit le MIE, ici LXR (Figure 1a), et la structure du DBN correspondant (Figure 1b)

Du point de vue mathématique, un AOP peut être représenté par une chaîne dirigée. Tous les AOPs partagent une structure commune consistant en un MIE, une série d'événements clés (KE : Key event) reliés par des relations d'événements clés (KER : KE relationship) et un AO. Le MIE, les KEs et l'AO constituent un ensemble d'états

à différents niveaux d'organisation biologique et sont représentés par des nœuds dans un AOP. Les KERs encodent des relations causales (e.g. dose-réponse ou cause-effet) et sont représentées par les arêtes dirigées de la chaîne.

Un réseau d'AOPs est un ensemble d'AOPs qui conduisent au même AO et dont certains nœuds et arêtes sont partagés. La Figure 1a illustre un mini-réseau d'AOPs pour la stéatose hépatique qui contient deux AOPs.

Un tel réseau représente qualitativement des connaissances pathologiques (causales) concernant un AO, mais il ne possède pas de capacité de prédiction quantitative. C'est pourquoi sa variante, qAOP (et les réseaux de qAOPs) est apparue ces dernières années. En particulier, on peut citer (Gust *et al.*, 2015) et (Jaworska, 2016) qui considèrent une modélisation de qAOPs par des réseaux bayésiens (BN) et par conséquent, négligent la dynamique de ces AOs.

En revanche, afin de tenir compte de l'aspect temporel, nos travaux s'appuient sur la famille des modèles DBN pour modéliser un réseau d'AOPs en un réseau de qAOPs à partir de données de types dose-réponse et dose-temps-réponse. La modélisation des qAOPs par DBN présente l'avantage de pouvoir prédire, avec une mesure d'incertitude, le niveau d'AO à partir de données manquantes, i.e., des données récoltées à des temps différents et sur un nombre partiel des nœuds MIE ou KEs.

1.2. Réseaux bayésiens (BN) et réseaux bayésiens dynamiques (DBN)

Un réseau d'AOPs étant un graphe dirigé, il est naturel de modéliser sa version quantitative par un réseau bayésien. De plus, afin de prendre en compte la dynamique des AOPs, nous avons retenu les DBN qui semblent appropriés dans notre contexte ; en effet, 1) pour un état fixé du nœud parent, l'état du nœud enfant peut évoluer dans le temps. 2) les processus biologiques sous-jacents peuvent contenir des rétroactions. Par conséquent, le réseau d'AOPs n'est pas un graphe dirigé acyclique (DAG). Mais, sous l'hypothèse que ces mécanismes de rétroaction ne sont pas instantanés, la structure de DBN reste un DAG avec des arêtes, représentant les rétroactions, qui, en partant de la tranche de temps t , ne pointent que vers des nœuds de la tranche de temps $t + 1$.

1.2.1. Types et dépendances des nœuds

Intrinsèquement à la définition, les AOPs sont constitués de nœuds qui sont soit constants au cours du temps, soit dépendants du temps. C'est exactement ce que peuvent décrire les DBNs. En particulier, Schäfer et Weyrath (Schäfer, Weyrath, 1997 ; Mihajlovic, Petkovic, 2001) ont considéré trois types de nœuds selon leur dépendance au temps et les différentes dépendances possibles entre eux.

- un nœud est dit statique (SN : static node) si sa distribution est constante au cours du temps (e.g., le nœud X dans la Figure 1b).

- un nœud est dit dynamique (DN : dynamic node) si la distribution de son état présent est conditionnée par lui-même au temps passé (e.g., les nœuds LXR, CD36, FAS et TG dans la Figure 1b).

– un nœud est dit temporaire (TN : temporary node) si à un temps fixé, la distribution de son état n'est conditionnée que par celui de ses parents (e.g., le nœud Stea dans la Figure 1b).

Dans les DBNs, toutes les dépendances temporelles entre SN, DN et TN ne sont pas autorisées. En effet, on y distingue trois types de dépendances impliquant au moins un DN: $SN \rightarrow DN$, $DN \rightarrow DN$ et $DN \rightarrow TN$. Nous nous focalisons ici sur les dépendances de types : $SD \rightarrow DN$ et $DN \rightarrow DN$, qui, elles, ne sont pas instantanées.¹

1.2.2. Hypothèses classiques des modèles DBN

Dans cet article, nous considérons deux hypothèses qui correspondent à des hypothèses classiques minimales sur les DBNs, à savoir :

HYPOTHÈSE 1. — *Les observations sont équi-espacées temporellement.*

HYPOTHÈSE 2. — *La propriété markovienne d'ordre 1 est satisfaite :*

$$P(C_t | P_t, C_u, u < t) = P(C_t | P_t, C_{t-1}) \text{ où } P \text{ est le nœud parent du nœud } C.$$

1.3. Problème de données manquantes

1.3.1. Données expérimentales

Les données idéales pour mener à bien l'inférence sur les qAOPs devraient avoir la structure suivante : pour chaque réplication, on dispose d'une série temporelle d'un vecteur complet de mesures d'état du système $(X, \text{MIE}, \text{KE}_1, \text{KE}_2, \dots, \text{KE}_n, \text{AO})_{t; t \geq 1}$

Actuellement, nous travaillons sur la modélisation de deux réseaux d'AOPs : i) la stéatose et ii) la dégénérescence rénale chronique. Pour ces deux AOPs, nous disposons de données incomplètes sur les différents KEs. En effet, les données sont partielles puisqu'elles sont souvent issues de différentes expérimentations bibliographiques (données de toxicologie) réalisées indépendamment les unes des autres et dont les objectifs diffèrent de la modélisation de qAOPs. Par conséquent, les pas de temps de collecte des données peuvent ne pas être les mêmes, ou n'être pas constants. Dans ce cas, une mise en congruence de pas est nécessaire par imputation. Ceci est facilement réalisable dans le cadre de l'inférence bayésienne sur le modèle DBN linéaire généralisé (2) en introduisant les données manquantes comme paramètres supplémentaires à calibrer. Un modèle DBN linéaire gaussien classique (1) aurait plus de difficultés (cf Section 3.3 Discussion).

1. Les dépendances instantanées, e.g., $DN \rightarrow TN$ et $TN \rightarrow TN$, sont aussi autorisées mais leur modélisation revient à la situation des réseaux bayésiens non dynamiques. Elles ne sont donc pas dans le cadre de notre étude.

Exemple illustratif

Voici un exemple pour illustrer le type de données incomplètes que l'on peut avoir dans un contexte réel.

Soit l'AOP très simple :

$$X_1 \rightarrow X_2 \rightarrow X_3$$

On dispose 2 jeux de données incomplètes de forme suivante :

- $\mathcal{D}_1 := \{(X_1, X_2)_{n_{t1} * \Delta t_1, n_{r1}}\}$ pour X_3 non mesurés.
- $\mathcal{D}_2 := \{(X_1, X_3)_{n_{t2} * \Delta t_2, n_{r2}}\}$ pour X_2 non mesurés.

avec

- Δt . délais entre deux observations consécutives : $\Delta t. \in \mathbb{R}^+$
 - n_r . et N . : indice et nombre total de réplifications : $n_r. \in \{1, \dots, N.\}$ avec $N. \in \mathbb{N}^+$
 - n_t . et T . : indice et nombre total d'observations par réplification : $n_t. \in \{1, \dots, T.\}$
- avec $T. \in \mathbb{N}^+$

Si $\Delta t_1 \neq \Delta t_2$, alors l'imputation est nécessaire pour la modélisation de $X_2 \rightarrow X_3$.

2. Réseaux bayésiens dynamiques linéaires

2.1. Réseaux bayésiens dynamiques : Dépendance linéaire classique

La méthode de référence consiste à modéliser les dépendances causales dynamiques de type $SN \rightarrow DN$ et $DN \rightarrow DN$ par un modèle linéaire gaussien (il s'agit d'un modèle d'espace-état linéaire-gaussien à temps discret, (Durbin, Koopman, 2012)).

Notations :

- C_t, P_t : activités respectives du nœud enfant et du nœud parent en temps t ,
- ϵ : bruit de transition,
- $\alpha, \beta_{prev}, \beta_{curr}, \sigma$: paramètres.

Le modèle linéaire gaussien classique est donné par

$$C_{t+1} = \alpha + \beta_{prev} C_t + \beta_{curr} P_{t+1} + \epsilon \quad (1)$$

avec $\epsilon \sim N(0, \sigma^2)$

REMARQUE 3. — Le cas de $\beta_{prev} < 0$ ne nous intéresse pas, car il rend le problème d'estimation dépendant du pas d'observation.

Lorsque $\beta_{prev} = 0$, le nœud enfant devient un nœud temporaire et on se ramène à un modèle BN puisqu'on a alors une relation de type DN ou $SN \rightarrow TN$.

Pour $\beta_{prev} \geq 1$, le système diverge, il est donc instable au cours du temps. Dans notre application cette divergence ne présente pas d'intérêt, ainsi on se focalise par la suite uniquement sur le cas $\beta_{prev} \in]0, 1[$. \square

2.2. Réseaux bayésiens dynamiques : Dépendance linéaire généralisée

Dans cet article, nous proposons une forme généralisée pour la modélisation de ces dépendances conditionnelles qui sont exprimées explicitement en fonction de l'intervalle temporel h .

Notations :

- C_t, P_t : activités respectives du nœud enfant et du nœud parent en temps t ,
- h : intervalle temporel entre deux temps d'observation consécutifs,
- $S_C[P_t]$: état vers lequel évolue C_t lorsque l'état de P_t est connu; c'est une fonction déterministe de P_t ,
- ϵ : bruit de transition,
- $\beta, \beta_0, \nu, \sigma$: paramètres.

Notre modèle DBN linéaire généralisé est donné par le système d'équations (2)

$$S_C[P_{t+h}] - C_{t+h} = (S_C[P_{t+h}] - C_t) e^{-\nu h} + \sqrt{\frac{1 - e^{-\nu h}}{1 - e^{-\nu}}} \epsilon$$

avec $S_C[P_{t+h}] = \beta P_{t+h} + \beta_0$ (2)

$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$

$$\nu > 0, h \in \mathbb{R}^+$$

REMARQUE 4. — Pour $h = 1$, le modèle DBN linéaire généralisé (2) est équivalent (cf annexe A) au modèle linéaire gaussien (1) dont le paramètre $\beta_{prev} \in]0, 1[$. \square

REMARQUE 5. — La linéarité du modèle DBN linéaire généralisé (2) a un double sens :

1. en fixant $h = 1$, son équivalence avec le modèle (1) ayant une forme linéaire ;
2. la dépendance linéaire de S_C par rapport à P_t .

\square

2.3. Avantages du modèle DBN généralisé (2)

Un modèle mieux interprétable

Au regard de notre contexte biologique, face à une perturbation, un système biologique présente des mécanismes complexes de régulation pour se stabiliser, c'est-à-dire retrouver, au fur et à mesure, un nouvel état d'équilibre.

Nous faisons l'hypothèse que l'état d'équilibre, encodé par S_C , dépend instantanément de l'intensité de la perturbation P_t . L'état du système C_t converge vers $S_C[P_t]$ à une vitesse exponentielle contrôlée par le paramètre ν .

Un modèle plus facile à généraliser dans le cas $S_C[P_t]$ non linéaire.

Nous pouvons par exemple, considérer un modèle non linéaire (3) qui contient un paramètre supplémentaire α par rapport au modèle DBN linéaire généralisé (2) :

$$S_C[P_{t+h}] - C_{t+h} = (S_C[P_{t+h}] - C_t) e^{-\nu h} + \sqrt{\frac{1 - e^{-\nu h}}{1 - e^{-\nu}}} \epsilon \quad (3)$$

avec $S_C[P_{t+h}] = \alpha(1 - e^{-\beta P_{t+h}}) + \beta_0$

$$\nu > 0, \alpha > 0, h \in \mathbb{R}^+$$

La flexibilité du modèle (2) facilite l'intégration des connaissances des experts du domaine (les toxicologues dans notre contexte). En effet, les toxicologues experts ont souvent une idée de l'évolution de S_C selon les KEs et le fait que dans le modèle (2) S_C s'exprime directement au travers de P_t rend possible l'intégration de ces connaissances. Cela n'est en revanche pas faisable dans le modèle (1).

Invariance des paramètres du modèle (2) par rapport à l'intervalle h

Contrairement au modèle classique (1), les paramètres du modèle généralisé (2) sont invariants quel que soit h pour les dépendances temporelles de types $SN \rightarrow DN$ (cf Section 3.3).

3. Application à l'inférence paramétrique sur les qAOPs

3.1. Données simulées

Pour valider la robustesse du modèle généralisé (2) face au problème des données manquantes (voir la Section 1.3).

Nous proposons de comparer le comportement de deux modèles au travers d'un jeu de données simulées :

$$\mathcal{D}_1 := \{X_{t,1}, C_{1t,i}, C_{2t,i}\}_{t \in \{1, \dots, 72\}, i \in \{1 \dots 8\}}$$

pour l'AOP suivant :



où

- X représente un produit chimique introduisant un MIE au système,
- $C_{1t,i}$ représente l'activité du MIE pour la répllication $i \in \{1 \dots 8\}$ mesurée au temps $t \in \{0 \dots 72\}$, et où l'unité de t est l'heure et le pas de temps est égal à 1 heure,

– $C_{2t,i}$ représente l'activité de l'AO dont les indices ont la même signification que ceux de $C_{1t,i}$

Les données \mathcal{D}_1 sont simulées à partir du modèle (2) à un pas de temps égal à 1 heure avec les paramètres suivants:

- $\beta = 1.5, \beta_0 = 2.0, \nu = 0.08, \sigma = 0.15$ pour $X \rightarrow C_1$
- $\beta = 0.3, \beta_0 = 1.5, \nu = 0.05, \sigma = 0.07$ pour $C_1 \rightarrow C_2$

Au vu de l'équivalence entre les deux modèles (voir Annexe A), le jeu de données \mathcal{D}_1 correspond à la réalisation de données issues du modèle DBN linéaire classique (1) en utilisant les paramètres suivants :

- $\alpha = 0.1538, \beta_{prev} = 0.9231, \beta_{curr} = 0.1153, \sigma = 0.15$ pour $X \rightarrow C_1$
- $\alpha = 0.0732, \beta_{prev} = 0.9512, \beta_{curr} = 0.0146, \sigma = 0.07$ pour $C_1 \rightarrow C_2$

Ensuite, on construit deux jeux de données \mathcal{D}_2 et \mathcal{D}_3 , tels que :

- $\mathcal{D}_2 \subset \mathcal{D}_1, \mathcal{D}_2 = \{X_{t,i}, C_{1t,i}, C_{2t,i}\}_{i \in \{1 \dots 8\}, t=3n; n \in \{1, \dots, 24\}}$
- $\mathcal{D}_3 \subset \mathcal{D}_2 \subset \mathcal{D}_1, \mathcal{D}_3 = \{X_{t,i}, C_{1t,i}, C_{2t,i}\}_{i \in \{1 \dots 8\}, t=6n; n \in \{1, \dots, 12\}}$

3.2. Inférence bayésienne et implémentation

L'inférence des paramètres est effectuée via l'approche bayésienne par des méthodes MCMC avec

- pour le modèle DBN classique (1), des distributions a priori non informatives (loi uniforme impropre) sur tous les paramètres : $\alpha, \beta_{prev}, \beta_{curr}$ et σ .
- pour le modèle DBN généralisé (2), des distributions a priori non informatives (loi uniforme impropre) sur les paramètres β, β_0, σ et une distribution a priori gaussienne faiblement informative ($N(0, 10^2)$) sur ν .

Le programme d'inférence est codé en langage mc-Stan et manipulé à travers l'interface R sur la plateforme MacOS. La machine de calcul possède un CPU de modèle i5 double cœur à 3.1 GHz et une mémoire de 8 GB à 2133 MHz.

3.3. Résultats

Avec 4000 itérations, dont 2000 préliminaires, les estimations des paramètres par la moyenne *a posteriori* des deux modèles pour les deux types de dépendances : $SN \rightarrow DN$ et $DN \rightarrow DN$ sont fournies dans les tableaux 1 et 2.

Tableau 1. Paramètres caractérisant la dépendance entre X et C_1

TrainSet	Δt	Modèle	$\hat{\alpha}$	$\hat{\beta}_{prev}$	$\hat{\beta}_{curr}$	$\hat{\sigma}$	RunTime (s)
\mathcal{D}_1	1	DBNLC	0.1501	0.9239	0.1142	0.1455	42
\mathcal{D}_2	3		0.4263	0.7892	0.3169	0.2324	10.5
\mathcal{D}_3	6		0.7452	0.6211	0.5702	0.3059	5
Vrai	1	DBNLC	0.1538	0.9231	0.1153	0.15	
TrainSet	Δt	Modèle	$\hat{\beta}$	$\hat{\beta}_0$	$\hat{\nu}$	$\hat{\sigma}$	RunTime (s)
\mathcal{D}_1	1	DBNLG	1.504	1.974	0.0789	0.1454	41
\mathcal{D}_2	3		1.504	1.975	0.0788	0.1408	12.5
\mathcal{D}_3	6		1.505	1.967	0.0792	0.1406	6.5
Vrai	1	DBNLG	1.5	2	0.08	0.15	

Tableau 2. Paramètres caractérisant la dépendance entre C_1 et C_2

TrainSet	Δt	Modèle	$\hat{\alpha}$	$\hat{\beta}_{prev}$	$\hat{\beta}_{curr}$	$\hat{\sigma}$	RunTime (s)
\mathcal{D}_1	1	DBNLC	0.0715	0.9511	0.0151	0.0716	52
\mathcal{D}_2	3		0.1926	0.8661	0.0421	0.1164	16
\mathcal{D}_3	6		0.3217	0.7728	0.0732	0.1496	9
Vrai	1	DBNLC	0.0732	0.9512	0.0146	0.07	
TrainSet	Δt	Modèle	$\hat{\beta}$	$\hat{\beta}_0$	$\hat{\nu}$	$\hat{\sigma}$	RunTime (s)
\mathcal{D}_1	1	DBNLG	0.3108	1.4643	0.0496	0.0716	75
\mathcal{D}_2	3		0.3155	1.4426	0.0472	0.0697	19
\mathcal{D}_3	6		0.3242	1.4161	0.0424	0.0667	9.25
Vrai	1	DBNLG	0.3	1.5	0.05	0.07	

Discussions

On peut voir que, les paramètres du modèle DBN linéaire classique (1) (DBNLC) varient en fonction de taille de pas de temps des \mathcal{D}_i . Cela suggère que les paramètres estimés à partir d'un jeu de données à un pas de temps 3 heures ne peuvent pas être utilisés directement pour l'imputation des données à un pas de temps 1 heure.

À l'inverse, les estimations des paramètres du modèle DBN linéaire généralisé (2) (DBNLG) sont plus robustes à ce changement de pas de temps.

Il faut remarquer que le pas de temps est néanmoins fixé au sein du même jeu de données simulées. Ceci peut ne pas être le cas dans la réalité, si les observations de la même expérience biologique sont effectuées à des pas de temps non réguliers (e.g. des délais entre les observations qui alternent entre 8 heures et 12 heures.)

4. Conclusions et perspectives

Nous avons présenté dans cet article un modèle linéaire généralisé (2) qui correspond à une extension du modèle linéaire gaussien classique. En particulier, dans le contexte de $SN \rightarrow DN$, on peut s'affranchir de l'hypothèse d'observations équi-espacées temporellement, grâce à l'incorporation explicite de l'intervalle de temps dans le modèle qui dissocie les paramètres du facteur temporel. Néanmoins, dans le cas $DN \rightarrow DN$, cette hypothèse doit être satisfaite mais elle est moins cruciale que pour le modèle linéaire gaussien. Le modèle généralisé (2) est également plus flexible au sens où il peut être facilement étendu à des familles de modèles non linéaires (cf. modèle (3)).

Les propriétés du modèle généralisé sont importantes pour son application aux qAOPs car i) les données expérimentales peuvent être récoltées à des pas de temps non réguliers et ii) même s'ils le sont, la taille du pas peut varier d'une source de données à une autre.

5. Remerciements

F. Bois est financé par le projet EU-ToxRisk du programme Horizon 2020 de la Commission Européenne (Contrat No. 681002) et par le PNREST Anses, 2014/1/093

Bibliographie

- Durbin J., Koopman S. J. (2012). *Time series analysis by state space methods* (2^e éd.). Oxford University Press.
- Gust K. A., Collier Z. A., Mayo M. L., Stanley J. K., Gong P., Chappell M. A. (2015). Limitations of toxicity characterization in life cycle assessment: Can adverse outcome pathways provide a new foundation? *Integrated Environmental Assessment and Management*, vol. 12, n° 3, p. 580-590.
- Jaworska J. (2016). Integrated testing strategies for skin sensitization hazard and potency assessment—state of the art and challenges. *Cosmetics*, vol. 3, n° 2.
- Leist M., Ghallab A., Graepel R., Marchan R., Hassan R., Bennekou S. H. *et al.* (2017, 01 Nov). Adverse outcome pathways: opportunities, limitations and open questions. *Archives of Toxicology*, vol. 91, n° 11, p. 3477–3505.
- Mihajlovic V., Petkovic M. (2001). *Dynamic bayesian networks: A state of the art* (vol. TR-CTIT-34).
- Schäfer R., Weyrath T. (1997). Assessing temporally variable user properties with dynamic bayesian networks. In A. Jameson, C. Paris, C. Tasso (Eds.), *User modeling*, p. 377–388. Vienna, Springer Vienna.

Annexe A. Équivalence entre le modèle classique (1) et le modèle généralisé (2)

Pour $\beta_{curr} \in]0, 1[$, on a

$$\begin{aligned}
 C_{t+1} &= \alpha + \beta_{prev} C_t + \beta_{curr} P_{t+1} + \epsilon_t \\
 \Leftrightarrow C_{t+1} + \frac{\alpha + \beta_{curr} P_{t+1}}{\beta_{prev} - 1} &= \beta_{prev} \left(C_t + \frac{\alpha + \beta_{curr} P_{t+1}}{\beta_{prev} - 1} \right) + \epsilon_t \\
 \Leftrightarrow \frac{\alpha + \beta_{curr} P_{t+1}}{1 - \beta_{prev}} - C_{t+1} &= \beta_{prev} \left(\frac{\alpha + \beta_{curr} P_{t+1}}{1 - \beta_{prev}} - C_t \right) + \epsilon'_t
 \end{aligned}$$

avec $\epsilon_t \sim N(0, \sigma^2)$ et $\epsilon'_t = -\epsilon_t$.

En posant :

- $S[P_t] := \beta P_t + \beta_0$ avec $\beta = \frac{\beta_{curr}}{1 - \beta_{prev}}$ et $\beta_0 = \frac{\alpha}{1 - \beta_{prev}}$,
- $h = 1$,
- $e^{-\nu h} = \beta_{prev}$ avec $\nu > 0$ qui correspond à $\beta_{prev} \in]0, 1[$,

on obtient :

$$\begin{aligned}
 S[P_{t+1}] - C_{t+1} &= (S[P_{t+1}] - C_t) e^{-\nu h} + \epsilon'_t \\
 C_{t+1} &= S[P_{t+1}] - (S[P_{t+1}] - C_t) e^{-\nu h} + \epsilon_t
 \end{aligned}$$