

Modèles et modélisation :

Généralités

Définition

Les objectifs de la modélisation

Description des données

Prédiction – simulation

Comprendre le système secondaire (valeur heuristique d'un modèle)

La modélisation

1. Modèles et modélisation : définition

Dans le cadre de la modélisation, le système étudié est appelé **système primaire**. Un modèle est une représentation de la réalité et le modèle est appelé **système secondaire**. Par définition un modèle est une simplification de la réalité car s'il en avait tous les attributs, il s'agirait non plus d'un modèle mais de la réalité.

2 - Les objectifs de la modélisation

Pourquoi modéliser ?

Cette question doit avoir reçu une réponse, non seulement pour justifier la démarche, mais aussi pour sélectionner l'approche modélisatrice la plus appropriée. Schématiquement, la modélisation peut avoir trois objectifs :

- (i) **décrire** (résumer) les données
- (ii) **prédire** (simuler), c'est à dire faire des inférences
- (iii) **expliquer** (comprendre) c'est à dire avoir des objectifs mécanistiques.

Selon que l'on souhaite parvenir à l'un ou l'autre de ces objectifs, la modélisation sera réalisée selon des modalités différentes.

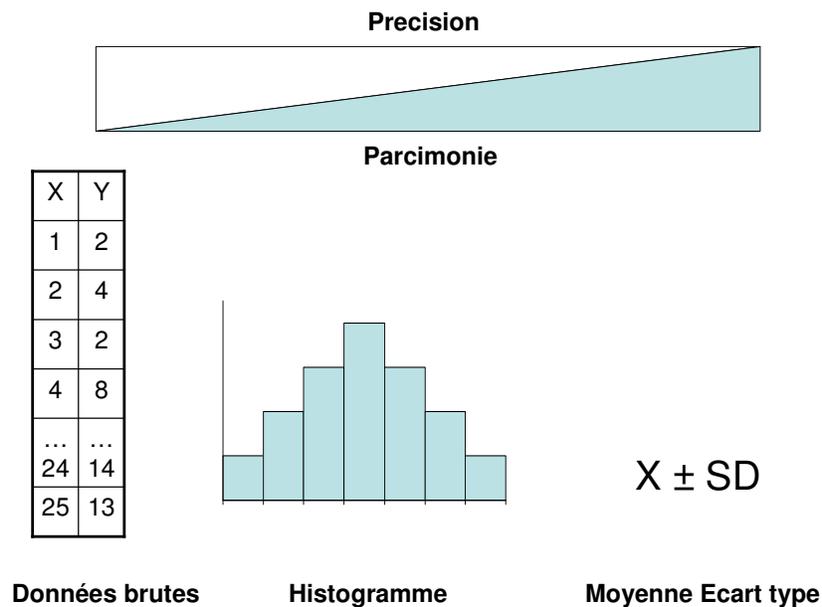
3 - Description des données

Cet objectif consiste à présenter les données en réalisant un compromis entre la précision de l'information ainsi résumée et la nécessité de le faire avec un maximum d'économie (**principe de parcimonie**).

Prenons pour premier exemple une série de 100 concentrations en fonction du temps ; pour présenter ces résultats, il y a trois présentations possibles : (fig. 1)

- (i) donner un tableau avec toutes les valeurs
- (ii) réaliser un graphique (histogramme)
- (iii) calculer la moyenne et l'écart type

Figure 1:



La parcimonie augmente de la présentation 1 à la présentation 3 alors que l'inverse est vraie pour la précision.

- Le tableau contient toute l'information
- Le graphique est un **codage** de l'information. Dans le cas d'un histogramme, une partie de l'information est perdue.
- Le calcul de la moyenne et de l'écart type implique le recours implicite ou explicite à un modèle. Ce dernier postule la normalité de la distribution

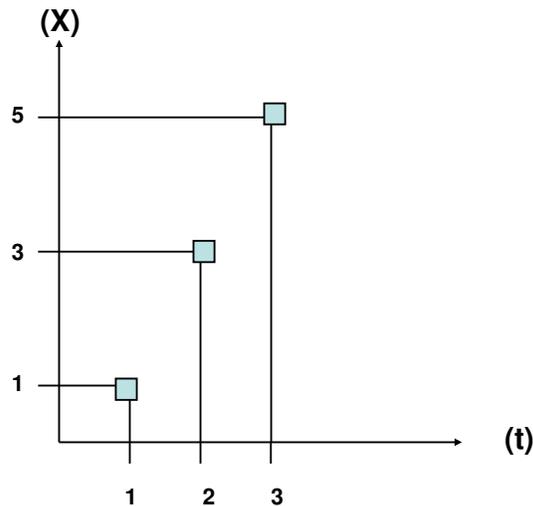
Prenons un deuxième exemple. Le système primaire est une quantité X mesurée au cours du temps (t). Les données sont présentées dans le tableau suivant (tableau 1).

Tableau 1 : jeux de données illustrant les différentes modalités de présentation des données

temps	1	2	3
Y	1	3	5

Le tableau 1 peut être représenté par le graphique suivant (fig. 2)

Figure 2 : graphique représentant les données du tableau 1



Ce graphique n'a perdu aucune information par rapport au tableau 1.

Ces mêmes données peuvent être exprimées par l'équation 1 :

$$y = 2t - 1 \quad \text{Eq.1}$$

avec $t = 1, 2$ et 3

Ici encore cette présentation des données a gardé toute l'information contenue initialement dans le tableau 1.

On peut encore présenter ces données par l'équation 2

$$X = 2t - 1 \quad \text{Eq. 2}$$

avec $t > 0$

Avec cette présentation on a perdu l'information initiale mais on a gagné en généralité. En effet l'information de l'équation 2 est différente de l'information de l'équation 1. Avec l'équation 2 on perd l'information relative aux points effectivement

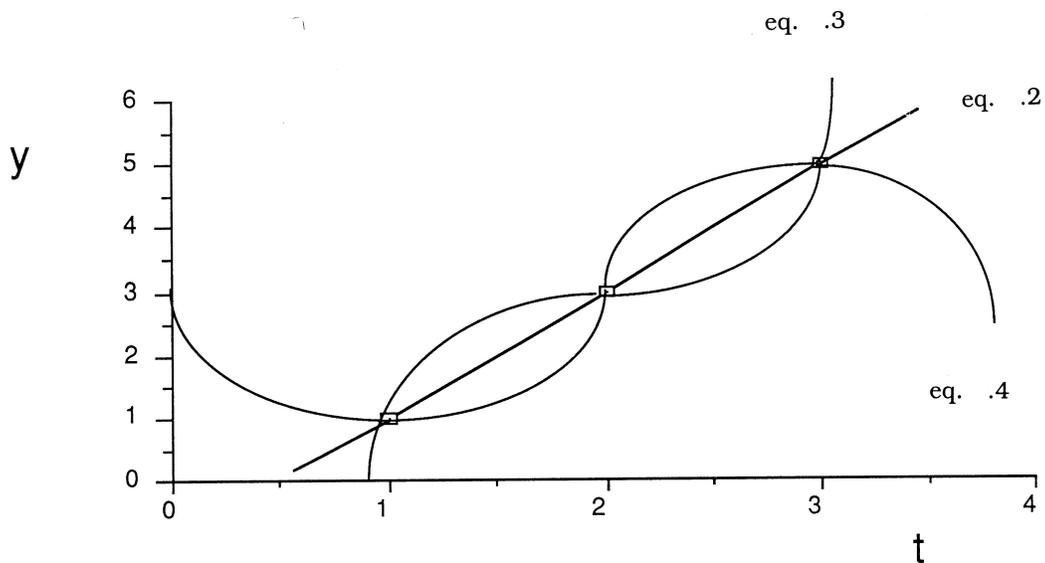
mesurés ; au contraire, on gagne de l'information car on peut estimer les valeurs de X pour les temps non mesurés. Cette nouvelle information implique une hypothèse faite *a priori* par l'analyste : X varie de façon linéaire avec t. Cet *a priori* n'est pas forcément justifié. En effet, les données du tableau 1 peuvent être aussi bien ajustées par les équations 3 et 4 :

$$X = 3 - 2\sin\left(\frac{\Pi t}{2}\right) \quad \text{Eq.3}$$

$$X = t^3 - 6t^2 + 13t - 7 \quad \text{Eq.4}$$

La figure 3 donne une représentation graphique des 3 équations capables d'ajuster de façon parfaite les données du tableau 1

Figure 3 : représentation graphique des 3 équations ajustant de façon parfaite les données du tableau 1



Eq. 2 : $X = 2t - 1$

Eq. 3: $X = 3 - 2\sin\left(\frac{\Pi t}{2}\right)$

Eq. 4 : $X = t^3 - 6t^2 + 13t - 7$

Nous avons donc trois modèles concurrents pour présenter les données du tableau 1. Leurs hypothèses sous-jacentes sont totalement différentes et avec les équations 3 et 4, les estimations des valeurs de X pour des temps autres que ceux qui ont été mesurés, sont différentes de celles qui ont été obtenues avec la droite.

En définitive, au niveau de la description des données on doit distinguer le codage de la modélisation : le codage est une technique de représentation généralement destinée à dégager une impression visuelle. Avec cette approche, on peut perdre une partie de l'information initiale sans jamais gagner une information nouvelle. Avec la modélisation, on perd obligatoirement une partie de l'information initiale mais on gagne en généralité. Cette généralisation implique des hypothèses. En pharmacocinétique, l'aspect descriptif est majeur et de nombreuses séries de données (concentrations en fonction du temps) sont décrites par des équations exponentielles de la forme (Equation 5)

$$C = \sum_{i=1}^n Y_i \exp(-\lambda_i t) \quad \text{Eq. 5}$$

Une première **erreur courante** lorsque l'on rédige un rapport ou un article, consiste à écrire : les données sont décrites par un modèle mono- ou pluri-compartimental et elles ont été ajustées à une somme d'exponentielles. Hormis le cas rare où l'on a une connaissance a priori du modèle, c'est exactement l'inverse qu'il conviendra d'écrire. En effet, **c'est parce qu'on peut décrire les données par une somme d'exponentielles que l'on peut postuler un modèle à un ou plusieurs compartiments.**

Il convient de remarquer que le fait de pouvoir ajuster les données à une somme de n exponentielles ne veut pas forcément dire que le modèle est à n compartiments. Le tableau 2 illustre cette notion.

Tableau 2 : exemple d'une série de concentrations pouvant être ajustées par deux équations différentes.

t	0	5	10	15	20	25	30	35	40	45
X	1	0.16	0.0909	0.0628	0.0475	0.0386	0.0319	0.0278	0.0245	0.0217

Ces données peuvent être ajustées par l'équation 6 :

$$C=0.7616\exp(-0.6749t)+0.1815\exp(-0.1412t)+0.0569\exp(-0.0219t) \quad \text{Eq.6}$$

mais elles peuvent aussi bien être ajustées par une hyperbole :

$$C = 1.0006/(t+1.0006) \quad \text{Eq.7}$$

4 - Prédiction - Simulation

La simulation consiste à utiliser le système secondaire pour générer de nouvelles valeurs. Les objectifs de la simulation peuvent être au nombre de trois.

- (i) compléter les informations, c'est à dire connaître les valeurs intermédiaires (**interpolation**) ou une valeur extérieure à l'intervalle observé (**extrapolation**). L'exactitude et la précision des résultats obtenus seront fonction du modèle retenu.
- (ii) faire des inférences c'est à dire prédire ce que sera le comportement du système primaire lorsqu'on modifie les paramètres de structure du modèle ou encore son entrée.
 - la simulation peut répondre à des questions du genre
 - on double la dose du médicament (modification de l'entrée) : quelles sont les nouvelles concentrations, le nouvel effet au temps t ?
 - on réalise plusieurs administrations du médicament quelle est l'évolution des concentrations plasmatiques en fonction du temps ?
 - quel sera la conséquence d'une réduction de la constante d'élimination (modification d'un paramètre de structure) sur la cinétique plasmatique de l'analyte ?

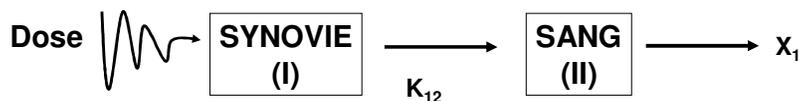
Signalons pour terminer que la simulation peut permettre d'évaluer certains paramètres inaccessibles à la mesure expérimentale (voir le chapitre sur les modèles physiologiques).

5 - Comprendre le système secondaire (valeur heuristique d'un modèle)

On cherche, à l'aide du modèle, à comprendre la réalité anatomique, biochimique, physiologique, ... du système primaire.

Pour illustrer cette possibilité, nous allons prendre l'exemple d'une administration intra-articulaire d'un corticoïde. Après une administration intra-articulaire du principe actif, on mesure la disparition du médicament à partir du liquide synovial, ainsi que sa vitesse d'apparition dans le sang ; a priori le modèle le plus simple est le suivant (fig. 4):

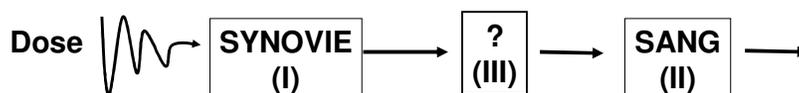
Figure 4 : modèle construit *a priori* pour rendre compte de la disposition d'un corticoïde administré par voie intra-articulaire.



Ce modèle est formé de deux compartiments : la synovie (I) et le sang (II).

L'administration s'effectue dans la synovie et les observations se font à la fois dans la synovie et dans le sang. Après avoir ajusté au mieux les données avec ce modèle, on constate que la vitesse de disparition de l'analyte à partir de la synovie est beaucoup plus rapide que la vitesse d'apparition de l'analyte dans le sang. Cela invalide le modèle de la figure 4 et conduit à tester un modèle plus compliqué, c'est à dire un modèle dans lequel on a intercalé un nouveau compartiment (fig.5) :

Figure 5 : modèle construit *a priori* pour rendre compte de la disposition d'un corticoïde administré par voie intra-articulaire :



Ce modèle est formé de 3 compartiments : la synovie (I), le sang (II) et un compartiment intermédiaire (III). Avec ce nouveau modèle on rend compte de façon appropriée des données. Cela conduit le biologiste à postuler l'existence d'une entité anatomique (ex.: membrane) jouant un rôle de tampon entre la synovie et le sang. Si effectivement cela est vrai, c'est à dire vérifié expérimentalement, ce modèle aura eu une valeur heuristique.

Cette approche est très largement utilisée en biochimie, en physiologie...

Figure 6 : les grandes étapes de la modélisation (exemple de la construction d'un modèle physiologique)

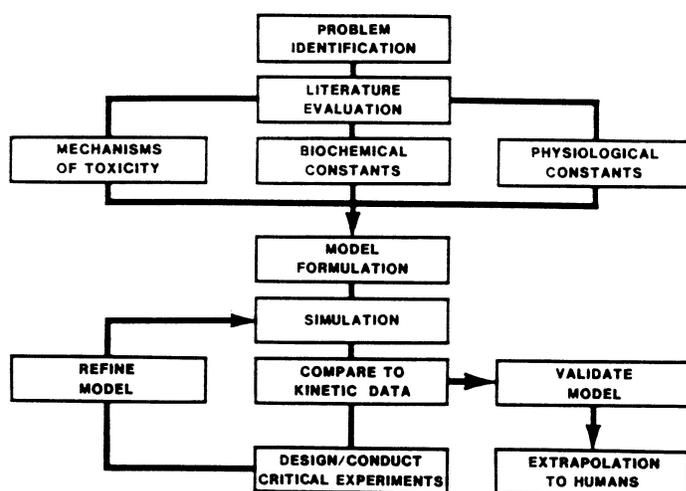


FIGURE 2. Flow-chart of the biologically motivated modeling approach to chemical hazard assessment. The process begins with the discovery of a particular toxic effect in a particular organ of a particular species. Literature evaluation involves the integration of available information about the mechanism of toxicity, the pathways of chemical metabolism, the nature of the toxic chemical species, the tissue binding characteristics, and the physiological parameters of the target species. From these data, a model is developed to estimate the appropriate measure of tissue exposure for a wide variety of exposure conditions. Before the model can be used in human risk assessment it has to be validated against kinetic, metabolic, and toxicity information and, in many cases, refined based on comparison with experimental results. The model itself can frequently be used to design critical experiments to collect data needed for final validation.

6 - LA MODELISATION

6.1 - Les grandes étapes de la modélisation

La modélisation se déroule en 5 grandes étapes (fig.6):

- (i) on se fixe les objectifs de la modélisation
- (ii) on fait l'inventaire des informations
 - données externes (loi, théorie, bibliographie)
 - axiomes (vérités indémontrables)
 - hypothèses
 - données expérimentales
- (iii) on réalise un modèle (spécifications, construction en fonction des axiomes et hypothèses)
- (iv) on identifie les paramètres du modèle
- (v) on valide en confrontant le modèle à la réalité
- (vi) on tire des conclusions

Compte tenu des interactions entre ces différentes étapes, le processus est itératif.

6.2 - Les objectifs

(déjà vus au chapitre 2)

6.3 - L'inventaire des données

Le modèle sera formulé en tenant compte de l'ensemble des connaissances relatives au système. Si on réalise un modèle biochimique, il faut reconnaître les cascades des réactions.

Si on bâtit un modèle physiologique, on devra tenir compte des données anatomiques, physiologiques... Il faut prendre en compte ses propres données expérimentales (ex.: si on ne dispose que de 3 couples de valeurs, il sera impossible d'avoir un modèle bi-compartmental).

6.4 - Formulation du modèle

La formulation du modèle pourra se réaliser selon trois approches :

- (i) **empirique**
- (ii) **théorique** (modèles dits **physiologiques**)
- (iii) **mixte (empirico-théorique)**

6.4.1 - Modèle empirique ou modèle de données

Ce sont les données qui suggèrent le modèle et ces modèles empiriques ne postulent aucune connaissance concernant la structure interne (réelle) du modèle. Ces modèles ne prennent en compte que **l'entrée** et la **sortie** et ils tentent de les relier entre elles. Entre l'entrée et la sortie, il y a une **boîte noire**.

Les modèles empiriques sont essentiellement descriptifs. Leur valeur prédictive est faible sauf pour prédire ce que sera la réponse en fonction d'un stimulus survenant dans les mêmes conditions (cela implique la linéarité).

Les modèles empiriques exponentiels peuvent permettre de passer à une formulation empirico-théorique du modèle (modèles compartimentaux).

Les modèles descriptifs les plus utilisés en pharmacocinétique sont les modèles exponentiels ; on peut aussi faire appel à des polynômes (ex.: splines). Cette approche offre l'avantage de bien localiser les pics des concentrations plasmatiques et de permettre des calculs précis des AUC.

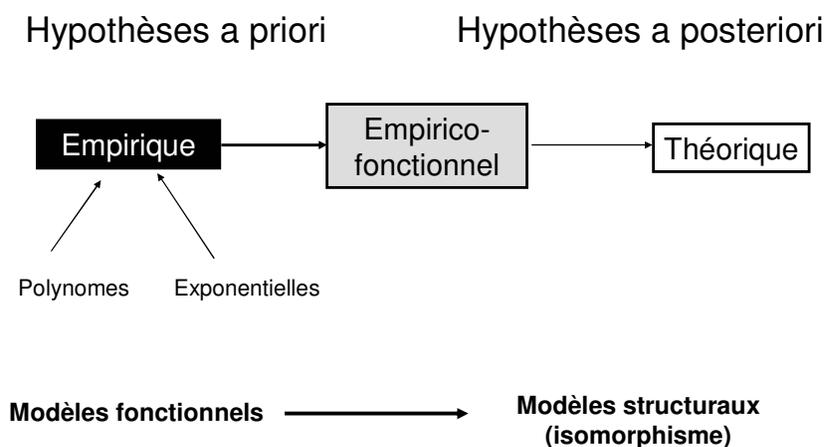
6.4.2 - Les modèles empirico-théoriques

Ce sont des modèles situés à mi-chemin entre les modèles empiriques et les modèles théoriques (physiologiques).

Dans les modèles empirico-théoriques (exemple : les modèles compartimentaux), on postule une structure simplifiée de la réalité physiologique (compartiment). On peut parler d'une structure en boîte grise par rapport au modèle empirique pour lequel on avait une boîte noire entre l'entrée et la sortie du système.

La figure 7 montre le degré de complexité des différents modèles. Les modèles compartimentaux font l'objet d'une présentation spécifique ([accéder au chapitre](#)).

Figure 7 : les différents types de modèles utilisés en pharmacocinétique.



6.4.3 - Les modèles théoriques (physiologiques)

Avec ces modèles, on a une connaissance réelle et *a priori* du système ; Ils impliquent le recours aux données anatomiques, physiologiques, biochimiques...

Ces modèles sont beaucoup plus compliqués que les précédents. Leur objectif essentiel est de réaliser des **simulations** afin de tester des **hypothèses biologiques**. Ils sont un outil privilégié pour réaliser des **extrapolations interspécifiques** ou encore pour anticiper ce que sera la répercussion d'une altération anatomique ou physiologique sur la cinétique d'un produit (exemple : hémorragie). La puissance de ces modèles est limitée par leur validité (liée à leur complexité). Ils seront présentés en détail dans [un autre chapitre](#).

Certains modèles **physiologiques partiels** sont très utilisés en pharmacocinétique ; c'est ainsi que nos conceptions en matière de clairance hépatique font très largement appel à un modèle proposé par Rowland et faisant intervenir le débit sanguin hépatique (Cl_H) :

$$Cl_H = \frac{\dot{Q} \cdot fu \times Cl_{int}}{\dot{Q} + fu \times Cl_{int}}$$

avec \dot{Q} , le débit sanguin hépatique, fu , la fraction libre de l'analyte et Cl_{int} , la clairance intrinsèque libre de l'analyte.

6.5 - Identification du modèle (notion d'identifiabilité)

Le but de l'identification est de calculer les paramètres des modèles à partir de la connaissance que l'on a de son comportement dynamique. Sur un plan pratique, on compare une série de données obtenues expérimentalement avec celles que prédit le modèle ; ensuite on fait varier les paramètres du modèle jusqu'à l'obtention de la meilleure adéquation possible entre les données observées et les données calculées.

L'identifiabilité est un problème théorique qui étudie la **possibilité de déterminer de façon unique les paramètres** du modèle lorsque sa structure est connue et que l'on connaît son entrée et sa sortie sans erreur. Ce problème sera repris lors de l'analyse des données.

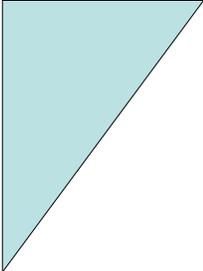
6.6 - Valeur et validité du modèle

La valeur et la validité sont deux concepts différents.

La **valeur** d'un modèle est liée à son utilité. Cette dernière dépend des objectifs (en d'autres termes un modèle peut être "faux" mais utile).

La **validité** exprime la conformité structurale à la réalité. La figure 8 explicite ces deux notions.

Figure 8 : valeur et validité d'un modèle. Soit un sujet ayant à faire trois déplacements (1) de Paris à Toulouse (2) du bois de l'Ecole Vétérinaire au bois de l'Ecole des CRS (3) du bâtiment de physiologie de l'ENVT aux animaleries du même service.

	Valeur en fonction du déplacement			Validité
	1	2	3	
Carte d'état major	0	+++	0	
Carte Michelin	+++	0	0	
Croquis à main levée	0	0	+++	

Dans le cadre des modèles pharmacocinétiques, les modèles compartimentaux représentent des simplifications extrêmes de la réalité mais leur utilité peut être grande pour proposer un schéma posologique. En revanche, leur valeur heuristique est très limitée, voire nulle. Par opposition, la validité des modèles physiologiques est meilleure que celle des modèles compartimentaux mais leur valeur sera très différente selon les objectifs poursuivis. Elle sera grande pour réaliser des extrapolations d'une espèce à l'autre, ou encore pour prévoir ce que sera la cinétique du médicament dans des conditions physiopathologiques. Par contre, la valeur de ces modèles est mauvaise pour décrire les données et elle est généralement moins bonne que celle des modèles compartimentaux pour déterminer un schéma posologique (exemple : prévoir le profil des concentrations plasmatiques lors d'une dose multiple à partir des résultats obtenus avec une dose simple).