

# Chapitre 3

## Paramétrer un modèle dans un contexte de co-construction interdisciplinaire

Version 2018-02-26

Le modèle, tel qu'il a été présenté dans le chapitre précédent, était un « état », c'est-à-dire que les mécanismes, paramètres et les valeurs de ceux-ci correspondent à une étape d'un modèle amené à évoluer pour répondre aux problèmes soulevés dans la dernière partie ([Ref dernière section chap 2](#)). Dans ce chapitre, nous nous attacherons à présenter le travail de paramétrage réalisé à la suite, ayant abouti à une version plus adaptée aux questions des thématiciens. Le descriptif technique de cette version « finale » se trouve dans l'annexe *n* ([Ref à l'annexe contenant le descriptif technique de la dernière version du modèle.](#)) Par paramétrage, nous entendons ici le processus visant à doter le modèle de paramètres (empiriques, « commensurables » et techniques (encadré 3.2)) lui permettant de mieux répondre aux objectifs fixé en termes de comportements attendus ou d'objectifs quantitatifs. Nous nous attacherons dans un premier temps à expliciter et spécifier ce sens.

### 3.1 Paramétrer ? Quoi et quand ?

Avant de préciser le sens du terme « paramétrage », il semble important de définir précisément ce qu'est un paramètre. C'est en particulier nécessaire en ce que ce terme recouvre de nombreux sens selon les champs disciplinaires qui l'emploient, mais aussi, au sein même de ceux-ci, par les différents chercheurs.

#### 3.1.1 Différents points de vue sur la définition d'un paramètre

Au plus général, le nouveau petit Robert définit un paramètre en ces mots :

- « 1. MATH. Quantité à fixer librement, maintenue constante, dont dépend une fonction de variables indépendantes, une équation ou une expression mathématique. — Variable en fonction de laquelle on exprime chacune des variables d'une équation.
  2. FIG. et DIDACT. Élément important dont la connaissance explicite les caractéristiques essentielles de l'ensemble d'une question.
  3. PAR EXT. Élément nécessaire pour juger, évaluer, comprendre (qqch.). »
- (ROBERT, REY-DEBOVE et REY 1993, **Paramètre**)

Seule la première définition correspond, à grands traits, à ce que l'on attend ici, mais elle est très généraliste, bien que ne correspondant pas pour autant à tous les usages du terme employés dans la littérature.

### En mathématiques, une définition univoque

L'acceptation mathématiques d'un paramètre est sans doute celle qui souffre le moins d'ambiguïté : il s'agit des termes fixes d'une équation simple, par opposition aux variables qui en constituent les éléments qui seront amenés à évoluer. Par exemple, dans la formulation d'une fonction affine,  $f(x) = ax + b$ , la valeur de  $f(x)$  dépend de la variable  $x$  et des paramètres fixés  $a$  et  $b$ . Quelles que soient les valeurs empruntées par  $x$ , ces paramètres demeurent constants. Dans une famille d'équations plus complexes, par exemple le modèle de croissance logistique de population<sup>1</sup>, l'accroissement de population au cours du temps ( $\frac{dy}{dt}$ ) dépend de la valeur de la variable  $y$ , population à cet instant, ainsi que de deux paramètres,  $\alpha$ , le taux de croissance, et  $K$ , la « capacité d'accueil », c'est-à-dire un potentiel maximum de population vers laquelle tend – et ne peut donc dépasser – le système modélisé. La définition mathématique est donc assez universelle et convient à la quasi-totalité des systèmes d'équations<sup>2</sup>. Il convient toutefois de noter que la différence entre variable et paramètre est une affaire de point de vue, une inversion de perspective menant à échanger les paramètres et variables, tel que décrit dans l'exemple suivant :

« For the function  $f(x) = ax^2 + bx + c$ , could we learn something by leaving the expression in terms of the symbols  $a$ ,  $b$ , and  $c$  and seeing how  $f(x)$  depends on the parameters? Maybe we could fix  $x = 2$  and look how  $f(2)$  changes as we let  $a$  vary.

If we do such manipulations and look at how the output of a function depends on varying a parameter, then we are treating the function as though the parameter were a input variable. But that's OK, as the difference between variables and parameters is really just a matter of perspective. »

(NYKAMP 2015a)

On peut donc en retenir qu'en mathématiques, ce qui différencie la variable du paramètre est l'aspect fixe de ce dernier, au moins pendant la durée d'exécution d'une fonction :

« A parameter is a quantity that influences the output or behavior of a mathematical object but is viewed as being held constant. [...] Variables are viewed as changing while parameters typically either don't change or change more slowly. »

(NYKAMP 2015b)

On propose donc une définition sans doute plus spécifique que celle du nouveau petit Robert, mais aussi plus tournée vers l'usage : un paramètre est une variable maintenue constante durant l'ensemble de l'utilisation d'une fonction.

### Une vision duale en statistiques

En statistiques, quand bien même cette discipline fait un large usage des formalismes mathématiques, les paramètres recouvrent un ensemble assez différent :

1.  $\frac{dy}{dt} = \alpha y \times (1 - \frac{y}{K})$  d'après (VERHULST 1838)

2. A l'exclusion notable des systèmes d'équation paramétriques, où, à l'inverse, le terme de paramètre désigne alors les variables indépendantes.

il s'agit d'une « grandeur mesurable qui permet de présenter de façon plus simple, plus abrégée les caractéristiques essentielles d'un ensemble statistique. » (IMBS et QUEMADA 1992, Paramètre, STAT. (calcul des probabilités))

*In statistics*, the most common use of « parameter » is for a characteristic of a population, or of a distribution of scores, described by a statistic such as a mean or a standard deviation. For example, the mean (average) score of on the midterm exam in Psychology 201 is a parameter. It describes the population composed of all those who took the exam.

(VOGT 1993, p. 164)

Ainsi, pour les statisticiens, la moyenne, l'écart-type ou encore le coefficient d'asymétrie sont des paramètres, que l'on symbolise alors au moyen de lettres grecques (VOGT 1993, *ibid.*). On peut toutefois y retrouver une logique commune avec les paramètres mathématiques quand on décrit une loi statistique avec ces valeurs, qui deviennent alors les éléments permettant de caractériser une variable. Ainsi, pour décrire les caractéristiques de la distribution théorique – normale – d'une variable  $X$ , on fera appel aux paramètres théoriques de cette loi que sont l'espérance<sup>3</sup> ( $\mu$ ) et l'écart-type ( $\sigma^2$ ) :  $X \hookrightarrow \mathcal{N}(\mu, \sigma^2)$ .

### Une absence de consensus en informatique

Dans le domaine informatique, la définition est bien plus floue et inclusive que dans les champs décrits auparavant, sans doute en raison d'une hétérogénéité bien supérieure dans les pratiques et construits informatiques. Comme dans les fonctions mathématiques, les paramètres sont les arguments des fonctions informatiques. Ces fonctions recouvrent toutefois un ensemble extrêmement vaste, bien plus hétérogène que dans les domaines mentionnés ci-dessus. On parle ainsi de fonction<sup>4</sup> pour toute suite d'instructions informatique ayant pour vocation – ou pour capacité – à être répétée, ré-exécutée. Ces fonctions appliquent ainsi, le plus fréquemment, cette suite d'instruction sur des données en entrée, et produisent ainsi des données en sortie, résultantes du traitement effectué sur les entrées<sup>5</sup>. Certaines fonctions sont donc très simples, et peuvent s'apparenter à des fonctions mathématiques, telles que par exemple la fonction arrondi (`round()` dans sa version la plus courante). Cette dernière prend en entrée un nombre décimal, et retourne un nombre entier. Le nombre décimal passé en entrée est alors nommé paramètre. Notons que cette fonction accepte le plus souvent un second paramètre, sous forme d'un nombre entier, qui permet de définir le nombre de décimales (*digits*) que l'on souhaite conserver. Par exemple, l'arrondi de 2,551 à une décimale renverra la valeur 2,6. Il est important de noter qu'en informatique, les paramètres ont souvent une « valeur par défaut », c'est-à-dire une valeur qui

3. Il s'agit de la moyenne théorique. On réserve ainsi le terme de moyenne à une valeur calculée depuis des valeurs empiriques.

4. On utilisera aussi indistinctement les termes de procédure, de (sub)routine ou encore de méthode. Notons que chacun de ces mots a normalement un sens précis. Par exemple, une méthode désigne une fonction qui peut être exécutée par une instance de classe en programmation orientée objet. Certaines définitions plus précises sont toutefois souvent utilisées à mauvais escient, provoquant alors des inversions de sens. Ainsi, une procédure est le plus souvent décrite comme une suite d'instructions ne retournant pas de valeur, au contraire d'une fonction. Mais, par exemple dans le langage SAS (INSTITUTE 1990), toute fonction est nommée procédure, à l'instar de la moyenne par exemple (`PROC MEANS`).

5. Notons que certaines fonctions produisent des « effets de bord », c'est-à-dire ne renvoient pas de données en sortie, mais effectuent une action à partir des données en entrée. La fonction imprimer par exemple, ne renvoie aucune sortie informatique, ce sont ses effets de bord qui déclenchent l'impression d'un document.

sera utilisée si le paramètre n'est pas explicitement spécifié. D'où une confusion fréquente entre paramètre et valeur initiale, en particulier dans le domaine de la simulation à base d'agents. Dans le cas de la fonction `arrondi`, si le premier paramètre n'a pas de valeur par défaut, ce qui n'aurait aucun sens, le second paramètre est souvent proposé avec une valeur par défaut de 0. Si on ne la précise pas, le nombre renvoyé par la fonction sera ainsi un entier, soit 3 dans le cas précédent : `round(2.551) = 3`, mais `round(2.551, digits = 1) = 2,6`.

Ce premier exemple est quasiment en tout point assimilable à l'acceptation mathématique d'une fonction, et dès lors, ses paramètres ressemblent fortement à ceux que ce domaine définit, à l'exception que le premier paramètre de la fonction `arrondi` serait défini comme une variable en mathématiques. Pour autant, l'informatique fait usage de nombreuses fonctions bien moins comparables, car formulées de manière algorithmique et non mathématique.

Prenons l'exemple d'une fonction simple de conversion d'image, permettant par exemple de convertir une image du format JPEG au format PDF. Cette fonction, que l'on nommera `convert`<sup>6</sup>, requiert au moins deux « paramètres » : l'emplacement informatique (le « chemin », ou *path*) du fichier image (JPEG) d'origine, et le chemin du PDF en sortie. Il ne s'agit plus dès lors de paramètres numériques comme en mathématiques ou en statistiques, mais d'éléments nécessaires à une fonction pour être exécutée. De plus, cette fonction accepte aussi d'autres « paramètres », permettant entre autre de redimensionner l'image pendant cette conversion, d'en modifier la résolution, ou encore d'en transformer les couleurs, par exemple en la convertissant en nuances de gris. Ces paramètres, facultatifs, agissent alors comme autant de nouvelles fonctions. Ils ne servent plus uniquement à « paramétrer » le but premier de la fonction de conversion, mais en fait à y ajouter des fonctionnalités, des mécanismes de transformation de l'image source.

On ne peut donc plus véritablement parler de variables qui seraient affectées ou transformées par des valeurs statiques de paramètres. On utilisera donc, dans ce contexte, davantage le terme de paramètre d'entrée ou encore d'argument dans ce cas. Notons qu'en anglais, la différence entre *parameter* et *argument* est plus formalisée qu'en français : ils se définissent par le lieu de leur utilisation. Lors de la définition d'une fonction, on fait appel à des paramètres qui seront utilisés au sein de la fonction. Lors de l'utilisation de cette fonction, l'utilisateur fournira des arguments, dont les valeurs seront alors utilisés en remplacement des paramètres dans la fonction.

« The terms parameter and argument are sometimes used interchangeably, and the context is used to distinguish the meaning. The term parameter (sometimes called formal parameter) is often used to refer to the variable as found in the function definition, while argument (sometimes called actual parameter) refers to the actual input passed. For example, if one defines a function as `def f(x): ...`, then `x` is the parameter, while if it [is] called by `a = ...; f(a)` then `a` is the argument. »

(*Parameter (Computer Programming)*)

Il apparaît donc que si les définitions mathématiques et statistiques d'un paramètre sont assez largement précises et explicites, il en est tout autre dans le champs disciplinaire informatique. Peut-être parce que ce champs est composé de bien plus de praticiens (les développeurs) que de chercheurs, on constate que

6. Cette fonction est disponible dans le logiciel ImageMagick (IMAGEMAGICK STUDIO 2008).

les termes de variables, de paramètres, d'arguments ou encore d'entrées (*inputs*) y sont assez régulièrement intervertis. Afin de préciser l'emploi que nous ferons de ces termes dans le cadre de la modélisation à base d'agents présentée dans le chapitre précédent, il convient donc de s'intéresser plus spécifiquement aux usages de ces termes dans le domaine de la simulation informatique en sciences humaines.

### 3.1.2 Les paramètres dans les modèles agents

#### L'approche classique

Un premier point est à noter : nous n'avons trouvé que très peu de définitions spécifiques de ce qu'est un paramètre dans le champs de la simulation à base d'agents. C'est pourtant un terme employé dans la quasi-totalité de la littérature existante. Cela ne relèverait que d'un problème de jargon non explicité s'il y avait consensus que le sens donné à ce mot, mais au contraire, les acceptations, qui doivent être comprises par le contexte en l'absence de définitions formelles, varient fortement selon les auteurs. Par exemple, la définition que l'on peut extraire de l'un des manuels de référence en modélisation agent (TREUIL, DROGOUL et ZUCKER 2008) s'éloigne fortement de ce que l'on a pu décrire ci-dessus :

Un modèle dynamique renferme en effet deux composants distincts : une représentation de la structure du système de référence (exprimée dans le langage du méta-modèle), et une représentation des lois régissant sa dynamique. Ces deux représentations sont habituellement pourvues de données ou d'éléments d'information souvent numériques (le minimum pour un modèle dynamique étant d'être pourvu d'un élément représentant le temps) appelés **paramètres**. La « perturbation » d'un modèle par simulation va donc signifier la modification contrôlée de la valeur de certains de ces paramètres, que l'on appellera **entrées** du modèle. Inversement, ce que l'on pourra mesurer dans une simulation sera décrit sous la forme d'autres paramètres qui seront appelé **sorties**.

Les *entrées* d'un modèle dynamique sont des paramètres dont la valeur est définie en dehors du modèle et qui représentent ce que le simulateur peut perturber. Les *sorties* d'un modèle dynamique sont également des paramètres qui expriment ce que l'on cherche à mesurer en réponse à ces perturbations.

(TREUIL, DROGOUL et ZUCKER 2008, p. 8)

Nous trouvons plusieurs problèmes à cette définition. En premier lieu, l'acceptation très globale de ce qu'est un paramètre rappelle celle d'une variable en informatique. Il semble s'agir d'une vision plus orientée techniquement que conceptuellement. Ainsi, définir le temps — ou la variable informatique permettant de le mesurer — comme un paramètre nous semble bien trop à contre-courant des définitions de paramètres issues des autres champs scientifiques. De plus, le fait que les sorties d'un modèle soient considérées comme des paramètres est en opposition avec l'ensemble de l'usage courant de ce terme, y compris dans le domaine spécifique de la modélisation agent. Nous ne pouvons donc souscrire à cette définition, ni à la vision qu'elle dépeint de ce qu'est un paramètre.

Dans le présent ouvrage, nous donnerons donc à ces termes des sens différents, voire opposés, qui nous semblent plus fréquents dans le champ de la modélisation en sciences humaines et se retrouvent partiellement dans le schéma de Balci

(fig. 3.1).

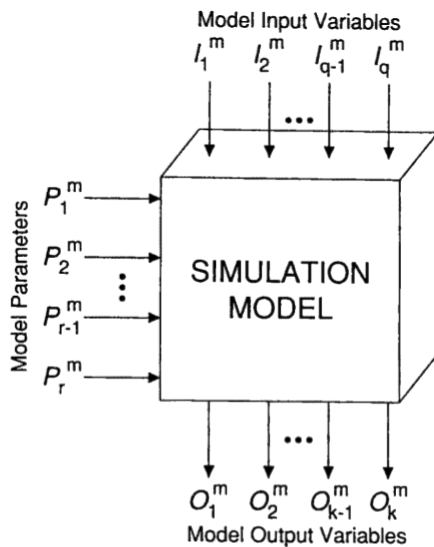


FIGURE 3.1 – Les variables d’un modèle de simulation selon Balci (BALCI 1994, p. 122).

*Nota bene* : ce schéma est tronqué, ne présentant que la partie « modèle de simulation » alors que celle-ci est mise en miroir à une partie « système » représentant l’empirique.

Dans ce schéma et l’article associé, et bien qu’il n’en définisse nulle part explicitement le sens, Balci esquisse la composition d’un modèle de simulation.

Décrire le schéma de Balci point par point. Faire un schéma pour Treuil, puis introduire les miens en les explicitant complètement.

### Paramètres, variables, indicateurs : un essai de définition

Nous considérerons donc plutôt les paramètres comme le sous-ensemble des entrées, qui peuvent s’exprimer sous forme numérique (excluant donc de fait certaines entrées telles que les configurations spatiales initiales), et que l’on fera varier, aussi bien dans la calibration du modèle que pour l’exploration de scénarios.

Faire un schéma avec organisation des variables/paramètres/entrées/-sorties telles que proposée dans ma thèse.

Ne pas oublier de préciser différence entre paramètres et situation initiale. Ex. de situation initiale : des variables déjà allouées, par ex. pop des villes. Il n’y en a pas vraiment dans mon modèle pk tout est endogénéisé : pas de situation initiale, mais des paramètres qui définissent la création de celle-ci.

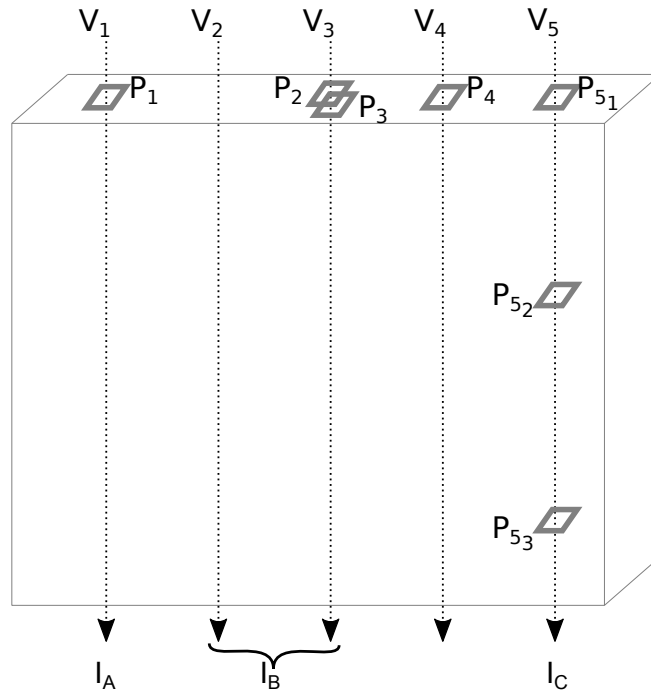


FIGURE 3.2 – Schématisation des définitions de variables ( $V$ ), indicateurs de sortie ( $I$ ) et paramètres ( $P$ ).

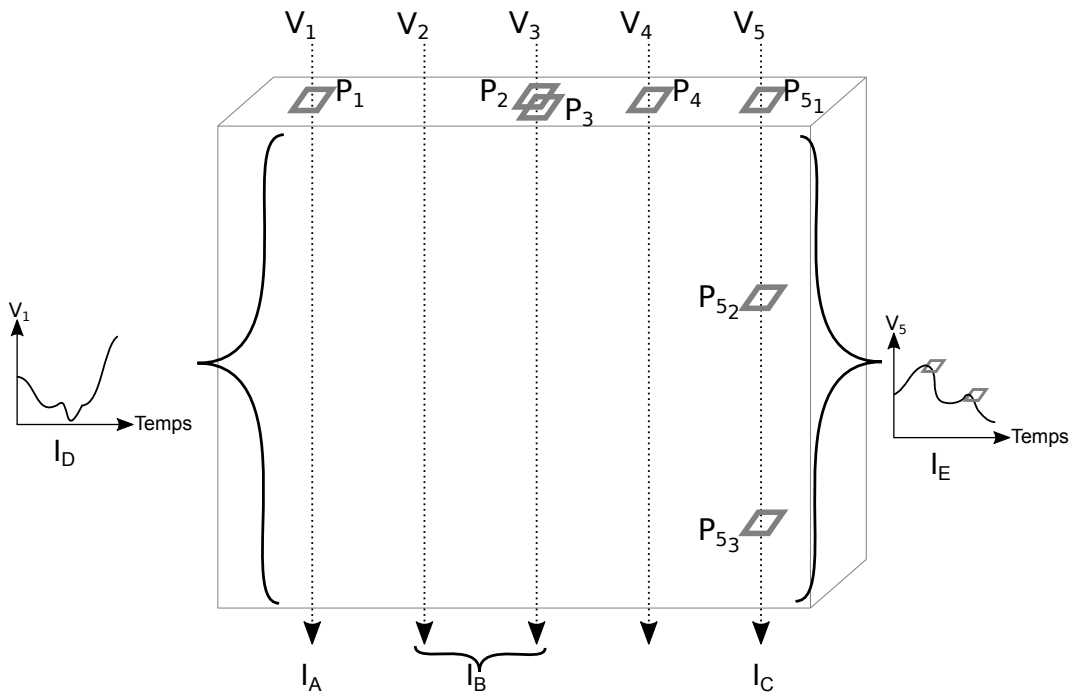


FIGURE 3.3 – Schématisation des définitions de variables ( $V$ ), indicateurs de sortie ( $I$ ) et paramètres ( $P$ ) incorporant l'aspect temporel de certains indicateurs ( $I_D$  et  $I_E$ ).

### Paramètres et variables

*Encadré 3.1 : Variables, entrées, paramètres...*

Au sein des **variables**, constituées par l'ensemble des éléments d'un modèle ayant la capacité de changer, que ce soit au cours d'une simulation ou au travers de plusieurs expérimentations, nous distinguerons trois types spécifiques.

**Entrées**

Les **entrées** constituent l'ensemble des informations nécessaires à l'exécution d'un modèle en ce qu'elles en alimentent les mécanismes. Elles peuvent être constantes au cours d'une simulation (la date de fin d'exécution de notre modèle par exemple), ou évoluer de par l'action du modèle (le nombre d'églises paroissiale du modèle, qui est une entrée, et augmente tout au long de la simulation).

**Paramètres**

Certaines de ces entrées sont des **paramètres** : ce sont les entrées qui peuvent être mobilisées afin de modifier le comportement du modèle, et qui « représentent ce que le simulateur peut perturber »<sup>6</sup>. Dans cet ouvrage, nous distinguerons plusieurs types de paramètres (encadré 3.2) selon leur ancrage dans l'empirique. Notons toutefois dès maintenant que le nombre de paramètres d'un modèle peut être bien supérieur au nombre de paramètres que l'on fera réellement varier dans les expérimentations.

**Sorties**

Les sorties d'un modèle représentent ce que l'on va observer après l'exécution du modèle afin d'observer son déroulement et aboutissement.<sup>6</sup> Ces sorties peuvent être très nombreuses, et concerner autant les variables individuelles (une liste des Foyers Paysans et de leur satisfaction au cours du temps par exemple) que des variables agrégées (Quartiles de satisfaction des foyers paysans par exemple), moins coûteuses en temps d'écriture ainsi qu'en espace de stockage. De la même manière que pour les paramètres, sur l'infinité potentielle de sorties d'un modèle, seules quelques-unes seront effectivement générées et enregistrées afin de minimiser les données produites à celles que le modélisateur compte explorer.

6. Pour reprendre les mots de Treuil et al. qu'ils appliquent aux entrées (TREUIL, DROGOUL et ZUCKER 2008, p. 8).

6. Pas clair

**Différents types de paramètres**

Les paramètres sont donc un sous-ensemble des entrées d'un modèle qui ont vocation à varier dans les différentes exécutions de ce modèle. Pour autant, c'est un sous-ensemble assez divers, et tous les paramètres n'apportent pas les mêmes connaissances, que ce soit par leur valeur ou par la manière dont ils varient. Il convient donc d'identifier différents types de paramètres, non selon leur caractéristique propre (c'est-à-dire leurs propriétés, qui permettent de les différencier des sorties par exemple), mais selon leur usage et donc la manière dont on pourra les mobiliser. A ce titre, quand la littérature en sciences humaines et sociales<sup>7</sup> distingue souvent (**refs**) deux types de paramètres : les paramètres

7. En physique ou en mathématiques par exemple, une telle distinction n'existe pas, la définition d'un paramètre prenant une considération plus simple d'élément rendu constant dans



empiriques, dont la valeur a une correspondance directe dans le domaine empirique (on peut donc l'observer), et les paramètres techniques<sup>8</sup>, pour lesquels cette correspondance n'existe pas. Comme dans la plupart des typologies ayant trait à la catégorisation de valeurs numériques, de nombreux paramètres peuvent se trouver à l'interface de ces deux classes. Nous en proposons donc une troisième, que nous avons choisi de nommer « paramètres commensurables »<sup>9</sup>, qui nous semble à même caractériser des paramètres partiellement inscrits dans l'empirie, c'est-à-dire sans valeur empirique propre, mais permettant tout de même d'être mobilisés comme les paramètres empiriques, c'est-à-dire à même d'être comparés, que ce soit par les fluctuations de leurs propres valeurs ou en comparant des paramètres commensurables les uns avec les autres. Avec cette typologie des paramètres basées non sur la nature de ceux-ci mais sur leur utilisation dans un modèle, nous nous inscrivons dans une vision fonctionnaliste et donc très subjective, rappelant la définition d'un modèle de Minsky<sup>10</sup> (MINSKY 1965). Pour différents modélisateurs, un même paramètre pourra donc être considéré comme empirique ou commensurable, ou encore comme commensurable ou technique.

Ne pas oublier de préciser que selon le point de vue du modélisateur, un paramètre peut passer de empirique à commensurable, ou de commensurable à technique (et vice-versa), mais pas « franchir » toutes les catégories.

#### *Encadré 3.2 : Trois types de paramètres*

Regarder s'il y a des enquêtes sur les seuils de satisfaction de Schelling dans la littérature.

**Paramètres empiriques** Les paramètres empiriques sont les paramètres qui peuvent être fixés à l'aide d'une ou de valeur(s) empiriquement connue(s), ou à défaut, dont la valeur se réfère à une quantité directement identifiable par une personne ayant des connaissances thématiques sur ce qui est modélisé.

Dans le modèle de Schelling (3.1.4), le paramètre **S** est un paramètre empirique : la valeur qui lui est donnée peut être interprétée empiriquement, quand bien même il serait difficile d'obtenir, par exemple d'un enquêteur, une valeur exacte à utiliser.

**Paramètres techniques** Il s'agit des paramètres dont la valeur, ou même son ordre de grandeur, n'apportent aucune connaissance thématique

une exécution d'un modèle. **pas clair, prendre une définition quelconque reprenant cette idée.**

8. Mathian et Tannier (MATHIAN et TANNIER 2015) utilisent le terme de paramètre mécanique pour définir ce type de paramètres. Nous trouvons l'usage du mot technique plus approprié, en ce qu'il se différencie plus de l'empirique sur le pan de l'usage qui en est fait. De plus, chaque paramètre, empirique, technique ou autre, a un impact sur les mécanismes du modèle, de par la définition que nous empruntons des paramètres encadré 3.1. Nous préférons donc ce terme de « technique » pour éviter les confusions.

9. La définition du dictionnaire « Le Petit Robert » pour ce terme est celle-ci : « Se dit d'une grandeur qui a, avec une autre grandeur, une commune mesure. » (ROBERT, REY-DEBOVE et REY 1993).

Cf. aussi : « Les mondes n'apparaissent commensurables ou incommensurables qu'à ceux qui s'attachent aux mesures mesurées. Or, toutes les mesures, en science dure comme en science souple, sont aussi des mesures mesurantes et celles-là construisent une commensurabilité qui n'existait pas avant leur mise au point. » (LATOUR 2013, p. 153)

10. **Cf. sans doute le chapitre 1, à faire.**

au modélisateur. Ces paramètres ont pour raison d'être de permettre à d'autres types de paramètres de s'exprimer en valeurs compréhensibles et exploitables. Dès lors, leurs valeurs sont propres à chaque modèle formalisé, et une comparaison de ces valeurs entre les différents modèles n'apporte pas de connaissance.

Dans le modèle gravitaire (section 3.1.4),  $k$  est un paramètre technique : c'est un ratio qui varie afin d'ajuster l'ordre de grandeur des masses ( $M_i$  et  $M_j$ ) à celui des flux échangés ( $F_{ij}$ ). C'est une valeur d'ajustement, qui ne peut être comparée entre des modèles portant sur différents espaces et/ou thématiques<sup>11</sup>.

**Paramètres « commensurable »** Cette catégorie, forgée dans le cadre de cet ouvrage, permet de représenter les paramètres qui ne sont pas fondamentalement issus de l'empirique mais qui peuvent toutefois être interprétés, que ce soit par rapport à une valeur canonique, définissant alors un repère, ou alors par rapport à d'autres valeurs de ce même paramètre, créant dès lors une échelle d'analyse. Outre la calibration d'un modèle, les valeurs de ces paramètres ont donc un intérêt propre et peuvent être analysées. pour elles-mêmes.

Dans le modèle gravitaire (section 3.1.4), le frein de la distance (paramètre  $\alpha$ ) peut être considéré comme un paramètre commensurable. Sa valeur ne renvoie ainsi à aucune dimension empirique, mais son usage répandu permet de comparer les valeurs qu'il peut prendre à des valeurs bien connues. Par exemple, on utilise fréquemment des valeurs de  $\alpha$  comprises entre 1 et 2 pour exprimer la friction de la distance dans les navettes domicile-travail. Une valeur inférieure à 1 peut donc être interprétée comme laissant une part plus faible à la distance, par exemple dans le cas de déplacements de loisirs. A contrario, un paramètre  $\alpha$  supérieur à 2 montrera une plus forte friction de la distance, par exemple dans les choix d'une école élémentaire à proximité du foyer d'un ménage.

11. Dans le cas où  $\alpha$  vaut 0, ce paramètre peut être vu comme commensurable dans la mesure où il devient alors la part de la masse qui devient un flux.

## Comment choisir les valeurs de paramètres ?

Ici, faire un court paragraphe de transition pour présenter rapidement l'approche générale du paramétrage : peut-être redondant avec ce qui suit, en particulier dans les présentations des paramétrages du modèle gravitaire et de Schelling.

### 3.1.3 Le paramétrage, un processus d'amélioration du modèle.

Le paramétrage d'un modèle est souvent réduit à l'un de ses aspects, le « calibrage », étape finale de la construction d'un modèle qui cherchera à reproduire autant que possible des données empirique en faisant varier les valeurs des paramètres jusqu'à ce qu'une combinaison de celles-ci soit satisfaisante.

De nombreux auteurs ont montré que le paramétrage d'un modèle ne pouvait se réduire à cette étape, chacun employant des termes différents pour désigner le processus de paramétrage.

Gilbert et Troitzsch (GILBERT et TROITZSCH 2005) restent très flous en utilisant les termes de *checking* et de *debugging*, qu'ils inscrivent dans une démarche plus large de *verification* correspondant à la **validation interne**<sup>12</sup>. Osman Balci (BALCI 1994) crée et emploie l'acronyme « *VV&T* »<sup>13</sup>, à portée plus large puisqu'il s'applique à chacune des étapes de la modélisation

Le plus souvent (**ref Seb, Clem...**), une fois le modèle construit, le modélisateur s'attache à son « calibrage », en cherchant pour chaque paramètre la ou les valeurs qui permettront au modèle de s'approcher des données empiriques devant être reproduites. Cette étape, que l'on nomme généralement calibration, peut se faire de manière manuelle, par approximations successives (**C. Cottineau et/ou S. Rey, d'après Hermann**), par semi-automatisme, exemple en effectuant des analyses de sensibilité (**C. Schmitt, J. Hirtzel**), ou encore de manière entièrement automatique (**C. Schmitt, S. Rey, C. Cottineau avec PSE**). L'approche souvent défendue, notamment dans les travaux les plus récents (**Rimbault ?**), voudrait que cette étape soit obligatoire pour toute modélisation, permettant par une exploration systématique de comprendre l'entière du comportement d'un modèle, et d'en faire dès lors un outil complètement maîtrisé rendant possible l'établissement d'une loi.

Lena : trouver d'autres exemples plus classiques et lointains

Lena : expliciter

Dans notre travail, nous souhaitons revenir sur cette approche de la modélisation, ancrant le paramétrage comme étape ultime de la construction d'un modèle, en particulier en ce que nous considérons que cette pratique de recherches de valeurs optimales est un exercice qui devrait s'effectuer tout au long de la construction du modèle, de manière plus itérative que conclusive.

Lena : penser à bien l'introduire avant.

### 3.1.4 Qu'est-ce que le(s) paramétrage(s) ?

#### Définition

Paramétrer (informatique) : « Programmer (un appareil complexe), en définissant les paramètres assurant son fonctionnement optimal. Ex. *Paramétrer une imprimante* »(ROBERT, REY-DEBOVE et REY 1993).

J'aime bien cette définition, c'est la seule à avoir le sens que j'entends dans paramétrage/paramétrer. Et elle introduit une logique d'optimalité.

Le terme recouvre deux sens différents, dont la distinction peut se faire selon qu'on l'utilise pour définir un processus ou pour caractériser une configuration. Ici, nous emploierons plutôt le premier cas, définissant dès lors le paramétrage comme le processus, manuel ou automatique, visant à constituer cette configuration de paramètres. Dans ce deuxième cas, le paramétrage désigne un ensemble de valeurs de paramètres, par exemple quand on mentionne un paramétrage par défaut, ou un paramétrage optimal. Pour ne pas risquer de contre-sens, nous préférons le terme de configuration de paramètres.

On tend à distinguer le paramétrage – passage obligé ne nécessitant pas d'être évoqué – de la calibration, processus systématique qui inscrirait le modèle comme un outil scientifique et incontestable. Nous choisissons ici de confondre ces approches, non pas en considérant le paramétrage comme un outil d'évaluation du modèle, mais comme une composante inhérente à la construction d'un modèle, quelque soient les formes et les temporalités que le paramétrage adopte. Le paramétrage est en effet une pratique utile dans la construction du modèle, car les résultats auxquels il aboutit, c'est-à-dire les valeurs de paramètres qui semblent

12. **Définir !**

13. *Validation, Verification and Testing*

mieux adaptés, renseignent aussi bien sur les biais des mécanismes adoptés que sur leur efficacité réelle. Par exemple, quand, après avoir ajouté un mécanisme, on se rend compte que des variations dans les valeurs de paramètres ne changent pas réellement les sorties du modèle, cela peut être l’occasion de repenser le mécanisme dans son ensemble, ou plus souvent, la manière dont le paramètre est mobilisé dans ce mécanisme. On retrouve cette logique dans l’exploration par Clara Schmitt du modèle SimpopLocal (SCHMITT 2014), qui a permis de réaliser que la variation de l’un des paramètres (*InnovationLife*) n’avait que peu d’impact sur les sorties du modèle, tout en rendant sa calibration plus complexe et instable :

Au dessous du seuil des 150 pas de simulation pour le paramétrage de *InnovationLife*, le calibrage du modèle est très difficile voire impossible. Au dessus de ce seuil, le mécanisme associé au paramètre *InnovationLife* n’a plus d’effet sur le calibrage du modèle. Dans un soucis de parcimonie du nombre et de la complexité des mécanismes simulés dans le modèle SimpopLocal, il est justifiable de retirer du modèle ce mécanisme qui n’est pas nécessaire à la simulation de la dynamique de croissance recherchée.

(SCHMITT 2014, p. 224)

Afin d’illustrer ces propos, on peut s’appuyer sur des modèles bien connus de la littérature, issus de deux champs disciplinaires différents.

### Illustration : le modèle gravitaire.

**Description** Ce modèle statistique<sup>14</sup>, dans la formulation qu’en a faite Stewart (STEWART 1948), vise à prédire des flux (démographiques, marchands etc.) potentiels entre des lieux à partir d’une analogie avec la loi physique de la gravitation. Dans le formalisme le plus simple, « sans contrainte » (PUMAIN et SAINT-JULIEN 2001), on peut l’exprimer ainsi :

$$F_{ij} = \frac{k \times M_i \times M_j}{d_{ij}^\alpha}$$

Ce modèle présente trois valeurs empiriques ( $M_i$ ,  $M_j$  et  $d_{ij}$ ) et deux paramètres ( $k$  et  $\alpha$ ).  $k$  est un paramètre technique (encadré 3.2), puisque bien que basée sur l’empirique, sa valeur – et les fluctuations de celle-ci – ne donne aucune information sur le système étudié.  $\alpha$  est un paramètre commensurable pour les raisons explicitées dans l’encadré 3.2.

À partir des valeurs empiriques  $M_i$  et  $M_j$  qui caractérisent les masses (populations, ou stocks de marchandise par exemple) des lieux  $i$  et  $j$ , et de la distance qui les sépare  $d_{ij}$ , on peut ainsi prédire  $F_{ij}$ , le flux entre ces lieux.

**Objectif** Afin que ce modèle donne des ordres de grandeur réalistes quand aux quantités échangées, il faut le paramétrer en définissant une valeur de  $k$ , permettant dès lors d’obtenir un rapport entre les masses d’origine et les quantités échangées. La valeur de ce paramètre est conditionnée par celle de  $\alpha$ , que l’on nomme fréquemment « frein de la distance », en ce qu’il permet de quantifier l’impact qu’aura un éloignement plus ou moins important sur la quantité de flux échangés.

---

14. Pour une description plus complète, se référer à l’ouvrage de Pumain et Saint-Julien (PUMAIN et SAINT-JULIEN 2001).

**Paramétrage** Pour « ajuster »<sup>15</sup> le modèle, il faut donc en réaliser le paramétrage en s'appuyant sur les éléments connus de l'équation (les valeurs empiriques).  $\alpha$  et  $k$  étant liés dans l'équation, la valeur de chacun de ces paramètres dépend de celle choisie pour l'autre, et un changement dans l'un des paramètres entraînera la nécessité de modifier l'autre.

Le paramétrage le plus simple consiste à réaliser une régression linéaire sur les logarithmes décimaux des distances et des flux observés,  $k$  prenant alors la valeur du logarithme de l'ordonnée à l'origine et  $\alpha$  celle du coefficient directeur de la courbe. Le modèle est alors calibré (voir encadré 3.3), mais si l'on souhaite donner une valeur spécifique à l'un des paramètres, par exemple pour utiliser une valeur classique de 2 au frein de la distance ( $\alpha$ ) (PUMAIN 2004), il faudra alors le re-soumettre à paramétrage pour adapter  $k$ . De même si l'on modifie les lieux et/ou les masses sur lesquels il s'applique.

### Illustration : le modèle de ségrégation de Schelling.

**Description** Le modèle de Schelling<sup>16</sup> est un modèle de simulation décrit par Thomas Schelling (SCHELLING 1971) qui vise à montrer comment un espace peut passer d'un état intégré à un état ségrégué ethniquement à travers une succession de comportements individuels de mobilité résidentielle. Il montre en particulier qu'on peut parvenir à un état ségrégué même quand les comportements individuels sont majoritairement tolérants.

L'espace du modèle est défini comme une grille carrée composée de  $N$ <sup>17</sup> cellules. Au début de la simulation, des agents de deux types, en proportions similaires, sont distribués aléatoirement dans l'espace du modèle. Chaque agent occupe une cellule, et le nombre total de cellules occupées (et donc d'agents) dépend d'un paramètre de densité  $d$  qui représente la part (entre 0% et 100%) du nombre de cellule qui sera occupé<sup>18</sup>. Chaque agent est défini par une satisfaction. Celle-ci correspond à la part de cellules voisines occupées par des agents « étrangers », c'est-à-dire d'un autre type : plus le voisinage contient d'étrangers, moins l'agent est satisfait. Si cette satisfaction est inférieure à un paramètre de tolérance  $S$ , l'agent est considéré comme insatisfait et se déplace aléatoirement dans une cellule non occupée. A terme, Schelling montre que même en considérant des valeurs de  $S$  assez élevées, c'est-à-dire un comportement plutôt tolérant, la succession de choix individuels entraîne la mise en place d'une distribution spatiale très ségréguée. On considère que le modèle a convergé quand l'ensemble des agents sont satisfaits, ce que ne permettent pas toutes les combinaisons de valeurs de paramètres.

Dans ce modèle, on peut donc dénoter trois paramètres,  $N$ ,  $d$  et  $S$ . Les deux premiers sont des paramètres techniques (encadré 3.2), en ce que, purement théoriques, ils ne se réfèrent à aucune quantité empirique, ni ne peuvent être utilisés afin de se référer à un ordre de grandeur connu qui en ferait des paramètres commensurables.  $S$  est un paramètre empirique. Bien que le modèle soit théorique, ce paramètre se réfère toutefois à un comportement ayant un sens empirique fort.

15. C'est le terme employé par (PUMAIN et SAINT-JULIEN 2001), que l'on peut ici retenir comme équivalent de paramétrer. François Durand-Dastès (DURAND-DASTÈS 1995, p. 298) utilise pour sa part le terme de calibration à propos du même modèle.

16. Une description plus poussée accompagnée d'une description de l'exploration du modèle peut être lue dans (DAUDÉ et LANGLOIS 2006)

17. Nous reprenons ici la notation  $M(N, d, n, S)$  proposée par (DAUDÉ et LANGLOIS 2006, p. 433) en n'explicitant toutefois pas le paramètre  $n$  décrivant le type de voisinage (4 ou 8) utilisé.

18. Il y a donc  $N \times d$  agents, et donc  $\frac{N \times d}{2}$  agents de chaque type.

**Objectif** Le paramétrage de ce modèle consiste à fixer un  $N$  constant et à faire varier  $d$  et  $S$  afin d'en trouver des valeurs permettant la convergence vers une situation stable. Quand  $d$  est très faible (30% dans l'exemple de (DAUDÉ et LANGLOIS 2006)), quelles que soient les valeurs de  $S$ , le modèle converge rapidement : quand l'espace du modèle dispose d'une faible densité d'agents, il est facile pour ceux-là de créer des agrégats homogènes distants d'agrégats de l'autre type d'agents. Quand  $d$  est plus important ( $\geq 66\%$ ), l'espace disponible étant limité, toutes les valeurs de  $S$  ne permettent pas la convergence du modèle. Le paramétrage du modèle aura donc pour objectif de trouver la valeur maximale possible – pour un  $N$  et un  $d$  donné – que peut prendre le paramètre de tolérance  $S$  tout en laissant le modèle converger.

**Paramétrage** Contrairement au modèle gravitaire, le modèle de Schelling est stochastique : les agents se déplacent aléatoirement quand non satisfaits. Dès lors, deux exécutions du modèle avec le même jeu de paramètres n'entraîneront pas forcément la même configuration spatiale. Qui plus est, pour certains jeux de paramètres, seules certaines exécutions convergeront. Le paramétrage de ce modèle ressemble donc à celui qui est réalisé pour le modèle SimFeodal. La manière « traditionnelle »<sup>19</sup> consiste à fixer un  $d$ , puis à essayer d'augmenter le  $S$  tant que le modèle converge. Si le modèle converge pour tout  $S$ , on augmente la valeur de  $d$  et on recommence à chercher la valeur maximale possible pour  $S$ . De part la nature stochastique du modèle, chaque jeu de paramètres doit être simulé plusieurs fois, le nombre de ces répliques dépendant de la part d'aléa dans le comportement du modèle. Un modèle de Schelling calibré, c'est-à-dire dont le paramétrage est achevé, donnera pour un  $N$  fixé les valeurs maximales de  $d$  et de  $S$  atteignables.

A travers ces deux exemples ayant traits à des méthodes différentes, on retrouve deux approches de paramétrage que l'on souhaite ici voire confondues : dans le premier cas, le paramétrage du modèle gravitaire tend à sa calibration, à la recherche d'une solution optimale, c'est-à-dire meilleure que toute autre. Pour un phénomène donné (un jeu de données précis par exemple) et un formalisme donné (l'expression la plus simple du modèle, ici sans contrainte), seule un couple de paramètres  $\alpha$  et  $k$  permet ainsi d'obtenir des flux modélisés proches des flux observés. Dans le cas du modèle de Schelling, il n'y a pas de recherche d'optimum : le paramétrage permet de comprendre le modèle en en définissant les limites en matière de convergence. Une configuration de paramètres  $S$  et  $d$  ne sera pas meilleure qu'une autre, mais pour un  $S$  donné, on saura quel est la valeur maximale de  $d$  permettant cette convergence (et réciproquement).

### Paramétrer un ou des modèles ?

Cottineau, Reuillon et Rey ont montré qu'on construisait plus souvent des familles de modèles que des modèles uniques. Ces auteurs plaident pour une construction modulaire de modèles, chacun des mécanismes et complexifications devant être autonomes afin de pouvoir les assembler en autant de combinaisons que nécessaire à une exploration d'ensemble de leurs interactions.

Nous reprendrons de leur discours une vision moins ambitieuse, en considérant (ref Rey) simplement qu'en fait d'un modèle, la modélisation passe par plusieurs modèles, ayant chacun différentes implémentations d'un mécanisme conceptuellement identique et chacun se devant dès lors d'être adapté par un paramétrage dédié.

Lena : trop endogame, prendre exemples modèles modulaires dans JASSS

à chercher il en parlait à sa soutenance je crois

19. C'est-à-dire manuelle, au contraire de méthodes automatiques plus rigoureuses et récentes.

Pour illustrer ce propos, on peut prendre l'exemple d'une modification qui a eu lieu sur notre modèle au cours de son développement. On considérait jusque là que pour prendre en compte la population rurale (hors Tours donc), 1000 foyers paysans donnaient une bonne idée de la situation en 800 pour l'espace considéré. Lors d'une réunion, les thématiciens se sont rendus compte que ce nombre sous-estimait très largement la population réelle, et qu'il valait mieux passer ce paramètre à 4000 foyers paysans. Les autres paramètres, fixés en bonne partie empiriquement, n'avaient visiblement aucune raison d'évoluer du fait de ce changement. Les mécanismes et effets de seuil étaient en effet conçus de manière à être relatifs aux masses manipulées, et le changement attendu était une augmentation linéaire des indicateurs de sortie.

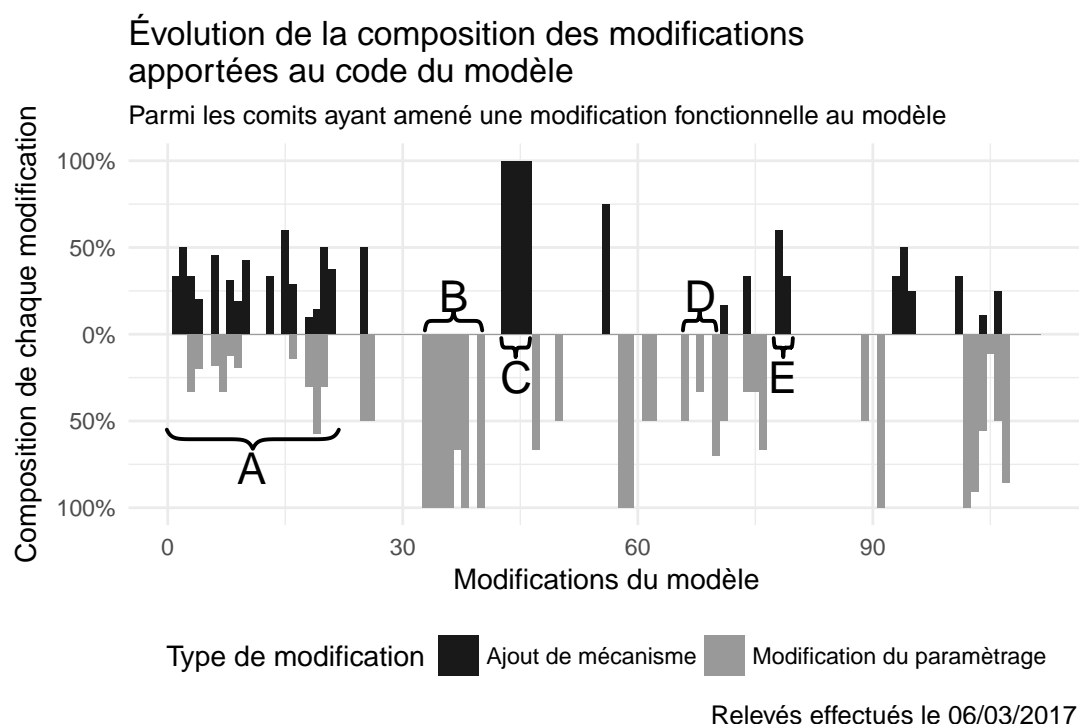


FIGURE 3.4 – Temporalité du paramétrage du modèle.

Chaque « enregistrement » (*comit*) correspond à une version modifiée et sauvegardée du modèle, contenant un ou plusieurs changements. On a ici rapporté le nombre de changements de chaque type (modification de mécanisme ou de valeur de paramètre) au nombre total de changements de chaque *comit* afin de figurer l'évolution des types de modification au cours de la construction du modèle.

Dans les faits, cette légère modification a entraîné une obligation de repenser la quasi-totalité des autres paramètres et d'ajuster une bonne part des règles. La concentration des foyers paysans était en effet bien trop rapide avec autant d'individus, et le modèle convergeait en quelques pas de temps vers une configuration presque statique et très concentrée. Tous les mécanismes de régulation de ce comportement étaient dès lors rendus ineptes, et il a fallu changer en profondeur la manière dont chaque paramètre et mécanisme interagissait avec les autres. C'est un changement majeur, que l'on peut constater dans la figure 3.4 (1ère modification de la période B), ayant donc entraîné des adaptations de tous les plans du modèle (paramètres en B, mécanismes en C). La première version avait été paramétré aussi correctement que possible, mais ce paramétrage était entièrement à refaire avec la nouvelle version (modifications de la période B, fig. 3.4). On pourrait dès lors différencier le modèle pré-existant et celui qui a suivi ce changement, et les considérer comme deux modèles puisque réagissant de

Lena : Là on ne suit pas, mais ce sera explicité dans le chap. 2

Lena : à discuter plus profondément



manière extrêmement différente.

Cet exemple renforce le caractère nécessaire du paramétrage, et qui plus est, de la continuité de cette étape, qui ne peut être pensée que comme une calibration finale d'un modèle abouti, auquel on ne peut en fait parvenir que par des paramétrages réguliers de modèles successifs moins aboutis.

Lena : Pas clair et pas assez logique

## Désambiguïsation

De nombreux termes sont utilisés dans la littérature, souvent sans réelle distinction, pour désigner cette opération qui consiste à choisir un jeu de paramètres pour un modèle. Pêle-mêle, on y retrouve le paramétrage, la validation, l'évaluation ou encore la calibration. Nous définissons dans l'encadré 3.3 le sens donné à chacun de ces termes dans le cadre de ce manuscrit.

### *Encadré 3.3 : Calibration, évaluation, validation...*

#### Calibration

On réserve souvent, et nous nous y tiendrons, ce terme à la dernière étape dans l'aboutissement d'un modèle. Une fois les mécanismes fixés et des objectifs définis, on peut procéder à la calibration, c'est-à-dire à une exploration de l'espace des paramètres ayant pour but de stabiliser les paramètres afin de se rapprocher autant que possible de ces objectifs. Cette étape, quelques soient les moyens employés, s'approche de la résolution utilisée dans le cadre de systèmes d'équations. Le contexte des systèmes complexes, et donc d'une non-linéarité des effets des paramètres, rend toutefois difficile l'obtention d'une unique configuration de **paramètres optimale**<sup>19</sup>, et la calibration aboutit donc souvent à un ensemble de configurations possibles, constituant par exemple un optimum de Pareto (**Trouver ref**).

#### Évaluation

On emploie majoritairement ce mot pour décrire les méthodes permettant de comprendre le comportement du modèle et sa réaction aux différents paramètres. Là où la calibration cherche une configuration de paramètres optimale, l'évaluation tend surtout à caractériser la stabilité du modèle face à l'aléa ou à des configurations exceptionnelles. On vise ainsi à s'assurer que le modèle reproduise **les faits stylisés**<sup>19</sup> voulus quelque soient son réglage, ou au moins à quantifier les intervalles de paramètres qui y satisfont.

#### Validation

La validation (refs Seb) tire son origine de sciences plus nomothétiques, et correspond donc à la démonstration qu'un modèle reproduit correctement ce qu'il représente. Au delà de l'évaluation, le terme amène une logique de preuve formelle que le modèle réagit bien ainsi et pas autrement quelque soient les conditions d'exécution. Cette démonstration formelle peut être effectuée sur des modèles à faible nombre de paramètres et mécanismes, mais le terme n'est jamais (à vérifier) employé dès lors que les modèles se complexifient.<sup>19</sup>

19. Lena : il faut partir d'une def plus basique en intro

19. Lena : pas introduits avant

19. Lena : cf. JASSS : Faire un rappel du sens classique, par ex. en stats.



### 3.1.5 Comment paramétrer ?

#### Visual validation

(Trouver ref dans thèse Clémentine, sans doute Hermann encore)

#### Indicateurs

#### L'importance de la réplication

## 3.2 Paramétrage et ajustement du modèle Sim-Feodal

A changer : le chap 2 présentera la version 0 (Base) du modèle, et on vient donc ici montrer comment on a répondu aux limites identifiées à la fin du chapitre, cf. les conclusions/résultats du chapitre TransMonDyn.

Comme préconisé ci-dessus, le modèle SimFeodal a été paramétré tout au long de sa construction. Dans cette partie, nous reviendrons en détail sur chacune des étapes de ce paramétrage, en retraçant les choix et questionnements qui ont orienté ces modifications du modèle. L'énumération est en ordre chronologique, et l'on a essayé d'organiser ces étapes en grands blocs-type de paramétrisation. Pour autant, le paramétrage et la construction d'un modèle, et du notre en particulier, sont un travail constant d'allers-retours entre l'identification et la résolution de problèmes. La structure réelle de développement, que l'on peut retrouver dans l'historique des modifications du modèle (les « *commits* »), ne correspond donc pas exactement à la chronologie organisée qui suit.

Les deux § suivants ne fonctionnent plus, vu qu'ils ne sont plus suivis du tableau récapitulatif des étapes.

Le tableau suivant (table 3.1) se veut être un point de repère autant qu'une table des matières synthétique pour la compréhension des étapes qui suivent. Si la description de chacune de celles-ci revêt une forme assez factuelle et systématique, nous pensons toutefois que l'explicitation de chacune de ces étapes permet de saisir la démarche proposée ainsi que l'intérêt de celle-ci.

Notons que l'étape 0 ici présentée correspond à une première version « exploitable » du modèle, c'est-à-dire contenant *a minima* l'ensemble des agents et mécanismes identifiés dans un premier temps par l'équipe de modélisation (TANNIER et al. 2014). Nous ne détaillerons pas ici les différentes phases de construction précédentes. Nous reviendrons toutefois sur plusieurs d'entre elles, en guise d'exemples illustratifs, dans les chapitres 6 et 7 (Partie 3).

### 3.2.1 SimFeodal : pourquoi et quoi paramétrer ?

Description des « résultats » du modèle à l'étape v0, (c'est-à-dire reprise et complétion du chap 11 de TMD).

==> Point par point, qu'est-ce qu'on observe, qu'est-ce qui va et qu'est-ce qui ne va pas.

Pour transition avec : sur quels leviers jouer pour améliorer/paramétrer le modèle ?

Ajuster le chapeau introductif pour prendre en compte tout ce qui vient après, par exemple avec :

1. Indices empiriques
2. Indicateurs
3. Hiérarchie des indicateurs

Le modèle SimFeodal présenté dans le chapitre 2 correspond à une « version 0 » du modèle souhaité, c'est-à-dire qu'il en constitue une première pré-version – implémentant l'ensemble des mécanismes décidés dans le modèle conceptuel –, mais pas encore stabilisée dans les liens, interactions et valeurs de paramètres de ceux-ci. Il ne répond donc pas nécessairement aux attentes que l'on peut en avoir.

On peut caractériser ces dernière, de manière très générale, comme permettant de reproduire les grands traits de cette transition. La reproduction d'un fait stylisé peut s'entendre de multiples manières, tout au long du gradient qualitatif-quantitatif. Dans notre cas, en l'absence d'une connaissance quantitative totale des phénomènes mobilisés et observés, on peut s'appuyer sur deux grands ensembles d'« indices ».

#### Ordres de grandeur

Le premier est constitué d'**ordres de grandeurs** empiriques estimés – avec une précision plus ou moins importante (cf. tableau 3 p. 317 du chap. TMD, à reproduire dans chap 2). Certaines valeurs sont ainsi connues, que ce soit d'après des sources primaires ou secondaires, et peuvent ainsi constituer des « indicateurs ». Par exemple, on connaît avec quasi-certitude le nombre d'églises paroissiales de la région Touraine en 1100. Certaines de ces valeurs sont toutefois plus proches de l'estimation, comme le taux de foyers paysans isolés en fin de période, que l'on ne peut qu'estimer à partir de sources secondaires et en menant des extrapolations. Ces différents indicateurs observés ou estimés peuvent cependant tous trouver une correspondance presque exacte dans le modèle de simulation, et dès lors, mener à comparaison entre données observées/estimées et données simulées. Ils peuvent ainsi jouer le rôle d'éléments d'évaluation du comportement du modèle simulé.

#### Faits et formes stylisés

Le second type d'indice est moins précis, ne reposant pas sur une valeur estimable, mais plutôt sur la connaissance experte d'un phénomène. Il s'agit des « **faits stylisés** »<sup>20</sup>, plus tendanciels que les indicateurs. On fait un large usage de ces faits stylisés en économie, mais aussi en géographie, par exemple quand on estime que les systèmes de peuplement tendent à se hiérarchiser, la pente de leur rang-taille tendant vers 1 à mesure que le système perdure (Trouver ref, sans doute Pumain/Saint-Julien). De la même manière, le modèle de transition démographique d'Adolphe Landry est un fait stylisé, énoncé à partir de l'observation

de nombreuses récurrences de l'évolution des populations d'un pays en fonction des taux de natalité et de mortalité. Ces exemples montrent qu'au sein des faits stylisés, on retrouve aussi une diversité quant à la précision de leurs énoncés. Dans notre cas d'étude, on retiendra cette logique de faits stylisés concernant des allures de courbes (par exemple l'évolution dans le temps d'un indicateur) ou de répartition spatiale. On pourra dès lors parler de « **formes stylisées** », aussi bien temporelles – courbe logistique estimée pour la polarisation des foyers paysans –, spatiales – différence dans l'occupation de l'espace par les agrégats entre le début et la fin de la période –, que correspondant à une agrégation à l'échelle du système, comme dans l'observation des hiérarchies grâce aux courbes rangs-tailles.

### Indicateurs en sortie

Les ordres de grandeur et formes stylisées évoquées relèvent du domaine empirique. Afin de pouvoir les mobiliser, il est nécessaire de définir des **indicateurs en sortie** dans le modèle de simulation, c'est-à-dire des variables informatiques que l'on enregistrera durant l'exécution du modèle et que l'on pourra alors comparer aux indices empiriques définis. On pourrait définir une typologie différenciant les indicateurs en sortie, de formes numériques simples (des nombres), et des indicateurs plus complexes, multidimensionnels, à même de permettre la confrontation du modèle de simulation avec les formes stylisées identifiées. Ici, on se contentera de regrouper ces différents types de production informatiques du modèles sous la dénomination d'« indicateurs en sortie », ou, plus simplement, d'**indicateurs**.

### Hiérarchiser les indicateurs

SimFeodal s'appuie sur une dizaine d'indicateurs numériques, ainsi que sur plus d'une trentaine d'indicateurs multidimensionnels. Tous ces indicateurs ne présentent pas le même degré de certitude, la même échelle d'observation, et surtout, le même niveau de précision sur les phénomènes modélisés. A chaque changement dans le modèle, pour une évaluation complète de la capacité de cette version à reproduire les indices empiriques, il faudrait donc observer et analyser chacun de ces nombreux et divers indicateurs. Dans le contexte du paramétrage d'un modèle s'appuyant sur une logique itérative et incrémentielle (voir encadré 3.4), on imagine bien que cela n'est pas possible : le nombre d'indicateurs est bien trop élevé pour avoir rapidement une vision globale de la qualité de représentation du modèle. Il faut dès lors, comme pour toute analyse synthétique, concevoir une hiérarchie d'observation et d'utilisation des indicateurs : il ne sera pas nécessaire d'analyser chacun des indicateurs dans la plupart des cas, seuls les indicateurs jugés plus importants pourront être analysés. Les indicateurs de moindre importance ne seront mobilisés que pour départager des situations dont la différence ne serait pas suffisamment explicitée par l'usage des indicateurs principaux.

---

20. Définis ainsi par (LIVET, PHAN et SANDERS 2014) : « Un “fait stylisé” est une présentation simplifiée (i.e. taux, ratio ou écart, structure spatiale) d'une régularité empirique sur l'observation de laquelle il y a un large accord. Le terme a été popularisé en économie par Nicholas Kaldor (1961). [Les] faits stylisés peuvent être construits de la manière suivante : 1) en partant du domaine empirique, on identifie des relations saillantes ; 2) on opère quelques simplifications qui permettent d'inclure formellement ces relations dans des modèles ; 3) une fois admis que ces simplifications ne faussent pas trop les choses, on érige ces relations à la fois simplificatrices et formalisables au rang de “ faits stylisés”, dont les concepts théoriques doivent rendre compte. »

**Incertitude** Dans le modèle de simulation, les indicateurs sont à analyser en tenant compte de la précision du fait stylisé qu'ils représentent. Il ne faudra ainsi pas étudier la croissance du nombre d'agrégats au cours de la simulation de manière fine, par exemple en étudiant le coefficient directeur de la courbe, quand l'empirie ne donne quasiment aucune information à ce sujet si ce n'est qu'il y a bien plus d'agrégats (cette quantité étant elle-même analysable avec plus de précision) en fin de période qu'au début. On peut vouloir quantifier la précision de ces données, par exemple à l'aide des méthodes développées dans le champ des observations floues et/ou incertaines (voir par exemple le travail de Cyril de Runz sur les données « imparfaites » (DE RUNZ 2008)). Cette quantification de l'incertitude pourrait alors servir de base à l'établissement d'une hiérarchie des indicateurs : on analyserait en premier lieu l'écart entre les ordres de grandeurs et les indicateurs qui les représentent, avant d'étudier les indicateurs plus incertains (augmentation de la charge fiscale entre 800 et 1100 par exemple) et de finir par les formes stylisées et leur représentations dans le modèle. Toutefois, SimFeodal se caractérise d'une part par une très forte hétérogénéité dans les niveaux de connaissance des ordres de grandeurs et faits stylisés modélisés, et d'autre part, se voulant un modèle théorique (**A dire spécifiquement dans le chapitre 2 ; y faire une ref ici**), il n'y a pas de volonté de « coller aux données » : la vraisemblance d'ensemble du modèle compte bien plus que la précision de chacune de ses composantes.

**Indicateur synthétique** Une autre piste, fréquemment utilisée dans les modèles théoriques, consiste à créer un indicateur synthétique – ou un faible nombre d'entre eux –, par exemple une moyenne de plusieurs indicateurs. Ces indicateurs synthétiques résultent d'une quantification des autres indicateurs (excluant donc les formes stylisées qui sont plus libres d'interprétation), et permettent en particulier de mener des explorations rapides et simples du comportement d'un modèle selon les valeurs de paramètres. SimFeodal n'est pas adapté à de tels indicateurs : une large partie des faits stylisés et ordres de grandeur mobilisés proviennent de connaissances expertes, et les thématiciens qui les ont consolidées rechignent à créer de tels indicateurs composites, en ce que cela demande de pondérer précisément l'importance de chacun des indicateurs par rapport aux autres. Si l'on pourrait créer quelques indicateurs synthétiques, ceux-là ne prendraient en compte qu'une faible proportion du comportement attendu du modèle, résultant en une forte perte de pouvoir explicatif ou descriptif du modèle.

**Définir des dimensions d'analyse** En présence de plus d'une quarantaine d'indicateurs, il est toutefois nécessaire, a minima, d'organiser leur analyse. On a vu qu'il n'était pas justifié de mener cet ordonnancement à partir des propriétés intrinsèques des indicateurs du modèle. Au contraire, et cela porte bien plus de sens vis-à-vis du rôle d'un modèle, la hiérarchisation des sorties du modèle doit suivre la hiérarchie implicite qui structure les hypothèses et objectifs du modèle en lui-même. Ces hypothèses et objectifs sont multiples dans SimFeodal, et dès lors, une hiérarchie ne pourra être établie qu'en leur sein. Il convient donc d'identifier des ensembles d'indices permettant de caractériser les différentes dimensions à analyser, avant de définir, dans chacune de ces dimensions, une hiérarchie d'indicateurs. Dans le chapitre précédent (**ref chap 2**), nous présentions les principales dynamiques du modèle : (1) polarisation, (2) hiérarchisation et (3) fixation des foyers paysans. En postulant que ces dynamiques sont caractéristiques du modèle, on peut s'appuyer sur cette triade pour caractériser les sorties du modèle, c'est-à-dire mener la confrontation entre indices et indicateurs. Ces trois ensembles,

et les indicateurs qui les caractérisent dans le modèle, sont dès lors considérés comme les trois dimensions d'analyse des sorties de SimFeodal.

**Hiérarchiser dans les dimensions** Chacune de ces dimensions s'appliquent à plusieurs types d'agents du modèle. Pour définir notre hiérarchie, on retiendra les agents les plus impactés par les dynamiques correspondant à ces dimensions : la polarisation, par exemple, peut être observé depuis le point de vue de ce qui polarise (les attracteurs) tout autant que de ce qui est polarisé (les foyers paysans). On aura alors tendance à examiner d'abord un résultat agrégé, représentatif de la structure dans son ensemble, avant d'observer les indicateurs représentatifs des dynamiques ayant mené à cette structure. Dans cet exemple, on analysera donc d'abord le résultat effectif de la polarisation, c'est-à-dire la concentration des foyers paysans en agrégats, avant d'observer la répartition et les diversité des attracteurs ayant entraîné ce phénomène. On peut dès lors définir des « indicateurs principaux » pour chaque dimension, représentatifs des grands traits des structures auxquelles on souhaite aboutir en sortie de simulation, et des « indicateurs secondaires », permettant d'affiner l'évaluation de chacune de ces dimensions.

**Une hiérarchie mouvante** Notons que l'analyse des indicateurs de sortie suit une hiérarchie parfois mouvante, et en tous les cas, assez peu quantifiable : si l'ordre d'observation est plutôt stable, l'importance que l'on portera à chacun des indicateurs peut varier. Les indicateurs principaux de chaque dynamique sont ainsi « incontournables », c'est-à-dire qu'un résultat trop loin de celui des indices empiriques est disqualifiant. Parmi les indicateurs secondaires, il n'est pas toujours possible, d'après les connaissances des experts sur le sujet, d'établir une priorité ou une pondération de chaque indicateur. L'évaluation de la polarisation par exemple (page 23), se définit principalement par rapport à un indicateur principal – le taux de foyers paysans dispersés –, mais selon les résultats des autres indicateurs de sortie, on leur attribuera une importance variable. L'étude de la dispersion des agrégats et pôles peut ainsi se révéler plus importante que celle de l'évolution du nombre d'agrégats selon les paramètres que l'on souhaite ajuster, ou se montrer tout au moins plus différenciante selon l'état du paramétrage.

#### *Encadré 3.4 : Incrémentalité des indicateurs*

De la même manière que les paramètres et mécanismes d'un modèle de simulation tendent à évoluer<sup>21</sup> au cours du temps de la construction, souvent afin d'affiner un comportement observé, les indicateurs de sortie sont amenés à évoluer aussi.

Ainsi, en cas de modifications fines du modèle, il est fréquent que les indicateurs initialement choisis ne suffisent plus à départager des versions du modèle quant à un phénomène spécifique. Par exemple, quand on observe le phénomène de polarisation dans les sorties de SimFeodal, l'indicateur du nombre d'agrégats est extrêmement synthétique et informatif jusqu'à ce que l'objectif soit atteint ou que les modifications ne parviennent plus à le faire évoluer. À partir de ce moment, afin d'améliorer la vraisemblance de la situation simulée par le modèle, on peut se focaliser sur la distribution spatiale de ces agrégats, par exemple pour vérifier qu'ils sont bien répartis de manière homogène dans l'espace, et non trop concentrés.

L'observation de la répartition spatiale requiert certes de nouvelles analyses, mais surtout, par exemple, d'enregistrer les positions des agrégats au cours du temps. Si cet indicateur de sortie n'était pas utile avant cela, il n'y

avait aucun intérêt à l'enregistrer. Il faut donc adapter l'implémentation du modèle pour enregistrer ce nouvel indicateur. Dès lors, on pourra composer un nouvel indicateur synthétique, qui, dans cet exemple, pourrait prendre la forme d'un indice de concentration spatiale).

Ce procédé incrémental dans la construction des indicateurs est très fréquent, mais pose toutefois un problème majeur : sauf à adapter chacune des anciennes versions du modèle implémenté pour y ajouter l'enregistrement des nouveaux indicateurs nécessaires, on ne pourra rendre strictement comparable les sorties de toutes les itérations du modèles informatique. Et même alors, il faudrait ré-exécuter des répliques de chaque version du modèle implémenté à chaque ajout d'indicateur, quand bien même les indicateurs présent initialement étaient jugés suffisants. Un dernier obstacle est plus gênant : certains indicateurs sont spécifiques à des mécanismes, et en cas de changement de ces derniers, ils peuvent ne plus être calculables ou simplement comparables. Par exemple, des versions antérieures du modèle enregistreraient les comportements individuels des foyers paysans quant à leur « choix » de déplacement, selon qu'ils étaient à l'origine localisés dans un agrégat ou dispersés. Une simplification du modèle a abouti à la modification des règles différenciant les possibilités de déplacement : on n'observe plus si le foyer paysan est dans un agrégat, mais plutôt s'il est dans un agrégat doté d'un pôle d'attraction. Dès lors, les analyses basées sur les choix de déplacement des foyers paysans selon leur origine ne sont plus comparables avec celles des versions antérieures au changement dans le modèle, quels que soient les détails d'implémentation de ce dernier.

Ces éléments expliquent que dans les résultats de chaque étape du paramétrage du modèle, on ne présente pas systématiquement l'ensemble des indicateurs, y compris quand ceux-ci pourraient être plus pertinents que les indicateurs présentés.

---

21. De manière incrémentielle et itérative, voir dans le chapitre x? et [http ://itsadeliverything.com/revisiting-the-iterative-incremental-mona-lisa](http://itsadeliverything.com/revisiting-the-iterative-incremental-mona-lisa)

### Évaluer la polarisation des foyers paysans

La polarisation des foyers paysans dans l'espace du modèle est sans doute le fait stylisé principal parmi ceux que l'on cherche à reproduire. Rappelons ici que l'on estime, à partir des connaissances d'experts, que les foyers paysans sont très majoritairement dispersés en 800, et concentrés au sein de villages et petites villes en 1100. Le modèle cherche à reproduire cette polarisation, par le biais d'une concentration des foyers paysans, initialement localisés aléatoirement dans l'espace mais n'en parsemant qu'une faible part, en des agrégats de foyers paysans répartis dans une plus large partie de l'espace modélisé.

Pour analyser la polarisation du système de peuplement, il est nécessaire de définir des indices permettant de caractériser ce phénomène. Ces indices doivent d'une part avoir une logique thématique, c'est-à-dire être appropriés à la description et à l'étude de la polarisation, mais doivent pouvoir être produits et enregistrés dans le modèle de simulation, formant des indicateurs. En analysant les valeurs que ces indicateurs prennent en sortie de simulation, on pourra alors juger de la capacité du modèle à reproduire le phénomène de polarisation.

Pour l'étude de la polarisation, il est nécessaire de faire appel à des indices hétérogènes, chacun devant être en mesure de décrire les différents aspects du phénomène de polarisation. En conséquence, on a choisi de faire appel à plusieurs indicateurs qui doivent permettre d'étudier aussi bien l'aspect structurel du système simulé en son état final que la forme et la tendance que prennent les changements qu'il subit.

L'indicateur principal est le taux de dispersion des foyers paysans. Si celui-ci est trop important (c'est-à-dire très supérieur aux valeurs estimées empiriquement), cela signifie que la polarisation générée par le modèle est insuffisante, et dès lors, obligatoirement insatisfaisante. A contrario, une valeur trop faible serait symptomatique d'un emballement des mécanismes simulés, figeant la situation dans une concentration absolue des foyers paysans, ne laissant dès lors plus de place à la diversification des situations locales et de la hiérarchisation d'ensemble.

Pour affiner ce constat, on fait appel à d'autres indicateurs : le nombre d'agrégats, de pôles, ou encore la dispersion spatiale de ces deux types d'entités. Ces indicateurs ne permettent pas, à eux seuls, de caractériser le succès de la dynamique de polarisation modélisée, mais ils aident à affiner l'analyse de cette dynamique telle que produite par le modèle de simulation. Ils éclairent ainsi le phénomène de polarisation sous des angles légèrement différents, ayant plus pour objet de diagnostiquer les problèmes potentiels qui mèneraient à une mauvaise polarisation plutôt que de qualifier celle-ci. Par exemple, la dispersion des agrégats et pôles peut renseigner, une fois le taux de foyers paysans dispersé jugé trop important, sur une des raisons probables de ce résultat non satisfaisant. Il s'agit donc d'indicateurs secondaires, permettant de préciser la capacité du modèle à reproduire les faits stylisés, alors que l'évaluation de cette capacité est surtout le rôle de l'indicateur principal.

**Taux de foyers paysans dispersés** Cet indicateur, et sa déclinaison temporelle, sont vraisemblablement les plus évidents : plus le taux de foyers paysans dispersés en fin de simulation est faible, plus le système de peuplement est polarisé. On a vu<sup>22</sup> qu'empiriquement, autour de 1160, on observe environ 20% de foyers paysans encore dispersés, alors qu'ils sont près de 95% au départ. Le modèle sera donc d'autant plus satisfaisant que cet indicateur s'approchera de 20%

22. Ne pas oublier de présenter les objectifs dans chapitre 2. Ou alors, on les présente ici au fur et à mesure qu'on mentionne les indicateurs



en fin de simulation.

La « déclinaison temporelle » mentionnée juste au dessus permet d’affiner légèrement la précision de l’information communiquée par cet indicateur : il faut certes atteindre un objectif quantifié (20%), mais les hypothèses empiriques permettent aussi de penser qu’il faut que l’évolution de cet indicateur au cours du temps présente une tendance stable à la diminution, diminuant ainsi plus ou moins, avec de faibles fluctuations à chaque pas de temps. Une configuration de paramètres atteignant l’objectif en fin de simulation mais fluctuant fortement ne sera ainsi pas valide, faisant apparaître un phénomène trop aléatoire dont on ne peut penser qu’il s’est déroulé historiquement.

Dans la version 0 de SimFeodal, on atteint en moyenne (moyenne des répliques) **57%** de foyers paysans isolés en fin de simulation. C’est un taux bien trop important, illustrant dès lors une polarisation trop faible. Qui plus est, alors que cet indicateur est en diminution constante (fig. 3.5) pendant la première moitié de la période, on le voit se stabiliser puis augmenter, résultant ainsi en ce taux très éloigné de l’objectif.

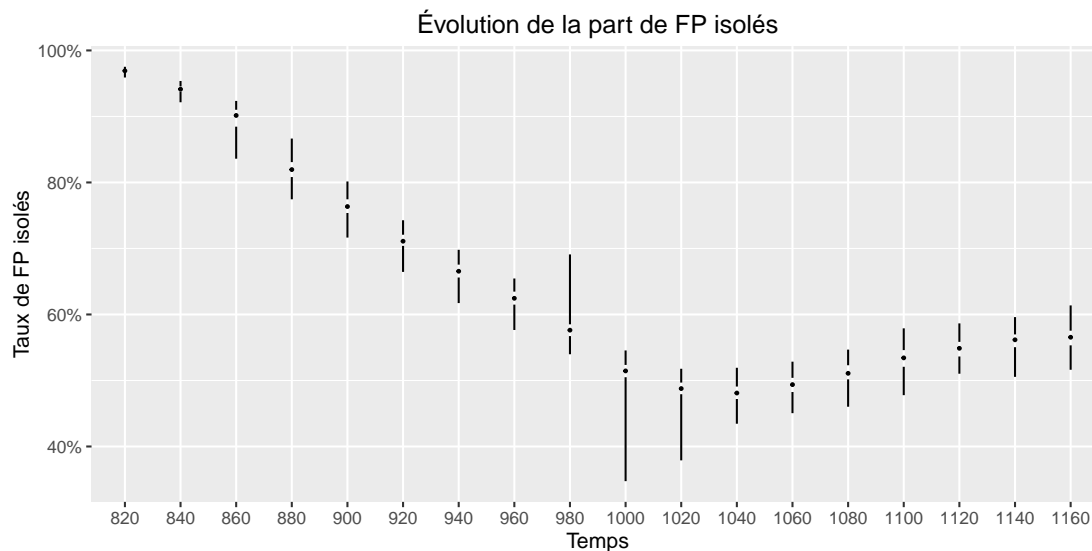


FIGURE 3.5 – Évolution de la part des foyers paysans isolés.

**Nombre d’agrégats** Puisque les foyers paysans se concentrent au sein d’agrégats, il est logique d’observer l’évolution de ces derniers. Là aussi, (**objectif à définir**), on peut considérer qu’un nombre d’agrégats en fin de simulation proche de l’objectif, 200, permet de caractériser une polarisation réussie. Cet indicateur ne peut être lu seul, et c’est pour cela qu’il vient dans un second temps : en effet, un faible nombre d’agrégats peut aussi bien être révélateur d’une très faible polarisation des foyers paysans (ceux-ci restant dispersés) que d’une trop importante (un unique agrégat concentrant l’ensemble des foyers paysans par exemple). Une fois le taux de foyers paysans dispersés connu et ces potentiels biais pris en compte, le nombre d’agrégats et son évolution nous renseigne cependant sur les dynamiques de polarisation. On s’attend ainsi à ce que le nombre d’agrégats, très faible au départ (24 dans la version 0), suive trois phases : une première phase de croissance lente, le temps que les mécanismes agissent sur la polarisation, suivie d’une période de croissance plus rapide, une fois que tous les foyers paysans commenceront à être suffisamment attirés par les pôles pour y former des agrégats, et enfin, une nouvelle phase de croissance plus lente, une fois les



foyers paysans répartis dans les agrégats existants et qui se déplaceront vers des agrégats plus importants, hiérarchisant le système de peuplement. Cette allure d'évolution rappelle les fonctions logistiques connues par exemple pour les cycles de diffusion/adoption des innovations [ref Pumain ou MN Comin](#), et résulte des connaissances expertes des archéologues spécialistes de la période.

Le nombre d'agrégats est assez satisfaisant dans la version 0 du modèle. Il s'élève ainsi à **187** en moyenne, cette dernière étant d'ailleurs stable au regard des répliques. L'écart à l'objectif (200) est donc assez minime. On a toutefois vu que le taux de foyers paysans isolé était bien trop important, et dès lors, ces agrégats sont logiquement composés de trop peu de foyers paysans.

L'observation de cet indicateur au cours du temps (fig. 3.6) est quant à elle assez satisfaisante : on retrouve bien les trois phases attendues, quand bien même le début de la croissance plus importante (vers 1000 ici) est un peu trop tardive. Ce moment coïncide de plus avec la stagnation puis l'inversion de la courbe d'évolution de la part de foyers paysans isolés.

On peut quand même considérer que si la polarisation n'est pas assez importante, les structures qui en résultent semblent correspondre à l'empirie.

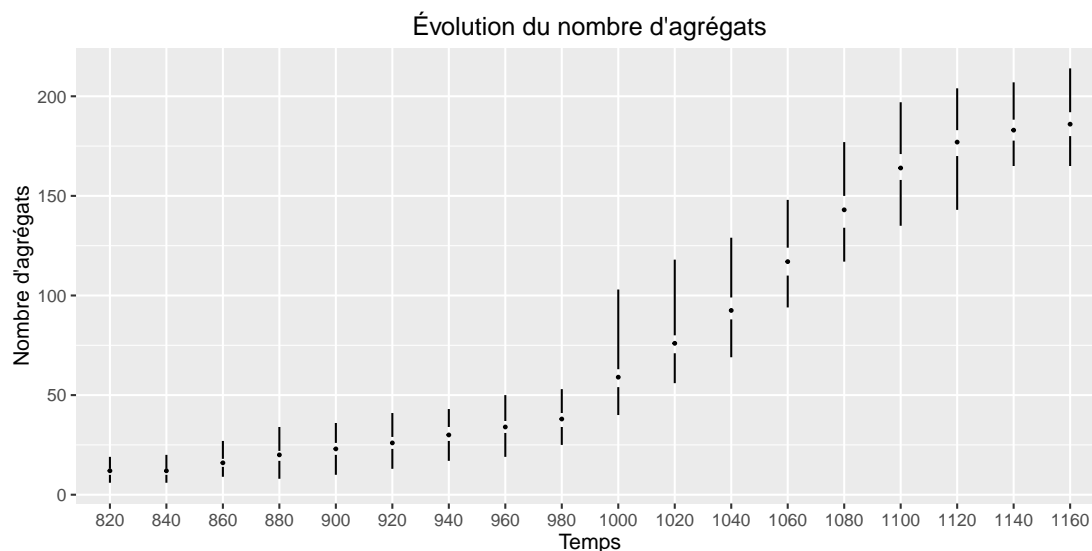


FIGURE 3.6 – Évolution du nombre d'agrégats.

Il y a un problème ici par rapport à ce qui a été mis dans le chapitre : dans le chapitre TMD, on mentionne 83 agrégats en moyenne, alors que dans l'analyse de Base, il y en a 187...

**Nombre de pôles** Dans le modèle SimFeodal, les foyers paysans sont polarisés par des pôles d'attraction. Pour une polarisation efficace, il est donc nécessaire que les pôles soient suffisamment nombreux, c'est-à-dire, *a minima*, autant que d'agrégats attendus (200). Contrairement aux agrégats, un nombre trop important de pôles ne constitue pas un problème : en considérant que 20% des foyers paysans doivent demeurer isolés en fin de simulation, il est vraisemblable qu'une partie des pôles, par exemple composés d'une église paroissiale, n'aient pas vocation à voir la constitution d'un agrégat autour d'eux.

Par ailleurs, afin de renforcer la polarisation par l'action de l'attraction différenciée – selon le nombre de foyers paysans contenus dans chaque agrégat<sup>23</sup> –, il faut que le taux de pôles contenant un agrégat soit important, et surtout croissant au cours du temps : comme dans l'empirie, cela est alors le marqueur que de petits pôles d'attractions, comme les églises paroissiales, parviennent à polariser suffisamment de foyers paysans de leur voisinage pour aboutir à la création d'un petit agrégat, un village par exemple.

Pour un résultat satisfaisant, il faut donc que le nombre de pôles augmente régulièrement au cours de la durée de la simulation, et que le taux de pôles contenant un agrégat augmente lui aussi de manière continue.

L'évolution du nombre de pôles est assez représentative de l'empirie dans la version 0 de SimFeodal. On aboutit ainsi sur **190** pôles en moyenne en fin de simulation, ce qui est dans le bon ordre de grandeur, quoi qu'un peu trop faible. Ce nombre est en croissance constante à partir de 940 (fig. 3.7). Ce départ quelque peu tardif peut expliquer le relatif manque de pôles en fin de simulation.

L'analyse de la constitution de ces pôles en matière d'agrégat est toutefois plus décevante : la croissance n'est pas linéaire, stagnant entre 940 et 1000, puis à partir de 1040. On constate ici un mauvais comportement du modèle par rapport à l'empirie : il y a création de nombreux pôles, mais ceux-ci ne suffisent pas à polariser les foyers paysans. On peut donc penser qu'il s'agit de nombreux pôles ruraux, par exemple composés d'une unique église paroissiale, ne suffisant dès lors pas à concentrer les foyers paysans alentours.

Dans cette dimension, comme pour l'indicateur principal qu'en est le taux de foyers paysans dispersé, l'effet de la polarisation apparaît nettement trop faible.

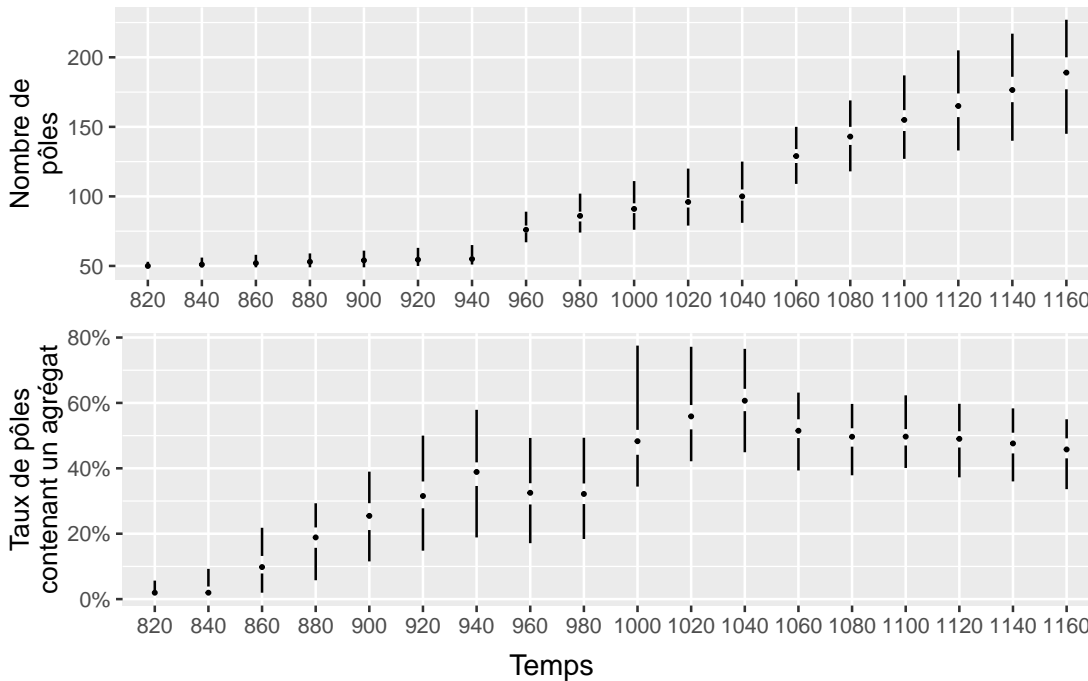


FIGURE 3.7 – Évolution du nombre de pôles et du taux de pôles contenant un agrégat.

23. L'attraction qu'exercent les pôles contenant des agrégats de population est ainsi plus forte lorsque ces derniers sont constitués d'un nombre de foyers paysans important, et moindre quand ce nombre est faible. Cela place ce mécanisme dans une logique proche de celle de l'attachement préférentiel identifié par (YULE 1925) et (SIMON 1955), cités par (SCHMITT 2014, p. 93)

**Dispersion des agrégats et pôles** Les indicateurs présentés ci-dessus avaient en commun d’être des indicateurs de sortie, générés automatiquement, et présentés sous forme d’agrégations des réplifications. Pour préciser l’analyse de la polarisation, il est toutefois nécessaire d’aborder une dimension qui n’est fondamentalement pas agrégable dans les indicateurs de sortie : l’espace du modèle. Comme celui-ci est théorique et aléatoire, il n’y a aucun sens à agréger des entités différentes, par exemple des agrégats, sachant que ceux-ci occupent des localisation aléatoires et ne sont pas identifiables en tant que tels.

On ne peut cependant pas se passer d’une analyse spatiale de la répartition des pôles et agrégats afin de comprendre les dynamiques effectivement simulées par le modèle. La distribution spatiale des agrégats et des pôles est en effet un facteur majeur de la polarisation : s’ils sont très concentrés, les foyers paysans non présents alentours ne trouveront pas d’attracteurs à proximité, et ne seront de plus pas particulièrement affectés par l’augmentation des droits (banaux etc.) et des contraintes spatiales (proximité à une église, à un château etc.). A l’inverse, des agrégats entièrement dispersés ne favoriseraient pas la structure spatiale hiérarchisée que l’on cherche à faire émerger.

Afin que le comportement du modèle soit satisfaisant, il faut donc que les pôles et agrégats occupent l’ensemble de l’espace du modèle, tout en présentant des zones de concentration relatives plus importantes. Comme on ne peut agréger les représentations spatiales, il convient, pour cette analyse, de regarder individuellement un échantillon de configurations spatiales générées.

La lecture des cartes de répartition des agrégats et des pôles de deux réplifications de la version 0 (fig. 3.8) va, dans l’ensemble, dans le sens de l’empirie. On peut ainsi remarquer que ces entités ont bien tendance, au cours du temps, à se disperser dans l’espace modélisé. En fin de simulation, tout l’espace est occupé, et on remarque même que certaines zones voient une forte concentration en agrégats, reproduisant les faits stylisés connus. La répartition spatiale des agrégats – et des pôles – est donc bonne, et ne semble pas opposer d’obstacle à la polarisation attendue du système de peuplement.

Ce résultat peut cependant être nuancé par l’observation de la dynamique de cette dispersion : on constate ainsi que la dispersion s’effectue rapidement (entre 800 et 940), et n’évolue plus vraiment après. Ce phénomène est donc plus rapide dans cette version 0 du modèle que dans les connaissances empiriques de la région Touraine.

On peut enfin remarquer, en préalable aux dynamiques de hiérarchisation du système que l’on s’apprête à étudier, que ces sorties illustrent un manque criant de hiérarchie quant à la composition des agrégats : on ne remarque, visuellement, que peu de différence entre les agrégats, et surtout, cette hiérarchie semble faiblir entre 1040 et 1160 (les agrégats les plus importants ont vu leur nombre de foyers paysans diminuer).

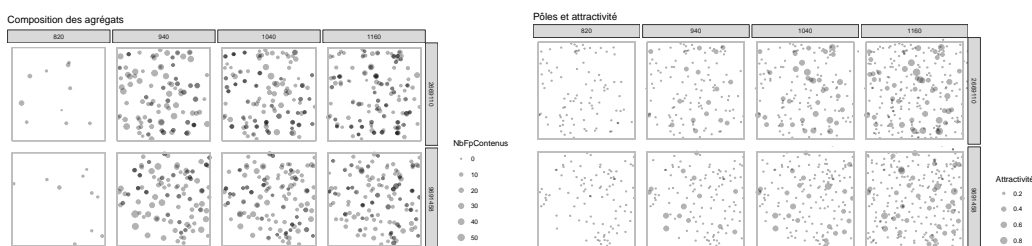


FIGURE 3.8 – Évolution de la répartition spatiale des agrégats et pôles, pour deux réplifications.

### Évaluer la hiérarchisation du système de peuplement

Le second grand fait stylisé que l'on cherche à reproduire dans SimFeodal correspond à une forte hiérarchisation du système de peuplement et plus généralement, des entités présentes. On passe ainsi, en 800, d'un habitat dispersé dans lequel coexistent quelques agrégats de taille uniforme, à un habitat concentré dans des agrégats de taille très hétérogènes. La distribution des tailles des agrégats est estimée par les connaissances expertes, toutes proportions gardées, comme assez proche des distributions observées aujourd'hui dans les systèmes de peuplement. On souhaite ainsi que les agrégats modélisés suivent une distribution approchant la distribution log-normale.

Par extension, et là encore d'après les connaissances thématiques, l'ensemble des entités doit aussi suivre le même type de forme. Par exemple, les pôles, tant en terme d'attractivité que de composition, doivent aussi montrer une hiérarchie du même ordre, ainsi que les seigneurs – à travers leur puissance, au moins pour les petits seigneurs –, ou encore les paroisses, par le nombre de paroissiens qu'elles desservent.

Comme pour l'étude de la polarisation, on peut définir un indicateur principal de cette hiérarchisation du système de peuplement : la forme de la distribution de la composition en foyers paysans des agrégats.

De la même manière que pour la polarisation, les indicateurs secondaires ont aussi pour but de préciser cet indicateur principal, et en particulier d'analyser les moteurs de cette hiérarchisation du peuplement. On a en effet choisi d'observer plutôt la hiérarchisation des autres types d'entités – pôles, seigneurs, paroisses –, pour vérifier qu'elles accompagnent et/ou entraînent bien la hiérarchisation des agrégats. La hiérarchie des pôles, par exemple, a une influence directe sur l'attraction effectuée sur les foyers paysans (polarisation) et sur la hiérarchisation des agrégats : par effet d'attraction différenciée (voir la note de bas de page 23 page 26), des agrégats plus importants se constituent autour des pôles les plus importants.

Comme pour la polarisation, l'analyse de la capacité du modèle à reproduire la hiérarchisation du système de peuplement se fait donc en deux temps : en premier lieu, on évalue cette capacité à l'aide de l'indicateur principal, puis on précise cette qualification et on essaie de l'expliquer à l'aide des indicateurs secondaires.

**Hiérarchie des agrégats** L'indicateur principal est un indicateur agrégé, correspondant à la forme de la distribution des agrégats mesurés par le nombre de foyers paysans qui les composent. Cet indicateur est classique dans l'analyse des systèmes de peuplement, et il est courant de l'observer par le biais d'un indicateur agrégé simple, correspondant à la loi rang-taille. On observe pour cela le modèle statistique, ou sa représentation graphique tout du moins, mettant en relation le logarithme de la taille des individus (le nombre de foyers paysans composant chaque agrégat ici) et le logarithme du rang de cet individu. Comme pour toute régression linéaire, on peut alors quantifier l'ajustement du modèle grâce au coefficient de détermination ( $R^2$ ), et spécifier la pente de la courbe, représentant le degré de hiérarchie, à travers le coefficient directeur ( $a$  dans la formule  $y = ax + b$ ).

Dans le cas de SimFeodal, le faible nombre d'agrégats ainsi que la variabilité de leurs tailles rend difficile cette analyse quantifiée, le coefficient directeur, par exemple, étant très sensible aux faibles effectifs. On utilise toutefois la représentation graphique décrite comme un indicateur majeur de la hiérarchie des agrégats.

Du point de vue des connaissances empiriques, la courbe doit ainsi voir sa

pente augmenter avec le temps, tout en devenant plus convexe, ce qui représente la « longue traîne » des petits agrégats, empiriquement observée dans toutes les distributions de systèmes de peuplement.

Avec une autre représentation graphique du même phénomène, en discrétisant les agrégats selon leur taille et en dénombrant le nombre d'agrégats de chaque classe, on peut aussi avoir une vision plus synthétique (car moins exhaustive) de la forme de la distribution. On doit alors obtenir un nombre décroissant d'agrégats à mesure que la classe représente un nombre élevé de foyers paysans.

Ça vaudrait peut-être le coup, pour chaque indicateur présenté sous forme graphique (évolution ou forme), de faire un graphique théorique (une sorte de courbe parfaite) de ce que l'on souhaite observer.

La version 0 de SimFeodal présente une hiérarchisation des agrégats nettement trop faible. La courbe rang-taille (fig. 3.9) est ainsi trop faiblement pentue, et présente une forme trop linéaire : la convexité due à la longue traîne n'est pas assez visible, en raison sans doute de trop faibles valeurs en haut de la hiérarchie, qui ne « tire » alors pas assez la distribution. Ce résultat de simulation est d'autant plus perturbant que son évolution est éloignée de l'empirie : on remarque en effet que la distribution se hiérarchise bien entre 800 et 1020, présentant à cette date une allure très satisfaisante. Pourtant, après cette période, les agrégats voient leur hiérarchie diminuer nettement et les agrégats les plus peuplés diminuer en taille.

On constate le même décrochage dans la discrétisation des agrégats (fig. 3.10), où on remarque de plus que la hiérarchie la plus proche de l'attendu, où la position des classes suivrait une ligne droite, semble se dessiner entre 940 et 1040. En 1040, et plus encore en 1160, la proportion d'agrégats de taille moyenne et haute (de 30 à 100, et de plus de 100 foyers paysans) est trop faible, et pas assez hiérarchisée.

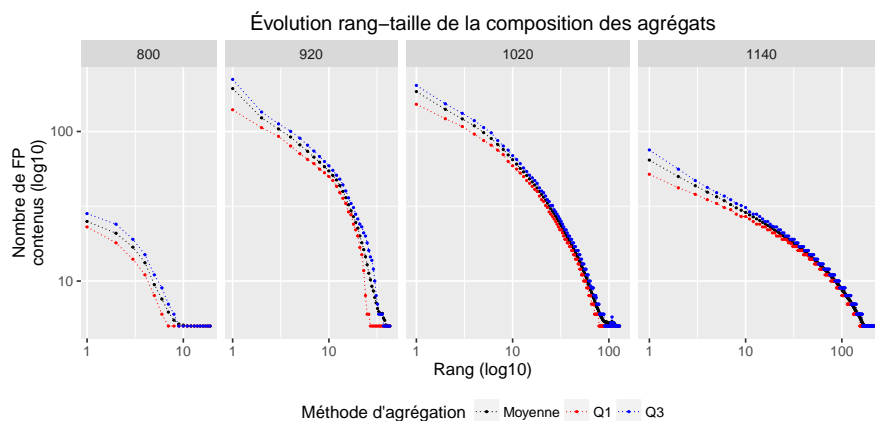


FIGURE 3.9 – Évolution de la courbe rang-taille des agrégats.

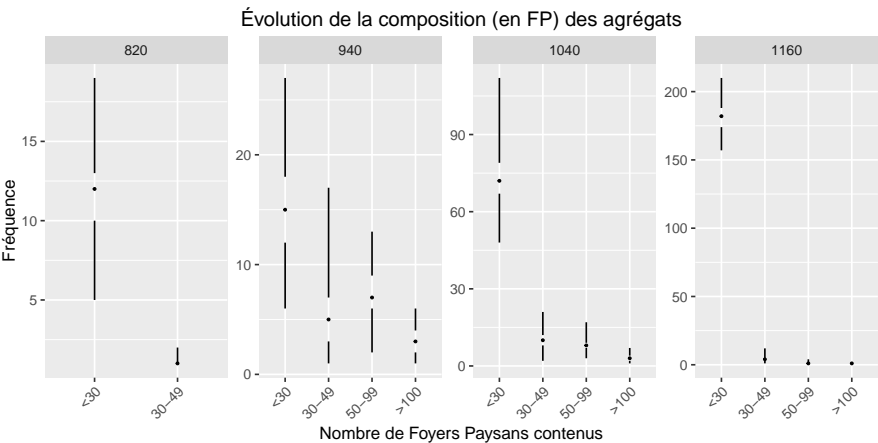
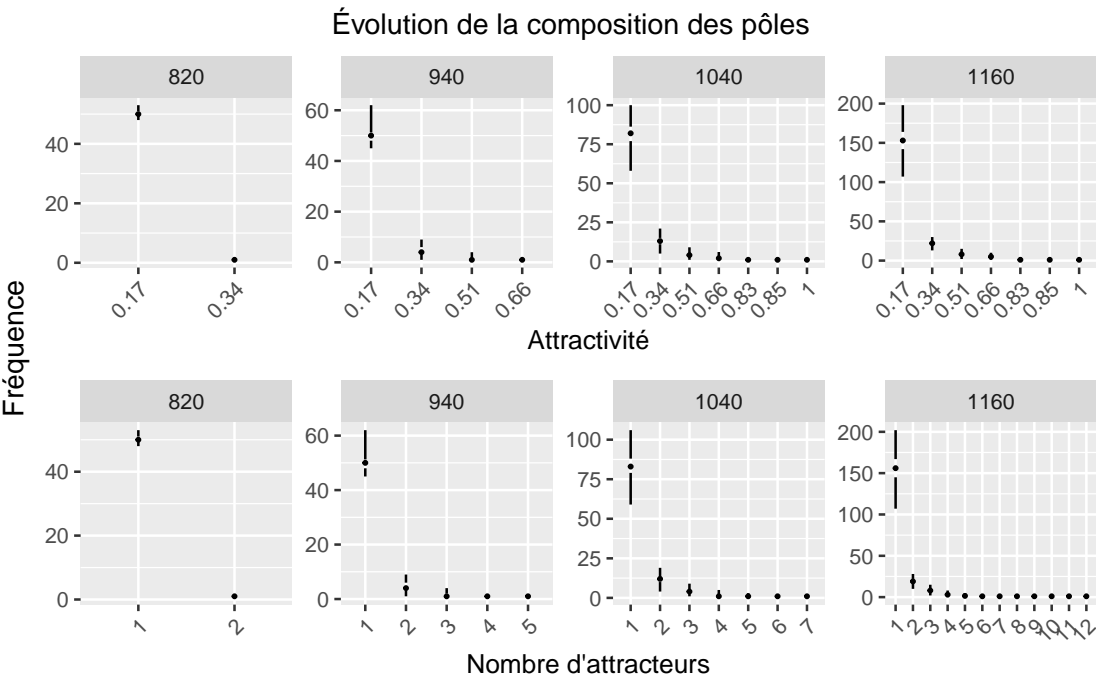


FIGURE 3.10 – Évolution de la composition des agrégats.

Hiérarchie des pôles

(fig. 3.11)



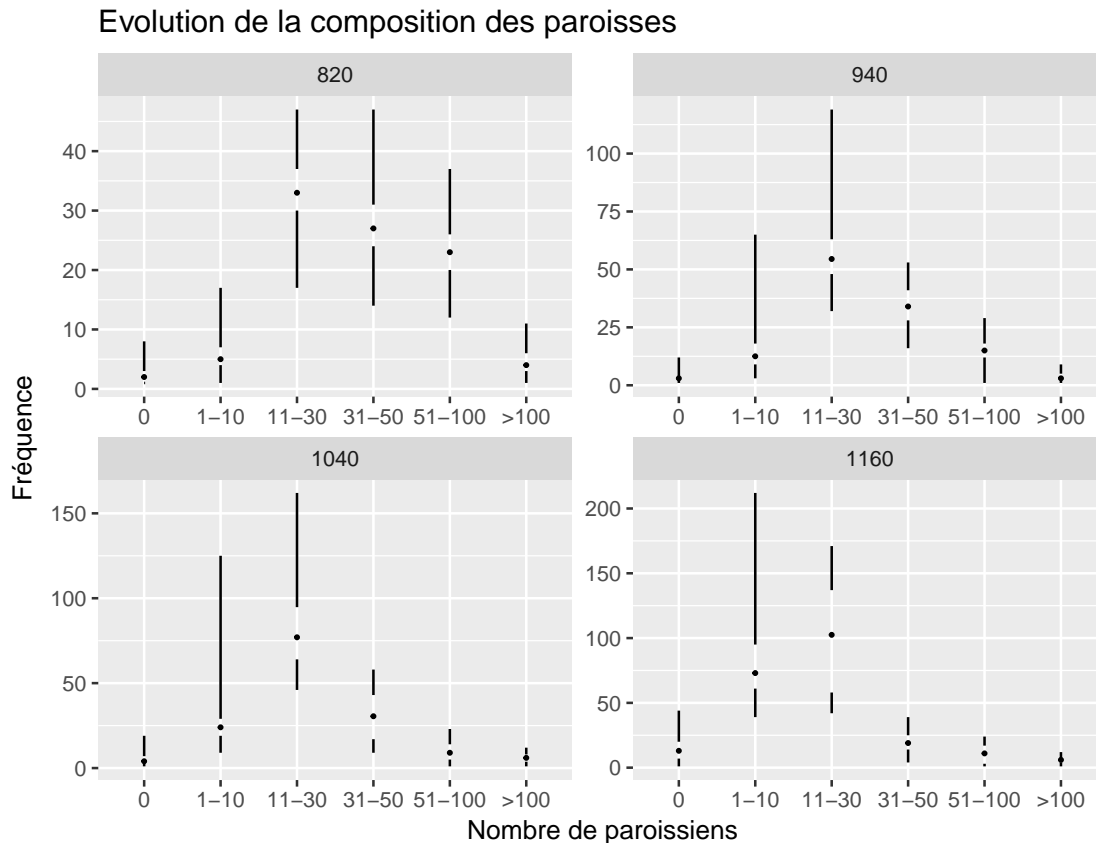


FIGURE 3.12 – Évolution de la composition des paroisses.

### Évaluer la fixation et la dissémination du peuplement

**Stabilité** On se focalise sur les agrégats :

- **Polarisation** : il faut qu'ils rassemblent les FP (taux isolés faible),
- **Hiérarchisation** : qu'ils soient hiérarchisés (composition des agrégats en FP doit tendre vers log-normale),
- **Fixation et dispersion** : et qu'ils soient dispersés dans l'espace de manière stable (fixation par paroisses).

3 leviers différents pour jouer sur ces dimensions :

- **Polarisation** : Améliorer le déplacement des FP (Agent concerné : FP) :
  - ils doivent se déplacer vers un petit agrégat/pôle et ne pas rester isolés
  - une fois dans un agrégat, ils doivent y rester ou aller vers un agrégat plus « important » : en terme de nb de FP, d'attractivité du pôle ou de présence de CA
  - Ils ne doivent pas faire des allers-retours incessants entre les agrégats.
- **Hiérarchisation** : Améliorer la hiérarchisation des attracteurs (Agent concerné : Attracteurs) :
  - La différence entre les pôles d'attraction doit être franche,
  - pour que les petits (1 église par ex.) attirent de petits agrégats,
  - et que les gros (châteaux, multi-pôles etc.) polarisent des agrégats importants et stables
- **Fixation et dispersion** : Améliorer la desserte des églises paroissiales (Agent concerné : Paroisse) :
  - Les paroisses doivent couvrir tout le territoire de manière à peu près régulière (logique d'équité),

- Mais aussi avec des zones de plus forte densité proche des gros pôles/agrégats (petites villes) (logique d'efficacité),

### 3.2.2 Avant propos sur le paramétrage d'un modèle complexe

Pour introduire le schéma des étapes, rappeler ici que les paramètres et mécanismes d'un modèle complexe sont en interaction, ce qui produit des effets non linéaires.

Principe des vases communicants !

Mentionner aussi la logique des attracteurs étranges de Lorenz comme paroxysme de cette complexité : ici, moins de variation, mais des répercussions importantes tout de même qu'il faut donc « corriger » par de nombreuses itérations de modifications.

De manière générale

### 3.2.3 Les étapes du paramétrage de SimFeodal



TABLE 3.1 – Tableau récapitulatif des étapes de construction du modèle. Conception : C. Tannier (Tannier 2017, HDR) et R. Cura, 2017

Étape	Version	Type	Modifications	Section
0	Base	Création	Création du modèle. Cf. tableau 14 de Tannier 2017	Chap 2 <sup>24</sup>
1	Base2	paramétrage	Hiérarchisation des valeurs d'attraction des attracteurs. Modification des critères de création de nouvelles paroisses. Réduction de la distance maximale de déplacement local. Assouplissement des règles de promotion d'un château en gros château.	Chap3-x.y
2	Base2 compo5 2bis	modélisation et paramétrage	Assouplissement des règles de création d'une paroisse. Réduction de la distance maximale de déplacement local. Modification des règles de calcul de la probabilité de construire un château. Modification du calcul de la satisfaction matérielle des foyers paysans.	Chap3-x.y
3	Base2 compo5 2quat	modélisation et paramétrage	Hiérarchisation des valeurs d'attraction des attracteurs. Facilitation de la construction de châteaux. Rationalisation du déplacement des foyers paysans. Ajout d'un nouveau type d'attracteur, les communautés.	Chap3-x.y
4	Base3 2	modélisation et paramétrage	Hiérarchisation des valeurs d'attraction des attracteurs. Modification des critères de création de nouvelles paroisses. Augmentation de la distance maximale de déplacement local. Modification de la procédure d'identification des agrégats et des pôles.	Chap3-x.y
5	Base4 1	modélisation et paramétrage	Modification de la répartition initiale des foyers paysans. Simplification de la procédure d'identification d'héritage des agrégats. Amélioration de la définition des pôles d'attraction et de leur enveloppe.	Chap3-x.y
6	Base4 2	modélisation et paramétrage	Hiérarchisation des valeurs d'attraction des attracteurs. Modification de l'ordonnancement des actions dans le modèle. Modification de la procédure d'identification des agrégats.	Chap3-x.y
7	Base4 3ter	modélisation et paramétrage	Dynamisation de la distance maximale de déplacement local. Modification du calcul de satisfaction des foyers paysans.	Chap3-x.y
8	Base4 4A	modélisation	Modification de l'ordonnancement des actions dans le modèle. Modification du mécanisme de déplacement local des foyers paysans.	Chap3-x.y

## Étape 0

TABLE 3.2 – Indicateurs synthétiques de l'étape 1

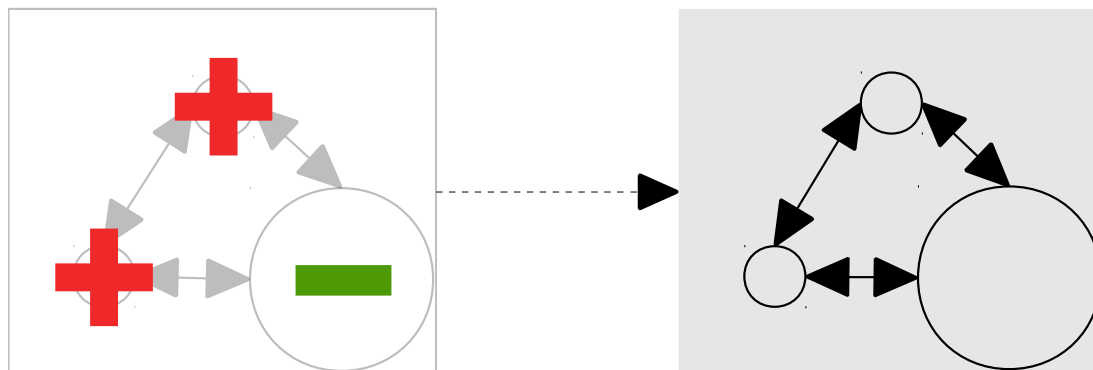
Indice	Objectif	Moyenne
Nb agrégats	200.00	145.30
Nb châteaux	50.00	69.78
Nb gros Châteaux	10.00	9.60
Nb egl. Par.	300.00	167.90
Dist entre egl.	3000.00	2944.00
FP isolés	0.20	0.49
Ratio Charge Fisc.	3.00	5.15

24. On s'appuie sur cette version « Base » dans le chapitre 2.

## Indicateurs généraux

## Étape 1

## Étape 1



## Version 1

FIGURE 3.13 – Étape 1 du paramétrage et résultats.

## Étape 2

## Étape 2

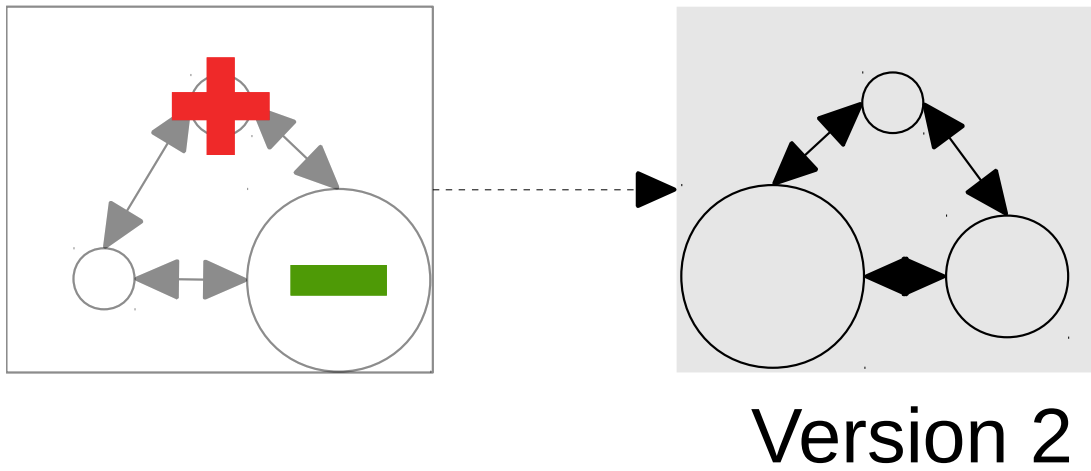
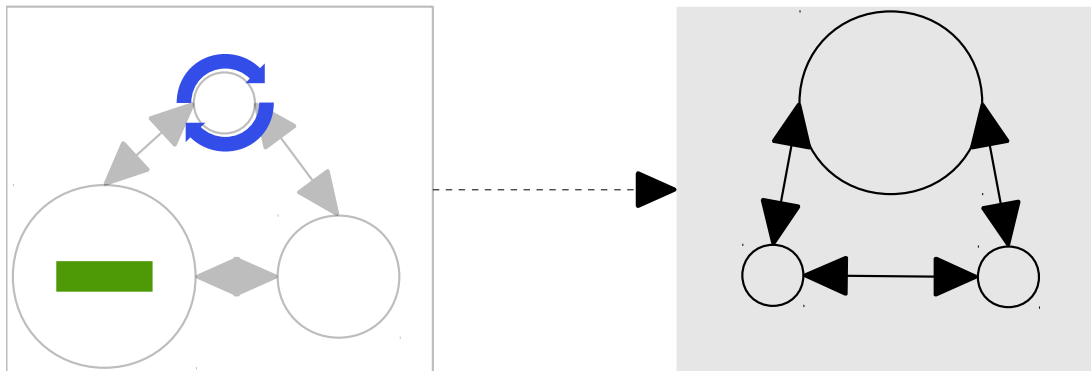


FIGURE 3.14 – Étape 2 du paramétrage et résultats.

## Étape 3

## Étape 3

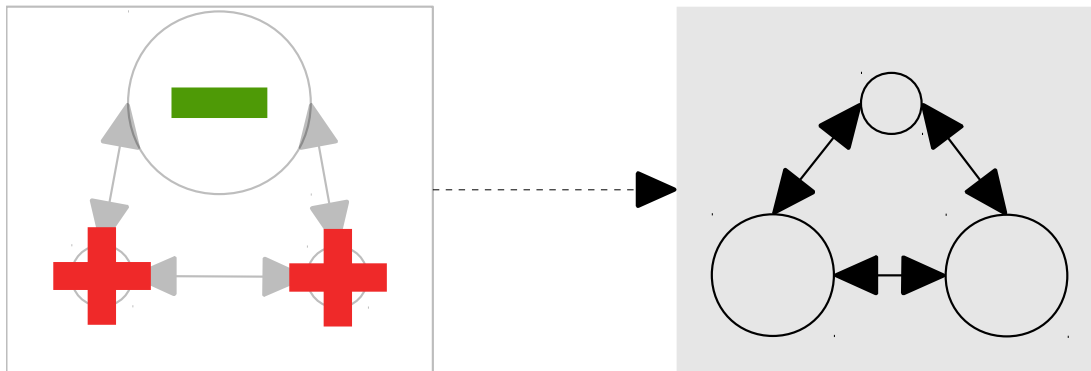


## Version 3

FIGURE 3.15 – Étape 3 du paramétrage et résultats.

## Étape 4

## Étape 4



## Version 4

FIGURE 3.16 – Étape 4 du paramétrage et résultats.

## Étape 5

## Étape 5

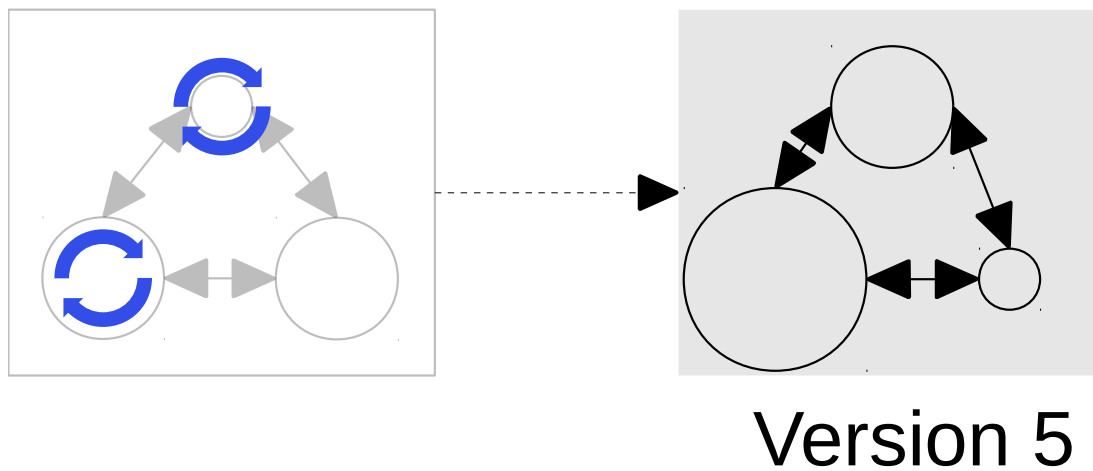
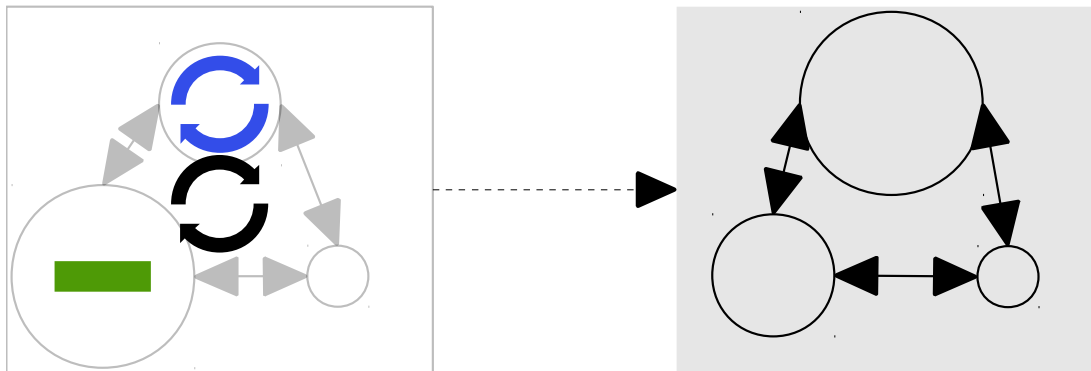


FIGURE 3.17 – Étape 5 du paramétrage et résultats.

## Étape 6

## Étape 6



## Version 6

FIGURE 3.18 – Étape 6 du paramétrage et résultats.

## Étape 7

## Étape 7

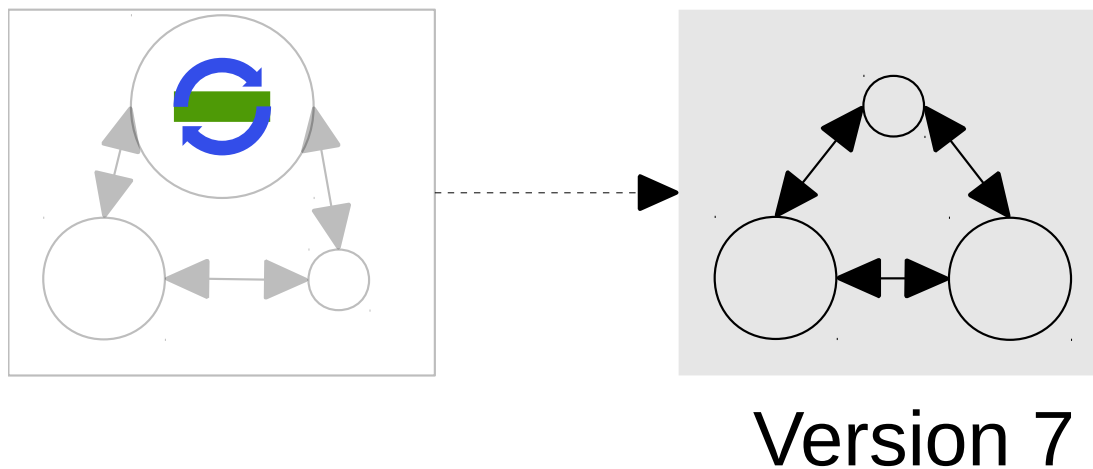


FIGURE 3.19 – Étape 7 du paramétrage et résultats.



## Étape 8

## Étape 8

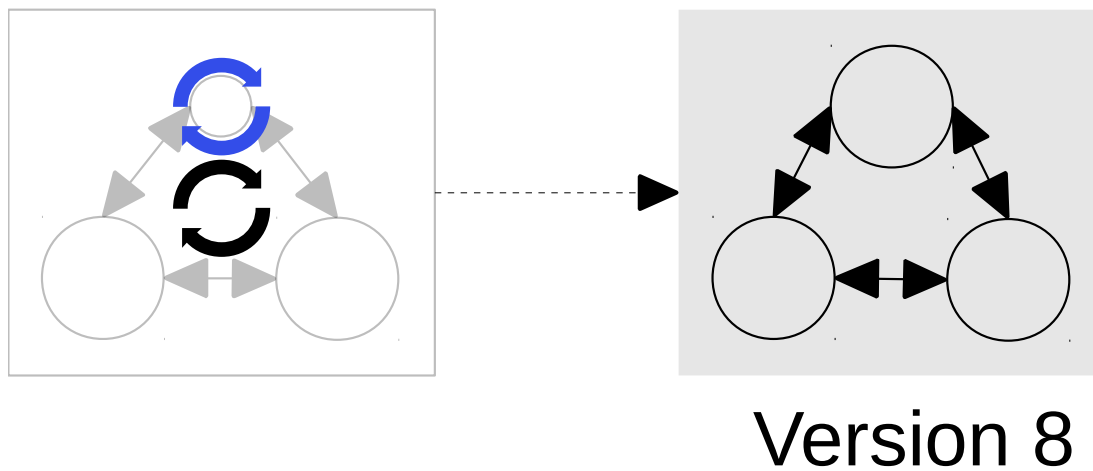


FIGURE 3.20 – Étape 8 du paramétrage et résultats.

### 3.2.4 Un bilan des changements majeurs à l'issu du paramétrage

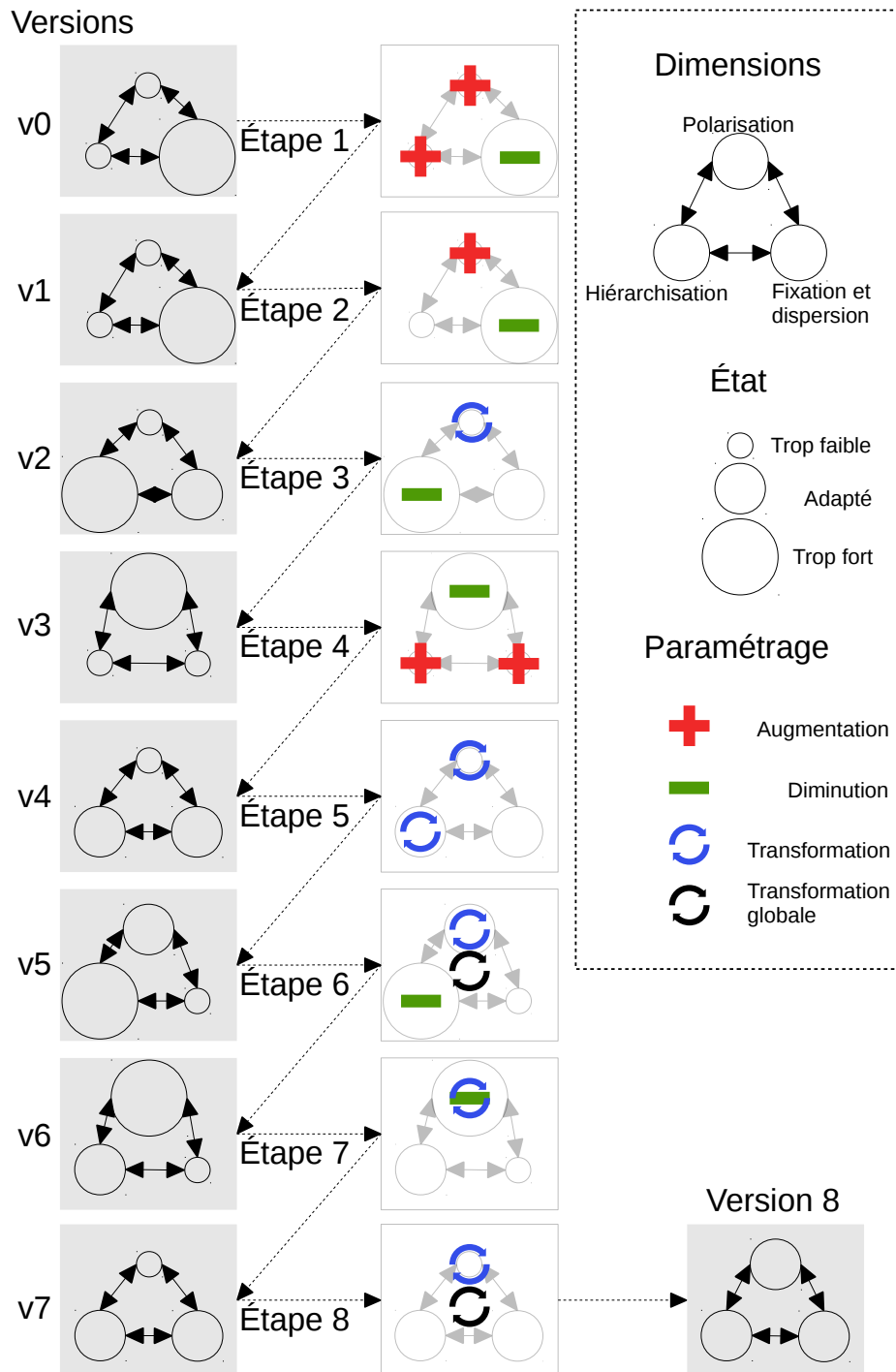


FIGURE 3.21 – Schéma récapitulatif et synthétique des étapes du paramétrage de SimFeodal.

## **3.3 Comment traiter les sorties du modèle ?**

### **3.3.1 Variabilité**

- Quelle variabilité du modèle ? Pour quelles raisons (aléa init. vs aléa mécanique) ?

### **3.3.2 Nombre de sortie**

- Sorties graphiques, rapports, rapports ++

### **3.3.3 Masse des sorties**

- Comment aller plus loin ?

# Bibliographie

- BALCI, Osman (1994). « Validation, Verification, and Testing Techniques throughout the Life Cycle of a Simulation Study ». In : *Annals of operations research* 53.1, p. 121–173.
- DAUDÉ, Eric et Patrice LANGLOIS (2006). « Comparaison de Trois Implémentations Du Modèle de Schelling ». In : *Modélisation et Simulation Multi-Agents. Applications Pour Les Sciences de l'Homme et de La Société*. Sous la dir. d'HERMÈS, p. 411–441. URL : <https://halshs.archives-ouvertes.fr/halshs-01082655>.
- DE RUNZ, Cyril (2008). « Imperfection, Temps et Espace : Modélisation, Analyse et Visualisation Dans Un SIG Archéologique ». Université de Reims-Champagne Ardenne.
- DURAND-DASTÈS, François (1995). « Les Modèles En Géographie ». In : *Bailly A., Ferras R., Pumain D., Encyclopédie de La Géographie*. 2ème édition. Chapitre 16, p. 293–307.
- GILBERT, Nigel et Klaus TROITZSCH (2005). *Simulation for the Social Scientist*. McGraw-Hill Education (UK).
- IMAGEMAGICK STUDIO, LLC (2008). « ImageMagick ». In :
- IMBS, Paul et Bernard QUEMADA (1992). *Trésor de La Langue Française*. Éditions du Centre national de la recherche scientifique.
- INSTITUTE, SAS (1990). *SAS/STAT User's Guide : Version 6*. T. 2. Sas Inst.
- LATOUR, Bruno (2013). *Nous n'avons Jamais Été Modernes*. La découverte.
- LIVET, Pierre, Denis PHAN et Lena SANDERS (2014). « Diversité et Complémentarité Des Modèles Multi-Agents En Sciences Sociales ». In : *Revue française de sociologie* 55.4, p. 689–729.
- MATHIAN, Hélène et Cécile TANNIER (2015). « Formalisation des interactions spatiales dans les modèles multi-agents - Essai d'analyse comparative ». In : *Revue Internationale de Géomatique* 25.3, p. 393–414. DOI : 10.3166/RIG.25.393-414. URL : <http://dx.doi.org/10.3166/RIG.25.393-414> (visité le 23/05/2017).
- MINSKY, Marvin (1965). « Matter, Mind and Models ». In :
- NYKAMP, Duane Q. (2015a). *Function Machine Parameters*. URL : [http://mathinsight.org/function\\_machine\\_parameters](http://mathinsight.org/function_machine_parameters) (visité le 18/12/2017).
- (2015b). *Parameter Definition*. URL : <http://mathinsight.org/definition/parameter> (visité le 18/12/2017).
- Parameter (Computer Programming)*. In : *Wikipedia*. Page Version ID : 815191973. URL : [https://en.wikipedia.org/w/index.php?title=Parameter\\_\(computer\\_programming\)&oldid=815191973](https://en.wikipedia.org/w/index.php?title=Parameter_(computer_programming)&oldid=815191973) (visité le 18/12/2017).
- PUMAIN, Denise (2004). *Modèle Gravitaire - Hypergéométrie*. URL : <http://www.hypergeo.eu/spip.php?article76> (visité le 11/01/2018).
- PUMAIN, Denise et Thérèse SAINT-JULIEN (2001). « Les Interactions Spatiales ». In : *Armand Colin, Paris*, p. 19–31.
- ROBERT, Paul, Josette REY-DEBOVE et Alain REY (1993). *Le Nouveau Petit Robert*. le Robert.

- SCHELLING, Thomas C (1971). « Dynamic Models of Segregation ». In : *Journal of mathematical sociology* 1.2, p. 143–186.
- SCHMITT, Clara (2014). « Modélisation de la dynamique des systèmes de peuplement : de SimpopLocal à SimpopNet. » Thèse de Doctorat. paris : Université Paris I - Panthéon-Sorbonne. URL : <https://tel.archives-ouvertes.fr/tel-01077891/document>.
- SIMON, Herbert A (1955). « On a Class of Skew Distribution Functions ». In : *Biometrika* 42 (3/4), p. 425–440.
- STEWART, John Q (1948). « Demographic Gravitation : Evidence and Applications ». In : *Sociometry* 11 (1/2), p. 31–58.
- TANNIER, Cécile et al. (2014). « Une Ontologie Pour Décrire Les Transformations Du Système de Peuplement Européen Entre 800 et 1100 ». In : *Ontologies et Modélisation Par SMA En SHS*. Sous la dir. de Denis PHAN. Hermès - Lavoisier, p. 289–310. URL : <https://hal.archives-ouvertes.fr/hal-01021976> (visité le 31/12/2014).
- TREUIL, Jean-Pierre, Alexis DROGOUL et Jean-Daniel ZUCKER (2008). « Modélisation et Simulation à Base d’agents ». In : *Dunod Editions*, p. 1–13.
- VERHULST, Pierre-François (1838). « Notice Sur La Loi Que La Population Suit Dans Son Accroissement. Correspondance Mathématique et Physique Publiée Par A ». In : *Quetelet* 10, p. 113–121.
- VOGT, W Paul (1993). « Dictionary of Statistics and Methodology : A Nontechnical Guide for the Social Sciences. » In :
- YULE, G Udny (1925). « II.—A Mathematical Theory of Evolution, Based on the Conclusions of Dr. JC Willis, FR S ». In : *Phil. Trans. R. Soc. Lond. B* 213 (402-410), p. 21–87.