

### Introduction

<sup>1</sup>Dans les chapitres précédents ont été présentés <sup>2</sup>le modèle SimFeodal (<sup>3</sup>chapitre 2), la manière de l'évaluer (<sup>4</sup>chapitre 3), la méthode suivie pour son paramétrage (<sup>5</sup>chapitre 4) et les outils développés pour mener à bien ces étapes de construction et d'évaluation (<sup>6</sup>chapitre 5). Avec cet outillage, théorique, méthodologique et technique, nous sommes désormais en mesure d'explorer le modèle SimFeodal. Par « explorer le modèle », on entend ici l'exploration des sorties produites par le modèle, sous toutes leurs formes, afin de gagner en connaissance sur ce qui est modélisé, mais aussi sur la manière dont le modèle est construit et implémenté. Il s'agit autant d'analyser les « résultats » de SimFeodal que d'en explorer le fonctionnement et la robustesse. Ce chapitre est construit en deux parties assez indépendantes mais qui se concentrent sur différents aspects de l'évaluation d'un modèle.

La première partie concerne l'analyse des « résultats » de SimFeodal, c'est-à-dire les réponses produites par le modèle aux questionnements énumérés dans le chapitre 3 : le modèle permet-il bien de générer une polarisation des foyers paysans ? Dans quelles conditions ? Le système de peuplement généré par le modèle est-il hiérarchisé <sup>7</sup>tel qu'on l'attendait ? <sup>8</sup>Et sa distribution est-elle <sup>9</sup>proche des connaissances empiriques ? Observe-t-on une fixation du peuplement <sup>10</sup>dans un espace plus disséminé que dans les configurations initiales ? Pour répondre à ces questions, on présente d'abord les étapes de calibrage qui ont permis d'aboutir à une version « définitive » du modèle. On pourra alors introduire les résultats de cette version, en <sup>11</sup>en présentant les conclusions les plus saillantes, c'est-à-dire sans s'attacher à une présentation exhaustive de l'ensemble des indicateurs de sortie issus de cette version.

La deuxième partie de ce chapitre s'attachera à une exploration du comportement du modèle, c'est-à-dire à sa sensibilité. Cette sensibilité peut être entendue au sens de robustesse du modèle face aux différentes valeurs de paramètres (analyse de sensibilité classique). Comment les différents paramètres – et leurs valeurs associées – jouent-ils sur la variabilité des indicateurs de sortie ? Un premier questionnement concerne <sup>12</sup>Les valeurs de paramètres testées <sup>13</sup>elles un effet sur les valeurs des indicateurs de sortie ? Une seconde question concerne la robustesse du modèle à l'aléa. Observe-t-on des différences de variabilité répliquative du modèle selon les différents paramètres et valeurs de paramètres testés ? Dans le cas de SimFeodal, une analyse de ce type pose des questions méthodologiques complexes, en raison notamment du grand nombre de paramètres impliqués et de « l'explosion combinatoire » qui en découle.

# Résumé des commentaires sur Robin Cura - Thèse

---

Page : 2

	Nombre : 1	Auteur : Paul Gourdon	Sujet : Barrer	Date : 08/10/2019 18:22:35 +02'00'
	Nombre : 2	Auteur : Paul Gourdon	Sujet : Barrer	Date : 08/10/2019 18:22:45 +02'00'
	Nombre : 3	Auteur : Paul Gourdon	Sujet : Barrer	Date : 08/10/2019 18:22:49 +02'00'
	Nombre : 4	Auteur : Paul Gourdon	Sujet : Barrer	Date : 08/10/2019 18:23:23 +02'00'
	Nombre : 5	Auteur : Paul Gourdon	Sujet : Texte de remplacement comme attendu / tel qu'attendu	Date : 08/10/2019 18:27:45 +02'00'
	Nombre : 6	Auteur : Paul Gourdon	Sujet : Barrer	Date : 08/10/2019 18:28:36 +02'00'
	Nombre : 7	Auteur : Paul Gourdon	Sujet : Texte surligné pas très élégant. "une fixation du peuplement plus dispersée que dans les configurations initiales ?"	Date : 11/10/2019 17:32:27 +02'00'
	Nombre : 8	Auteur : Paul Gourdon	Sujet : Barrer	Date : 11/10/2019 17:32:47 +02'00'
	Nombre : 9	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 11/10/2019 17:33:30 +02'00'
	Nombre : 10	Auteur : Paul Gourdon	Sujet : Texte de remplacement :	Date : 11/10/2019 17:35:35 +02'00'

## 6.1 Calibrage du modèle et premiers résultats

Dans le chapitre 3 (??), on indiquait que SimFeodal s'inscrivait dans la lignée des modèles dont le but est d'« assister la construction de théories », ou encore « à utilité de développement », c'est-à-dire permettant à des chercheurs d'explicitier et de vérifier la cohérence de leurs hypothèses plutôt qu'à les infirmer ou confirmer.

En tant que tels, les « résultats » de SimFeodal ne nous semblent pas revêtir d'enjeu confirmatoire majeur : ils sont là pour participer à l'évaluation du modèle, tant sur un plan interne qu'externe (cf. chap3), mais ne constituent pas pour autant des éléments objectifs et quantifiables de validation.

Dans cette partie, nous rappelons les objectifs du modèle, nous décrivons les étapes de calibrage qui ont été nécessaires afin d'en approcher, et nous présentons enfin les résultats de la version actuelle (6.6<sup>1</sup>) du modèle, c'est-à-dire les indicateurs de sortie de simulation issus de cette version, commentés selon la perspective de leur correspondance aux objectifs.

### 6.1.1 Calibrage du modèle

La phase de calibrage diffère des nombreuses étapes de paramétrage décrites dans le chapitre 4. Il ne s'agit ainsi plus d'ajuster les mécanismes et paramètres pour obtenir une cohérence d'ensemble dans la manière dont le modèle réagit, mais de se concentrer sur quelques paramètres, contextuels, dont on va régler finement les valeurs afin que le contexte de déroulement des autres mécanismes du modèle soit aussi contrôlé et fidèle aux connaissances expertes que possible.




Ces paramètres contextuels sont étroitement liés à certains indicateurs de sortie. Conceptuellement, paramètres et indicateurs sont très largement différents, situés de part et d'autre de la simulation dans une typologie comme celle de Balci (ref chap 4). Pourtant, dans le cas de ces éléments de contexte, les paramètres sont étroitement liés, de manière presque déterministe, aux indicateurs de sortie qui en découleront. Pour « ajuster » la valeur d'un indicateur de sortie, on pourra donc tester différents ensembles de valeurs pour le ou les paramètres qui agissent directement sur ces indicateurs. De plus, pour le calibrage des paramètres, on ne s'attachera pas à l'observation de l'ensemble des sorties possibles, mais seulement des indicateurs directement liés aux paramètres. Par exemple, les paramètres agissant sur le nombre de châteaux en fin de simulation ont un effet global sur le modèle, mais, pour l'aspect contextuel qui nous intéresse dans leur calibrage, seuls les indicateurs relatifs au nombre et au type de châteaux seront mobilisés.

Le calibrage d'un paramètre peut résulter d'un ajustement obtenu de manière expérimentale. Dans ce cas, on cherche les valeurs de paramètres qui assureront un écart minimal entre l'indicateur de sortie de simulation simulé et sa correspondance empirique. Le calibrage peut aussi être thématique, par exemple

---

1. Faire référence à tableau avec historique des versions du modèle, sans doute chap 3-4.

## Page : 3

 Nombre : 1 que de	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 11/10/2019 17:39:26 +02'00'
 Nombre : 2	Auteur : Paul Gourdon	Sujet : Barrer	Date : 11/10/2019 17:39:28 +02'00'
 Nombre : 3 aussi	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 11/10/2019 17:42:24 +02'00'

en menant une recherche plus approfondie sur les valeurs que peuvent prendre, thématiquement, certains paramètres du modèle au regard des connaissances expertes sur lesquels ils s'appuient. Ce second type de calibrage pourrait sembler évident, préalable à une démarche rigoureuse et scientifique, et indispensable à réaliser sur chacun des paramètres d'un modèle qui peuvent être estimés ou ajustés de manière empirique et thématique. Dans un modèle descriptif, exploratoire et surtout basé sur des dynamiques passées, <sup>[2]</sup> ~~pourant~~, la <sup>[3]</sup> ~~tâche de~~ recherche précise de valeurs pour des dizaines de paramètres semble irréalisable en dehors de projets de modélisation de très large ampleur, ce qui n'est pas le cas de SimFeodal.

Dans les paragraphes suivants, nous donnons des exemples de paramètres qui ont été calibrés dans les dernières phases de <sup>[4]</sup> **paramétrage** du modèle. Ces exemples reprennent la distinction établie ci-dessus entre les types de calibrages.

- Le calibrage des paramètres d'*inputs*, c'est-à-dire des paramètres agissant sur l'initialisation du monde simulé, est entièrement thématique ;
- le calibrage des paramètres liés à l'émergence des paroisses, rurales et urbaines, est mixte : il fait appel à des éléments de calibrage thématiques et expérimentaux ;
- Le calibrage des paramètres liés à la création de châteaux par les seigneurs est pour sa part purement expérimental : les paramètres calibrés ne reposent pas sur des données empiriques, au contraire des objectifs quantitatifs <sup>[6]</sup> fin de simulation ~~poursuivis~~.

#### 6.1.1.1 Calibrage des *inputs*

**Taille du monde simulé.** Depuis sa conception (en 2014) jusqu'à la version 5.1 (Novembre 2018) du modèle, on avait choisi de simuler les agents du modèle dans un monde théorique, de forme carrée, de 100 km de côté. Ces dimensions donnaient au monde du modèle l'étendue des grands départements français contemporains (ou des petites régions), et constituaient ainsi l'échelle à laquelle on souhaitait modéliser les phénomènes décrits dans SimFeodal. Les différentes quantités empiriques relatives à cette dimension y étaient nécessairement fortement liées : population, quantité d'églises, de petites villes etc. On s'appuyait sur une vision assez large et englobante de la Touraine historique, en y incluant par exemple quelques éléments du duché d'Anjou voisin. La Touraine est en effet un espace qui a défini plusieurs ensembles régionaux au cours du temps, du diocèse de Tours dès le III<sup>e</sup> siècle au département d'Indre-et-Loire de la Révolution Française, en passant par la province historique de Touraine, successivement comtat (VI<sup>e</sup>) et duché (XIV<sup>e</sup>). <sup>[8]</sup> **Lors de chacune de ces étapes – et au cours d'entre elles –** les frontières ont assez largement fluctué, notamment vis-à-vis de la province d'Anjou.

En avançant dans le calibrage de SimFeodal, on a choisi de préciser l'aire d'étude, en la calant sur une étendue plus restreinte correspondant au diocèse médiéval de Tours, plus stable en matière de délimitation<sup>2</sup>. Cette redéfi-

2. Il n'est ainsi pas mentionné de changement relatif au diocèse de Tours dans MIROT (1950,

## Page : 4

 Nombre : 1	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 11/10/2019 17:42:44 +02'00'
Pourtant,			
 Nombre : 2	Auteur : Paul Gourdon	Sujet : Barrer	Date : 11/10/2019 17:42:26 +02'00'
 Nombre : 3	Auteur : Paul Gourdon	Sujet : Barrer	Date : 11/10/2019 17:43:08 +02'00'
 Nombre : 4	Auteur : Paul Gourdon	Sujet : Barrer	Date : 11/10/2019 17:45:41 +02'00'
 Nombre : 5	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 11/10/2019 17:45:37 +02'00'
le calibrage est donc une étape du processus de paramétrage ? Ce n'est pas ce qui est suggéré dans la première phrase de cette sous-partie			
 Nombre : 6	Auteur : Paul Gourdon	Sujet : Barrer	Date : 11/10/2019 17:46:31 +02'00'
 Nombre : 7	Auteur : Paul Gourdon	Sujet : Texte inséré	Date : 11/10/2019 17:46:44 +02'00'
poursuivis			
 Nombre : 8	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 11/10/2019 17:50:17 +02'00'
Au cours de cette longue période, les frontières...			

inition de l'espace d'étude permettait en effet de pouvoir déterminer plus strictement quels éléments empiriques y inclure<sup>1</sup> (châteaux, églises paroissiales...) afin d'avoir des chiffres empiriques mieux spécifiés.

En faisant ce choix, il a fallu en premier lieu revoir à la baisse la taille du monde simulé. Celui-ci, pour rappel, est une surface de forme carrée dont les côtés faisaient initialement 100 km de longueur. Le diocèse de Tours a une superficie proche de 6 200 km<sup>2</sup>, qui est ainsi inférieure de plus d'un tiers à la superficie originellement choisie (~~100 × 100 km~~, soit 10 000 km<sup>2</sup>). On a choisi, pour conserver un aspect théorique, de conserver la forme carrée tout en en réduisant la superficie : la taille des côtés du monde simulé a été réduite à 80 km, dont seuls 79 sont utilisables dans le cadre du modèle (voir le monde restreint, chapitre 2), soit une surface de 6 240 km<sup>2</sup>, équivalente à celle de la Touraine<sup>3</sup>.

Dans SimFeodal, la plupart des mécanismes ont une importante composante spatiale. Dès lors, avec la diminution de la superficie du monde, il a été nécessaire de modifier certains paramètres de contexte et une large partie des paramètres techniques. Par exemple, un paramètre technique (coef\_redevance) permet d'ajuster le seuil de satisfaction matérielle des foyers paysans en fonction du nombre de redevances dont ils doivent s'acquitter. En diminuant la taille du monde du modèle, et sans diminuer proportionnellement le nombre ou les surfaces des zones de prélèvement, le nombre de redevances des foyers paysans augmente mécaniquement.

**Population.** Le nombre de foyers paysans aurait aussi pu être affecté par la taille du monde du modèle, mais on le considère comme un *input* guidé par les connaissances empiriques plus que comme un élément contextuel. C'est un *input* un peu particulier en ce qu'il est extrêmement difficile, sinon vain, de parvenir à une estimation de la population d'une région française durant la période étudiée. On peut tout de même obtenir des indices sur des ordres de grandeur de population à des moments clés de l'histoire. Sur la Touraine, l'expertise des archéologues et historiens (ref à EZR + Monique Bourrin) permet par exemple d'avoir une idée de la population au début du XVII<sup>e</sup> siècle, mais les différentes sources préalables<sup>3</sup> présentent des écarts majeurs.

Pour SimFeodal, on a repris une hypothèse d'historien, qui semble réunir un certain consensus dans la communauté. Cette hypothèse consiste à penser qu'un optimum démographique a été atteint au début du XIII<sup>e</sup> siècle, et que la population a ensuite diminué considérablement. Les historiens estiment que les niveaux de populations du XIII<sup>e</sup> siècle n'ont été rattrapées qu'au début du XVII<sup>e</sup>. Il est de ce fait possible d'estimer que les niveaux de populations à l'issue de la période étudiée sont proches de celles, mieux connues empiriquement, du XVII<sup>e</sup> siècle. On a choisi de fixer un objectif de population, en fin de simulation (en 1200), à 40 000 foyers paysans, soit une densité d'environ 6.5 feu / km<sup>2</sup>,


p. 309–326) jusqu'en 1790, où Tours perd son statut d'archidiocèse (avant de le retrouver quelques années plus tard, en 1801).

3. Dans la suite de ce chapitre, nous référons au diocèse de Tours lors des mentions de la Touraine.


---

 Nombre : 1	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 11/10/2019 17:51:18 +02'00'
---	-----------------------	------------------------	------------------------------------

---

 Nombre : 2	Auteur : Paul Gourdon	Sujet : Barrer	Date : 11/10/2019 17:52:12 +02'00'
---	-----------------------	----------------	------------------------------------

---

 Nombre : 3	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 11/10/2019 17:55:12 +02'00'
---	-----------------------	------------------------	------------------------------------

---

"sources préalables" à la définition de l'input population ? ou sources concernant des périodes antérieures au XVIIème siècle ?



ou encore de 30 habitants/ km<sup>2</sup>.

La population initiale est bien plus difficile à estimer. Selon les sources<sup>4</sup>, certains présentent le IX<sup>e</sup> siècle comme un « nadir » démographique (c'est-à-dire un minimum dans l'ensemble du moyen-âge), dont la population aurait été multipliée par plus de 7 pour atteindre son niveau maximal – relativement à la période médiévale – à la fin du XIII<sup>e</sup>. Pour d'autres historiens et archéologues, rien ne permet de penser qu'il y ait eu une croissance démographique significative entre ces périodes (voir avec EZR et SL).

Dans SimFeodal, nous avons choisi l'hypothèse la plus prudente et qui a le moins d'implications : considérer que la population est relativement stable entre le début et la fin de la période simulée. Cette « hypothèse nulle » n'a pas d'implication thématique forte ici, et nous permettra par la suite de tester des scénarios où l'on ajoute de la croissance démographique dans le déroulement du modèle (voir encadré 6.2).

#### 6.1.1.2 Calibrage des paroisses


Le calibrage du nombre et de la distribution spatiale paroisses a posé des questions tant thématiques que méthodologiques. Dans SimFeodal, il existe deux mécanismes distincts de création ou promotion de paroisses (voir ?? dans le chapitre 2) : un mécanisme dédié aux églises paroissiales situées dans des agrégats (les paroisses « urbaines ») et un mécanisme pour la promotion de nouvelles paroisses en zone peu dense (les paroisses « rurales »). Ces mécanismes servent à faire émerger, de manière guidée, un maillage paroissial-similaire, ~~structurellement~~, à celui que l'on peut reconstituer à partir des connaissances empiriques. L'apparition et la densification de ce maillage, est une conséquence recherchée, dans le modèle, des migrations individuelles des foyers paysans. De plus, ce maillage influe à son tour sur les futures migrations de ces mêmes foyers paysans : la création de nouvelles paroisses et leur localisation agissent sur les choix de migration des foyers paysans, en constituant l'un des éléments polarisant.

Les paroisses sont ainsi autant des marqueurs spatiaux, témoins de la distribution spatiale des foyers paysans de l'époque, que des catalyseurs à la polarisation et à la fixation de cette même population. Au sein du modèle, la transformation du maillage paroissial joue un rôle contextuel, conditionnant les migrations des foyers paysans, et un rôle émergent, que l'on observe à une échelle plus agrégée (densification ou étalement, rythmes de changements etc.)<sup>3</sup>. C'est le rôle contextuel que l'on a calibré, par l'intermédiaire des paramètres de contexte agissant sur la création et la localisation des paroisses, et non des paramètres régissant l'influence des paroisses sur les foyers paysans.


**Calibrer le nombre et l'espacement des paroisses « rurales ».** Pour le calibrage des paroisses « rurales », on se base surtout sur des aspects thématiques issus de connaissances archéologiques. On connaît ainsi, au moins

4. Voir avec les archéos. Par exemple, dans le nord de la France (Picardie, IdF etc.) : « La population du Nord au Moyen Age. I : avant 1384 » de Alain Derville (1998).

---

 Nombre : 1	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 11/10/2019 18:05:22 +02'00'
---	-----------------------	-------------------------------	------------------------------------

---

 Nombre : 2	Auteur : Paul Gourdon	Sujet : Barrer	Date : 11/10/2019 18:04:53 +02'00'
---	-----------------------	----------------	------------------------------------

---

 Nombre : 3	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 11/10/2019 18:14:05 +02'00'
---	-----------------------	------------------------	------------------------------------

---

ça c'est pas très clair pour moi.

Dans la phrase précédente je comprends que le maillage paroissial joue en rôle contextuel, c-a-d influe sur les migrations des foyers.

Dans cette phrase (surligné), je comprends le début avec les paramètres de contexte : la création et la localisation des paroisses, c'est bien ce qui change le maillage. Par contre, je comprends pas pk les paramètres qui régissent l'influence des paroisses ne sont pas aussi contextuels.

En effet, le rôle émergent, il semble que ça soit à un autre niveau d'observation => si j'ai bien compris on observe le rôle de l'ensemble des paroisses sur le système de peuplement.

Du coup je comprends pas pourquoi, le niveau plus micro de l'influence des paroisses n'est pas "contextuel".

pour la fin du XII<sup>e</sup> siècle, le nombre et la répartition spatiale de la plupart des paroisses de Touraine (en s'appuyant sur ZADORA-RIO (2008, p. 31), on considère dans le modèle qu'il y a environ 300 églises paroissiales). Celles-ci correspondent majoritairement à des milieux ruraux et ont en partie perduré dans le maillage communal actuel. La fin du XII<sup>e</sup> correspond à la date d'arrêt du modèle, et cette quantité empirique de paroisses constitue ainsi un objectif à atteindre pour SimFeodal en fin de simulation.

La répartition spatiale de ces églises paroissiales, en fin de période, peut être dérivée d'estimations empiriques sur l'espacement moyen entre églises paroissiales : « Entre 900 et 1200, l'augmentation importante du nombre de lieux de cultes attestés par les sources écrites se traduit par une diminution nette de l'espacement observé entre les sites : on passe d'une distance moyenne d'un peu plus de 4 km entre deux églises en 900, à une distance d'environ 2,8 km en 1200. » (CHAREILLE, RODIER et ZADORA-RIO 2008, p. 261). Ces espacements empiriques permettent de calibrer le modèle via le calcul d'un indicateur dédié à la mesure de la moyenne des distance à la plus proche église paroissiale.

Le calibrage porte sur le paramètre associé aux mécanismes de création/promotion de paroisses rurales. Ces mécanismes sont complexes (voir chap2, section 2.7.2.3) et le paramètre qui les régit, `seuil_nb_paroissiens_insatisfaits`, a donc une influence importante et difficilement prévisible sans expérimentation. Dans l'ensemble, ce paramètre agit comme un seuil de foyers paysans au delà duquel une nouvelle église paroissiale est créée ou promue en zone rurale. Pour le calibrage, on a fait varier le paramètre : diminuer la valeur de ce seuil poussait à la création de plus de paroisses rurales, et l'augmenter limitait le nombre final. Dans l'état actuel du modèle, le nombre de paroisses rurales est encore trop important au regard des connaissances empiriques (380 au lieu de 300, voir le tableau 6.1 page 13), mais on n'a pas augmenté le seuil afin qu'il garde du sens sur le plan thématique. Le seuil est fixé à 20 foyers paysans, et thématiquement, on estime que cette quantité pouvait suffire à la création d'une nouvelle paroisse.

**Calibrer le nombre <sup>1</sup>et la nombre et la hiérarchie des paroisses « urbaines ».** On dispose de moins de données empiriques pour les paroisses urbaines que pour les paroisses rurales. <sup>2</sup>On sait que le nombre de paroisses d'une ville est à peu près corrélé à sa population (mais aussi à l'ancienneté de la ville par exemple). On estime aussi que dans les plus grosses villes de la région (Tours, Loches...), le nombre de paroisses ne dépasse pas la dizaine. On ne peut pas reconstituer la hiérarchie du nombre de paroisses par villes en fonction des tailles de celles-ci, mais ces éléments empiriques nous fournissent toutefois des cadres pour le calibrage du modèle.

Dans SimFeodal, un paramètre (`ponderation_creation_paroisse_agregat`) contrôle seul la création de paroisses au sein des agrégats. C'est un « paramètre de mécanisme », quand bien même sa valeur est assez éloignée de l'empirie : elle définit le seuil de foyers paysans <sup>3</sup>par paroisse urbaine, <sup>4</sup>c'est-à-dire pondéré par le nombre de paroisses présentes dans l'agrégat) <sup>5</sup>à partir duquel la probabilité de créer une nouvelle paroisse dans l'agrégat atteint 1. On a procédé par

	Nombre : 1	Auteur : Paul Gourdon	Sujet : Barrer	Date : 11/10/2019 18:18:03 +02'00'
	Nombre : 2	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 11/10/2019 18:20:41 +02'00'
	Attention à la construction du paragraphe. Peut-être atténuer cette phrase. Car sinon on ne comprend pas le "On ne peut pas reconstituer la hiérarchie du nombre de paroisses en fonction de la taille..."			
	Nombre : 3	Auteur : Paul Gourdon	Sujet : Barrer	Date : 11/10/2019 18:21:29 +02'00'
	Nombre : 4	Auteur : Paul Gourdon	Sujet : Barrer	Date : 11/10/2019 18:21:41 +02'00'
	Nombre : 5	Auteur : Paul Gourdon	Sujet : Barrer	Date : 11/10/2019 18:21:47 +02'00'
	Nombre : 6	Auteur : Paul Gourdon	Sujet : Texte inséré	Date : 11/10/2019 18:21:47 +02'00'
	(			

calibrage expérimental, manuel, en testant différentes valeurs pour ce seuil, tout en restant dans des ordres de grandeur acceptables d'un point de vue empirique. Il n'aurait par exemple pas été souhaitable de placer ce seuil à 10 foyers paysans, ce qui aurait impliqué que des paroisses secondaires soient générées dans chaque petit agrégat, résultant thématiquement en une paroisse par hameau par exemple).

Une difficulté particulière du calibrage a porté sur une des spécificités du mécanisme : il consiste à pondérer le nombre de paroissiens par le nombre de paroisses de l'agrégat. Dès lors, la définition des « paroisses de l'agrégat » revêt une importance considérable et est difficile à stabiliser : un agrégat qui change légèrement d'emprise spatiale entre deux pas de temps peut « exclure » une église paroissiale de son emprise, parfois à quelques dizaines de mètres près seulement. Le calibrage du paramètre de pondération de création de paroisse au sein d'agréats a donc été mené conjointement à des ajustements sur les mécanismes de définition des agrégats, et donc à un calibrage aussi des paramètres techniques et de mécanismes impliqués dans la définition spatiale des agrégats.

#### 6.1.1.3 Calibrage des châteaux

Le calibrage des châteaux a été plus simple du point de vue de son <sup>1</sup>préalable, c'est-à-dire de la recherche de données empiriques en vue de l'estimation de la situation initiale et finale. La documentation est précise quant au nombre et aux périodes d'apparition de ces monuments dans l'espace d'étude, et leur nature massive leur a le plus souvent assuré une forte pérennité, atout rare dans l'étude de monuments anciens. Dans l'ensemble, on sait que le nombre de châteaux<sup>5</sup> est très faible, si ce n'est nul, au début de la période. On voit apparaître des châteaux dès le milieu du X<sup>e</sup> siècle, sans doute en réaction au climat de violence qui s'établit à ce moment. En 1200, on estime<sup>6</sup> le nombre de châteaux, en Touraine, à une cinquantaine.

Dans SimFeodal, les châteaux apportent une protection nécessaire aux foyers paysans, mais en contre-partie permettent aussi aux seigneurs de prélever des droits supplémentaires. Leur quantité a donc une importance certaine sur le déroulement des simulations, et l'ajustement des châteaux revêt donc un intérêt majeur en termes d'élément structurant du contexte spatial dans lequel évoluent les foyers paysans.

**Nombre de châteaux.** Dans les premières versions du modèle, des seuils fixes, dépendant de la puissance des seigneurs, avaient ainsi été définis pour caractériser la probabilité de chaque seigneur de construire un château. Dès

---

5. Comme pour la définition des villes, il y a un débat important en archéologie et en histoire sur la définition de ce qu'est un château. Faut-il y inclure, par exemple, les *castra* antiques ? Les mottes castrales ? Dans le cadre de ce modèle, nous avons considéré les « châteaux forts », par nécessité d'établir un référentiel accessible pour la collaboration entre thématiciens et modélisateurs.

6. L'approximation ne porte pas sur le nombre concret de châteaux construits au total dans la région, mais sur leur date de construction : on ne peut alors que mener une estimation du nombre de châteaux déjà existants à cette date.

c'est un nom ?

Si non, il manque le nom.

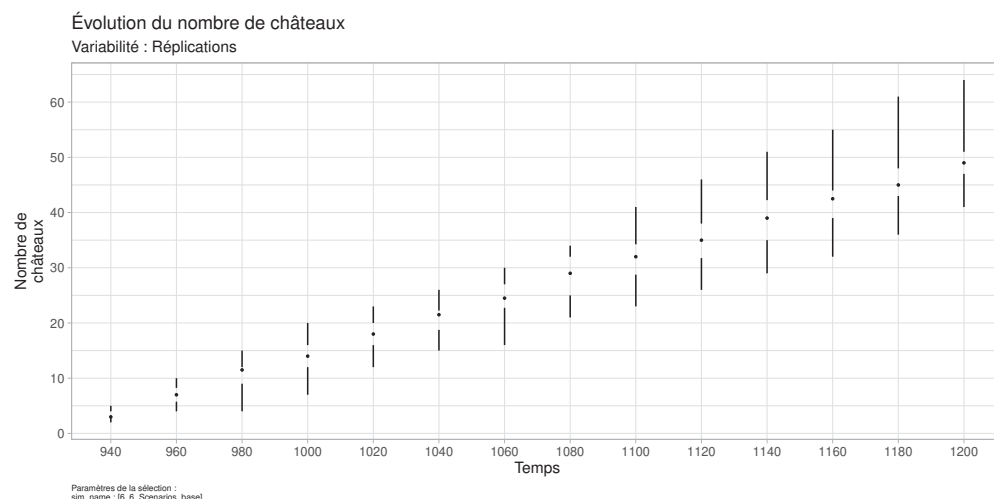


FIGURE 6.2 – Évolution du nombre de châteaux simulé par le modèle calibré.

**Détail des types des châteaux.** Le nombre de châteaux n'est toutefois pas la seule valeur empirique sur laquelle on essaie d'ajuster le contexte. En effet, dans SimFeodal, on distingue plusieurs catégories de châteaux, d'une part selon leur importance (gros et petits châteaux), et d'autre part selon le type de seigneurs qui les ont construits (petits ou grands seigneurs). L'importance des châteaux joue sur leur attractivité vis-à-vis des foyers paysans (un gros château contribue davantage à l'attractivité du pôle d'attraction qui le contient qu'un petit château). Le type de propriétaire joue quant à lui sur le prélèvement des droits associés : un château construit par un grand seigneur a des zones de prélèvement associées plus larges que celles d'un petit seigneur.

Ces distinctions dans le modèle, d'un niveau de détail supérieur à celui de nombreux mécanismes, sont possibles car elles s'appuient sur des typologies empiriques connues sur la région d'étude. Leur prise en compte permet d'affiner le contexte spatial dans lequel les foyers paysans évoluent, aussi bien en matière de répulsion (*push*, par le type de constructeur) que d'attraction (*pull*, par les attractivités différenciées).

À la fin de la période, en Touraine, on estime à une dizaine le nombre de « forteresses »<sup>7</sup>. Dans SimFeodal, on a donc pu calibrer les paramètres liés à la promotion de châteaux de manière à obtenir 10 gros châteaux (correspondants aux forteresses empiriques) et 40 (50 - 10) petits châteaux en fin de simulation.

On connaît de plus, empiriquement, les seigneurs à l'initiative de la construction des châteaux. Dans la plupart des cas (de 40 à 45 châteaux sur les 50), ce sont les seigneurs les plus importants : ducs et comtes d'Anjou et de Touraine, représentés dans SimFeodal par les grands seigneurs. Les 5 à 10 châteaux restant sont issus de seigneurs de moindre importance qui ont toutefois acquis une puissance symbolique et militaire bien supérieure à celles des autres petits seigneurs. Dans SimFeodal, on a donc calibré les paramètres régissant les mécanismes de création de châteaux, différenciés pour les grands et petits seigneurs, afin que les valeurs obtenues par simulation soient similaires aux valeurs em-

7. Trouver ref. avec Samuel pour distinguer les « forteresses importantes » et les « points fortifiés secondaires »





piriquement connues.

Les graphiques de la figure 6.3 présentent les résultats obtenus dans la dernière version de SimFeodal. Ils ne sont pas entièrement satisfaisants, mais résultent d'un compromis entre le calibrage des trois indicateurs que sont le nombre et le type des châteaux, ainsi que le statut de leurs constructeurs.

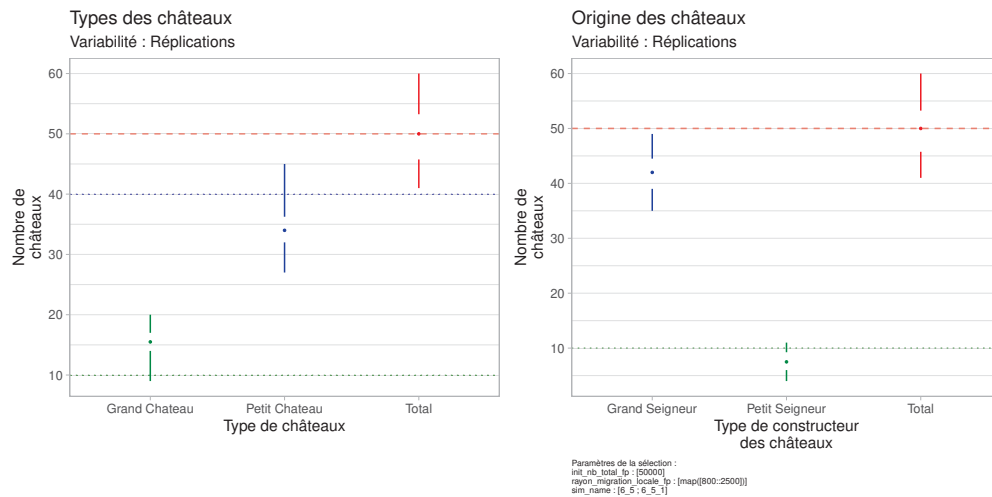


FIGURE 6.3 – Détail de la composition des châteaux en fin de simulation à l'issue du calibrage de SimFeodal.

Les lignes horizontales en pointillés représentent les objectifs à atteindre, définis selon les connaissances empiriques.



Indicateur	Valeur attendue <sup>8</sup>	Moyenne	Médiane	Q1	Q3	<sup>1</sup> Écart-type
<i>Agrégats</i>	200	249	248	244	253	10.45
<i>Gros châteaux</i>	10	15	15	13	17	2.87
<i>Églises paroissiales</i>	300	348	348	338	359	12.96
<i>Distance moyenne entre églises</i>	3 000 m	1 459 m	1 456 m	1 391 m	1 537 m	97 m
<i>Part de foyers paysans isolés</i>	20 %	30 %	30 %	30 %	30 %	0.8 %
<i>Augmentation de la charge fiscale des foyers paysans</i>	x 3	x 2.4	x 2.4	x 2.4	x 2.5	x 0.03

TABLEAU 6.1 – Valeurs des indicateurs numériques en fin de simulation.

On y constate en premier lieu que les ordres de grandeur sont plutôt respectés, à l'exception peut-être de la distance entre églises paroissiales<sup>9</sup>.

Concernant les autres indicateurs, on peut noter qu'ils <sup>2</sup>sont un peu trop importants, ce qui <sup>3</sup>est rendu plus significatif par des variabilités faibles (les médianes et quartiles sont assez proches de la moyenne, et l'écart-type est faible relativement aux grandeurs considérées). Le nombre d'agrégats et d'églises paroissiales simulés dépasse <sup>4</sup>ainsi dans les deux cas d'une cinquantaine les valeurs empiriques correspondantes. La part de foyers paysans isolés en fin de simulation est trop importante de 10%, quand bien même l'intervalle de confiance empirique est assez flou.

Le nombre de gros châteaux dépasse significativement l'objectif, comme on a pu le constater dans la figure 6.3 (gauche), mais <sup>5</sup>on expliquait alors que nous ne parvenions à ajuster davantage cette valeur aux quantités empiriques sans déstabiliser d'autres éléments du modèle.

Le dernier indicateur, l'augmentation de la charge fiscale des foyers paysans, est quant à lui assez satisfaisant : il est certes plus faible que l'objectif empirique fixé (augmentation de 2.4 au lieu de 3 de la charge fiscale moyenne entre le début et la fin de la simulation), mais dans le cas de cet objectif empirique difficile à estimer, cette valeur nous paraît suffisante.

Dans la suite de cette partie, nous menons une analyse plus fine des résultats du modèle, en observant de manière plus précise les valeurs des indicateurs de sortie de simulation qui caractérisent les « trois familles » d'objectifs thématiques : polarisation, hiérarchisation et fixation-dissémination.

8. Objectif en fin de simulation

9. La distance moyenne simulée est ainsi de 1459m, c'est-à-dire deux fois moindre que les estimations empiriques. Notons que cela s'explique notamment par le nombre trop élevé d'églises paroissiales, et par la difficile constitution de cet indicateur : les données empiriques concernent surtout les paroisses rurales, alors que l'on tient ici compte de l'ensemble des églises paroissiales. Les églises paroissiales urbaines, très proches les unes des autres, tirent ainsi considérablement la moyenne des distances à la baisse. Il est difficile, dans le modèle, de différencier les églises paroissiales rurales et urbaines, et on ne peut donc obtenir un indicateur de sortie directement comparable aux données empiriques.

## Page : 13

	Nombre : 1	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 11/10/2019 18:44:52 +02'00'
question très con : pourquoi ya pas le coef de variation en plus ? Pour qu'on puisse comparer la dispersion des différents indicateurs (qui ont des ordres de grandeurs très différents)				
	Nombre : 2	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 11/10/2019 18:47:33 +02'00'
sont en général supérieurs aux valeurs attendues				
	Nombre : 3	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 11/10/2019 18:48:42 +02'00'
ce qui est d'autant plus significatif que la variabilité de ces indicateurs est faible				
	Nombre : 4	Auteur : Paul Gourdon	Sujet : Barrer	Date : 11/10/2019 18:51:41 +02'00'
	Nombre : 5	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 11/10/2019 18:54:11 +02'00'
mais ajuster davantage cette valeur aux quantités empiriques déstabilise d'autres éléments du modèle				

paysannes). Pour savoir si le modèle parvient bien à reproduire le fait stylisé qu'est la polarisation du peuplement, on mobilise donc des indicateurs relatifs à la quantité de pôles et à leur localisation.

Dans cette version du modèle, on constate bien une croissance du nombre de pôles (figure 6.6-a), assez semblable à celle des agrégats en termes de rythme et de valeur. Les valeurs atteintes (environ 250 pôles en fin de simulation) sont très satisfaisantes, particulièrement au regard des résultats de la version 0 du modèle (ref chap 3 ou annexe).

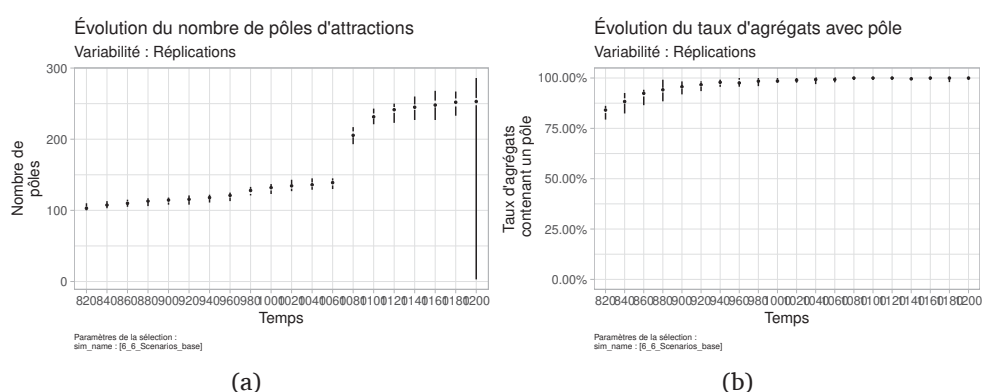


FIGURE 6.6 – Nombre de pôles et part des agrégats comprenant un pôle.

La figure 6.6-b présente elle-aussi des résultats satisfaisants, qui viennent préciser l'analyse précédente. Cette figure présente l'évolution du taux d'agrégats qui sont situés dans un pôle d'attraction. On remarque que ce taux augmente très rapidement pour ensuite se maintenir autour de 100% : cela signifie que tous les agrégats sont situés dans un pôle, et donc que les agrégats se sont bien formés autour de pôles d'attraction plutôt que de manière purement dispersée. Cette figure constitue ainsi un des indices montrant que SimFeodal parvient bien à reproduire le processus de polarisation tel qu'observé empiriquement.

Les cartes de la figure 6.7 permettent de noter que le semis des pôles se confond spatialement avec celui des agrégats, ce qui amène un élément d'interprétation supplémentaire : les agrégats sont bien constitués dans des pôles (paragraphe précédent), mais en plus, tous les pôles semblent contenir un ou des agrégats. Le modèle fait donc émerger une quasi-équivalence entre pôles et agrégats, quasi-équivalence que l'on retrouve chez LE BRAS et LE BRAS-FOLAIN (1976, p. 27-28), en assimilant les pôles à leurs seules églises : « Le village appelle l'église. [...] L'église fait naître le village. ».

c'est quoi l'énorme variabilité sur le dernier pas de temps ?

sur la forme des distributions. Ces sources s'accordent en revanche sur une nette hiérarchisation, avec une distribution qui doit tendre vers les formes log-normales dont l'on retrouve l'existence dans les sociétés contemporaines.

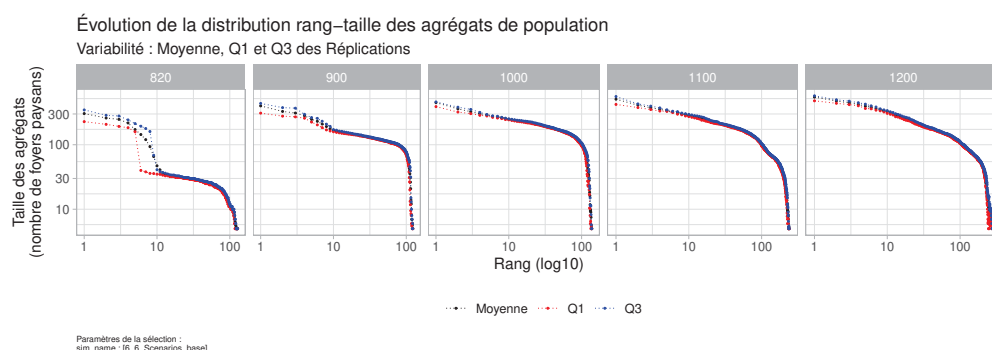


FIGURE 6.8 – Organisation hiérarchique des agrégats.

La figure 6.8 montre une claire hiérarchisation des agrégats : la courbe se « redresse » au cours de la simulation, marque d'une pente croissante. Les valeurs absolues augmentent aussi : les plus gros agrégats voient leur population croître. Le « coude » dans la courbe, qui correspond à la longue traîne de petits agrégats, se réduit.

En parallèle de cette nette hiérarchisation des agrégats, les graphiques de la figure 6.9 permettent de constater une toute aussi nette hiérarchisation des pôles. Cela n'est pas surprenant dans la mesure où on a vu qu'agrégats et pôles se confondaient, ce qui constitue en soi un résultat satisfaisant.

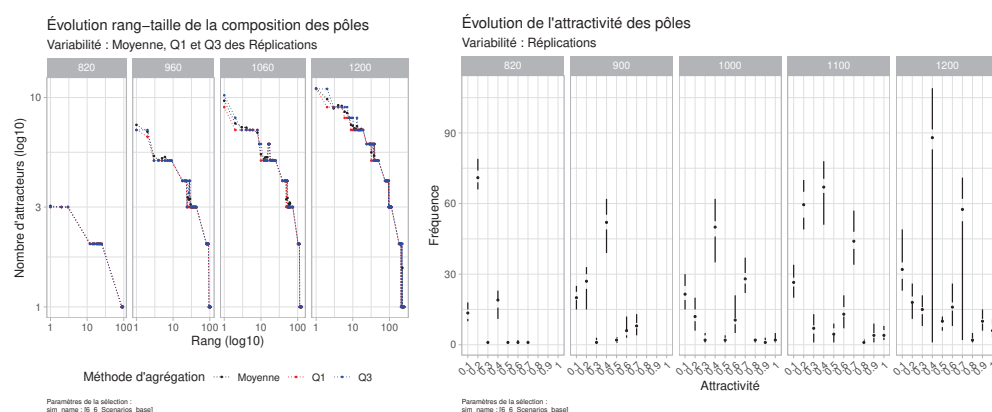


FIGURE 6.9 – Organisation hiérarchique des pôles.

La mesure représentée dans la figure de gauche – le nombre <sup>1</sup>d'attracteurs de chaque pôle – est beaucoup plus « discrète » que le nombre de foyers paysans des agrégats. Il y a moins de modalités (de un à une dizaine d'attracteurs, contre de 5 à plus de 500 foyers paysans), donc moins d'effets possibles de coudes. Ce graphique communique donc plus lisiblement, visuellement, la forte hiérarchisation des pôles. Comme le nombre maximum d'attracteur augmente régulièrement (environ 3 en début de simulation contre plus de 10 à la fin), on peut dire que dans le modèle, cette hiérarchisation est due à une croissance des pôles les plus importants et non à une croissance équitablement répartie. Par exemple, en fin de simulation, sur les 250 pôles, moins de la moitié est composée de plus de trois attracteurs.

---

Nombre : 1	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 15/10/2019 18:57:59 +02'00'
------------	-----------------------	------------------------	------------------------------------

---

là je suis perdu. J'ai bien compris que les pôles attiraient les agrégats et donc les foyers paysans, mais je ne vois pas à quoi corresponds ce nombre d'attracteurs. En fait plus haut, tu parles d'agents attracteurs (églises, châteaux etc.) mais aussi du rôle attracteur des poles. Bref, j'imagine que l'attractivité d'un pole dépend du nombre d'agents attracteurs, mais c'est pas clair pour un non spécialiste. Peut-être préciser dans une phrase dans le paragraphe précédent (celui juste au dessus de la figure 6.9) comme on mesure la hiérarchisation des pôles



50 à 200 paroissiens (51-100 et 101-200) qui deviennent en fin de simulation le mode principal de la distribution. Cet intervalle correspond aux paroisses rurales qui contiennent quelques agrégats ruraux de taille moyenne à faible (figure 6.8).

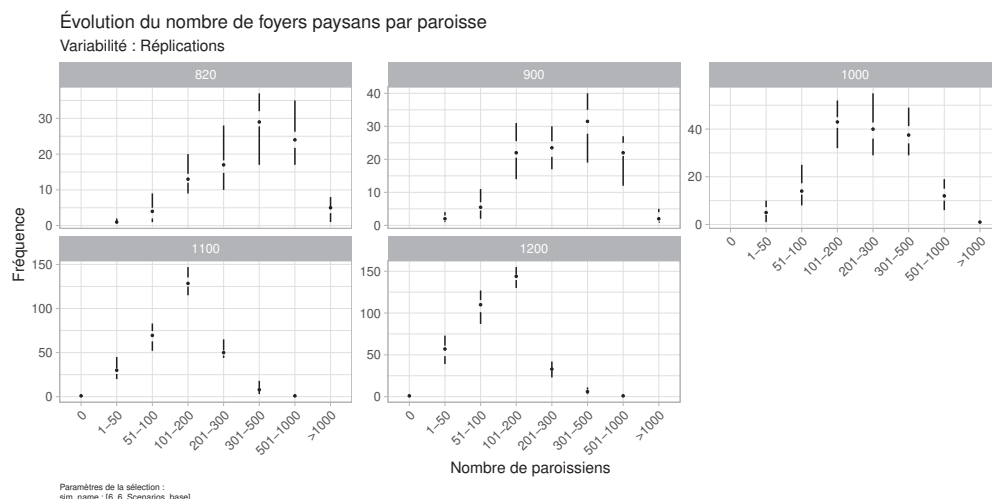






FIGURE 6.11 – Détail de la distribution hiérarchique des paroisses.

SimFeodal parvient donc bien à reproduire la double-densification du maillage paroissial. En milieu rural, le nombre de paroisses augmente jusqu'à assurer une desserte équitable des foyers paysans qui dans le même temps ont tendance à se concentrer, et en milieu urbain, le nombre de paroisses augmente aussi jusqu'à uniformiser le nombre de foyers paysans par paroisse entre 200 et 300.

Dans l'ensemble, SimFeodal montre une bonne capacité à reproduire la hiérarchisation du système de peuplement. Avec les informations dont l'on dispose pour évaluer le modèle, on ne peut qu'être satisfait des tendances présentes dans cette version calibrée de SimFeodal. Le modèle reproduit en effet bien les faits stylisés estimés, même si ces derniers sont formalisés de manière plus floues que par exemple le phénomène de polarisation. Cette incertitude, ~~ou au moins de manque de détail~~, est due à la faiblesse de la documentation empirique sur ces questions thématiques : il est déjà difficile d'avoir une estimation des populations à cette période féodale, ~~alors le détail de la~~ forme de distribution de ces populations est une connaissance encore plus dure à estimer. À ce stade de maturité du modèle, il faudrait sans doute collecter de nouvelles sources historiques pour pouvoir raffiner le comportement du modèle, ou au moins, départager des simulations présentant de légères variations au niveau des indicateurs analysés dans cette sous-partie.

#### 6.1.2.4 Capacité du modèle à simuler la fixation et la dissémination du peuplement

Dans ce dernier objectif thématique, on cherche à vérifier si le modèle parvient bien à reproduire le double processus de fixation des foyers paysans et de dissémination des peuplements dans l'espace.

	Nombre : 1	Auteur : Paul Gourdon	Sujet : Barrer	Date : 15/10/2019 19:06:08 +02'00'
	Nombre : 2	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 15/10/2019 19:06:56 +02'00'
	Nombre : 3	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 15/10/2019 19:07:38 +02'00'
	Nombre : 4	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 15/10/2019 19:09:10 +02'00'

il s'avère difficile

ainsi la forme précise de distribution des populations est encore plus difficile à estimer

Dans le modèle, ces processus devraient s'exprimer sous la forme d'un accroissement des migrations (restructuration) suivi d'une diminution nette (fixation). Au niveau d'observation des agrégats, on devrait observer une couverture croissante, de plus en plus dense, du monde simulé.

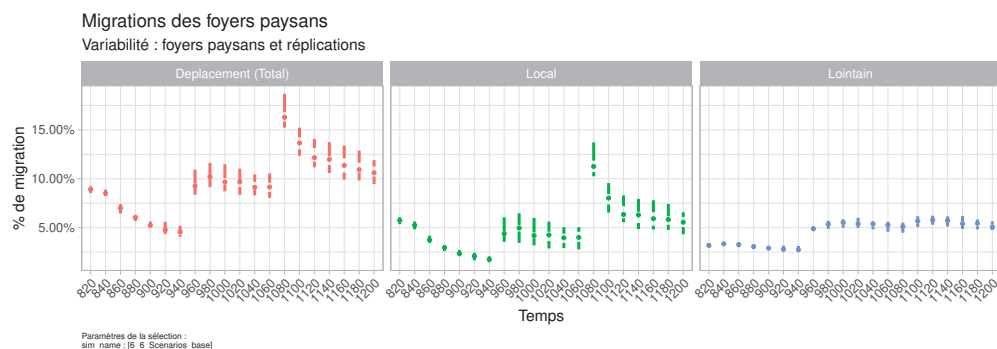


FIGURE 6.12 – Migration des foyers paysans.

**Fixation des foyers paysans.** La figure 6.12 illustre l'évolution des migrations, selon leur type (locale ou lointaine, cf. [chap 2, règles migrations](#)) au cours du temps. On y constate que le modèle produit un motif spécifique, composé de trois phases :

- Avant 960, les migrations diminuent de manière régulière, depuis près de 10% de migrations (total) jusqu'à 5%. Dans le modèle, cette période correspond aux premiers regroupements de foyers paysans. Ces derniers, initialement isolés, rejoignent des pôles locaux (églises rurales par exemple) et y constituent ainsi de petits agrégats ruraux locaux. Ils se fixent dans ces agrégats car il n'y pas encore véritablement de motif d'insatisfaction (religieuse ou de protection du moins).
- En 960, deux éléments exogènes viennent perturber le système : l'augmentation de la pression religieuse (diminution des seuils de distance aux églises) et de la pression de protection (nécessité de s'approcher des châteaux). En conséquence, le nombre de migrations bondit et retrouve son niveau de début de simulation (environ 10%). Le besoin de protection augmente régulièrement à cette période, et les foyers paysans sont donc sans cesse amenés à migrer, ce qui explique que ce niveau de migration semble stable jusqu'à la fin de cette deuxième phase<sup>10</sup>.
- Une nouvelle rupture survient en 1060, là encore en raison d'une augmentation, exogène, de la pression religieuse (les seuils acceptables de distance à l'église paroissiale diminuent encore jusqu'à devenir très restreints). À nouveau, les migrations (uniquement locales cette fois-ci) augmentent de manière abrupte, et, comme dans la première phase, tendent ensuite à diminuer : le niveau maximal d'exigences (religieuse, de protection) est atteint, et les migrations des foyers paysans y pallient.

10. Si de nouveaux éléments exogènes ne venaient pas à nouveau perturber le système après 1060, le niveau de migration diminuerait à son tour, comme avant 960.

---

 Nombre : 1	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 15/10/2019 19:12:36 +02'00'
---	-----------------------	------------------------	------------------------------------

---

 Nombre : 2	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 15/10/2019 19:14:17 +02'00'
---	-----------------------	------------------------	------------------------------------

---

après, les foyers paysans isolés représentent encore une part substantielle de la population (50% d'après la figure 6.4), mais leur poids relatif dans les migrations est devenu bien plus faible que celui des migrations entre agrégats de population. Après une première période de concentration arrive une période de choix hiérarchique pour les foyers paysans, où les différences d'attractivité des agrégats jouent alors un rôle prépondérant. Cela indique aussi qu'à partir de cette période, les agrégats sont pour la plupart pérennes et voient alors se mettre une compétition en place.

**Dissémination du peuplement.** La partition des paroisses, dans le modèle, constitue un observable intéressant pour évaluer la dissémination du peuplement. Comme les paroisses ont comme vocation de desservir la population des foyers paysans, elles constituent un proxy de sa répartition tout au long du temps. Les indicateurs liés (figures 6.15 et 6.16) donnent une lecture satisfaisante du processus de dissémination.

En premier lieu, on note que le nombre d'églises paroissiales augmente de manière régulière au cours du temps, avec un saut entre 1060 et 1080 (comme pour de nombreux indicateurs vu auparavant (figure 6.15). Par rapport aux logiques de création et de promotion, on remarque que le nombre d'églises non paroissiales chute fortement à la même période. Ces églises se voient attribuer les droits paroissiaux, et on peut dès lors affirmer que l'accroît de paroisses de 1080 correspond surtout à des églises rurales puisque ce sont elles qui sont susceptibles d'être promues par le mécanisme (voir chap2, ??).

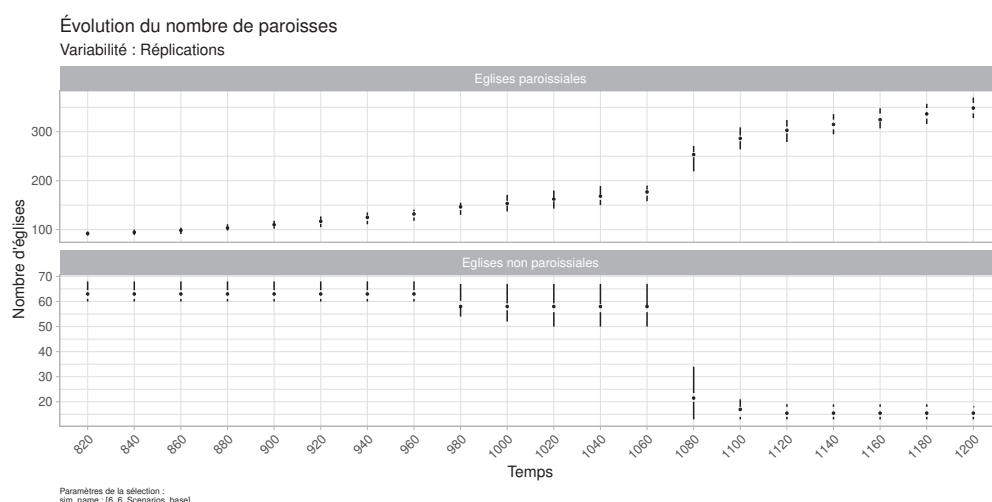





FIGURE 6.15 – Nombre de paroisses.

On constate nettement dans la figure 6.16a une densification généralisée du maillage paroissial. Cette densification est visible à deux niveaux. En premier lieu, le maillage est de plus en plus dense globalement : la superficie moyenne des paroisses diminue (visible aussi dans la figure 6.16b), et visuellement, on constate une certaine homogénéisation et normalisation des paroisses. Les très grandes paroisses initiales (plus de 50 km<sup>2</sup> dans la figure 6.16b), surtout situées dans les marges de la région, disparaissent progressivement à mesure qu'elles sont subdivisées par le mécanisme de création/promotion d'églises paroissiales rurales (ref méca paroisse, chap2).

	Nombre : 1	Auteur : Paul Gourdon	Sujet : Barrer	Date : 15/10/2019 19:22:01 +02'00'
	Nombre : 2	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 15/10/2019 19:22:19 +02'00'
	Dans le modèle, la répartition...			
	Nombre : 3	Auteur : Paul Gourdon	Sujet : Barrer	Date : 15/10/2019 19:22:44 +02'00'
	à			

simulé et très variable localement comme celui de SimFeodal se prête en effet mal à une agrégation de données. Les différentes répliques d'un modèle ne partagent ainsi pas véritablement de référentiel spatial commun en dehors des limites du monde simulé : les positions des agrégats, des églises, des paroisses, sont en effet incomparables d'une réplique à l'autre. Sans données empiriques susceptibles de préciser les taux d'occupation attendus, les mesures de dispersion des semis d'églises, de villages etc., on ne peut pousser l'évaluation visuelle plus avant.

### 6.1.3 Après le calibrage, comment affiner le modèle ?

Dans l'ensemble, l'analyse des résultats montre que cette version calibrée de SimFeodal répond aux attentes en matière de reproduction des faits stylisés identifiés d'après les connaissances expertes. Le modèle aboutit bien sur une polarisation de l'habitat, autour de pôles d'attractions très hiérarchisés et disséminés dans l'espace simulé. Cette polarisation débouche sur une structure spatiale pérenne, due à la fixation des foyers paysans au sein d'agréats qui composent le système de peuplement.









Le modèle SimFeodal est donc satisfaisant vis-à-vis de ce pour quoi il a été conçu, développé et calibré. En comparant les dizaines de versions successives du modèle, ce constat est d'autant renforcé : chaque version a permis des améliorations notables, depuis une version 0 qui parvenait tout juste à une concentration du peuplement jusqu'aux versions plus récentes, préalables à la version 6.6 ici présentée, qui ont renforcé l'ajustement aux connaissances expertes chacune sur des points plus spécifiques. Certains indicateurs, toutefois, et en particulier les indicateurs quantitatifs qui donnent un aperçu global et synthétique des résultats du modèle (section 6.1.2.1), demeurent toutefois relativement éloignés des objectifs fixés. On pourrait donc chercher désormais à prolonger et affiner le calibrage du modèle afin que sa capacité à reproduire les faits stylisés soit améliorée.

Il nous semble toutefois que cette démarche ne vaudrait pas le temps et l'effort qu'il serait nécessaire de lui consacrer. Nous appuyons ce sentiment sur deux raisons principales, la première liée au domaine de l'empirie et la seconde au domaine du modèle.

**Calibrer sans sur-ajuster.** Dès le calibrage des paramètres d'input (section 6.1.1.1), on mentionnait les difficultés du recueil de sources empiriques – matérielles ou littéraires – pour des éléments qui paraîtraient pourtant triviaux et indispensables à tout géographe contemporain. La population de la région et son évolution sur la période considérée, par exemple, ne peut donner lieu qu'à une estimation très grossière. De plus, on ne peut avoir aucune garantie, ni même d'espoir, que de futures sources permettent de mieux les spécifier.

Dans le même temps, on cherche à améliorer l'ajustement des sorties du modèle à des objectifs fixés sur ces sources empiriques floues et lacunaires. L'expertise des thématiciens du groupe de modélisation de SimFeodal a permis de définir ces objectifs, mais ceux-ci présentent un niveau de précision parfois as-

## Page : 25

	Nombre : 1	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 15/10/2019 19:31:20 +02'00'
Oui, mais tu n'as pas vraiment dit pourquoi les mesures synthétiques de l'analyse spatiale n'étaient pas pertinentes. Tu reviens dessus plus tard ?				
	Nombre : 2	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 15/10/2019 19:33:25 +02'00'
s'affirme avec plus de force : (proposition)				
	Nombre : 3	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 15/10/2019 19:35:12 +02'00'
chacune ayant renforcé, sur des points spécifiques, l'ajustement aux connaissances expertes.				
	Nombre : 4	Auteur : Paul Gourdon	Sujet : Barrer	Date : 15/10/2019 19:36:04 +02'00'
assez éloignés				
	Nombre : 5	Auteur : Paul Gourdon	Sujet : Barrer	Date : 15/10/2019 19:36:07 +02'00'
	Nombre : 6	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 15/10/2019 19:36:40 +02'00'
Il serait donc envisageable de chercher...				
	Nombre : 7	Auteur : Paul Gourdon	Sujet : Barrer	Date : 15/10/2019 19:38:49 +02'00'
ce sentiment s'appuie sur deux raisons distinctes : la première liée au domaine de l'empirie...				
	Nombre : 8	Auteur : Paul Gourdon	Sujet : Barrer	Date : 15/10/2019 19:37:48 +02'00'
qu'elle appelle/ qu'elle nécessite (la démarche)				



sez faible. Par exemple, on cherche à ce que SimFeodal puisse atteindre 20% de foyers paysans isolés en fin de simulation. Les versions successives du modèle ont permis de faire descendre ce taux d'environ 60% à environ 20%. En considérant que les connaissances empiriques donnent une estimation plausible située entre 20% et 30%, quel sens y aurait-il à départager des modèles selon qu'ils atteignent 25% ou 27% ?




Ce problème, qui se pose pour chacun des objectifs et attendus du modèle, est un cas typique de sur-ajustement, ou « overfitting ». Sur le plan thématique, chercher à coller absolument à un objectif sans tenir compte du degré d'incertitude de ce dernier, bien plus élevé que celui que le modèle est en mesure de produire, n'a ainsi aucun sens. Sur le plan méthodologique, on peut aussi noter qu'à mesure que le calibrage du modèle progresse, les « gains » de chaque étape se font plus faibles. Chercher à améliorer le calibrage du modèle reviendrait alors à différencier des modèles sur des critères de plus en plus précis, alors même que le niveau de précision de ces critères est en lui-même statique et d'un ordre de grandeur inférieur.

Sans connaissances empiriques supplémentaires, il n'est pas possible de véritablement mieux calibrer SimFeodal, qui a déjà atteint un niveau de détail qui dépasse presque celui des connaissances expertes par le prisme desquelles tout résultat doit être étudié pour porter un sens.

**Calibrer un modèle (très) complexe.** Un autre élément concourt aussi à la difficulté d'améliorer le modèle, cette fois-ci directement lié aux choix de conception de SimFeodal. Comme décrit dans les deux premiers chapitres, SimFeodal est un modèle exploratoire, plus descriptif (KIDS) que parcimonieux. Les paramètres y sont très nombreux, de même que les mécanismes qu'ils ajustent. Les agents du modèle interagissent, et, avec eux, les mécanismes qui les caractérisent sont aussi entrecroisés et imbriqués. Si l'on a une connaissance parfaite de chaque mécanisme (puisqu'on les a conçus et implémentés), il est difficile d'avoir une connaissance de même ordre sur l'aboutissement de leurs interactions, c'est la raison d'être de la simulation.

Dans des modèles moins descriptifs, mêmes complexes, on peut avoir des idées relativement précises de la combinaison des mécanismes, et donc une bonne intuition des sorties de simulation qui <sup>1</sup> en aboutiront. Dans un modèle comme SimFeodal, les paramètres sont trop nombreux pour que ce soit réellement possible, et les résultats contre-intuitifs ont été fréquents dans les différentes étapes de paramétrage et de calibrage du modèle. Ces éléments contre-intuitifs, surprenants, sont extrêmement stimulants du point de vue de la connaissance des effets du modèles. Ils placent les modélisateurs et <sup>2</sup> évaluateurs dans une démarche abductive où l'identification de l'explication <sup>3</sup> de ces contre-intuitions permet de mieux comprendre le fonctionnement effectif du modèle et <sup>3</sup> à partir de là du système modélisé.

Dans la recherche d'un calibrage mieux adapté, cette approche abductive ne permet toutefois pas véritablement de progresser. À mesure que le calibrage se précise, la complexité du modèle apparaît renforcée. Comme dans le prin-

	Nombre : 1	Auteur : Paul Gourdon	Sujet : Barrer	Date : 15/10/2019 19:41:23 +02'00'
	Nombre : 2	Auteur : Paul Gourdon	Sujet : Texte surligné des diverses surprises (proposition pour alléger la phrase)	Date : 15/10/2019 19:42:39 +02'00'
	Nombre : 3	Auteur : Paul Gourdon	Sujet : Texte surligné et, à partir de là, du système modélisé	Date : 15/10/2019 19:43:06 +02'00'

cipe des vases communicants, une modification de valeur d'un paramètre est presque systématiquement assortie d'une réaction imprévu (et imprévisible) sur un indicateur différent de celui sur lequel le paramètre devait agir. Chaque modification de paramètre devrait donc donner lieu à de nouveaux ajustements d'autres paramètres, jusqu'à obtenir un ensemble de valeurs de paramètres stable.

Pour améliorer le calibrage, et éviter ces allers-retours imprévisibles entre paramètres, il manque alors une vision plus globale des réactions engendrées par les modifications de chaque paramètre. Une telle vision permettrait de réorganiser les paramètres, pour chaque indicateur de sortie, en ensembles thématiques dont l'on ferait systématiquement co-varier les valeurs testées afin d'approcher de l'objectif de manière harmonisée. Pour mieux comprendre les aboutissements du modèle, il est dès lors nécessaire de l'explorer de manière plus systématique, en essayant de comprendre l'influence réelle de chacun des paramètres plutôt que de tâtonner, de manière experte certes, en agissant sur les paramètres qui nous semblent intuitivement les plus importants. Ce sont ces objectifs de compréhension plus fine du comportement du modèle que nous allons poursuivre dans la suite de ce chapitre, au moyen d'analyses de sensibilité.



## 6.2 Analyser la sensibilité de SimFeodal

Parmi les nombreuses méthodes dédiées à l'évaluation de modèles, il en est une que l'on retrouve dans tous les manuels et dans la plupart des articles dédiés à la présentation de modèles. Il s'agit de l'analyse de sensibilité, catégorie en fait plurielle qui regroupe l'ensemble des méthodes vouées à tester ou à explorer la stabilité d'un modèle face aux éléments qui le composent : poids des *inputs* dans les résultats obtenus ; variabilité des résultats selon les valeurs de paramètres choisies ; variabilité des résultats due à l'aléa etc. Ces méthodes sont extrêmement nombreuses et constituent presque un champ scientifique entier, [1] é à l'évaluation de modèles, [2] soient ils à base d'agents ou statistiques.

Parmi celles-ci, les méthodes les plus classiques (CROOKS et al. 2019, p. 257) visent à « déterminer l'influence des paramètres sur les sorties du modèle. » (GINOT et MONOD 2005, p. 75). Il s'agit de faire varier les valeurs des paramètres et de mesurer les écarts résultant dans les sorties. Le plus souvent, cette mesure est quantitative, par exemple sous la forme d'un « indice de sensibilité » qui dépend des variations des sorties mais aussi de l'amplitude de la variation des paramètres<sup>13</sup>.

Les analyses peuvent être menées paramètre par paramètre, en conservant par exemple des valeurs fixes pour un jeu de paramètres (5) sus de calibrage) de base et en faisant varier un unique paramètre à la fois (analyse de type OFAT, « *one factor at a time* »). On peut aussi procéder en croisant des valeurs pour tous les paramètres, c'est-à-dire en analysant la sensibilité du modèle aux interactions entre paramètres.

Les analyses de sensibilités sont largement recommandées [6] comme une pratique indispensable à la validation de modèle : on les retrouve dans l'ensemble des schémas relatifs à la démarche d'évaluation du chapitre 3. Au-delà de la validation des modèles, ce type d'analyses constitue un outil extrêmement utile pour aider à la compréhension d'un modèle, sans nécessairement chercher à en quantifier la validité. [7] Cette méthode repose sur l'exploration d'un modèle par le prisme de ses réactions face aux paramètres choisis, et permet ainsi de mener une étude approfondie de l'influence des paramètres. Dans certains modèles, une analyse de ce type a par exemple permis de rendre plus parcimonieux un modèle KISS, en mettant en lumière le peu d'influence d'un paramètre sur l'ensemble des sorties d'un modèle. C'est le cas dans le travail de thèse de Clara SCHMITT, où une analyse de sensibilité a révélé la relative inutilité de l'un des 6 paramètres mobilisés dans le modèle SimpopLocal (SCHMITT 2014, p. 224-225).

Dans le cadre d'un modèle exploratoire, où les très nombreux paramètres com-

13. Cette prise en compte de la variation des valeurs de paramètre, par exemple dans l'indice proposé par CROOKS et al. (2019, p. 258) et dérivé de celui de HAMBY (1994) (in O'SULLIVAN et PERRY 2013, p. 201), permet de s'assurer, lors de la comparaison de la sensibilité des paramètres, que les valeurs comparées sont bien comparables. Pour prendre l'exemple du modèle de Schelling, il s'agit de s'assurer qu'on ne compare pas une variation de 0,1% du seuil de tolérance avec une variation de 20% dans la part d'espace vide. Dans ces indices, les amplitudes des valeurs testées sont en fait normalisées.

 Nombre : 1 s	Auteur : Paul Gourdon	Sujet : Texte inséré	Date : 19/10/2019 16:49:05 +02'00'
 Nombre : 2	Auteur : Paul Gourdon	Sujet : Barrer	Date : 19/10/2019 16:50:07 +02'00'
 Nombre : 3 à part entière	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 19/10/2019 16:49:59 +02'00'
 Nombre : 4 produits	Auteur : Paul Gourdon	Sujet : Barrer	Date : 19/10/2019 16:52:57 +02'00'
 Nombre : 5 s	Auteur : Paul Gourdon	Sujet : Texte inséré	Date : 19/10/2019 16:54:29 +02'00'
 Nombre : 6 souvent présentées	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 19/10/2019 16:55:20 +02'00'
 Nombre : 7 En effet, cette...	Auteur : Paul Gourdon	Sujet : Texte inséré	Date : 19/10/2019 16:56:00 +02'00'

portent vraisemblablement une part de redondance, l'ambition n'est pas de rechercher à réduire la masse de paramètres à tout prix, mais plutôt d'aider à comprendre lesquels ont la plus grande influence sur le modèle.

Dans le cadre de SimFeodal, l'analyse de sensibilité doit permettre de gagner en compréhension du modèle, et en conséquence, des dynamiques modélisées.

La nature exploratoire et descriptive de SimFeodal rend l'application des méthodes classiques de l'analyse de sensibilité assez difficile : les paramètres ne sont pas tous quantitatifs, certains fonctionnent par paires, par grappes etc. Dans cette partie, nous nous en tiendrons à une analyse de sensibilité **gros-sière**, orientée vers une évaluation visuelle, à l'instar des autres démarches d'évaluation du modèle mises en places. Nous nous inscrivons pleinement dans le raisonnement de Joanne HIRTZEL, d'autant plus que le dit raisonnement est tenu dans une thèse dont l'analyse de sensibilité de modèles descriptifs est l'un des enjeux principaux :

« Ces différents constats nous ont conduit à procéder à des analyses de sensibilité locales, avec la méthode OAT<sup>14</sup>, en modifiant les valeurs de chacun des paramètres les uns après les autres, toutes choses égales par ailleurs, c'est-à-dire tous les autres paramètres étant fixés à leur valeur par défaut [...]. Ce choix n'est pas unique dans la modélisation individu-centrée : la méthode OAT est utilisée dans plusieurs travaux en géographie ou en écologie (Ginot et al., 2006 ; Sanders et al., 2006 ; Laperrière et al., 2009 ; Schouten et al., 2014).

Nous n'avons pas jugé indispensable le calcul d'indices de sensibilité pour étudier la sensibilité des résultats de simulation aux différents paramètres du modèle. Une analyse graphique à la manière de Sanders et al. (2006a), Laperrière et al. (2009) ou encore Schouten et al. (2014) nous a paru suffisante, dans un premier temps. Ainsi, pour reprendre les termes évoqués dans les deux sous-parties précédentes, l'analyse de sensibilité présentée dans ce chapitre est une analyse locale (OAT), avec des évaluations qualitatives de l'impact de l'incertitude émanant des valeurs de paramètres sur différents résultats de simulation. »






HIRTZEL 2015, p. 251-252

### 6.2.1 Méthodologie - Analyse visuelle de sensibilité.

Dans cette partie, nous décrivons et justifions brièvement la méthodologie mise en place pour l'analyse de sensibilité de SimFeodal. Les sources informatiques, précises, de la démarche sont disponibles dans le dépôt du modèle (**mettre URL dossier**). Le détail des paramètres, les outputs du modèle, les traitements et les sorties graphiques sont disponibles quant à eux à cette adresse : **depot these / anasensib**

14. [C'est un autre acronyme de « one factor at a time », identique à OFAT que nous utilisons dans cette partie.]

## Page : 29

	Nombre : 1	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 19/10/2019 16:57:13 +02'00'
Ainsi, pour SimFeodal,				
	Nombre : 2	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 19/10/2019 16:57:30 +02'00'
'				
	Nombre : 3	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 19/10/2019 17:04:34 +02'00'
terme un peu fort je trouve. Par ailleurs, on découvre ici qu'il s'agira d'une évaluation visuelle, alors que les paragraphes précédent et la note de bas de page suggèrent plus le quantitatif.				
Comme c'est un point important de ta démarche, je me demande si en plus de l'aspect "difficile" il ne faut pas réinsister sur la démarche "compréhensive" de cette approche de l'analyse de sensibilité. Et annoncer la question de l'évaluation visuelle plus franchement dans le début de ce paragraphe.				
	Nombre : 4	Auteur : Paul Gourdon	Sujet : Texte inséré	Date : 19/10/2019 17:05:27 +02'00'
A ce tire,				
	Nombre : 5	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 19/10/2019 17:07:25 +02'00'
notamment sur la question de l'analyse de sensibilité des modèles descriptifs.				



### 6.2.1.1 Calcul de la sensibilité.

Nous avons choisi de mener l'analyse de sensibilité de SimFeodal en empruntant l'approche *OFAT*, c'est-à-dire en faisant varier les paramètres un par un. Il convient en effet de rappeler que ce modèle est caractérisé par un nombre important de paramètres : près de 70. En ne choisissant que 5 valeurs pour chaque paramètre et pour croiser tous les paramètres, une analyse de type plan complet demanderait l'exécution de  $5^{70} \times 20_{[\text{réplifications}]}$ , soit environ  $10^{50}$  simulations... Une approche plus simple, de type *OFAT*, nous paraît dès lors la seule à être applicable dans le cas de SimFeodal.

**Paramètres.** Dans une analyse de sensibilité, le premier choix à faire concerne les paramètres à analyser. Dans les modèles statistiques ou KISS, l'analyse de chacun des paramètres est une évidence. Dans des modèles plus descriptifs tels que SimFeodal (ou les modèles analysés par HIRTZEL (2015)), on procède souvent à une sélection des paramètres afin de réduire l'ampleur de la tâche. Par exemple, quand certains paramètres ont un ancrage empirique fort, on peut considérer qu'ils se comportent ~~plus~~ comme des constantes que comme des paramètres, ~~ils~~ n'auront jamais à varier, et ne requièrent ainsi pas d'être analysés.

Dans SimFeodal, on aurait pu par exemple ~~se passer d'analyser~~ les paramètres de contexte les plus inscrits dans les connaissances empiriques. Pourtant, comme nous envisageons d'éprouver le modèle sur des scénarios susceptibles de faire varier tous les types de paramètres, ~~il~~ nous a semblé important de tester aussi ces paramètres. On a ~~opté~~ pour une analyse de sensibilité de l'ensemble des paramètres du modèle. SimFeodal comporte dans les faits 70 paramètres, mais en pratique, certains n'ont aucun sens quand ils sont mobilisés seuls. Par exemple, deux paramètres définissent le rayon minimum et maximum des zones de prélèvement. Il n'y a pas grand sens à faire varier l'un sans l'autre. Dans l'analyse de sensibilité, nous considérons ces deux paramètres comme un unique paramètre à analyser, qui ne prend pas la forme d'une valeur numérique simple, mais plutôt celle d'une étendue.

Pour cette analyse de sensibilité, nous avons au final analysé 57 « paramètres », certains de forme numérique simple, d'autres sous formes d'étendues, et enfin quelques-uns ayant des structures plus complexes (étendues changeant au cours du temps par exemple).

**Méthode.** La méthode choisie est simple : on définit un jeu de paramètres de base, issu du calibrage, et pour chacun des paramètres, on exécute un ensemble de simulations en faisant varier ce paramètre autour de sa valeur de base. Comme pour toute simulation d'un modèle stochastique, il est indispensable de procéder à des réplifications de ces simulations.

Comme le nombre de paramètres était déjà important, que le nombre de réplifications (20) l'était lui aussi, on a choisi de mener une analyse de sensibilité assez réduite, en ne testant, pour chaque paramètre, que 5 valeurs différentes. Ce nombre amène déjà à l'exécution de  $57_{[\text{paramètres}]} \times 5_{[\text{valeurs}]} \times 20_{[\text{réplifications}]} = 5700$

 Nombre : 1 davantage	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 19/10/2019 17:13:06 +02'00'
 Nombre : 2 qu'ils	Auteur : Paul Gourdon	Sujet : Texte inséré	Date : 19/10/2019 17:13:13 +02'00'
 Nombre : 3 écarter l'analyse	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 19/10/2019 17:14:12 +02'00'
 Nombre : 4 ---	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 19/10/2019 17:14:40 +02'00'
 Nombre : 5 ---	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 19/10/2019 17:15:10 +02'00'
 Nombre : 6 donc	Auteur : Paul Gourdon	Sujet : Texte inséré	Date : 19/10/2019 17:15:16 +02'00'

simulations, ce qui est une quantité importante de simulations au regard de toutes celles qui ont été menées dans les étapes de paramétrage et d'évaluation visuelle du modèle.

**Étendue.** Dans une analyse de sensibilité classique, on cherche à tester des variations de paramètres comparables, c'est-à-dire centrées autour des valeurs par défaut et présentant des variations relatives de même ampleur. Par exemple, pour un modèle dont le premier paramètre vaut 10 et le second 100, on cherchera à répartir les valeurs testées de manière comparable : le premier paramètre sera testé aux valeurs [0, 5, 10, 15 et 20], et le second pour [0, 50, 100, 150 et 200].

Dans un cas réel, cette règle est difficile à suivre : on a ici pris l'exemple de paramètres quantitatifs « de stock », qui ne sont comparables qu'entre eux. Dans SimFeodal, certains paramètres ont des structures bien plus difficilement comparables, à l'instar des étendues <sup>1</sup>~~plus haut~~. Comment rendre comparable cinq variations autour d'un stock de 10, et cinq variations autour d'une étendue de 1500 m à 5000 m ? L'étendue des possibles transformations pour ces étendues est bien plus importante : diminution, augmentation, translation... Contrairement aux variables de stocks qui ne comportent qu'une dimension, les étendues sont en effet composées de deux dimensions, et la quantité de variation possibles suit donc une loi de puissance.

Ce problème se pose ~~de manière plus importante encore~~ <sup>2</sup> pour les paramètres qui évoluent en fonction du temps. Par exemple, le paramètre qui agit sur la satisfaction protection des foyers paysans, a une valeur composite (de type *map*, voir ?? dans le chapitre 2) qui vaut « 0 entre 800 et 940 ; 0.2 en 960 ; 0.4 en 980 ; 0.6 en 1000 ; 0.8 en 1020 ; et 1 à partir de 1040 ». C'est donc un paramètre à trois dimensions : une valeur (dimension 1) qui évolue selon des intervalles de temps (dimensions 2 et 3). Pour l'analyser dans son entièreté, il faudrait supprimer la variation en testant plusieurs valeurs, mais aussi changer le rythme de cette variation, en décaler l'étendue temporelle, en changer l'intensité etc. Sous bien des aspects, ce paramètre peut être considéré comme qualitatif, et les variations qu'on lui appliquera dans l'analyse de sensibilité ne peuvent être que très subjectives et intrinsèquement différentes.

Face à ces difficultés, nous avons adopté une position générale acceptant la subjectivité, la non-comparabilité, mais cherchant à explorer des valeurs « caractéristiques » pour ce type de paramètres : activation et désactivation du mécanisme associé, valeur de base, et augmentation et diminution de l'ampleur de la valeur de paramètre. Dans l'ensemble, les valeurs de paramètres choisies (**mettre les tableaux en annexe**) ne sont que peu comparables de manière numérique, mais elles apportent un éclairage précieux sur le comportement du modèle en fonction de ses paramètres.

**Outputs.** Le chapitre 5 donnait des ordres de grandeur de la masse des données produites par le modèle, autour de 10 Mo pour une simulation. Avec 5700 simulations, un enregistrement complet des données aurait représenté plus de 50 Go de données et plus de 50 milliards de lignes de données à pré-

---

 Nombre : 1	Auteur : Paul Gourdon	Sujet : Barrer	Date : 19/10/2019 17:18:48 +02'00'
---	-----------------------	----------------	------------------------------------

---

 Nombre : 2	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 19/10/2019 17:19:58 +02'00'
---	-----------------------	-------------------------------	------------------------------------

---

avec plus de force

traiter, intégrer et analyser dans la base de données.

Cette masse de données pose des problèmes techniques tant que méthodologiques. Sur le plan technique, les solutions de stockage et d'interrogation de données utilisées pour l'exploration des données (chap 3.3) ne sont pas adaptées à une telle masse, du moins sans avoir recours à des infrastructures physiques coûteuses (serveurs équipés de cartes graphiques extrêmement puissantes etc.). En termes de capacité d'archivage, il est donc obligatoire de réduire considérablement la masse des données.

Sur le plan méthodologique, l'exploration visuelle de milliers de simulations organisées selon les variations des valeurs d'une soixantaine de paramètres nous semble répondre à des caractéristiques très particulières. Les méthodes et outils d'exploration interactive mis en place jusque là (chap 5.4) ne sont pas adaptés à ce type d'usages. Nous avons donc choisi, par soucis de lisibilité, de réduire autant que possible la production d'indicateurs de sorties de simulations pour cette analyse de sensibilité.

Dans un premier temps, nous n'enregistrons l'état du modèle à son pas de temps final (en 1200) plutôt que tout au long de son déroulement. L'analyse de sensibilité doit se concentrer sur aussi peu d'objectifs que possible, et il n'est pas véritablement gênant de perdre l'aspect dynamique et la possibilité de comparer les rythmes du changement entre les simulations. Pour les mêmes raisons, on a décidé de n'extraire que des données très agrégées du modèle. L'analyse de sensibilité s'appuie sur des résultats à l'échelle globale du modèle, et il n'est pas utile, à ce stade de l'analyse, d'enregistrer les indicateurs individuels relatifs aux agents du modèle.

#### Encadré 6.1 : Différencier indicateurs contextuels et émergents.

Dans les chapitres précédents, on a présenté de manière extensive les objectifs – dans tous les sens du terme – poursuivis par le modèle. On mobilise des indicateurs de sortie de simulation pour rendre compte de l'état du modèle, et le but est que ces indicateurs soient proches des objectifs empiriques formulés. Ces indicateurs peuvent être qualitatifs (courbes d'évolution...) ou quantitatifs (taux atteints en fin de simulation). En définissant un ensemble d'objectifs, qualifiés (allures de courbes, ordres de grandeur...) sinon quantifiés (les indicateurs numériques), on esquisait une grille d'évaluation du modèle.

Au sein de ces indicateurs de sortie de simulation, il est une distinction que l'on n'a pas encore réalisée, qui nous semble indispensable pour l'analyse de la sensibilité du modèle, où il est nécessaire de réduire autant que possible la quantité d'indicateurs calculés : le distinguer entre des indicateurs que l'on pourrait caractériser comme « émergents », et d'autres pourraient être qualifiés de « contextuels ».

- Le premier type, les indicateurs émergents, sont les indicateurs les plus classiques en modélisation agent : on s'attache à leur évaluation parce qu'ils sont entièrement générés par la combinaison et

## Page : 32

	Nombre : 1	Auteur : Paul Gourdon	Sujet : Barrer	Date : 19/10/2019 17:37:06 +02'00'
	Nombre : 2 D'abord	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 19/10/2019 17:37:45 +02'00'
	Nombre : 3 que l'état final (1200) du modèle	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 19/10/2019 17:37:38 +02'00'
	Nombre : 4 phrase à reprendre et alléger je pense. "n'est pas véritablement gênant" , au-delà de la tournure, le lien n'est pas évident avec la première proposition de la phrase.	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 19/10/2019 17:35:23 +02'00'
	Nombre : 5 et non pas l'ensemble des états intermédiaire	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 19/10/2019 17:38:36 +02'00'
	Nombre : 6 En effet,	Auteur : Paul Gourdon	Sujet : Texte inséré	Date : 19/10/2019 17:38:45 +02'00'
	Nombre : 7 mobilisée	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 19/10/2019 17:39:50 +02'00'
	Nombre : 8 pour laquelle il	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 19/10/2019 17:40:25 +02'00'
	Nombre : 9	Auteur : Paul Gourdon	Sujet : Barrer	Date : 19/10/2019 17:40:28 +02'00'
	Nombre : 10 	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 19/10/2019 17:40:48 +02'00'
	Nombre : 11 ceux	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 19/10/2019 17:40:47 +02'00'

Indicateur de sortie de simulation	Valeur attendue	Type	Dépendances directes
Nombre d'agrégats	200	Émergent	–
Nombre de châteaux	50	Contextuel	probabilités de construction de châteaux (PS et GS)
Nombre de gros châteaux	10	Émergent	–
Nombre de seigneurs	200	Contextuel	paramètre dédié : (objectif_nombre_seigneurs)
Nombre d'églises paroissiales	300	Émergent	–
Distance moyenne entre églises	3 000 m	Émergent	–
Part de foyers paysans isolés	20 %	Émergent	–
Augmentation de la charge fiscale des foyers paysans	x 3	Émergent	–
Nombre de Foyers Paysans	50 000	Contextuel	nombre initial; taux de croissance
Densité de population	8 feux/km <sup>2</sup>	Contextuel	nombre de foyers paysans; taille du monde

TABLEAU 6.2 – Les indicateurs de sortie de simulation quantitatifs de SimFeodal. *Les lignes grisées désignent les indicateurs « contextuels ».*

On a choisi<sup>1</sup> de mener l'analyse de sensibilité sur les indicateurs numériques synthétiques, présentés dans le tableau 6.2. Parmi ceux-ci, on a conservé uniquement les 6 indicateurs « émergents » (encadré 6.1) : nombre d'agrégats; nombre de grands châteaux; nombre d'églises paroissiales; distance moyenne entre églises paroissiales; part de foyers paysans isolés; et augmentation de la charge fiscale.

Dans la construction et l'évaluation du modèle, ce sont ces indicateurs que l'on a le plus observés, et s'ils n'informent pas sur l'ensemble des objectifs thématiques du modèle, ils permettent au moins de qualifier sommairement son comportement d'ensemble.

**Computation.** Dans une analyse de sensibilité portant sur autant de simulations, la masse de données n'est pas la seule difficulté technique. Le temps de calcul représente ainsi un obstacle qui peut devenir un véritable point de blocage, tant il peut s'allonger jusqu'à devenir irréalisable. Sur un ordinateur individuel, avec les valeurs de paramètre issues du calibrage, chaque simulation demande en moyenne (basse) 5 minutes de calcul. Multiplié par les 5700 simulations requises, l'exécution des simulations nécessaires à l'analyse de sensibilité réclame alors un temps de calcul de près de 20 jours<sup>2</sup> sur un ordinateur individuel, sans véritable droit à l'erreur sous peine<sup>3</sup> d'avoir à recommencer ce quasi-mois de calcul.

Pour que cette analyse soit effectuée dans un délai plus raisonnable, nous avons utilisé les capacités de calcul d'un serveur du laboratoire ainsi que d'un ser-

	Nombre : 1	Auteur : Paul Gourdon	Sujet : Barrer	Date : 19/10/2019 17:59:05 +02'00'
	Nombre : 2	Auteur : Paul Gourdon	Sujet : Barrer	Date : 19/10/2019 18:01:34 +02'00'
	Nombre : 3	Auteur : Paul Gourdon	Sujet : Barrer	Date : 19/10/2019 18:01:41 +02'00'
	Nombre : 4 peine	Auteur : Paul Gourdon	Sujet : Texte inséré	Date : 19/10/2019 18:01:46 +02'00'



veur de calcul géré par la TGIR Huma-Num<sup>15</sup>. En distribuant les exécutions de simulation sur la quarantaine de processeurs que compte ce serveur, tout en laissant tourner les simulations uniquement de nuit pour ne pas gêner les autres utilisateurs de cet équipement collectif, toutes les simulations requises ont été exécutées en 3 jours, ce qui représente un investissement temporel bien plus raisonnable.

### 6.2.1.2 Analyse Quantitative - Filtrage des paramètres

Devant la masse de paramètres à analyser, nous avons réalisé qu'il était difficile de mener une analyse visuelle directe de la sensibilité des paramètres. Il a été <sup>1</sup>choisi de mener <sup>2</sup>la première étape, quantitative, pour permettre de trier et de filtrer un sous-ensemble de paramètres dont une étude plus approfondie serait alors possible.

Pour cette étape, on a raisonné à l'échelle agrégée du paramètre, en cherchant à mesurer la variabilité des indicateurs de sortie <sup>5</sup>choisis provoquée par la variation des valeurs de paramètres.

**Normalisation des objectifs.** Un premier problème est que les indicateurs de sortie de simulation choisis pour caractériser les objectifs <sup>6</sup>dont les référentiels sont très différents. Une variation de 5 grands châteaux (objectif de 10) montre une sensibilité plus forte qu'une variation de 5 agrégats (objectif de 200). Il est alors nécessaire de procéder à un « centrage » des données issues de la simulation. Ce centrage peut être effectué, classiquement, sur les valeurs obtenues, pour obtenir une moyenne valant 0, comme on le fait lors d'une normalisation classique de données.

Pour l'étude de SimFeodal, nous avons préféré centrer les données de chaque indicateur relativement aux objectifs numériques identifiés. Par exemple, une simulation produisant 250 agrégats, quand l'objectif numérique est de 200, aura la valeur de + 50 pour l'indicateur « nombre d'agrégats ». Cela permet de mesurer les comportements de manière plus thématique qu'un centrage statistique autour de 0. Pour chaque indicateur, on a soustrait à la valeur obtenue par simulation la « valeur attendue » <sup>7</sup>déterminée thématiquement comme objectif (voir le tableau 6.1). Après cela, ~~pour chaque indicateur pris individuellement,~~ les données générées par chaque simulation deviennent comparables.

Pour atteindre une comparabilité plus importante entre des données hétérogènes, on procède aussi usuellement à une « réduction » des données, c'est-à-dire à la normalisation de leur variabilité (écart-type). Classiquement, cette étape consiste à diviser chaque résultat par l'écart-type. On obtient ainsi, pour chaque série de donnée, un écart-type de 1, qui permet alors de comparer la variabilité de ces séries hétérogènes. Comme pour le centrage, on a choisi de réduire les données en fonction de données connues plutôt qu'autour d'une valeur abstraite de 1. On a préféré pour cela se référer à nos données simulées de référence, c'est-à-dire issues des simulations présentées dans la première partie

15. [www.huma-num.fr](http://www.huma-num.fr)

	Nombre : 1	Auteur : Paul Gourdon	Sujet : Barrer	Date : 19/10/2019 18:04:17 +02'00'
	Nombre : 2	Auteur : Paul Gourdon	Sujet : Barrer	Date : 19/10/2019 18:04:48 +02'00'
	Nombre : 3	Auteur : Paul Gourdon	Sujet : Texte inséré	Date : 19/10/2019 18:04:21 +02'00'
	Nombre : 4	Auteur : Paul Gourdon	Sujet : Texte inséré	Date : 19/10/2019 18:05:50 +02'00'
	un premier tri quantitatif afin de filtrer un sous..., rendant possible une étude plus approfondie.			
	Nombre : 5	Auteur : Paul Gourdon	Sujet : Barrer	Date : 19/10/2019 18:06:24 +02'00'
	Nombre : 6	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 19/10/2019 18:07:12 +02'00'
	???			
	Nombre : 7	Auteur : Paul Gourdon	Sujet : Barrer	Date : 19/10/2019 18:08:22 +02'00'

de ce chapitre, après calibrage. Les données centrées ont ainsi été ensuite divisées par l'écart-type mesuré sur les simulations de cette version de référence, pour chaque indicateur numérique, plutôt que par l'écart-type des simulations issues de l'analyse de sensibilité.

Au final, ce procédé de normalisation est très proche d'une normalisation statistique classique, mais se base sur des valeurs qui ont un sens dans le modèle plutôt que sur les valeurs « abstraites » que constituent une moyenne à 0 et un écart-type de 1 : le référentiel est ici fixé à partir objectifs numériques, et la dispersion autour des résultats des simulations de la version calibrée du modèle présentée auparavant (section 6.1).

$$\text{valeur\_normalisée}_{\text{indicateur}_i} = \frac{\text{valeur}_{\text{indicateur}_i} - \text{valeur\_attendue}_{\text{indicateur}_i}}{\sigma(\text{valeurs\_calibrées}_{\text{indicateur}_i})}$$

Ainsi conçue, la valeur normalisée permet de comparer <sup>[1]</sup> entre les valeurs des indicateurs puisque les ordres de grandeur sont désormais similaires.

**Calcul de la sensibilité globale des paramètres.** Comme les valeurs sont désormais normalisées, il est possible de mener des opérations conjointes sur les différents indicateurs. On définit un indice global, intitulé « sensibilité », qui correspond à une moyenne, pour chaque paramètre, de ses valeurs normalisées dans chacun des indicateurs. Après le centrage, les valeurs deviennent négatives ou positives selon qu'elles sont inférieures ou supérieures aux objectifs. Pour un paramètre qui verrait des variations importantes, négatives et positives, le risque d'une moyenne est alors que la sensibilité calculée soit faible, le très positif compensant le très négatif. Pour prévenir ce risque, on a choisi de calculer la sensibilité sur les valeurs absolues des valeurs normalisées plutôt que sur leur valeur brute. La sensibilité de chaque paramètre peut alors être définie comme suit :

$$\text{sensibilité}_{\text{paramètre}_\alpha} = \frac{\sum |\text{valeurs\_normalisées}_{\text{paramètre}_\alpha}|}{n_{\text{indicateurs}} \times n_{\text{valeurs\_paramètres}_\alpha} \times n_{\text{réplications}}}$$

**Sélection des paramètres à étudier.** Après avoir calculé la sensibilité des paramètres, on sera en mesure d'en sélectionner un sous-ensemble que l'on pourra analyser, visuellement, de manière plus approfondie. Pour isoler ce sous-ensemble, on cherchera à conserver une dizaine des paramètres à la plus forte sensibilité globale. Le nombre précis dépendra de la forme de la distribution de la sensibilité des paramètres : on cherchera avant tout à isoler des paramètres dont la sensibilité les différencie franchement des autres. On se basera pour cela sur une méthode subjective d'examen visuel des écarts entre les valeurs de sensibilité, tel qu'on peut le faire pour choisir un nombre de classes à étudier lors de l'exécution d'une classification ascendante hiérarchique.

Un <sup>[2]</sup> que <sup>[3]</sup> cette sélection à partir de la sensibilité globale est que certains paramètres, très influents sur un indicateur particulier, <sup>[4]</sup> auraient <sup>[5]</sup> la sensibilité globale assez moyenne ou faible, <sup>[6]</sup> effet de resserrement des moyennes oblige. Pour ne pas risquer de négliger de tels paramètres, on complètera la sélection

	Nombre : 1	Auteur : Paul Gourdon	Sujet : Barrer	Date : 19/10/2019 18:09:54 +02'00'
	Nombre : 2 des	Auteur : Paul Gourdon	Sujet : Texte inséré	Date : 19/10/2019 18:14:31 +02'00'
	Nombre : 3 s	Auteur : Paul Gourdon	Sujet : Texte inséré	Date : 19/10/2019 18:14:31 +02'00'
	Nombre : 4	Auteur : Paul Gourdon	Sujet : Barrer	Date : 19/10/2019 18:14:43 +02'00'
	Nombre : 5 aient	Auteur : Paul Gourdon	Sujet : Texte inséré	Date : 19/10/2019 18:14:58 +02'00'
	Nombre : 6 tournure bizarre	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 19/10/2019 18:15:12 +02'00'

précédente par une sélection des trois paramètres les plus sensibles sur chacun des 6 indicateurs.


$$\text{sensibilité}_{\text{paramètre}_\alpha \text{ indicateur}_i} = \frac{\sum |\text{valeurs\_normalisées}_{\text{paramètre}_\alpha \text{ indicateur}_i}|}{n_{\text{valeurs\_paramètres}_\alpha} \times n_{\text{répliques}}}$$

Une large partie de ces paramètres auront vraisemblablement déjà été isolés par le filtre sur la sensibilité globale, mais cette méthode devrait permettre de sélectionner quelques paramètres supplémentaires sur lesquels mener l'analyse visuelle. On évitera en cela de laisser de côté un paramètre ayant un effet important quoi que sur un nombre réduit d'indicateurs seulement.

### 6.2.1.3 Analyse visuelle

Une fois les paramètres à étudier sélectionnés, on pourra mener une analyse visuelle pour comprendre l'influence des paramètres sur les indicateurs de sortie du modèle. Le nombre de paramètres étant réduit, il sera possible d'analyser l'importance de cette influence (la sensibilité), mais aussi sa direction : en observant la réaction des indicateurs de sortie en fonction des valeurs de paramètre, on pourra noter le sens des corrélations (visuelles) entre valeurs de paramètres et valeurs des indicateurs obtenus.

Dans la figure 6.18, la sensibilité du paramètre  $\alpha$  sur l'indicateur  $n$  (haut droite) en constitue un exemple. On peut y voir que les valeurs de paramètres (modalités  $\alpha_1$  à  $\alpha_5$ ) ont un effet sur l'indicateur, et que cet effet montre une corrélation visuelle positive : les valeurs de l'indicateur  $n$  augmentent dans le même sens que celles des modalités testées. Ce même paramètre  $\alpha$ , produit un effet opposé sur l'indicateur 1 : la corrélation y apparaît négative.

Cette analyse visuelle de sensibilité s'inscrit dans une analyse à la fois technique et thématique de SimFeodal. Elle doit en effet permettre de gagner en compréhension sur les interactions entre paramètres et mécanismes. Cela devrait éclairer l'aspect technique, lié à la validation interne du modèle – tel paramètre ~~que l'on pense jouer~~ sur tel mécanisme se comporte-t-il bien comme attendu ? – et l'aspect thématique – est-ce que la taille du monde simulé a un effet sur le rythme de la polarisation générée ? .


**Visualisation.** L'objectif est d'étudier, visuellement, pour chaque paramètre, comment une variation de ce dernier influence chacun des indicateurs. Pour cela, on représente la variation due aux répliques de chaque paramètre au moyen de représentations visuelles dédiées à la variabilité, comme des *box-plots* ou *violin-plots*. La figure 6.18 présente un exemple de mise en page de telles représentations graphiques afin de faciliter l'analyse de sensibilité visuelle.

---

 Nombre : 1	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 19/10/2019 18:18:23 +02'00'
---	-----------------------	-------------------------------	------------------------------------

---

supposé jouer sur

 Nombre : 2	Auteur : Paul Gourdon	Sujet : Barrer	Date : 19/10/2019 18:18:31 +02'00'
---	-----------------------	----------------	------------------------------------

---

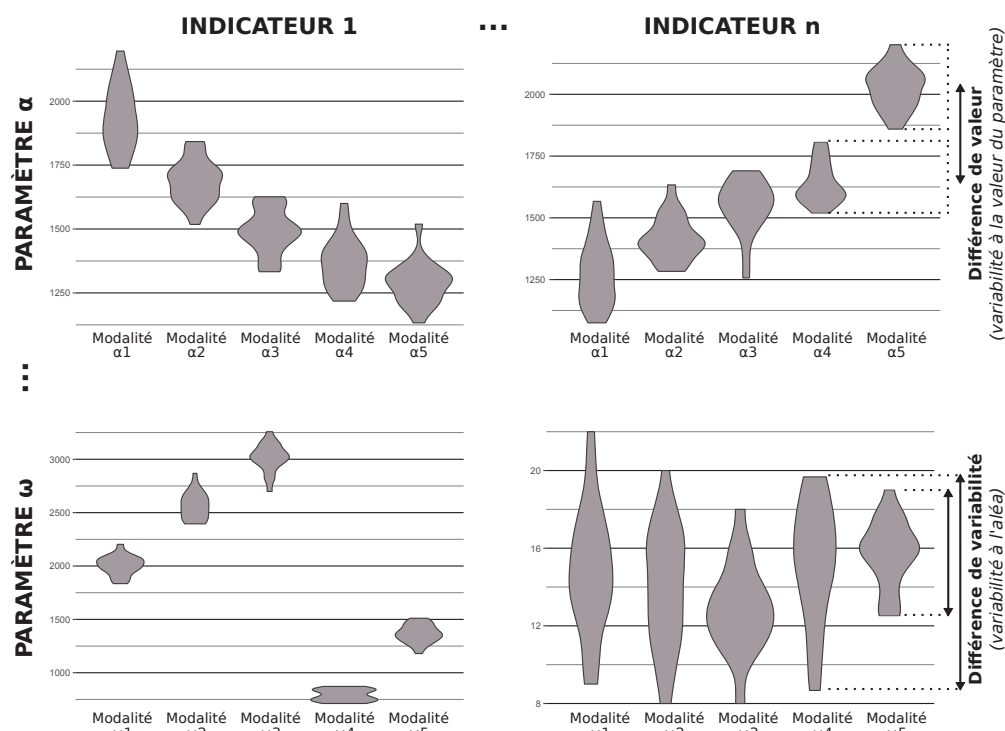


FIGURE 6.18 – Construction et mise en page de planches graphiques dédiées à l'analyse visuelle de la sensibilité de 5 modalités de  $\omega$  paramètres sur  $n$  indicateurs.

Ces planches permettent de lire une double information. En premier lieu, la position relative des valeurs correspondant à chaque modalité permet de remarquer d'éventuelles influences des valeurs de paramètres sur les valeurs d'indicateur. Dans le schéma, on représente ainsi les « différences de valeur » qui expriment une variabilité à la valeur du paramètre (exemple en haut à droite). Dans un second temps, et on y reviendra après (section 6.2.4), le choix des représentations en *violin-plots* permet aussi d'étudier la variabilité interne des valeurs de paramètres testées sur chaque indicateur (exemple en bas à droite). Les différences d'amplitude (sur l'axe des ordonnées) exprime ainsi les différentes variabilités dues à l'aléa : plus les « violons » sont grands, plus l'indicateur varie selon les réplifications pour la valeur de paramètre testée.

**Normalisation.** La normalisation des valeurs était indispensable lors du calcul de la sensibilité globale afin d'homogénéiser la variabilité des indicateurs. Pourtant, sur un sous-ensemble de paramètres cette normalisation devient un frein à l'interprétation et à la lisibilité des analyses : les valeurs des différents indicateurs n'étant plus exprimés dans l'unité d'origine, il est difficile d'en faire un commentaire thématique. L'analyse visuelle, sur les paramètres sélectionnés, sera ainsi présentée sur les valeurs brutes issues des simulations, et chaque graphique aura en conséquence un axe des ordonnées propre. La logique est encore une fois la même que dans les classifications ascendantes hiérarchiques couramment pratiquées en géographie quantitative : on effectue la classification sur des valeurs normalisées, mais on représente souvent les profils des classes en revenant aux valeurs brutes.

---

Nombre : 1	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 19/10/2019 18:28:06 +02'00'
------------	-----------------------	------------------------	------------------------------------

---

En fait, en voyant la figure 6.18, je suis pas sûr d'avoir bien saisi.

En haut on regarde l'influence du paramètre alpha sur les indicateurs de sortie.

Les modalités : se sont bien les 5 valeurs que prend la paramètre alpha ?

Et les valeurs en ordonnée ce sont bien celles de l'indicateur de sortie (pour les différentes répliques) ? Du coup je suis pas sûr de comprendre le graphique en bas à droite. C'est la variabilité pour chaque modalité du paramètre ? mais elle n'est pas en creux dans les figures du haut ?

Aussi la position des titres des axes (notamment les paramètres, ça m'a perdu je crois.) Je les pensais plutôt en dessous de l'axe des abscisses.

Ou alors j'ai rien capté



La figure 6.19 présente les « scores de sensibilité » de chacun des paramètres du modèle. Ces paramètres sont classés par ordre descendant selon leur sensibilité globale, et la couleur reprend la typologie des paramètres.

On constate tout d'abord que la sensibilité des paramètres est relativement peu dispersée : les valeurs abstraites de la sensibilité varient entre près de « 7 » et « 11 », soit un écart maximal de 65%, ~~là où on aurait pu imaginer que certains paramètres aient des sensibilités doublées, triplées... par rapport à d'autres.~~ Cela indique que les paramètres ont tous une influence sur les sorties du modèle, influence qui est située dans le même ordre de grandeur. Par extension, on note qu'aucun des paramètres testés ne présente une sensibilité véritablement faible. Ce constat est renforcé en considérant que le paramètre le moins sensible globalement (croissance\_demo) a une influence réelle sur le modèle, pour avoir notamment été testé plus extensivement en parallèle de cette analyse de sensibilité (voir ??). Pour le modélisateur, cette relative homogénéité des sensibilités est un résultat véritablement rassurant. Cela nous prouve qu'aucun paramètre n'est inutile, et ne devrait par conséquent être supprimé du modèle si on recherchait une parcimonie plus importante à cette étape.

Toujours à un niveau d'observation global, on peut remarquer que la sensibilité globale des paramètres ne semble pas liée aux types de paramètres. Visuellement, on constate en effet qu'il ne semble pas y avoir de motif particulier dans le positionnement des types représentés par les couleurs. En cela, le choix de mener l'analyse indistinctement sur tous les types de paramètres semble avoir été le bon.

En entrant dans le détail, on peut remarquer que trois paramètres se démarquent véritablement : les seuils de distance acceptables à l'église paroissiale la plus proche (dist\_minmax\_eglise), la dimension du monde simulé (taille\_cote\_monde) et le seuil de distance de détection des agrégats, en dessous duquel on considère que des foyers paysans peuvent former un agrégat (distance\_detection\_agregat). Il est intéressant de noter que les valeurs par défaut de ces paramètres ont toutes été choisies avec une précaution particulière.

- Le seuil de distance aux églises est un paramètre particulièrement complexe qui a donné lieu à de multiples allers-retours pour en fixer les valeurs. Sur la forme, c'est l'un des paramètres les plus « qualitatifs » : il désigne une étendue spatiale qui varie au cours du temps de la simulation. Comme la satisfaction religieuse que ce paramètre conditionne a une importance majeure sur les migrations des foyers paysans (voir section 6.1.2.4), on pouvait s'attendre à ce qu'il se montre particulièrement sensible.
- Les dimensions du monde simulé ont donné lieu à un calibrage, par ajustement empirique, dont l'on a discuté au début de ce chapitre (section 6.1.1.1). On décrivait alors toutes les répercussions que ce changement avait provoqué, et il n'est pas surprenant que ce paramètre soit parmi les plus sensibles dans le modèle.
- La distance de détection des agrégats a été l'un des premiers paramètres implémenté dans le modèle, ajusté selon des valeurs empiriques. La va-

	Nombre : 1	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 20/10/2019 20:54:31 +02'00'
précise				
	Nombre : 2	Auteur : Paul Gourdon	Sujet : Barrer	Date : 20/10/2019 20:56:35 +02'00'
je mettrai pas ça... c'est pas justifié. et le "relativement peu dispersée" montre déjà que c'est surprenant.				
	Nombre : 3	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 20/10/2019 20:59:27 +02'00'
je suis pas sûr de cette proposition. Je comprends l'idée, mais si on recherchait + de parcimonie on supprimerait quand meme des paramètres du modeles. C'est juste qu'on cherche pas la parcimonie.				
	Nombre : 4	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 20/10/2019 21:02:24 +02'00'
est-ce qu'inutile est le mot ? après tout si on a défini des paramètres c'est pour qu'ils soient utiles, non ? L'idée c'est plutôt que les paramètres ont un effet relativement équilibré sur le fonctionnement du modèle (et ça ça n'est pas prévisible en avance, j'imagine).				
	Nombre : 5	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 20/10/2019 21:05:43 +02'00'
c'est le paramètre qui est sensible ? c'est une tournure du domaine ? Car pour moi c'est le modèle (ses sorties) qui est sensible au paramètre				

leur fixée (100 m) n'a pas évoluée depuis les premières versions du modèle. C'en est presque une constante et il est très enrichissant pour le modélisateur de découvrir que ce paramètre a une influence importante sur l'issue des simulations.

Au sein des paramètres qui suivent immédiatement ce trio, les ruptures sont moins nettes et l'on découpera donc ce continuum entre les paramètres situés en 10<sup>e</sup> et 11<sup>e</sup> position (nb\_min\_fp\_agregat et taux\_prelevement\_zp\_chateau), où l'on discerne visuellement un nouveau saut dans les valeurs de sensibilité. Sur ces 57 paramètres, cette sélection en isole donc 10, auxquels on ajoute une sélection des trois paramètres ayant la plus forte sensibilité sur chacun des indicateurs. Comme ces derniers sont en bonne partie redondants avec les autres sélections, on obtient finalement un total de 14 paramètres à analyser visuellement par la suite. Ces paramètres, leur description complète ainsi que les valeurs testées sont présentées dans le tableau 6.3.

**Limites de cette approche de sélection.** Avant de passer à l'analyse détaillée des paramètres isolés, il nous paraît important de noter une limite substantielle de la méthode mise en place pour leur sélection. La mesure choisie, consistant en une sensibilité normalisée, ne tient pas compte de l'étendue des valeurs de paramètres testées. Ces étendues sont largement incommensurables, qui plus est quand il s'agit de comparer des paramètres numériques simples (la taille des côtés du monde simulé) à des paramètres plus qualitatifs (les seuils de distance évolutifs de la distance aux églises paroissiales). On peut par exemple noter que dans les 10 paramètres jugés les plus sensibles, 4 portent sur des valeurs testés « qualitatives » (étendues, variables au cours du temps..., voir p. 31), alors que ces paramètres ne représentent que 20% de l'ensemble des paramètres testés.

Lors du choix des valeurs de paramètres à tester pour cette analyse de sensibilité (disponibles dans l'annexe N), certains paramètres auront forcément été testés sur des valeurs plus ou moins éloignées des valeurs par défaut issues du calibrage. Cet éloignement se traduit forcément par des sensibilités mesurées très variables. On peut illustrer ce biais en comparant un paramètre de stock simple qui se montre très sensible (la taille du monde simulé à nouveau) et un paramètre de ratio qui se situe lui tout en bas du classement (le taux de croissance démographique, croissance\_demo).

- Pour le premier, les valeurs testées s'échelonne régulièrement entre 50 et 150 km. L'ordre de grandeur de ces valeurs est similaire, représentant une variation qui va environ de la division par deux à la multiplication par deux de la valeur par défaut (80 km).
- Pour le second, qui vaut 0% par défaut, les valeurs testées sont de 0%, 1.53%, 3.72%, 5.89% et 12.89%. Au regard du nombre de pas de temps, cela représente un doublement, un triplement, un quintuplement et enfin un décuplement de la population initiale, que l'on a fait covarier pour que la population finale soit systématiquement identique. Vis-à-vis du paramètre précédent, l'étendue interprétée est plus large, consistant même en une activation ou non du mécanisme de croissance démographique. En absolu, les 4 premières valeurs sont cependant très proches et on ne

10+ 3 = 14 ???

### 6.2.3 Évaluation visuelle de la sensibilité

Plutôt que de mener l'évaluation de la sensibilité des paramètres sélectionnés de manière linéaire, paramètre après paramètre, nous présentons ces derniers organisés par thématique c'est-à-dire selon les types d'agents concernés par chacun d'entre eux.

#### 6.2.3.1 Paramètres liés au monde simulé

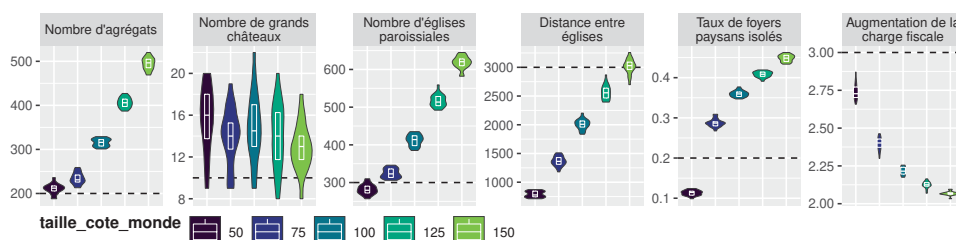



FIGURE 6.21 – Sensibilité à la taille du monde simulé.

Le paramètre régissant la taille du monde simulé est un paramètre majeur qui affecte la totalité des indicateurs étudiés, comme la figure 6.21 permet de le constater. Quand la taille du monde simulé augmente, le nombre d'agrégats, d'églises paroissiales, la distance moyenne entre églises paroissiales et le taux de foyers paysans augmente aussi. Au contraire, la pression fiscale varie de façon inverse : quand la taille du monde simulé augmente, elle diminue. On ne constate pas d'effet direct ou linéaire sur le nombre de grands châteaux, qui ne semblent pas être directement sensibles à la taille du monde. Les différences visuelles de position semblent surtout influencées par la forte variabilité liée à l'aléa que l'on constate sur cet indicateur de sortie de simulation.

Ces résultats sont assez intuitifs : plus le monde est restreint, plus les foyers paysans sont proches les uns des autres. Cela entraîne premièrement une concentration plus forte, mais aussi un nombre d'agrégats plus faible : au lieu d'être dispersés en une multitude de petits agrégats (500 en moyenne avec un monde de 150 km de côté), les foyers paysans se concentrent dans un nombre restreint d'agrégats (200 quand le côté est de 50 km), de superficie vraisemblablement supérieures par les effets des mécanismes de fusion des agrégats. En effet, quand la superficie d'ensemble est plus faible, toutes choses égales par ailleurs concernant le nombre et le rayon des zones de prélèvement, on assiste nécessairement à une superposition plus importante de ces zones. La charge fiscale des foyers paysans s'en retrouve fortement affectée.

Un effet de ce paramètre nous paraît légèrement contre-intuitif : avec une surface plus importante, il est entièrement logique que la distance entre les églises augmente, puisque celles-ci sont forcément plus dispersés dans un monde plus large. Pourtant, le nombre d'églises paroissiales croît aussi avec la superficie du monde simulé, ce qui ne nous semble pas découler directement des règles de création/promotion de nouvelles églises paroissiales. On peut interpréter ce comportement comme une conséquence imprévue et non linéaire de la dispersion des foyers paysans au sein de petits agrégats. De petites églises paroissiales sont mécaniquement créées ou promues en plus grand nombre dans les zones

---

	Nombre : 1	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 20/10/2019 21:28:28 +02'00'
--	------------	-----------------------	------------------------	------------------------------------

---

finalement tu ne parles que de taille cote monde dans ce paragraphe, non ?

	Nombre : 2	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 20/10/2019 21:19:54 +02'00'
--	------------	-----------------------	------------------------	------------------------------------

---

mettre le nom entre parenthèse (taille\_cote\_monde)

faiblement peuplées (petits agrégats proches du seuil minimal), là où la création d'une paroisse est moins coûteuse (en termes de nombre de foyers paysans requis) qu'au sein des agrégats plus importants. En renforçant le nombre d'agrégats, le paramètre de taille du monde simulé agit ainsi de manière indirecte sur le nombre d'églises paroissiales.

### 6.2.3.2 Paramètres liés aux foyers paysans et agrégats

Les foyers paysans (et les agrégats de population qui résultent de leur concentration) sont, par hypothèse, les agents les plus déterminants dans l'évolution des structures spatiales observées dans SimFeodal. À ce titre, il est attendu (et sécurisant en termes de validation interne) que les paramètres contrôlant leurs mécanismes propres aient une influence nette sur les indicateurs de sortie analysés.

**Foyers paysans.** Les trois paramètres, spécifiquement liés aux foyers paysans, les plus sensibles (figures 6.22 à 6.24), embrassent trois aspects bien différents des mécanismes gérant le comportement des foyers paysans.

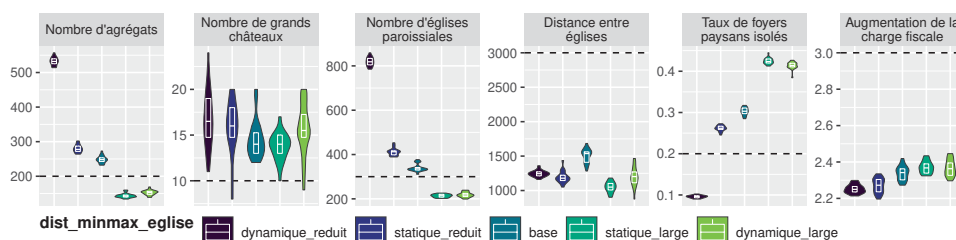


FIGURE 6.22 – Sensibilité du paramètre `dist_minmax_eglise`.

Le premier de ces paramètres ((figure 6.22), qui est aussi le plus sensible du modèle, joue sur la satisfaction religieuse des foyers paysans. Notons déjà que le calibrage de ce paramètre semble efficace : au moins sur les quatre premiers indicateurs, c'est la valeur par défaut dans le modèle qui approche le plus les objectifs fixés. En matière d'interprétation, le fait que ce paramètre ait une influence sur les indicateurs de sortie du modèle est attendu et intuitif. La satisfaction religieuse que ce paramètre conditionne, est en effet, intentionnellement l'un des éléments majeurs du modèle.

Pourtant, l'ampleur de cette influence est un peu plus forte qu'on aurait pu le penser au vu des mécanismes et résultats principaux du modèle. Dans la ?? qui présentait l'évolution de la satisfaction des foyers paysans, on remarquait que la satisfaction **protection** était la composante principale de la satisfaction globale. On aurait donc pu s'attendre, au regard des mécanismes implémentés, à ce que les paramètres liés à la satisfaction protection aient une sensibilité supérieure à celui-ci.

Pour ce paramètre qui régit les seuils de distance acceptables à l'église paroissiale la plus proche, on peut émettre l'hypothèse que les valeurs testées biaisent la mesure de la sensibilité globale. Ces valeurs sont qualitatives (étendues plus ou moins importantes, variables ou non au cours du temps, cf. le tableau 6.3), ont une large influence sur les aspects liés à la concentration des

---

 Nombre : 1      Auteur : Paul Gourdon      Sujet : Texte surligné      Date : 21/10/2019 08:59:12 +02'00'

---

entre guillemets.

Ou sinon " on remarquait que la composante "protection" de la satisfaction était la plus importante"



foyers paysans, et leur action sur les autres indicateurs est moins linéaire et prévisible.

Plus les étendues testées sont restreintes (modalités « dynamique\_reduit » et « statique\_reduit »), plus le nombre d'agrégats et de foyers paysans isolés est important. Pour ceux-là, la satisfaction religieuse joue à ce moment là le rôle limitant et « force » les foyers paysans à migrer à proximité d'églises paroissiales.

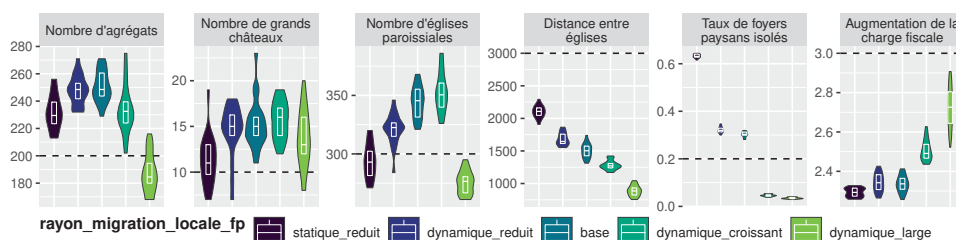


FIGURE 6.23 – Sensibilité du paramètre `rayon_migration_locale_fp`.

Cette migration, au moins pour l'aspect local, est largement influencée par le second paramètre de la sélection : le rayon maximal de migration locale des foyers paysans (figure 6.23). Celui-ci n'est pas une étendue, mais un rayon maximum qui peut (valeurs « dynamiques ») changer au cours du temps. Dans le modèle calibré, ce rayon est fixe au cours du temps (2 500 m, modalité « base »). Dans des versions précédentes du modèle, les valeurs de ce paramètre augmentaient au cours du temps<sup>16</sup>, et on a voulu tester l'influence de l'évolution de ces valeurs (modalités « dynamique\_reduit » et « dynamique\_croissant ») ainsi que d'augmentations et diminutions du rayon (modalités de type « réduit » par opposition à « croissant » et « large »).

Ce paramètre a notamment un rôle déterminant sur la concentration des foyers paysans, puisque selon les valeurs éprouvées, cet indicateur peut passer en moyenne de plus de 60% à moins de 5%. Les effets sur le nombre d'agrégats, de grands châteaux et d'églises paroissiales ne sont pas directement linéaires. C'est un exemple typique de paramètre qui régit un mécanisme dont les interactions avec les autres suit des effets de seuils. Sur les indicateurs de sortie, cela résulte en des tendances diverses selon l'ordre de grandeur des seuils empruntés.

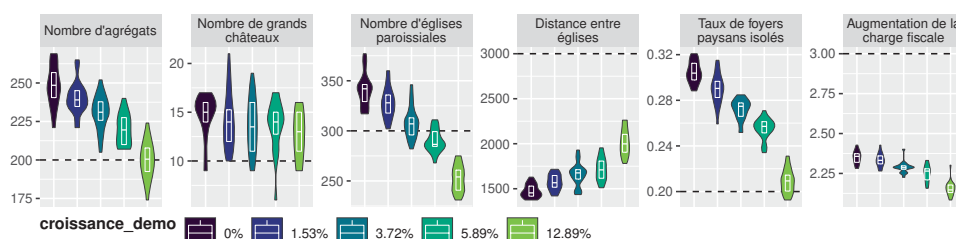


FIGURE 6.24 – Sensibilité du paramètre `croissance_demo`.

Le paramètre de croissance démographique régit un mécanisme particulier dans le modèle, puisque par hypothèse, on a choisi de considérer une popula-

16. Reproduisant ainsi *in silico* la diminution des coûts de distance-temps qui témoignent de l'intégration d'un système de peuplement.



tion stable dans le déroulement du modèle (ref chap 2, là où j'explique, sans doute inputs/situation initiale).

Ce paramètre est considéré comme relevant du contexte, mais aurait également entièrement sa place parmi les *inputs* tant son importance thématique est majeure. La très faible sensibilité globale de ce paramètre (figure 6.19), ne nous paraît pas représentative de son influence effective tant l'analyse visuelle de son influence sur les indicateurs de sortie (figure 6.24) montre un effet direct sur tous les indicateurs présentés (à l'exception du nombre de grands châteaux).

On peut constater que plus le taux de croissance est élevé, plus le nombre d'agréats est faible, de même que le taux de foyers paysans isolés. Sur les deux paramètres précédents (figures 6.22 et 6.23), de même que plus globalement, ces deux indicateurs sont pourtant le plus souvent opposés : quand le nombre d'agréat augmente, le taux de foyer paysans isolés diminue.

L'influence de ce paramètre est extrêmement intéressante pour le modèle tant ce paramètre permet d'approcher des objectifs de ces indicateurs qui sont tous deux dépassés alors qu'opposés. On notera par exemple que parmi les valeurs de croissance démographique testées, la plus forte (12.80%) permet un excellent ajustement du nombre d'agréats et du taux de foyers paysans isolés, deux objectifs majeurs dans l'évaluation de SimFeodal. L'encadré 6.2 illustre une démarche d'exploration plus poussée de ce paramètre.

*Encadré 6.2 : Un scénario thématique pour tester l'hypothèse de croissance démographique*

L'envie de mener des « scénarios thématiques » a émergé très tôt dans la conception du modèle SimFeodal. L'idée était de pouvoir tester des hypothèses différentes quant aux grands faits stylisés mobilisés dans le modèle. Ces versions alternatives permettraient d'éprouver la validité du modèle, et donc des logiques sur lequel il repose, en l'adaptant à d'autres régions où les processus étudiés se sont déroulés différemment <sup>1</sup> voire pas du tout.

À ce jour, une quinzaine de scénarios de tels types ont été conçus, implémentés et simulés. La temporalité plus large de ce projet de recherche par rapport à celle de ce travail de thèse fait que l'analyse des résultats de ces scénarios et des résultats thématiques qu'ils engendrent est encore en cours par l'ensemble de l'équipe de SimFeodal.

Dans cet encadré, nous présentons toutefois un scénario, qui permet de revenir sur un élément majeur du modèle tel qu'il a été conçu. Cet élément est l'absence de prise en compte de croissance démographique, quand bien même empiriquement, on sait qu'il y a eu une croissance. Pour tester les effets d'une telle croissance sur les sorties du modèle, on a procédé presque toute chose par ailleurs : la quasi-totalité des paramètres n'ont pas été modifiés, et le modèle n'a donc pas été re-calibré pour être adapté à ce fait stylisé, nouveau et majeur, qui a été implémenté. Seule

maladroit.

"où les processus sont différents voire inexistants"

exception : pour que l'état final du scénario soit comparable à l'état final de la version calibrée, il fallait que les populations soient comparables. De même que l'on a fait varier le taux de croissance démographique, on a donc aussi fait varier la population initiale du modèle, afin que la population atteigne 40 000 foyers paysans en fin de simulation.

Il y a eu trois scénarios conçus pour éprouver l'influence de taux variables de croissance démographique (D1, D2 et D3). On n'en présente ici qu'un seul, par soucis de brièveté. Ce scénario, intitulé « D2 », est celui dont les résultats sont les plus intéressants. Dans ce scénario, le taux de croissance démographique est fixé à 12.89%, ce qui permet de faire passer une population initiale de 4 000 foyers paysans à 40 000 en fin de simulation. Cela correspond donc à un décuplement de la population entre 800 et 1200, jugé non aberrant par les thématiciens.

Indicateur	Valeur attendue	Valeur de référence	Moyenne	Médiane
Agrégats	200	249	198	199
Gros châteaux	10	15	12	12
Églises paroissiales	300	348	252	253
Distance moyenne entre églises	3 000 m	1 459 m	1 988 m	1 975 m
Part de foyers paysans isolés	20 %	30 %	21 %	21 %
Augmentation de la charge fiscale des foyers paysans	x 3	x 2.4	x 2.6	x 2.6

TABLEAU 6.4 – Valeurs des indicateurs numériques du scénario D2 en fin de simulation. Les valeurs de référence correspondent aux moyennes obtenues dans la version calibrée de SimFeodal, sans croissance démographique donc.

Les résultats agrégés de ce scénario (tableau 6.4) sont globalement satisfaisants, et même **plus** que les résultats de la version calibrée : le nombre d'agrégats en fin de simulation et la part de foyers paysans isolés sont presque parfaitement ajustés aux valeurs attendues fixées empiriquement. Le nombre de grands châteaux et l'augmentation de la charge fiscale sont aussi plus proches des attentes, bien que d'une manière moins franche. Seul point négatif : les églises paroissiales sont beaucoup moins nombreuses dans ce scénario que dans la version de référence, et trop faibles désormais au regard des valeurs attendues.

En regardant le détail de la structure, et notamment relative aux indicateurs de sortie relatifs aux agrégats (figure 6.25), on constate qu'en matière de hiérarchisation du système, les résultats de ce scénario sont aussi meilleurs que ceux de la version de base (figures 6.5 et 6.8). La croissance du nombre d'agrégats y est assez régulière, contrairement à la version calibrée (figure 6.5) où l'évolution était très faible jusqu'en



1060. Plutôt que d'insuffler un changement de tendance, les mécanismes exogènes (960 et 1060) ne font qu'accélérer un processus déjà actif. On note aussi que l'organisation hiérarchique des agrégats, et son évolution, est bien plus proche des formes connues sur les sociétés contemporaines. Dans la version calibrée du modèle (figure 6.8), la courbe conservait un « coude » important, tandis qu'ici, on note que l'évolution tend vers un redressement de la pente de la courbe rang-taille. En cela, l'organisation hiérarchique tend vers le modèle log-normal observé empiriquement sur de très nombreux systèmes (CURA et al. (2017) par exemple).

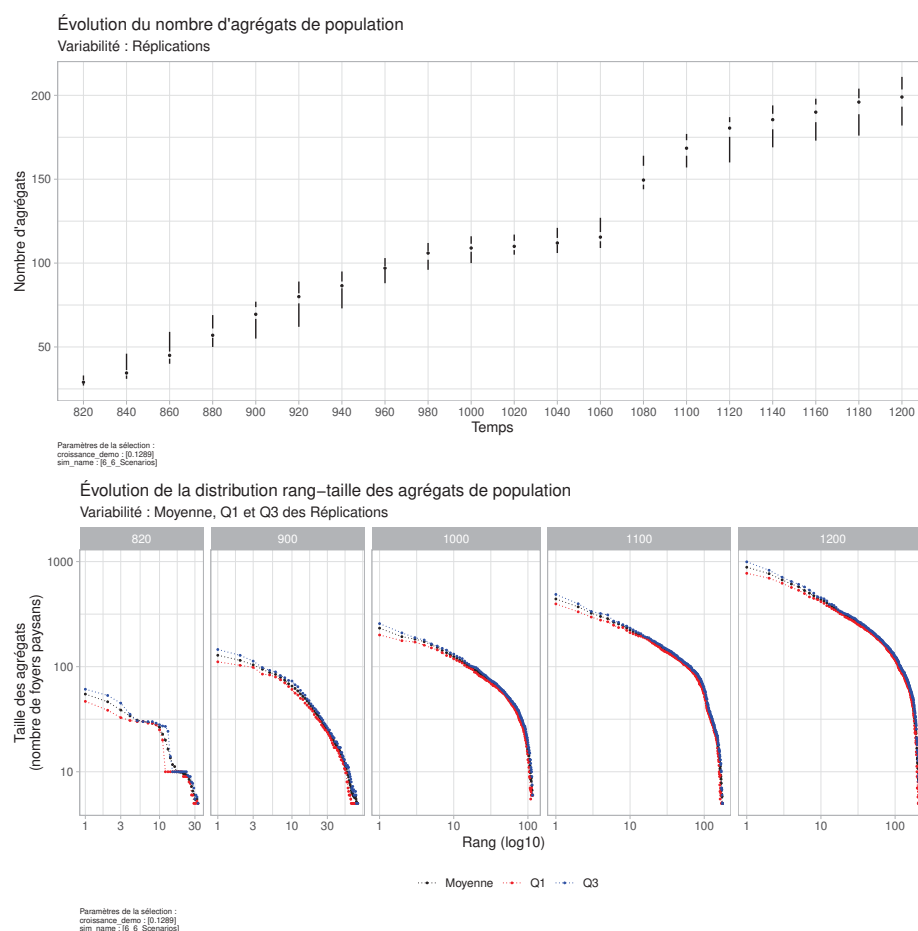


FIGURE 6.25 – Nombre et organisation des agrégats dans le scénario D2. Nous ne détaillerons pas les autres résultats de ce scénario D2, par ailleurs d'ores et déjà explorable dans la plateforme SimEDB ([mettre lien direct](#)), et nous contenterons d'affirmer que ce scénario répond mieux une majorité des attentes identifiées pour SimFeodal. C'est particulièrement étonnant dans la mesure où le modèle n'a jamais été conçu ~~en~~ <sup>en</sup> ayant l'arrière-pensée d'y implacer la stabilité démographique par de la croissance, et n'a pas non plus été calibré à cet effet. Sans perdre de vue les biais que l'implémentation informatique de mécanismes peut apporter vis-à-vis des processus modélisés, on notera tout de même qu'il est, thématiquement et méthodologiquement, très intéressant que les interactions entre mécanismes de SimFeodal s'ajustent mieux quand de la croissance démographique est mise en place. Cela nous semble constituer un





châteaux final sera important<sup>19</sup>. Au-delà des remarques précédentes, on notera que la date de construction des premiers châteaux influe sur l'augmentation de la charge fiscale, dans un sens qui nous semble à nouveau intuitif : plus il y a de châteaux, plus il y a de zones de prélèvement associées, et plus les foyers paysans voient leur charge fiscale augmenter.

De manière générale, la conception et le paramétrage des mécanismes liés aux châteaux ont demandé un travail conséquent (voir la partie dédiée à leur calibrage – section 6.1.1 – p. 8), qui nous a pourtant semblé, lors ces étapes, assez peu convaincant au regard de l'amélioration très modeste que cela a apporté en termes d'ajustement général du modèle. D'où de nombreux débats internes entre modélisateurs et thématiciens sur la nécessité de détailler autant les mécanismes liés à la construction des châteaux.

L'analyse du premier paramètre (*periode\_promotion\_chateaux*, figure 6.28) semble aller dans ce sens, puisque la complexité des mécanismes de promotion des châteaux ne semble influencer, sur ces indicateurs très synthétiques, que l'indicateur directement lié : le nombre de gros châteaux.

Les résultats de l'analyse de sensibilité des autres paramètres (figures 6.29 et 6.30) donnent tort à l'intuition du modélisateur-« calibreur », et gain de cause aux thématiciens pour lesquels les châteaux, et la justesse de leur implémentation, avaient une dimension thématique considérable. La portée réduite mais claire de ces paramètres sur le premier objectif thématique recherché (la concentration des foyers paysans) justifie de leur existence et de leur nécessité dans les processus modélisés au sein de SimFeodal.

#### 6.2.3.4 Paramètres liés aux églises et paroisses

Le dernier paramètre de cette analyse visuelle est aussi le seul qui porte sur les paroisses. Celles-ci constituent l'un des moteurs principaux de la fixation de la population dans les agrégats de taille réduite qui constituent la grande majorité des lieux de concentration des foyers paysans (la fameuse « longue traîne » de cette hiérarchie). En première analyse, cette sous-représentation des paramètres liés aux paroisses dans l'ensemble des paramètres sélectionnés nous apparaît donc contre-intuitif.

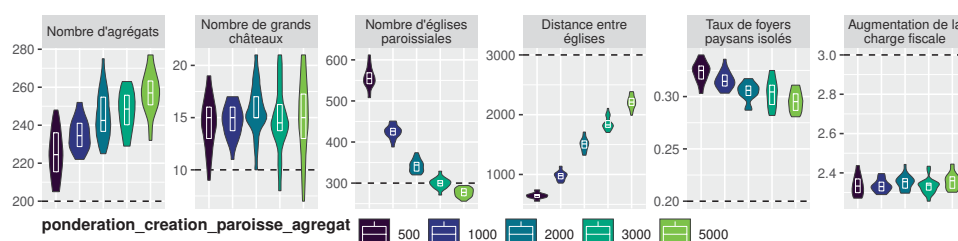


FIGURE 6.31 – Sensibilité du paramètre de pondération de création de nouvelles paroisses dans les agrégats.

La figure 6.31 présente les réponses des indicateurs à différentes valeurs du pa-

19. Au contraire du paramètre précédent, où une modalité plus forte impliquait plus de châteaux.



paramètre `ponderation_creation_paroisse_agregat`, qui agit sur le nombre de foyers paysans nécessaires à la création d'une nouvelle église paroissiale au sein des agrégats. De manière prévisible et évidente, ce paramètre influence nettement les deux indicateurs liés aux paroisses et églises : plus grand est le nombre de paroissiens nécessaire à la création d'une nouvelle paroisse dans un agrégat, plus faible est le nombre d'églises paroissiales conséquemment créées. Et moins il y a d'églises dans un monde à la superficie constante, plus large est la distance entre elles.

Passé ces éléments intuitifs, on notera avec intérêt la variation amenée par ce paramètre sur le nombre d'agrégats : plus le seuil est élevé, plus les agrégats sont nombreux, et cette corrélation apparaît visuellement significative et inverse à celle du nombre d'églises paroissiales. Cette tendance est inverse à celles observées dans les paramètres liés aux foyers paysans et agrégats (section 6.2.3.2, figures 6.22 à 6.24 et 6.26). Pour ces paramètres, les nombres d'agrégats et d'églises paroissiales varient dans le même sens face à une augmentation de la valeur des paramètres. Dans le cas du paramètre de pondération de la création de paroisses « urbaines », la relation est inverse (figure 6.31).

Cela nous paraît contre-intuitif au regard des mécanismes du modèle, et résulte vraisemblablement d'effets inattendus des interactions entre ces mécanismes. Une piste d'explication réside dans l'influence de ce paramètre de pondération des créations de paroisse sur la hiérarchisation des paroisses et des pôles. En créant moins de paroisses en zone dense, la distribution de l'attractivité des pôles d'attraction – qui dépend en large partie du nombre d'églises paroissiales qui les composent – tend vers plus d'uniformité. Quand les pôles ont une attractivité plus homogène, les migrations locales sont favorisées, et la concentration s'effectue alors vers des pôles moins attractifs constitués d'agrégats faiblement peuplés.



#### 6.2.4 Analyser la sensibilité à l'aléa

En menant l'analyse de sensibilité visuelle, on a pu remarquer que certains indicateurs présentaient une plus forte variation que d'autres. Une partie de l'explication tient à l'inégale amplitude des valeurs de paramètres testées, mais d'autres facteurs peuvent être en jeu.

De manière globale, on constate dans les résultats de la version calibrée de SimFeodal (tableau 6.1) que la variabilité des indicateurs émergents, mesurée à partir de leur écart-type, est forte. L'écart-type se lisant dans l'unité de l'indicateur, il peut être intéressant de le transformer en coefficient de variation ( $CV_{\text{indicateur}} = \sigma_{\text{indicateur}} / \mu_{\text{indicateur}}$ ) pour obtenir des valeurs comparables entre les indicateurs.



Indicateur	Moyenne	<sup>1</sup> Ecart-type	Coefficient de variation
<i>Agrégats</i>	249	10.45	0.042
<i>Grands châteaux</i>	15	2.87	0.191
<i>Églises paroissiales</i>	348	12.96	0.037
<i>Distance moyenne entre églises</i>	1459 m	97 m	0.066
<i>Part de foyers paysans isolés</i>	30%	0.8 %	0.027
<i>Augmentation de la charge fiscale des foyers paysans</i>	× 2.4	0.030	0.013
<b>Moyenne</b>			0.063

TABLEAU 6.5 – Mesures de dispersion des indicateurs de sortie de la version calibrée (6.6) de SimFeodal.

La comparaison des coefficients de variation tableau 6.5 montre que la variabilité des indicateurs due à l'aléa est assez faible, et d'ordre de grandeurs comparables d'un indicateur à l'autre<sup>20</sup>.

Pourtant, lors de l'analyse de sensibilité, on a pu constater des variations (étendue dans l'axe des ordonnées des *violin-plots*) bien différentes au sein des répliques des paramètres testés. Dans la figure 6.23 par exemple, on peut remarquer que la modalité « dynamique\_large » présente une variabilité à l'aléa (le « violon » est très étendu) différente des autres modalités : plus faible sur le nombre d'églises paroissiales et plus importante sur l'augmentation de la charge fiscale. De même, l'indicateur *ponderation\_creation\_paroisse\_agregat* (figure 6.31) présentait, sur le nombre de grands châteaux, une importante différence de variabilité à l'aléa.

Il nous paraît par conséquent utile de mener un bref complément d'analyse, dédié à l'étude de la variabilité au sein des répliques d'une expérience. De telles analyses systématiques nous paraissent peu fréquentes dans la littérature liée aux modèles de simulation en géographie. On trouve une définition de ce type d'approche chez GINOT et MONOD (2005), sous le terme d'« analyses d'incertitude » :

« Compte tenu des incertitudes sur les paramètres, de la variabilité naturelle des variables d'entrée et des composantes stochastiques qui peuvent être incluses dans la structure du modèle, il s'agit de calculer l'incertitude associée aux variables de sorties. Les analyses d'incertitude sont très liées aux analyses de sensibilité dans la mesure où l'on souhaite en général connaître non seulement cette incertitude, mais également son origine. C'est pourquoi ces deux types d'analyses sont souvent menées en parallèle, voire confondues. »

GINOT et MONOD 2005, p. 76

20. Le nombre de grands châteaux fait exception : son coefficient de variation (0.191) est nettement supérieur aux autres. Cela s'explique par l'ordre de grandeur du nombre de châteaux et grands châteaux, faible relativement aux autres, et où des variations de quelques unités jouent beaucoup sur l'écart-type en raison de la nature entière de ces variables.

---

Nombre : 1	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 21/10/2019 09:39:06 +02'00'
------------	-----------------------	------------------------	------------------------------------

---

là ça rejoint ma note sur la description des graphiques pour l'exploration de sensibilité.

J'ai pas saisi comment on isole ce qui est du à l'aléa... C'est la moyenne par indicateur des écarts liés à toutes les modalités des paramètres ?

Pour mesurer cette incertitude, nous avons repris l'ensemble des répliques correspondant à l'analyse de sensibilité de chaque paramètre et avons mesuré la variabilité (écart-type) sur chaque indicateur. Afin d'avoir des mesures comparables, on a ensuite procédé à une réduction (division par l'écart-type des valeurs de référence, présentées dans le tableau 6.5) de l'amplitude de ces valeurs.

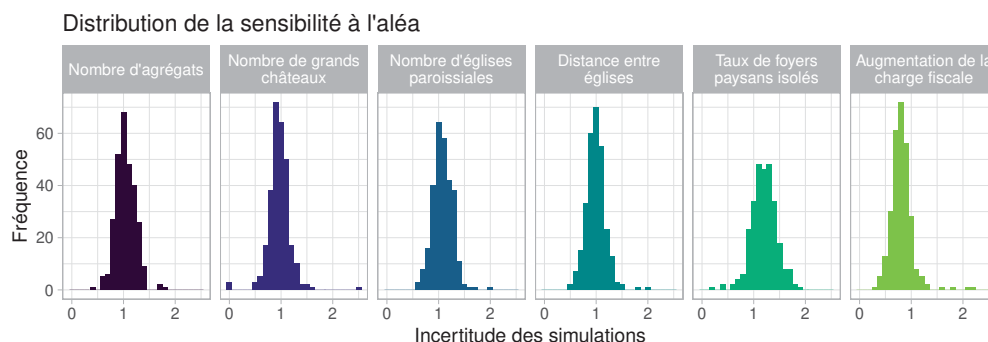


FIGURE 6.32 – Sensibilité à l'aléa selon les indicateurs.

*L'incertitude mesurée se lit en proportion de l'écart-type des résultats de la version calibrée de SimFeodal. Une valeur de 0.5, par exemple, signifie que la variabilité due à l'aléa est moitié moins importante que celle des valeurs de paramètres issues du calibrage.*

La distribution de ces valeurs d'incertitude relatives est présentée dans la figure 6.32. Dans ces histogrammes, la valeur de 1 correspond à une incertitude moyenne identique à celle des répliques de référence, c'est-à-dire des répliques de la version calibrée du modèle dont les résultats ont été commentés dans la première partie de ce chapitre (section 6.1.2). Une valeur de 2 peut être comprise ainsi : pour l'indicateur considéré, les différentes simulations exécutées lors de l'analyse de chaque valeur de chaque paramètre ont une variabilité deux fois supérieure à la variabilité de référence. Autrement dit, certaines valeurs de certains paramètres amènent une bien plus forte variabilité : ils laissent une part plus importante à la stochasticité du modèle.

On peut remarquer que la plupart des indicateurs présentent des *outliers*, c'est-à-dire des valeurs de paramètres pour lesquelles la variabilité due à l'aléa est nettement supérieure (ou inférieure) à la normale. Dans le cas du nombre de grands châteaux, il y a même des valeurs de paramètres qui montrent une absence presque totale de variabilité à l'aléa. Sans aller plus loin sur cet exemple, il s'agit en fait des valeurs de paramètre qui ont tendance à réduire très largement le nombre de châteaux, amenant alors à une variabilité extrêmement faible dans cette amplitude des possibles restreinte.

À partir de cet histogramme, nous avons isolés les *outliers* et en proposons une représentation graphique dans la figure 6.33.

---

	Nombre : 1	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 21/10/2019 09:40:56 +02'00'
--	------------	-----------------------	------------------------	------------------------------------

---

ok là c'est clair, mais ça vient un peu tard du coup.



### 6.2.5 Quels apports de l'analyse visuelle de sensibilité ?

1 Cette analyse de sensibilité, ~~même limitée et grossière~~, a permis de mettre en évidence un certain nombre de propriétés des paramètres et d'apporter des connaissances extrêmement utiles sur le fonctionnement du modèle face à divers types d'aléas en particulier.

~~Dans un premier temps~~, 2 3 ptons que la hiérarchie de la sensibilité des paramètres n'est pas véritablement proche de celle que l'on attendait. Le paramétrage et le calibrage ont ainsi été menés, en partie, sur des paramètres moins déterminants que notre intuition du comportement du modèle ne pouvait le laisser croire. Cette étape d'évaluation qu'est l'analyse de sensibilité gagnerait à être menée avant le calibrage du modèle afin d'obtenir une meilleure adéquation aux objectifs. Ce raisonnement est toutefois circulaire : sur un modèle moins calibré, l'analyse de sensibilité n'aurait pas forcément mis en avant les mêmes paramètres. Comme ceux-ci sont en interaction étroite, les résultats de cette analyse de sensibilité sont ainsi, eux-mêmes, extrêmement sensibles au paramétrage de base.

~~Dans un second temps~~, 4 l'analyse de sensibilité semble confirmer que le modèle est parcimonieux, relativement à son approche exploratoire du moins. On l'a dit lors de l'analyse quantitative, mais il nous semble important de le répéter tant ce résultat est rassurant au regard des craintes que l'on pouvait avoir : quand un modèle implique autant de mécanismes et paramètres, le risque est toujours important que certains soient entièrement redondants et donc superflus. Aucun des paramètres n'apparaît inutile, ce qui peut laisser entendre que les (très) nombreux mécanismes du modèle ne le sont pas non plus. Pour un modélisateur qui a des intuitions fortes sur les réactions d'un modèle aux différents changements de mécanismes et paramètres, 5 ~~c'est presque une surprise~~. Surprise d'autant plus positive que cet examen systématique des paramètres se révèle réellement une aide indéniable à la compréhension du modèle. Après plus de 5 ans à travailler régulièrement sur un modèle, il est très enrichissant d'y trouver encore des éléments inattendus voir nettement contre-intuitifs. De ce fait, l'analyse de sensibilité favorise elle aussi l'abduction (voir section 6.1.3).

Un autre point, classique, concerne les limites d'une telle analyse. Une approche entièrement quantitative, présentée dans l'introduction de cette partie, permet de s'affranchir des effets de faible comparabilité que l'on a pu constater dans l'analyse visuelle : certains paramètres ont une sensibilité globale importante, mais celle-ci se cantonne parfois à un unique indicateur, sans avoir de répercussions sur le reste du modèle. Un modèle comme SimFeodal, c'est-à-dire descriptif, exploratoire, est caractérisé par des paramètres nombreux, hétérogènes et parfois très qualitatifs. 6 Il ne serait pas pertinent de chercher à quantifier, 7 ~~juste pour mener~~ une analyse de sensibilité, tout ce qui n'a pas été quantifié dans le modèle en lui-même : objectivation des attendus dans les indicateurs graphiques, quantification de la pondération entre les différents objectifs etc.

La démarche mise en place pour cette analyse de sensibilité a été basée sur une volonté d'exploration graphique et s'inscrit, comme de nombreux aspects de

	Nombre : 1	Auteur : Paul Gourdon	Sujet : Barrer	Date : 21/10/2019 09:50:30 +02'00'
	Nombre : 2	Auteur : Paul Gourdon	Sujet : Texte inséré	Date : 21/10/2019 09:50:51 +02'00'
En dépit de certaines limites que nous avons détaillées, cette analyse				
	Nombre : 3	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 21/10/2019 09:52:06 +02'00'
D'abord / Premièrement				
	Nombre : 4	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 21/10/2019 09:52:13 +02'00'
Ensuite				
	Nombre : 5	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 21/10/2019 09:53:26 +02'00'
cela constitue une surprise				
	Nombre : 6	Auteur : Paul Gourdon	Sujet : Texte de remplacement	Date : 21/10/2019 09:55:38 +02'00'
Il				
	Nombre : 7	Auteur : Paul Gourdon	Sujet : Texte surligné	Date : 21/10/2019 09:56:53 +02'00'
uniquement pour				

ce travail de thèse, dans une approche d'analyse visuelle entièrement dédiée à l'exploration d'un modèle. Cette démarche paraît confirmer l'adéquation de ce type d'approches à la construction et à l'évaluation commune et interdisciplinaire de modèles. Les éléments de compréhension des interactions effectives du modèle et les retours auxquels ces découvertes incitent nous paraissent entièrement fructueux, et à mettre au crédit de cette approche exploratoire, visuelle, et fondamentalement très qualitative.

## Conclusion

Dans ce chapitre, nous avons présenté les dernières étapes de calibrage de SimFeodal et les résultats qui en émanent. À l'aune de ces résultats, on peut conclure que le modèle, une fois calibré, est globalement satisfaisant, sur chacune des dimensions d'objectifs poursuivis : concentration et polarisation, hiérarchisation et fixation-dissémination du peuplement. L'ajustement n'est certes jamais parfait, mais aussi bien en termes de valeurs <sup>1</sup> ~~finalem~~ obtenues que d'évolutions de ces valeurs au cours du paramétrage et du calibrage du modèle, cet ajustement nous apparaît tout à fait satisfaisant.


Dans la seconde partie de ce chapitre, nous partions du constat qu'il est difficile d'améliorer le calibrage du modèle à cause des effets d'interactions et des réponses inattendues de certains paramètres. Pour y remédier, nous avons mené une analyse de sensibilité de l'ensemble des paramètres de SimFeodal qui nous a permis de mieux comprendre l'influence de chaque paramètre et mécanisme lié sur les sorties du modèle. Cette analyse de sensibilité a aussi contribué à isoler certains paramètres qui constitueraient des leviers sur lesquels agir pour améliorer le calibrage du modèle : des paramètres sans effets de bord, qui n'agissent véritablement que sur les indicateurs auxquels ils sont directement liés.

L'analyse de sensibilité a aussi débouché sur l'identification de paramètres ayant une influence plus importante que nous ne le pensions, voir une influence inverse, de manière contre-intuitive. Ces paramètres, entre autres, nous semblent de très bons candidats à la réalisation de scénarios thématiques (voir encadré 6.2) qui permettraient de comprendre les répercussions thématiques de ces résultats simulés surprenants.


Un élément limite pourtant fortement d'éventuelles améliorations et exploitations plus fines des résultats du modèle : le risque d'*overfitting*. Au vu de la précision faible et hétérogène des connaissances empiriques par le prisme desquelles SimFeodal est évalué, il paraît difficile de chercher à comparer de nouvelles versions du modèle dont les variations sur les indicateurs sont de plus en plus fines à mesure que le calibrage progresse.

Cela n'empêche aucunement le modèle d'être utile, utilisable et utilisé. <sup>2</sup> ~~Par~~ ~~faute de temps~~, nous n'avons présenté qu'un des scénarios sur la quinzaine réalisée, mais celui-ci apporte déjà beaucoup de matière à réflexion, autant d'un point de vue restreint au modèle que, plus généralement, sur les processus thématiques que SimFeodal cherche à reproduire.

---

 Nombre : 1	Auteur : Paul Gourdon	Sujet : Barrer	Date : 21/10/2019 09:58:57 +02'00'
---	-----------------------	----------------	------------------------------------

---

 Nombre : 2	Auteur : Paul Gourdon	Sujet : Barrer	Date : 21/10/2019 10:00:50 +02'00'
---	-----------------------	----------------	------------------------------------

---