

# Exploration du comportement de SimFeodal

Version 2019-08-25

— 25/08/2019 : Rendu à Lena de tout sauf 6.3 (à faire après l'ECTQG).

## Sommaire

---

<b>Introduction</b>	<b>2</b>
<b>6.1 Calibrage du modèle et premiers résultats</b>	<b>3</b>
6.1.1 Différencier objectifs contextuels et émergents	3
6.1.2 Calibrage du modèle	5
6.1.3 Résultats du modèle	12
6.1.4 Après le calibrage, comment affiner le modèle?	22
<b>6.2 Analyser la sensibilité de SimFeodal</b>	<b>24</b>
6.2.1 Méthodologie - Analyse visuelle de sensibilité	25
6.2.2 Sélection des paramètres	32
6.2.3 Évaluation visuelle de la sensibilité	36
6.2.4 Analyser la sensibilité à l'aléa	42
6.2.5 Conclusion et apports de l'analyse visuelle de sensibilité	45
<b>6.3 Exploration de scénarios</b>	<b>47</b>
6.3.1 Tester l'hypothèse d'une croissance démographique	47
6.3.2 Modéliser la dépendance spatiale : le poids du servage	47
6.3.3 Quel rôle et importance des communautés paysannes dans la structuration du système de peuplement?	47
<b>Conclusion</b>	<b>47</b>

---

## Introduction

Les chapitres précédents ont présenté le modèle SimFeodal ([chapitre 2](#)), la manière de l'évaluer ([chapitre 3](#)), la méthode suivie pour le paramétrage du modèle ([chapitre 4](#)) et les outils développés pour mener à bien ces étapes de construction et d'évaluation ([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 trois parties, assez indépendantes les unes des autres, 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 aboutit-il bien à une polarisation des foyers paysans ? Le système de peuplement généré par le modèle est-il hiérarchisé tel qu'on l'attendait, et sa distribution est-elle proche des connaissances empiriques ? Observe-t-on une fixation du peuplement, dans un espace plus disséminé que dans les configurations initiales ? Pour répondre à ces questions, on présente les étapes de calibrage qui ont permis d'aboutir à une version « définitive » du modèle, ainsi que les éléments les plus marquants des indicateurs de sortie issus de cette version.

La deuxième partie de ce chapitre s'attachera à une exploration du comportement en lui-même, 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). Dans le cas de SimFeodal, une analyse de ce type pose toutefois des questions méthodologiques complexes, en raison notamment du grand nombre de paramètres impliqués et de « l'explosion combinatoire » qui en découle. Un autre perspective sur la sensibilité du modèle concerne sa robustesse face à l'aléa : dans quelle mesure les différents paramètres – et leurs valeurs associées – jouent-ils sur la variabilité des indicateurs de sortie ?

La dernière partie doit aussi permettre de gagner en compréhension du modèle en lui-même, et par cela de ce qui est modélisé. Il s'agira d'explorer des « scénarios », c'est-à-dire de tester des hypothèses, sous la forme d'ensemble de valeurs de paramètres ayant un sens empirique, pour lesquelles le modèle n'a pas été calibré. La réaction du modèle à ces scénarios sera l'occasion de comprendre plus profondément le comportement du modèle, mais aussi d'esquisser de premières pistes de réponses sur des questions empiriques que se posent les thématiciens spécialistes de la période étudiée. Ces scénarios peuvent de plus concourir à l'évaluation du modèle. Ils agissent comme des méthodes de « validation croisée » qualitative : si le modèle donne des résultats satisfaisants sur des éléments pour lesquels il n'a pas été construit, cela contribue à renforcer la vraisemblance des postulats sur lesquels il est conçu.

## 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) 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 Différencier objectifs contextuels et émergents

**N.B. :** Toute cette partie 6.1.1 (et sans doute 6.1.2) serait bien plus logiquement située dans le chapitre 4 sur le paramétrage. Il faudra notamment y mettre à jour la typologie des paramètres, et sans doute amener aussi le distinguo entre objectifs émergents et objectifs contextuels, qui en découle largement.

Dans les chapitres précédents, on a présenté de manière extensive les objectifs – dans tous les sens du terme – poursuivis par le modèle. Dans le chapitre 2, on décrivait l’objectif théorique du modèle, à savoir d’aider à la compréhension des phénomènes de restructuration spatiale du système de peuplement rural entre les IX<sup>e</sup> et XIII<sup>e</sup> siècles. Dans le chapitre 3, on présentait la formalisation, au moyen d’indicateurs de sortie de simulation, des résultats du modèle. En définissant un ensemble d’objectifs, qualifiés (allures de courbes, ordres de grandeur...) sinon quantifiés (les indicateurs numériques), on esquissait un crible d’évaluation du modèle.

Pourtant, même au sein des indicateurs de sortie de simulation, il est une distinction que l’on n’a pas encore réalisée : le distinguo entre des indicateurs « émergents » et des indicateurs « contextuels ».

- Les premiers 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 l’intrication des mécanismes du modèle. Parvenir à obtenir des valeurs proches des connaissances empiriques dans ces modèles est un critère de validation d’un modèle, et donc potentiellement un signe d’acquisition de connaissance thématique ou théorique. Par exemple, dans SimFeodal, le taux de foyers paysans dispersés en fin de simulation est un pur indicateur émergent : le modèle n’a nullement été paramétré

ou calibré pour lui faire atteindre une certaine valeur, et le taux obtenu donne des éléments d'interprétation thématique sur le modèle.

- Les indicateurs contextuels n'apportent aucune connaissance thématique. Ils remplissent un rôle de cadrage pour les indicateurs émergents : le modèle est paramétré et calibré pour que ces indicateurs parviennent à un résultat pré-décidé. Le calibrage des paramètres associés (c'est-à-dire des paramètres qui ont un impact majeur sur ces indicateurs) permet alors d'obtenir un contexte dans lequel les indicateurs émergents pourront être exploités. Pour illustrer ces indicateurs de contexte, on peut prendre l'exemple, dans SimFeodal, du nombre de foyers paysans en fin de simulation. Cette quantité est entièrement dépendante de deux paramètres (quantité initiale et taux de croissance), et peut ainsi être assimilée à un *input*, c'est-à-dire ici un contexte au sein duquel les autres indicateurs s'exprimeront.

Le tableau des objectifs présent dans SimEDB (figure 5.8 du chapitre 5) correspond à une partie des indicateurs de sortie (les indicateurs numériques ici) de ces deux types, tel qu'indiqué dans le tableau 6.1.

Indicateur	Valeur attendue <sup>1</sup>	Type	Dépendances directes <sup>2</sup>
<i>Agrégats</i>	200	Émergent	–
<i>Châteaux</i>	50	Contextuel	probabilités de construction de châteaux (PS et GS)
<i>Gros châteaux</i>	10	Émergent	–
<i>Seigneurs</i>	200	Contextuel	paramètre dédié : (objectif_nombre_seigneurs)
<i>Églises paroissiales</i>	300	Émergent	–
<i>Distance moyenne entre églises</i>	3 000 m	Émergent	–
<i>Part de foyers paysans isolés</i>	20 %	Émergent	–
<i>Augmentation de la charge fiscale des foyers paysans</i>	x 3	Émergent	–
<i>Nombre de Foyers Paysans</i>	50 000	Contextuel	nombre initial; taux de croissance
<i>Densité de population</i>	8 feux/km <sup>2</sup>	Contextuel	nombre de foyers paysans; taille du monde

TABLEAU 6.1 – Les objectifs numériques, contextuels et émergents, de SimFeodal.

Notons que si ces indicateurs ont une cohérence à être présentés ensemble au sein de SimEDB, dans le cas d'un tableau de synthèse dédié à l'exploration des résultats de simulation, leur traitement doit nécessairement être différent

2. Objectif en fin de simulation.

2. Paramètres influant directement sur la valeur de l'indicateur.

en matière d'évaluation du modèle. Cette distinction nous paraît importante en vue d'aborder le calibrage du modèle. En effet, étant donné la large quantité de paramètres dans le modèle, dont une combinaison judicieuse peut certainement aboutir à des résultats diamétralement opposés, il était nécessaire de restreindre le nombre de paramètres sur lesquels agir pour calibrer le modèle. Pour que ce calibrage ne soit pas guidé par une logique tautologique, en contraignant le modèle à produire des résultats attendus, on a choisi de mener le calibrage sur les éléments de contexte et sur les éléments techniques (cf. la typologie du chapitre 4.<sup>3</sup>).

### 6.1.2 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 plus importante dans la manière dont le modèle réagit, mais de se concentrer sur quelques paramètres et indicateurs de sortie<sup>4</sup>, contextuels, dont on va régler finement les valeurs de paramètre afin que le contexte de déroulement des autres mécanismes du modèle soit aussi contrôlé et fidèle aux connaissances expertes que possible.

Le calibrage d'un paramètre peut résulter d'un ajustement issu de simulations, par exemple en trouvant de manière expérimentale une valeur de paramètre qui assurera un écart minimal entre l'indicateur de sortie de simulation visé et sa correspondance empirique. Le calibrage peut aussi être thématique, par exemple 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, ce qui est d'ailleurs vraisemblablement fait dans la plupart des modèles. Dans un modèle descriptif, exploratoire et surtout basé sur des dynamiques passées, pourtant, la 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 suivant, nous donnons des exemples de paramètres et indicateurs qui ont été calibrés dans les dernières phases de paramétrage du modèle. Ces exemples reprennent la distinction établie ci-dessus entre les types de calibrages : le calibrage des *inputs* est entièrement thématique, celui des paroisses est mixte, entre thématique et expérimental, et le calibrage des

---

3. Le chapitre 4 est à reprendre, notamment en supprimant la typologie des paramètres commensurables, techniques etc. et en la remplaçant par la typologie utilisée dans le modèle, cf. tableau des paramètres : inputs, paramètres de contexte, paramètres de mécanismes, paramètres techniques.

4. 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.

châteaux est purement expérimental, guidé par un objectif thématique pré-défini clair.

## Inputs

**Taille du monde** 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 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 l'exemple de la Touraine historique, mais sans hésiter à inclure quelques éléments du duché d'Anjou voisin s'ils avaient été, à un moment ou à un autre, intégrés à notre région d'étude.

En faisant le choix de caler les dimensions du monde simulé plus fortement sur celles de la Touraine historique, il a fallu en premier lieu revoir à la baisse la taille du monde simulé. La Touraine historique a ainsi 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 carré et théorique, de diminuer la taille des côtés du monde simulé 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.

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 une large partie des paramètres techniques et certains paramètres de contexte. 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, 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, mais on le considère comme un *input* guidé par les connaissances empiriques plus que comme un élément contextuel. Il nous faut ici pondérer fortement ce recours aux connaissances empiriques : il est extrêmement difficile, sinon vain, de parvenir à une simple estimation de la population d'une région française durant la période étudiée. On peut avoir des indices sur des ordres de grandeur de population à des moments clés de l'histoire, mais d'une part, ils sont extrêmement incertains, et d'autre part, ces moments clés sont souvent différents des seuils temporels sur lesquels nous nous basons. 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 présentent des écarts ma-

jeurs.

Pour SimFeodal, on a repris une hypothèse d'historien, qui semble ne pas faire trop débat 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 alors diminué considérablement. Les historiens estiment que les populations du XIII<sup>e</sup> siècle n'ont été rattrapées qu'au début du XVII<sup>e</sup>. Il est de fait possible de considérer les populations à l'issu de la période étudiée comme 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>, ou encore d'environ 30 habitants/km<sup>2</sup>.

#### Mettre aussi nb de villes héritées de l'antiquité

La population initiale est bien plus difficile à estimer. Selon les sources<sup>5</sup>, certains présentent ainsi 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 optimal 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 porte 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 section 6.3.1).

### Paroisses

Le calibrage des paroisses a aussi demandé un calibrage important, et portant cette fois-ci autant sur des aspects 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 voulue et encouragée des migrations individuelles des foyers paysans. De plus, ce maillage influe à son tour sur les futures migrations de ces mêmes foyers paysans : les paroisses sont autant une résultante des changements de structure spatiale qu'une partie de leur explication.

Au sein du modèle, la transformation du maillage paroissial joue ainsi un rôle contextuel, conditionnant les migrations des foyers paysans, et un rôle

---

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



émergent, que l'on observe à une échelle plus agrégées (densification ou étalement, rythmes de changements etc.)

Pour que le contexte permette la meilleure réalisation des phénomènes émergents, il a été nécessaire de calibrer certaines propriétés structurelles du maillage paroissial.

**Paroisses « rurales »** Pour les paroisses « rurales », on se base plutôt sur des aspects thématiques issus de connaissances archéologiques. On connaît ainsi, au moins pour la fin du XII<sup>e</sup> siècle, le nombre et la répartition spatiale de la plupart des paroisses (301 églises selon ZADORA-RIO (2008, p. 31), que l'on extrapole par simplicité en 300 paroisses), qui correspondent majoritairement à des milieux ruraux et qui ont en partie perduré dans le maillage communal actuel. En termes de répartition spatiale, par soucis de généricité, on peut tirer du semis d'églises de l'Indre-et-Loire féodale des distances d'espacement moyennes entre églises paroissiales<sup>6</sup>, que l'on utilise ensuite pour calibrer le modèle.

Pour cela, on a joué sur le mécanismes de création/promotion de paroisses rurales et sur les paramètres associés, de manière à approcher des résultats empiriques. On a fait varier le paramètre `seuil_nb_paroissiens_insatisfaits` : 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 6.2 page 13), mais on n'a pas augmenté le seuil pour des raisons thématiques. Le seuil est fixé à 20 foyers paysans, et thématiquement, on estime que cette quantité pouvait suffire à la création d'une nouvelle paroisse.

**Paroisses « urbaines »** On dispose de moins de données empiriques pour les paroisses urbaines que pour les paroisses rurales. 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 seul paramètre (`ponderation_creation_paroisse_agregat`) contrôle la création de paroisses au sein des agrégats. Ce paramètre est un « paramètre de mécanisme », quand bien même sa valeur est assez éloignée de l'empirie : elle définit le seuil de foyers paysans (par paroisse urbaine, c'est-à-dire pondéré par le nombre de paroisses présentes dans l'agrégat) à partir duquel la probabilité de créer une nouvelle paroisse dans l'agrégat atteint 1. On a procédé par calibrage manuel, en testant différentes valeurs pour ce seuil,

---

6. On retrouve ainsi cette estimation dans CHAREILLE, RODIER et ZADORA-RIO (2008, p. 261) : « 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. »



tout en restant dans des ordres de grandeur acceptables d'un point de vue empirique (on n'aurait par exemple pas pu placer ce seuil à 10 foyers paysans, ce qui aurait signifié qu'on devait créer des paroisses secondaires dans chaque petit agrégat, soit une paroisse par hameau...).

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 de forme entre deux pas de temps peut « exclure » une église paroissiale de son emprise, parfois de quelques dizaines de mètres seulement. Pour minimiser la modification du paramètre de seuil, on a alors agi sur le critère de définition des paroisses d'un agrégat, en ajustant les distances et mécanismes impliqués.

## Châteaux

Le calibrage des châteaux a été bien plus simple d'un point de vue thématique : 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>7</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>8</sup> le nombre de châteaux, en Touraine, à une cinquantaine.

Dans SimFeodal, comme 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 une importance certaine, de manière contextuelle, sur le déroulement des simulations. Une exigence du paramétrage était que le nombre (et le type) de châteaux soit aussi stable que possible au regard des autres mécanismes du modèle et surtout des variations des autres paramètres.

**Nombre de châteaux** Dans les premières versions du mécanisme, des seuils fixes, dépendant de la puissance des seigneurs et donc techniques, avaient ainsi été définis pour caractériser la probabilité de chaque seigneur de construire un château. Dès lors que les modifications du modèle ont amené à des changements dans la population des foyers paysans (les scénarios démographiques notamment, voir section 6.3.1), le nombre de châteaux a été fortement impacté.

---

7. 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.

8. 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.

Afin que ce nombre reste stable, on a adapté le mécanisme et introduit de nouveaux paramètres techniques permettant de le contrôler. La figure 6.1 donne un exemple de test de l'un de ces paramètres (qui définit le nombre de tirages aléatoire qu'un grand seigneur peut réaliser pour éprouver sa probabilité de construire un château). Ce paramètre ayant une influence visuellement linéaire sur le nombre de châteaux en fin de simulation, on peut identifier dans ce graphique que pour obtenir 50 châteaux en fin de simulation (graphique de gauche), la valeur la plus adaptée du paramètre `nb_tirages` est de 3.

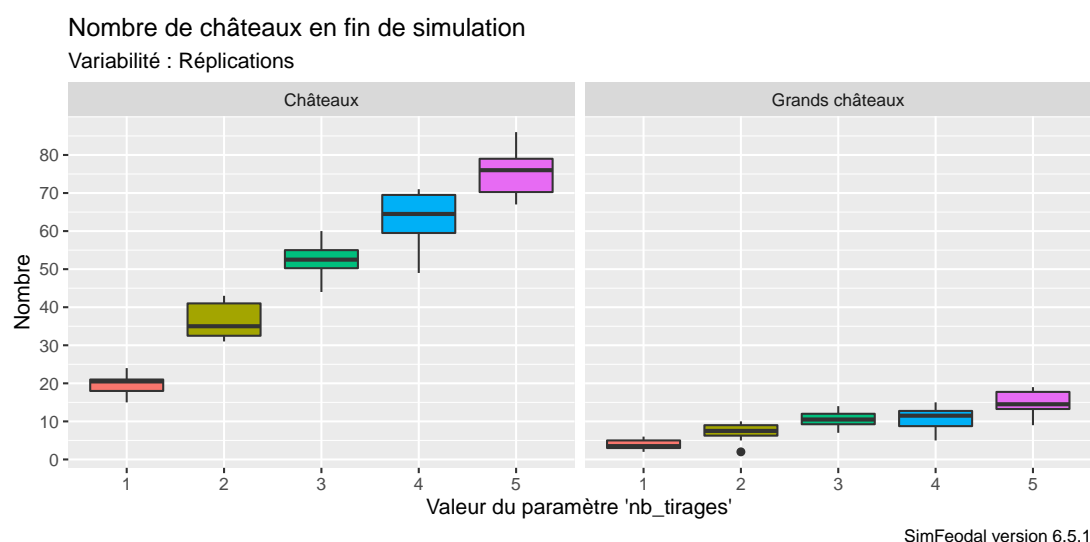


FIGURE 6.1 – Influence du paramètre `nb_tirages` sur le nombre de châteaux<sup>9</sup>.

En jouant sur les différents paramètres associés à la construction de château (nombre de tirages et pondérateurs de probabilité pour les grands seigneurs, paramètres liés à la probabilité pour les petits seigneurs...), on parvient à obtenir un résultat relativement stable dans le nombre de châteaux obtenus (figure 6.2).

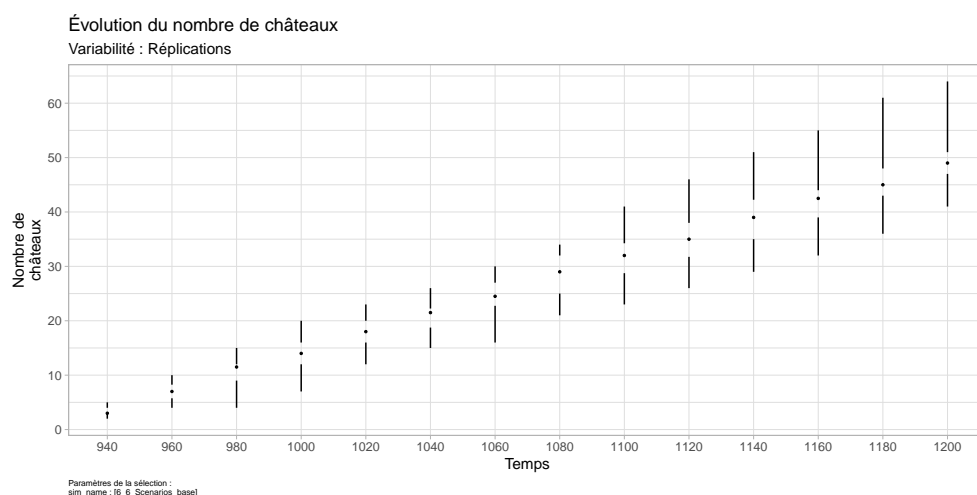


FIGURE 6.2 – Évolution du nombre de châteaux après calibrage. Version 6.6 de SimFeodal.

**Types et créateurs** Le nombre de châteaux n'est toutefois pas la seule valeur empirique sur laquelle on essaie d'ajuster le contexte. On peut ainsi

9. Ce graphique représente une version légèrement différente de la 6.5.1 présentée dans le reste du chapitre. Certaines valeurs de paramètres y sont adaptées au calibrage, fin, des effets de contexte liés aux châteaux.

estimer grossièrement la hiérarchie des châteaux, que nous avons choisi de discrétiser entre grands et petits châteaux. Dans l'ensemble, on estime à une dizaine le nombre de « grands châteaux », c'est-à-dire de « forteresses »<sup>10</sup>. 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), 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.

Pour que le contexte soit correct en matières de châteaux, il était nécessaire de prendre en compte ces particularités dans le calibrage des mécanismes de création et promotion de château. Le nombre de châteaux n'était pas suffisant, et les paramètres ont du être ajustés afin que ces éléments d'affinage (types des châteaux et types de leurs constructeurs) correspondent aux connaissances empiriques. 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.

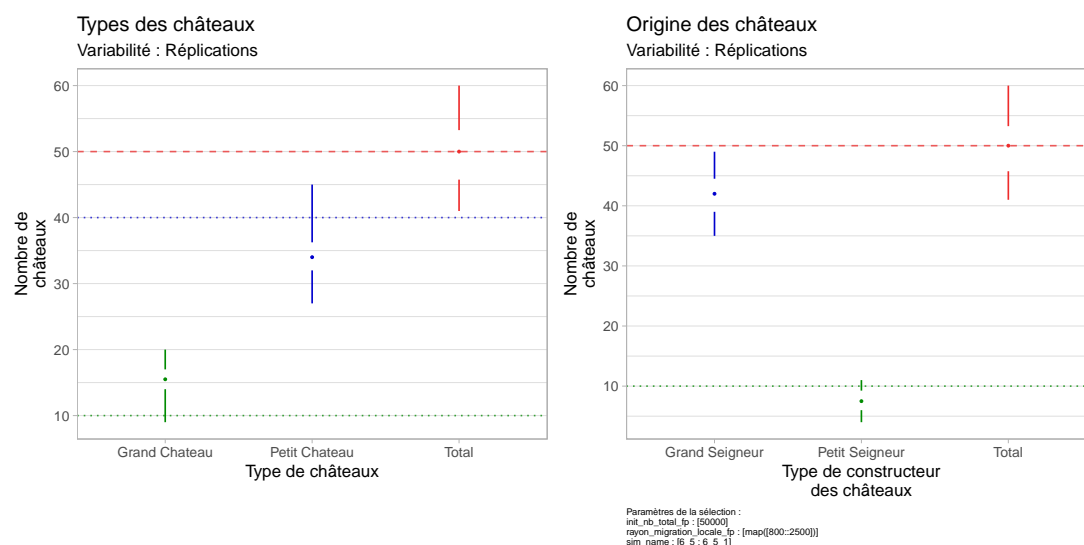


FIGURE 6.3 – Détail de la composition des châteaux en fin de simulation à l'issue du calibrage de SimFeodal.

*Les lignes horizontales en pointillés représentent les objectifs à atteindre.*

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

### 6.1.3 Résultats du modèle

Nous avons largement décrit, dans le [chapitre 3](#), les objectifs poursuivis par le modèle et les indicateurs de sortie de simulation employés pour les évaluer. Pour rappel, ces objectifs peuvent être catégorisés en trois familles, selon les objectifs thématiques qu'ils cherchent à reproduire :

- Polarisation du système de peuplement : le modèle parvient-il à faire émerger une structure polarisée et concentrée de l'habitat, où les foyers paysans sont concentrés dans des agrégats de population plutôt que dispersés comme dans la situation initiale ?
- Hiérarchisation du système de peuplement : depuis une situation initiale composée d'une faible hiérarchie dans les agrégats (des « agglomérations secondaires antiques » d'une trentaine de foyers et des « villages » d'une dizaine de foyers), parvient-on à une structure à la hiérarchie importante, proche du modèle log-normal identifié dans la majorité des systèmes de peuplement historiques et contemporains ?
- Fixation et dissémination du peuplement : on estime que la population, initialement assez mobile (sur la résolution temporelle modélisée) tend peu à peu à se fixer, c'est-à-dire à trouver une localisation dans laquelle leur satisfaction ne diminuera plus. Cette fixation, dans des agrégats, doit paradoxalement s'assortir d'une dispersion dans l'espace des agrégats : d'une occupation dispersée et quasi-aléatoire, l'objectif thématique est que les agrégats maillent progressivement l'ensemble du monde simulé. Observe-t-on bien ces deux processus dans le déroulement des simulations ?

Dans cette partie, nous allons synthétiquement commenter les indicateurs de sortie de simulation issus de la version calibrée (6.6) de SimFeodal, en analysant l'écart entre les objectifs attendus, thématiques, et les résultats du modèle.

Par soucis de parcimonie et de synthèse, les résultats présentés par la suite ne sont qu'une sélection de l'ensemble des résultats du modèle. Nous invitons le lecteur à les consulter directement dans l'application SimEDB d'où ces indicateurs sont extraits. Ce [lien direct](#) permet d'accéder aux résultats spécifiques à la version présentée ici : [LIEN direct](#)

#### 6.1.3.1 Résumé global des résultats

Avant de chercher à analyser de manière approfondie les résultats du modèle, le tableau 6.2 peut déjà synthétiser une bonne part des résultats agrégés, en fin de simulation.

On y lit en premier lieu une sur-représentation nette de la plupart des objectifs quantitatifs. Le nombre d'agrégat et d'églises paroissiales, entre autre, est trop important, et sa variabilité suffisamment faible pour que cette sur-représentation puisse être considérée comme significative. Le taux de foyer paysan isolés en fin de simulation est lui trop faible par rapport aux attentes, quand bien même l'intervalle de confiance empirique est assez flou.

Bien que ces résultats soient systématiquement éloignés des objectifs, on notera tout de même que les ordres de grandeur sont relativement respectés, à l'exception peut-être de la distance entre églises paroissiales, dont la concentration de certains au sein des agrégats ne peut que fausser la mesure.

Indicateur	Valeur attendue <sup>11</sup>	Moyenne	Médiane	Q1	Q3	Écart-type
<i>Agrégats</i>	200	249	248	244	253	10.45
<i>Châteaux</i>	50	49	49	47	51	5.82
<i>Gros châteaux</i>	10	15	15	13	17	2.87
<i>Seigneurs</i>	200	198	196	191	204	8.79
<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
<i>Nombre de Foyers Paysans</i>	—	40 000	40 000	40 000	40 000	0
<i>Densité de population</i>	—	6.25	6.25	6.25	6.25	0

TABLEAU 6.2 – Indicateurs numériques de la version 6.6 de SimFeodal.

### 6.1.3.2 Polarisation des foyers paysans

Les résultats d'ensemble ne sont pas entièrement satisfaisants, mais on y constate tout de même une forte concentration des foyers paysans, dont la part de foyers isolés diminue de manière importante, d'environ 90% à 30% en fin de simulation (figure 6.4).

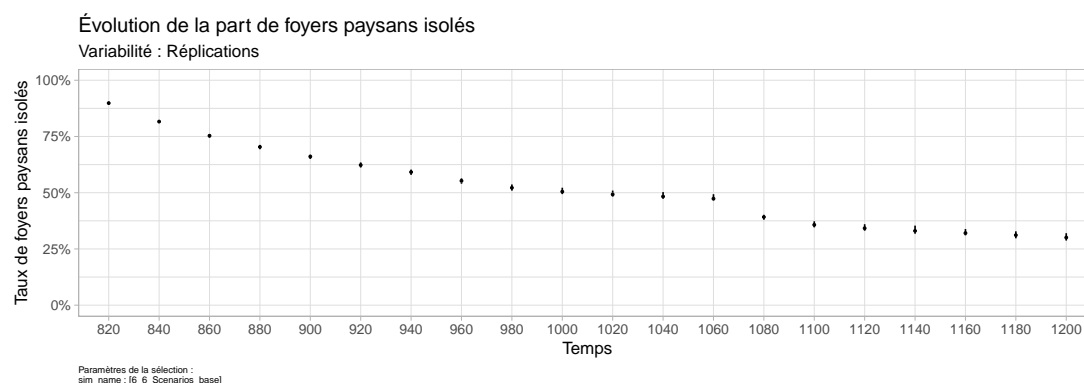


FIGURE 6.4 – Concentration des foyers paysans.

Cette diminution paraît assez régulière, en dépit d'une légère rupture de tendance entre 1060 et 1080, période qui correspond dans le modèle à une évo-

11. Objectif en fin de simulation

lution du seuil de satisfaction religieuse. Notons que parmi les 20 répliques étudiées, cet indicateur se montre extrêmement stable, les marques visuelles de variabilité étant à peine visibles.

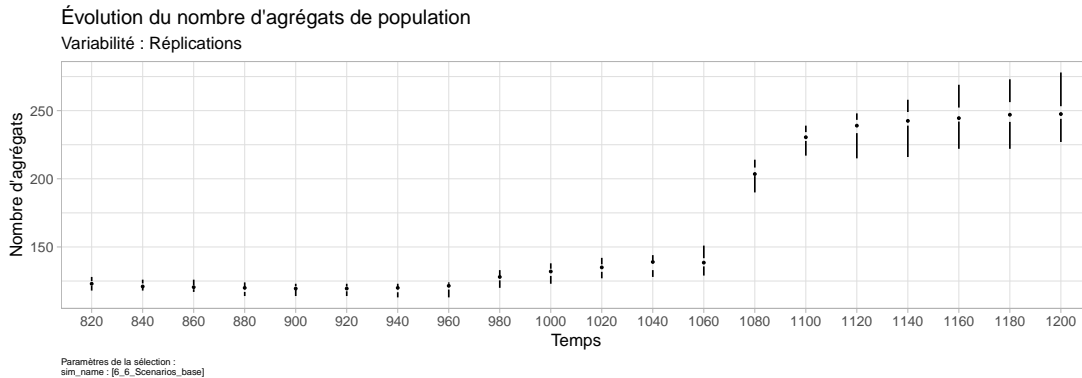


FIGURE 6.5 – Nombre d'agrégats.

La concentration des foyers paysans s'effectue à destination d'un nombre croissant d'agrégats (figure 6.5), nombre quant à lui bien plus affecté par l'effet de seuil en 1060 noté dans la concentration. On y retrouve les trois régimes estimés empiriquement, c'est-à-dire une augmentation lente, suivie d'une augmentation rapide et enfin une stabilisation du nombre d'agrégats.

Pour pouvoir parler de polarisation et pas seulement de concentration, il faut que le modèle soit en mesure non seulement de concentrer les foyers paysans dans des agrégats, mais aussi que ces agrégats se constituent autour des différents attracteurs du modèle (églises paroissiales, châteaux et agrégats dotés de communautés paysannes). Dans cette version du modèle, on constate bien une croissance des pôles (figure 6.6-a), assez semblable à celle des agrégats en termes de rythme et de valeurs. Par rapport à la version 0 commentée dans le [chapitre 3](#), les valeurs sont très satisfaisantes, qui plus est au regard de l'évolution de la répartition des pôles, qui parviennent très rapidement (peut-être même trop) à la création d'agrégats dans leur voisinage (figure 6.6-b).

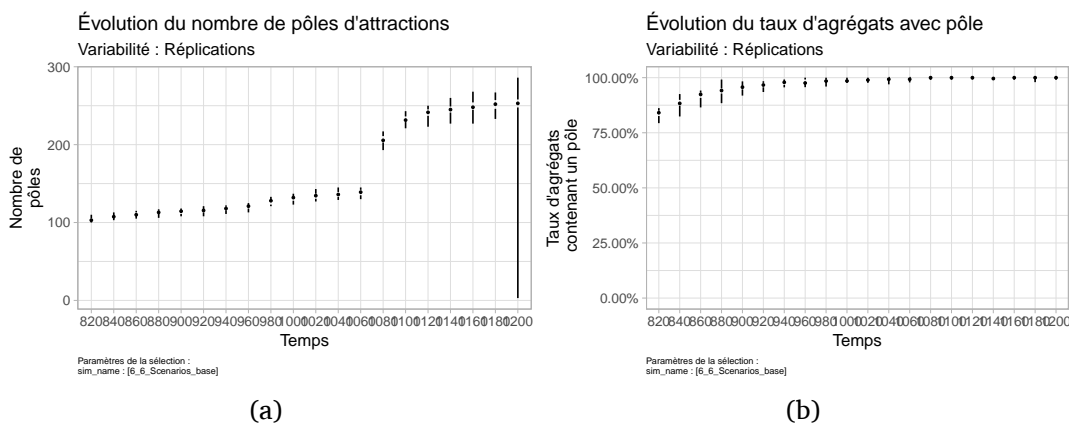


FIGURE 6.6 – Nombre de pôles et part des agrégats comprenant un pôle.

Ces agrégats et les pôles correspondant sont bien plus dispersés dans l'espace du modèle que dans la version 0, et on constate, au moins visuellement, que le semis des pôles se confond avec celui des agrégats (figure 6.7). Sur ces cartes, on constate une occupation importante de l'espace, largement due à une population bien supérieure en version 6.6 (40 000 foyers paysans, contre

4 000 en version 0), mais on remarque toutefois sur la troisième carte, lissée, qu'en dépit d'un semis visiblement homogène, des zones de plus forte densité sont présentes.



FIGURE 6.7 – Dispersion spatiales des agrégats et pôles.

Sur le plan de la polarisation d'ensemble des foyers paysans au sein d'agrégats de population, les résultats de SimFeodal répondent entièrement aux attentes. La valeur finale est certes légèrement inférieure à l'objectif, mais tous les ordres de grandeurs et surtout les rythmes estimés correspondent largement aux estimations issues des connaissances expertes.



### 6.1.3.3 Hiérarchisation du système de peuplement

Les dernières figures étudiées montraient une forte hétérogénéité dans la taille des agrégats, ce qui est constitué déjà un indice sur la hiérarchie de ces concentrations locales de foyers paysans.

Comme indiqué dans le chapitre 3, il est difficile d'avoir des mesures précises de la hiérarchie attendue dans le système de peuplement. Les différentes sources historiques divergent aussi bien sur les quantités absolues que sur la forme des distributions. Rappelons tout de même que les sources s'accordent sur une nette hiérarchisation qui doit tendre vers les formes log-normales que l'on retrouve dans les sociétés contemporaines.

	Nombre de foyers paysans					
	< 100	101-200	201-300	301-400	401-600	> 600
Nombre moyen	142	78	18	7	4	2
Taux moyen	56.8%	31.4%	7%	2.9%	1.6%	< 0.1%
<b>Objectif (taux)</b>	<b>?</b>	<b>?</b>	<b>?</b>	<b>?</b>	<b>?</b>	<b>?</b>

TABLEAU 6.3 – Distribution des agrégats par classe de taille en fin de simulation. **Ne conserver que si Cécile arrive à trouver, pour l'article anglais, une validation des objectifs par EZR/SL.**

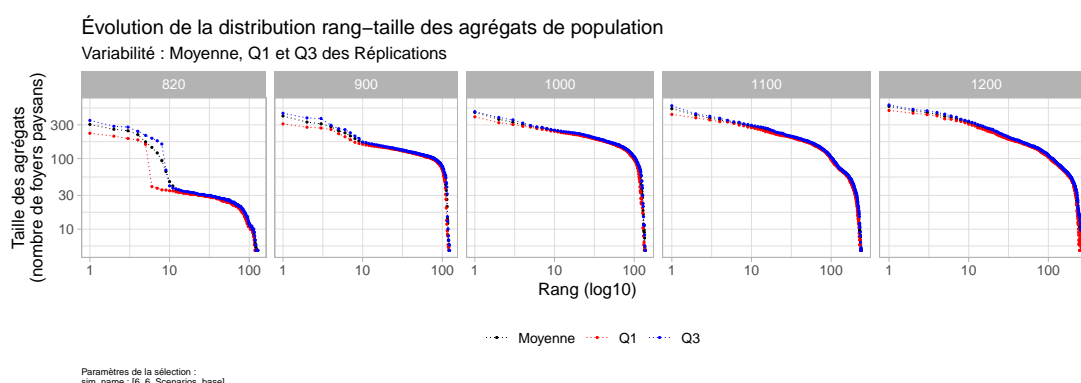


FIGURE 6.8 – Hiérarchie des agrégats.

La figure 6.8 montre une claire hiérarchisation des agrégats : la courbe se « redresse » au cours du temps simulé, marque d'une pente croissante. Les valeurs absolues augmentent aussi : les plus gros agrégats voient leur population croître. Le « coude » dans la courbe, correspondant à la longue traîne de petits agrégats, se réduit, mais on constate tout de même que le bas de la hiérarchie se renforce. Contrairement à la version 0 du modèle, la croissance de tous les agrégats semble constante, et on ne remarque pas, visuellement, les tendances à l'éclatement des gros agrégats qui caractérisaient la hiérarchie de cette version.

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'agréats et pôles se confondaient, ce qui constitue en soi un résultat satisfaisant. La mesure représentée dans cette première figure, le nombre d'attracteurs de chaque

pôle, est beaucoup plus discriminante et discrète que le nombre de foyers paysans des agrégats. La courbe en est donc d'autant plus évidente, montrant une forte hiérarchisation qui s'assortie d'une hausse importante du nombre maximum d'attracteurs des pôles les plus importants. Cela montre bien l'apparition d'une tête de hiérarchie urbaine dans le modèle, qui trouve une correspondance dans les villes (Amboise, Loches, Chinon...) de la région d'étude, organisées autour de châteaux et composées de multiples églises paroissiales. La figure 6.9-b montre aussi cette hiérarchisation : elle met en évidence un glissement des modalités d'attraction depuis une valeur de 0.2 (deux églises paroissiales) à un double mode à 0.4 (deux églises et une communauté) et 0.7 (plusieurs églises, un château, une communauté...).

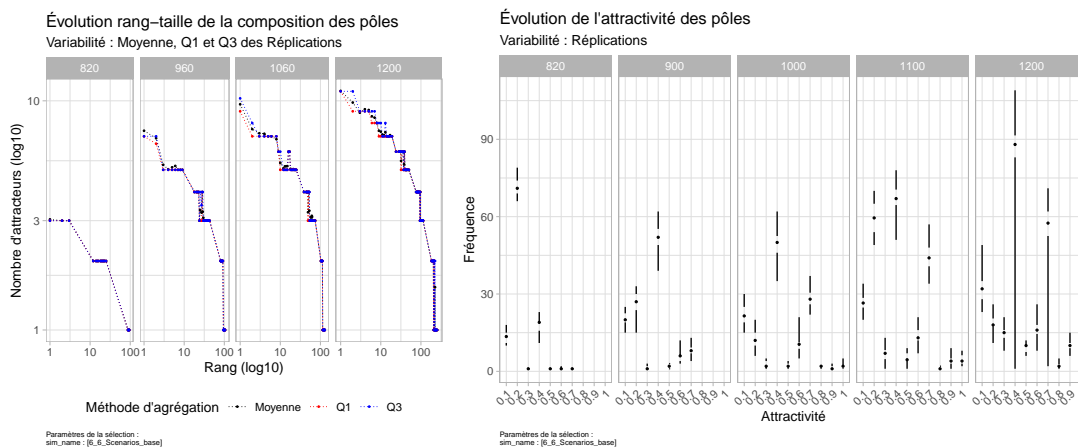
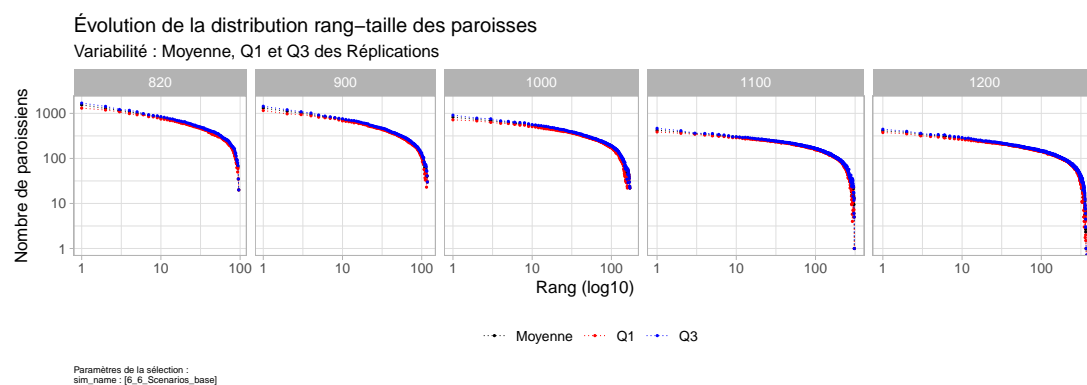


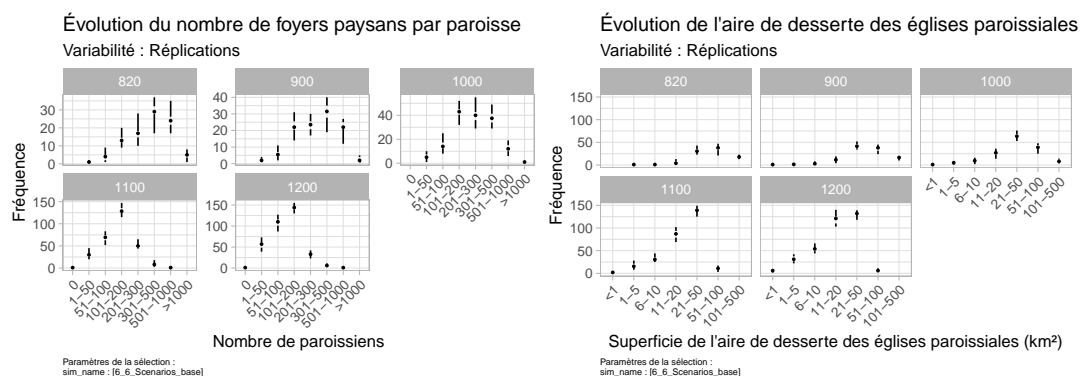
FIGURE 6.9 – Hiérarchie des pôles.

Dernier élément du modèle en lien avec la hiérarchisation du peuplement, les paroisses. Plutôt qu'une hiérarchisation, on cherche à y reproduire deux faits stylisés. En premier lieu, une large majorité des paroisses, que l'on pourrait nommer « rurales », doivent être peu fréquentées et viser surtout à une desserte équitable de la population. En second lieu, pour les paroisses « urbaines », on souhaite qu'elles aient un nombre de paroissiens bien supérieur à celui des paroisses rurales, tout en restant largement inférieur au millier de paroissiens. Dans les agrégats les plus peuplés, il doit en effet y avoir suffisamment de paroisses pour que les fidèles ne soient pas lésés.

Dans les graphiques de la figure 6.10, on constate un phénomène inattendu, à savoir une tendance exactement inverse à celle des agrégats. La convexité de la courbe augmente au cours du temps simulé, et les valeurs décroissent régulièrement. En regardant le détail (figure 6.10-b), on peut remarquer que cela correspond en fait à une forte homogénéisation dans l'intervalle de 50 à 200 paroissiens. Cet intervalle correspond certainement aux paroisses rurales qui contiennent quelques agrégats ruraux de taille moyenne à faible (figure 6.8). On assiste bien à une 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.



(a)



(b)

(c)

FIGURE 6.10 – Hiérarchie des paroisses.

**! Attention ! les seuils sont complètement différents du chapitre 3.**

Dans l'ensemble, la hiérarchisation est certaine mais moins connue empiriquement que la polarisation du 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 bien les hypothèses estimées, quand bien même ces dernières sont assez floues au regard d'autres éléments empiriques. 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.

### 6.1.3.4 Fixation et dissémination du peuplement

Le dernier objectif thématique est encore plus incertain que la hiérarchisation du peuplement. Il s'agit d'un double objectif à deux échelles différentes. À l'échelle des foyers paysans, on attend que les migrations croissent afin de restructurer le peuplement, puis diminuent, signes de la fixation des foyers paysans dans un nouvel agrégat qui leur apportera une satisfaction suffisante pour ne pas trop subir les évolutions négatives de leur environnement (distance à l'église, besoin de protection, hausse de la pression fiscale...). À l'échelle des agrégats en eux-même, on souhaite que ceux-ci, tel qu'observé empiriquement, soient dispersés dans l'ensemble de la région simulée.

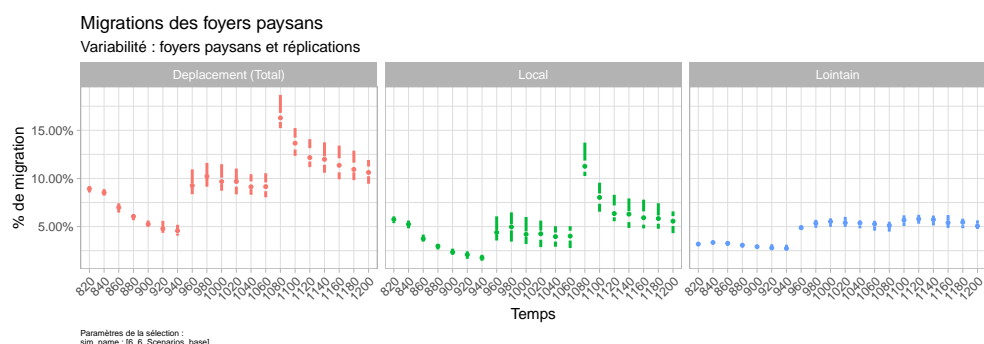


FIGURE 6.11 – Migration des foyers paysans.

Pour les foyers paysans, on peut observer ces tendances à l'aide d'indicateurs sur le nombre et le type de migrations au cours du temps (figure 6.11). Le décompte des migrations révèle un profil très intéressant. Au niveau des migrations lointaines, la tendance est globalement à la stabilité, à l'exception d'un décrochage très brusque entre 940 et 960. Celui-ci trouve peut-être une explication dans l'apparition du besoin de protection : d'un coup, les foyers paysans perdent beaucoup en satisfaction (figure 6.12) et le taux de foyers paysans effectuant des migrations augmente en conséquence. Au niveau des migrations locales, le comportement est plus explicable et attendu malgré son apparence surprenante. On y lit très clairement les deux périodes de bouleversement exogène amenées par les diminutions des seuils de distances aux églises. Ces événements amènent à des migrations assez massives, qui tendent après très nettement à se résorber : les foyers paysans isolés et ceux qui sont dans des agrégats peu stables rejoignent des agrégats plus peuplés et pérennes, et à mesure qu'ils y parviennent, le nombre de migrations locales diminue jusqu'à atteindre des niveaux assez faibles comparables à ceux des migrations lointaines (autour de 5% en 1200).

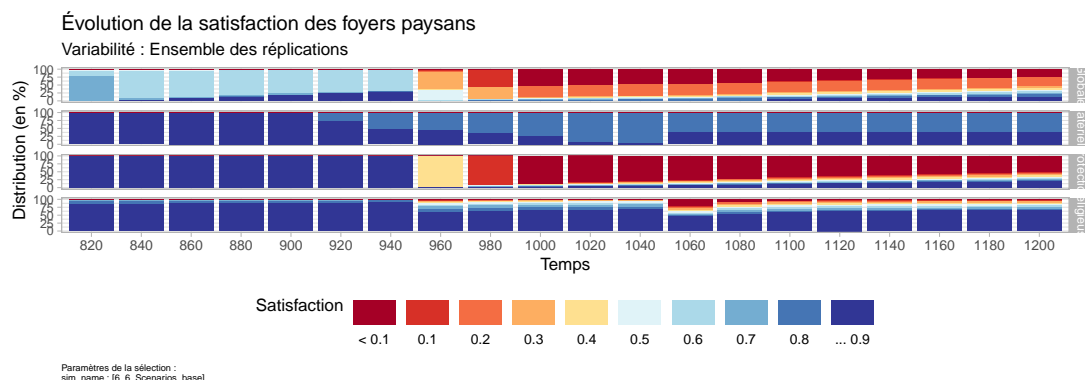


FIGURE 6.12 – Satisfaction des foyers paysans.

Si l'on observe le détail des migrations (figure 6.13), on peut noter deux périodes. En 820 et en 900, la plupart des migrations proviennent de foyers paysans isolés. Ces migrations, locales et lointaines, permettent aux foyers paysans de rejoindre un agrégat, quel qu'en soit la place dans la hiérarchie. En 1000 et 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.

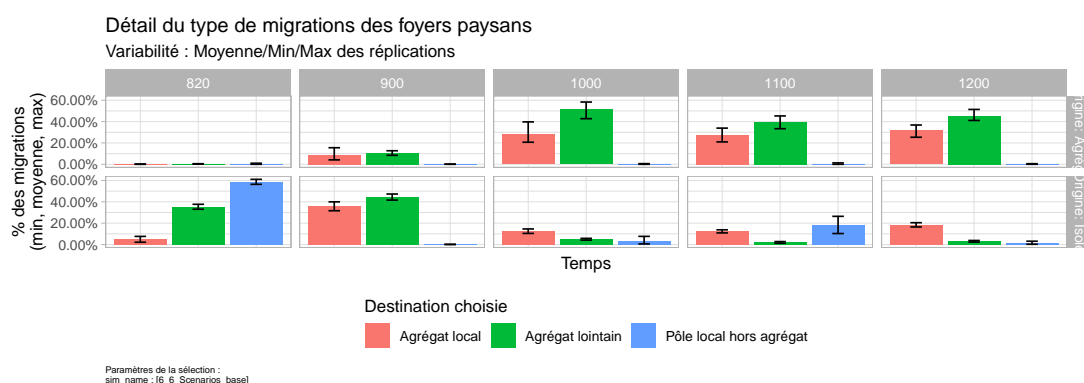


FIGURE 6.13 – Types de migration des foyers paysans.

Pour évaluer la dissémination du peuplement, on peut aussi observer la répartition des paroisses. Comme elles 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 figure 6.14 donnent une lecture satisfaisante là encore 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 6.14)). 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](#), ??).

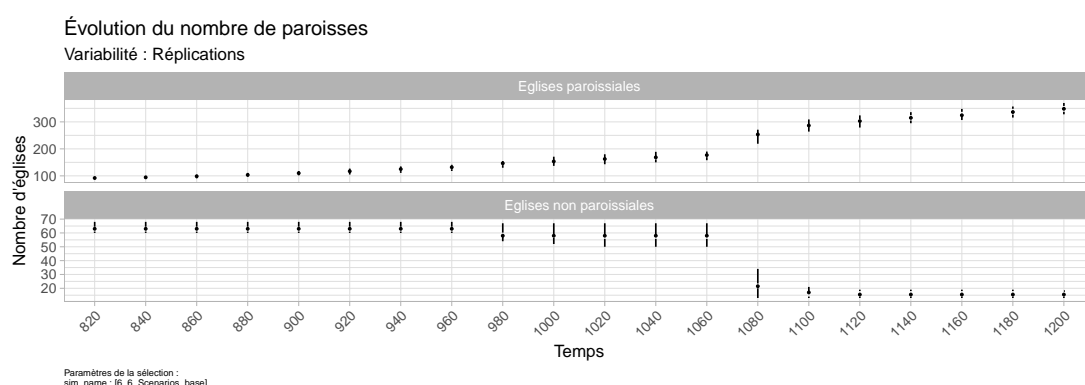


FIGURE 6.14 – Nombre de paroisses.

La carte de densité des paroisses (figure 6.15) n'est qu'un exemple puisqu'elle ne correspond qu'à deux répliques et non à une agrégation de données. Toutefois, on y constate nettement la densification généralisée du maillage paroissial : contrairement à la version 0 du modèle où les paroisses périphériques demeuraient très étendues, il n'y plus véritablement ici de paroisses marginalisées (on le note aussi dans la figure 6.10-c).

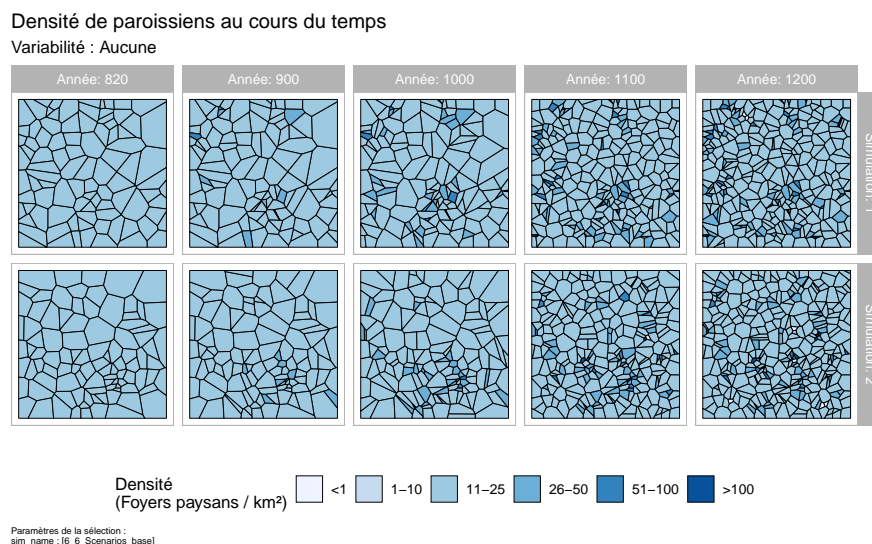


FIGURE 6.15 – Répartition des paroisses.

Un dernier indicateur (figure 6.16) apporte un complément généralisé. En tenant compte de la variabilité des répliques, l'indicateur représenté parvient à montrer que l'augmentation du nombre d'églises paroissiales est bien un phénomène réparti dans l'espace. La proportion du monde simulé située à plus de 1.5 kilomètres d'une église paroissiale y diminue en effet fortement au cours du temps.

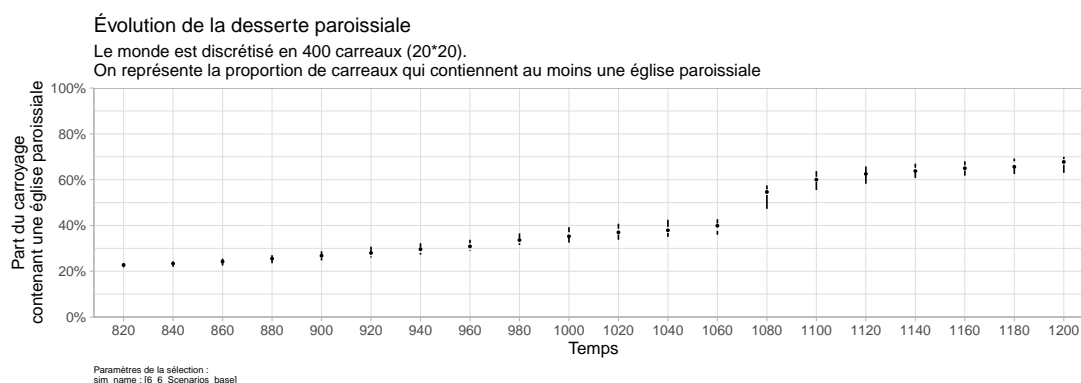


FIGURE 6.16 – Couverture de la desserte paroissiale.

Dans cette dimension d'analyse, SimFeodal parvient entièrement à reproduire le double fait stylisé recherché : fixation et dissémination du peuplement. Pourtant, plus encore que pour les dimensions précédentes, il nous paraît extrêmement difficile d'aller plus loin dans le calibrage du modèle. La raison en est principalement la difficulté de l'évaluation de ces phénomènes. Parmi les indicateurs de sortie de simulation étudiés, ceux de cette dimension sont les plus intrinsèquement spatio-temporels, et se prêtent très mal à des indicateurs synthétiques quantitatifs. Certes, le champ de l'analyse spatiale s'est en partie

constitué pour répondre à ce type de problème, en trouvant des mesures synthétiques et hétérogènes permettant de décrire et de comparer des semis de points, correspondent-ils à des entités ou à des époques différentes. Pourtant, et on y reviendra plus extensivement dans le chapitre suivant, on constate ici une limite importante de notre méthode d'évaluation des simulations. Un espace théorique, simulé et très variable localement comme celui de SimFeodal se prête en effet extrêmement mal à une agrégation de données, puisque les différentes réplifications d'un modèle ne partagent pas véritablement de référentiel spatial commun en dehors des limites du monde simulé.

Pour pousser l'évaluation et être en mesure de comparer de nouvelles simulations par le biais des indicateurs présentés ici, au moins pour ceux relatifs à la fixation et dissémination, il faudrait sans doute mener des études de cas plus approfondies. On pourrait par exemple suivre des ensembles de foyers paysans ou d'agréats tout au long des simulations pour analyser plus finement leur comportement. Mais même alors, le manque de sources empiriques se ferait forcément et fortement ressentir : on pourrait peut-être isoler la différence entre des simulations, mais serait-il vraiment possible de les départager ?

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

Le constat de la difficile amélioration des dimensions de fixation et de dissémination peut s'étendre à l'ensemble de la calibration du modèle. Quand on compare la version 0 à la version 6.6 de SimFeodal, les différences sont criantes. Les évolutions des mécanismes et paramétrages du modèle entre ces deux versions sont indiscutables, et amènent une amélioration évidente du modèle. Même au regard de versions intermédiaires, par exemple avant le calibrage du modèle (section 6.1.2), le modèle reproduit bien mieux les objectifs identifiés.

Dans un calibrage plus fin cependant, on atteint les limites de distinction possible entre des versions successives : les indicateurs de sortie deviennent plus proches les uns des autres, et la variabilité intrinsèque, due à la stochasticité du modèle, peut devenir prégnante par rapport à la variabilité des différentes configurations de paramètres. Le risque est alors d'entrer dans des erreurs d'*overfitting* (sur-ajustement) du modèle, simplement parce que les indicateurs de sortie de simulation choisis ne permettent plus véritablement de révéler l'essence des variations des résultats du modèle. Cet *overfitting* s'exprime de plus sur le terrain thématique. Les « objectifs » fixés sont ainsi porteurs d'un intervalle d'incertitude plus important que l'intervalle des variations du modèle. Départager des modèles selon qu'ils atteignent 25% ou 27% de concentration des foyers paysans n'a pas de sens quand l'intervalle jugé réaliste se situe entre 20% et 30%.

Un autre élément concoure aussi à la difficulté d'améliorer le modèle, en la matière de la complexité de ce dernier. Au niveau de précision où nous en sommes rendus, préciser le réglage d'un paramètre en vue d'un objectif précis se fait presque toujours au détriment d'un autre objectif. C'est le principe des vases communicants, où les variations de chaque paramètre doivent s'assortir



de variations d'autres paramètres. Cela rend la calibration plus difficile, car on manque d'une vision globale qui permettrait de comprendre l'influence réelle des paramètres sur les objectifs. Une telle vision permettrait de ré-organiser 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. L'analyse des résultats de la calibration et du modèle « final » montre qu'il y a encore beaucoup de surprise et de contre-intuitif dans notre compréhension du modèle.

Pour mieux comprendre ce modèle, sous sa forme actuelle, il serait alors 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. Une telle exploration, plus systématique, de grande ampleur et cherchant à dépasser l'intuition, ne peut se mener que sur une quantité réduite d'éléments du modèle (indicateurs de sortie, mais aussi intervalle de validité des paramètres), et cela requerrait dès lors de réussir à synthétiser de manière sensible les attentes que l'on a vis-à-vis des résultats du modèle.

## 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 voués 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, lié à l'évaluation de modèles, 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>12</sup>.

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

Les analyses de sensibilités sont largement recommandées comme une pratique indispensable à la validation de modèle, mais nous avons décrit dans le **chapitre 3** la démarche que nous avons préféré pour l'évaluation de SimFeodal. L'analyse de sensibilité est toutefois un outil extrêmement utile pour aider à la compréhension d'un modèle, sans chercher à en quantifier la validité. 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 comportent vraisemblablement une part de redondance, l'ambition n'est pas de réduire la masse de paramètres, mais plutôt d'aider à comprendre lesquels ont la plus grande influence sur le modèle.

---

12. 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 par une variation de 0,1% du seuil de tolérance avec une variation de 20% dans la part d'espace vide.

Dans le cadre de SimFeodal, une analyse de sensibilité doit permettre, comme l'évaluation visuelle, de gagner en compréhension sur le modèle, et en conséquence sur les 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. De manière plus générale, une analyse de sensibilité quantitative requiert à minima des objectifs quantitatifs synthétiques, hiérarchisés et parcimonieux, ce qui n'est pas la démarche empruntée dans SimFeodal.

Dans cette partie, nous nous en tiendrons à une analyse de sensibilité grossière, 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>13</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 quand à eux à cette adresse : [depot these / anasensib](#)

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

### 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. Dans un modèle complexe où les agents sont en interactions les uns avec les autres, cette approche est forcément biaisée et lacunaire au regard d'approches plus avancées comme les méthodes d'exploration de l'espace des paramètres. Pourtant, c'est la seule qui nous semblait applicable dans le cas de SimFeodal. 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 alors l'exécution de  $5^{70} \times 20_{\text{[replications]}}$ , soit environ  $10^{50}$  simulations...

**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, 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 hétérogènes, par exemple portant sur des régions et des périodes différentes, il nous a semblé important de tester aussi ces paramètres. On a opté pour une analyse de sensibilité de chacun des paramètres du modèle. SimFeodal comporte dans les faits 70 paramètres, mais en pratique, le nombre de paramètres sur lesquels on peut réaliser des analyse est plus faible. Ainsi, certains paramètres n'ont aucun sens quand 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 répliques de ces simulations.

Comme le nombre de paramètres était déjà important, que le nombre de

réplications (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 est assez faible, mais amène déjà à l'exécution de  $57_{[\text{paramètres}]} \times 5_{[\text{valeurs}]} \times 20_{[\text{réplications}]} = 5700$  simulations, ce qui est une quantité importante de simulations au regard de toutes celles qui ont été mené dans les étapes de paramétrage et d'évaluation visuelle du modèle.

**Étendue** Dans une analyse de sensibilité classique, on cherche à mener des variations de paramètres comparables, c'est-à-dire centrées autour des valeurs par défaut et présentant des variations relatives de 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 plus haut. Comment rendre comparable cinq variations autour de 10 et cinq variations autour de l'étendue [1500m; 5000m] ? L'étendue des possibles transformations pour ces étendues est bien plus important : diminution, augmentation, translation...

Ce problème se pose de manière plus importante encore 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 ». 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 cela n'empêche en rien qu'elles apportent un éclairage précieux sur le comportement du modèle en fonction de ses paramètres.

**Outputs** Le chapitre 5 donnait des ordres de grandeur correspondant à la masse de données produites par le modèle pour une simulation, autour de 10 Mo. 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-traiter, intégrer et analyser dans la base de données. Pour une analyse

qui n'est pas au cœur de notre démarche de co-construction, cela représentait un poids bien trop important. Nous avons choisi 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 qu'à son pas de temps final (en 1200) plutôt que tout au long de son déroulement. On perd certes l'aspect dynamique et la possibilité de comparer les rythmes entre les simulations, mais comme l'analyse de sensibilité doit se concentrer sur aussi peu d'objectifs que possible, ce n'est pas véritablement gênant. Pour les mêmes raisons, on a aussi 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 à ce moment nul besoin d'enregistrer les indicateurs individuels des agents du modèle.

La sélection des indicateurs agrégées a été assez rapide : on a choisit de reprendre les indicateurs numériques synthétiques, qui reprennent le tableau des objectifs présenté en tableau 6.1. Parmi ceux-ci, on a conservé uniquement les 6 indicateurs « émergents » : nombre d'agrégats ; nombre de grands châteaux ; nombre d'églises paroissiales ; distance moyenne entre églises paroissiales ; part de foyers paysans isolés ; 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 limite. Le temps de calcul l'est tout autant, sinon plus, tant il peut s'étendre rapidement. Sur un ordinateur individuel, avec les valeurs de paramètre décidées après calibrage, chaque simulation demande en moyenne 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 sur un ordinateur individuel, sans véritable droit à l'erreur sous faute de 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 serveur de calcul géré par la TGIR Huma-Num. En distribuant les exécutions de simulation sur la quarantaine de processeurs informatiques que cela représentait au total, tout en laissant tourner les simulations uniquement de nuit pour ne pas gêner les autres utilisateurs de ces serveurs, toutes les simulations requises ont été exécutées en 3 jours, ce qui représente cette fois-ci 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 choisi de mener une 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 choisis provoquée par la variation des valeurs de paramètres.

**Normalisation des objectifs** Un premier problème est que les indicateurs de sortie de simulation choisis, les objectifs, sont très hétérogènes en termes d'ordre de grandeur. Il était 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 pour indicateur « nombre d'agrégats » 50. 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 », identifiée thématiquement comme objectif (voir le tableau 6.2). 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 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.

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.

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

Ainsi conçue, la valeur normalisée permet une comparabilité au sein des indicateurs, mais aussi entre eux puisque les ordres de grandeur sont désormais similaires.



**Calcul de la sensibilité** Comme les valeurs sont désormais normalisées, il est possible de mener des opération 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 brut. La sensibilité de chaque paramètre peut alors être définie comme suit :

$$\text{sensibilité}_{\text{paramètre}_\alpha} = \frac{1}{n_{\text{indicateurs}} \times n_{\text{valeurs\_paramètres}_\alpha} \times n_{\text{réplications}}} \sum |\text{valeurs\_normalisées}_{\text{paramètre}_\alpha}|$$

**Sélection de paramètres** En calculant la sensibilité de chaque paramètre, on sera en mesure d'en sélectionner un sous-ensemble afin de procéder à une analyse visuelle de la sensibilité de ces derniers. Pour isoler ce sous-ensemble, on ne conservera que les paramètres à la plus forte sensibilité globale, en cherchant à en conserver une dizaine. Ce nombre est volontairement imprécis puisqu'il dépendra surtout 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 entièrement 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 risque de cette méthode est que certains indicateurs peuvent contribuer bien davantage que d'autres à la sensibilité d'ensemble, malgré la normalisation. Pour ne pas risquer de négliger des paramètres dont la sensibilité s'exprimerait uniquement sur certains indicateurs, on complètera la sélection de paramètres précédente par une sélection des trois paramètres les plus sensibles sur chacun des 6 indicateurs.

$$\text{sensibilité}_{\text{paramètre}_\alpha \text{ indicateur}_i} = \frac{1}{n_{\text{valeurs\_paramètres}_\alpha} \times n_{\text{réplications}}} \sum |\text{valeurs\_normalisées}_{\text{paramètre}_\alpha \text{ indicateur}_i}|$$

En considérant qu'une large partie de ces paramètres auront vraisemblablement déjà été isolés par le filtre sur la sensibilité globale, on devrait obtenir quelques paramètres supplémentaires sur lesquels mener l'analyse visuelle.

### 6.2.1.3 Analyse visuelle

Une fois les paramètres à étudier sélectionnés, on pourra alors 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 pos-

sible d'analyser l'importance de cette influence (la sensibilité), mais aussi sa cardinalité : en observant la réaction des indicateurs de sortie en fonction des valeurs de paramètre, on pourra noter les corrélations (visuelles).

Cette analyse visuelle de sensibilité s'inscrit ainsi 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

**Visualisation** L'objectif de la visualisation de la variation des paramètres est d'observer, pour chaque paramètre, comment ce dernier varie sur chacun des indicateurs. Pour cela, on représente la variation due aux réplifications de chaque paramètre au moyen de représentations visuelles dédiées à la variabilité, comme des *box-plots* ou *violin-plots*. La figure 6.17 présente un exemple de mise en page de telles représentations graphiques afin de faciliter l'analyse de sensibilité visuelle.

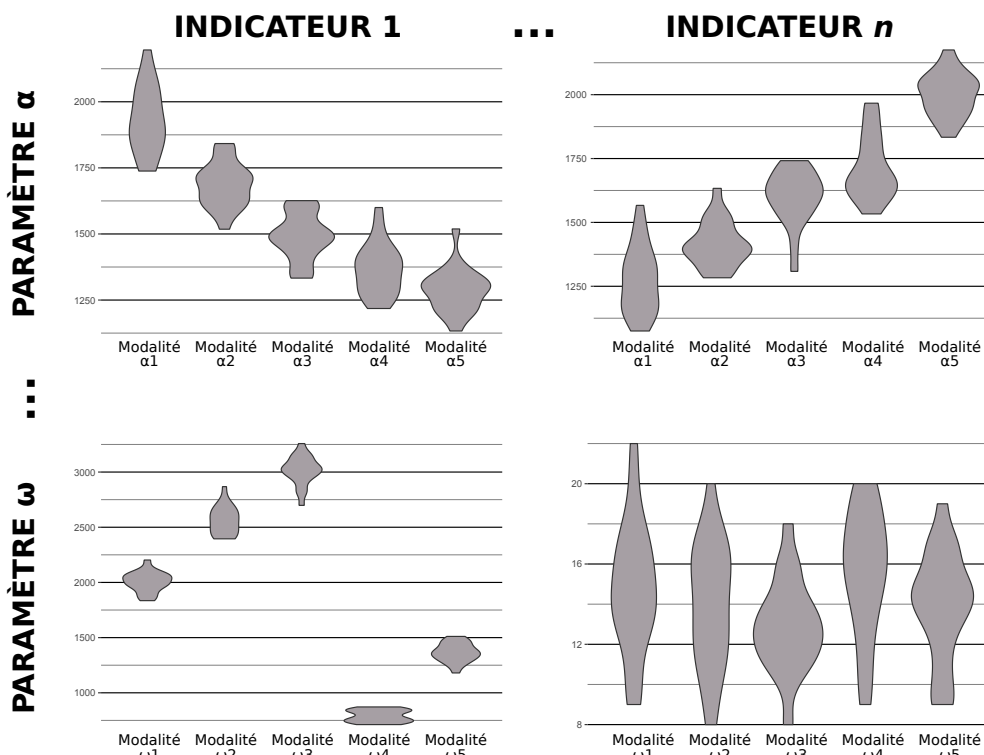


FIGURE 6.17 – Construction et mise en page de planches graphiques dédiées à l'analyse visuelle de la sensibilité des 5 modalités de  $\omega$  paramètres sur  $n$  indicateurs.

**Normalisation** La normalisation des valeurs afin d'homogénéiser indicateurs et variation des paramètres était indispensable afin de garantir une comparabilité acceptable lors de l'analyse de l'ensemble des paramètres. Pourtant, sur un sous-ensemble de paramètres qui doivent être étudiés plus précisément, cette normalisation est un frein à l'interprétation : les valeurs des différents indicateurs ne sont plus exprimés dans l'unité d'origine et il est alors difficile d'en faire un commentaire thématique. L'analyse visuelle sera ainsi présentée sur les valeurs brutes issues des simulations, et chaque graphique aura en conséquence un axe des ordonnées propre.

## 6.2.2 Sélection des paramètres

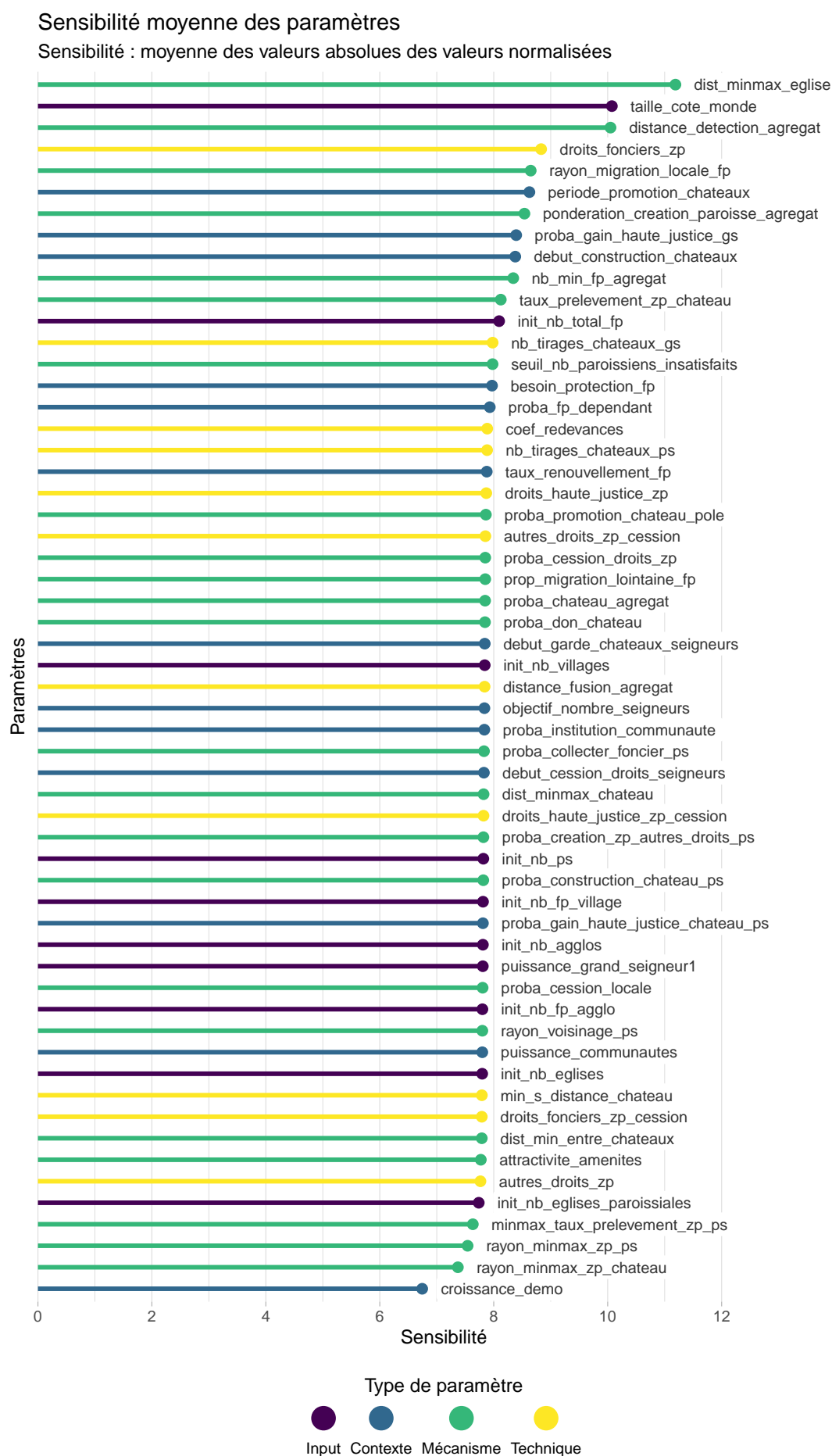


FIGURE 6.18 – Analyse de la sensibilité d'ensemble de tous les paramètres de SimFeodal.

La figure 6.18 présente les « scores de sensibilité » de chacun des paramètres du modèle.

Une première remarque est que de manière légèrement contre-intuitive, on ne peut discerner de motif particulier selon les types de paramètres : les *inputs*, paramètres de contexte, de mécanisme et enfin les paramètres techniques sont dispersés et entremêlés dans le graphique. On se serait vraisemblablement attendu à ce que les inputs et paramètres de contexte jouent un rôle plus important que les paramètres techniques par exemple.

Pondérons toutefois cette remarque en rappelant que malgré une homogénéisation des valeurs en sortie par leur normalisation, il n'y a ici aucune compensation de l'étendue très variable des valeurs de paramètre testées (**disponibles dans l'annexe N**). Par exemple, si l'on compare un paramètre très sensible (la taille du monde simulé, *taille\_cote\_monde*) et un paramètre peu sensible (le taux de croissance démographique introduit dans le modèle, *croissance\_demo*), on peut réaliser que l'ordre de grandeur de la variation est faible.

- Pour *taille\_cote\_monde*, dont la valeur de base est de 80 km, les valeurs testées sont 50, 75, 100, 125 et 150 km. Il y a donc à peu près une variation qui va de la moitié de la valeur de base à son double.
- Pour *croissance\_demo*, les seuils sont plus restreints. Pour ne pas déstabiliser les autres indicateurs, ce paramètre a été modifié en tenant compte d'une population finale stable de 40 000 foyers paysans, en adaptant à chaque valeur de croissance démographique une valeur différente de population initiale. De ce fait, la valeur de base de croissance démographique est de 0%, et les valeurs testées sont de 1.53%, 3.72%, 5.89% et 12.89%. Ces valeurs permettent de faire passer la population de la stabilité au décuplement. Au regard du paramètre précédent, l'étendue interprétée est large, mais en absolu, les 4 premières valeurs sont très proches et on ne s'attend ainsi pas à ce qu'elles aient un effet majeur sur les sorties du modèle.

Un autre exemple de ces étendues incomparables permet d'éclairer la première remarque quant au fait que les paramètres techniques ne sont pas relégués en bas de ce graphique. Par définition, les paramètres techniques ont des valeurs qui ne représentent strictement rien d'un point de vue thématique. Leur étendue acceptable est alors extrêmement difficile à évaluer, et on aura ainsi pu avoir tendance à effectuer de mauvais jugements sur les valeurs testées de ces paramètres, les conduisant soit à être sur-valorisés (le paramètre de caractérisation des droits fonciers, *droits\_fonciers\_zp* semble entrer dans cette catégorie avant examen spécifique), soit à sembler sous-valorisés (le paramètre *distance\_fusion\_agregat* par exemple, a été modifié à de nombreuses reprises lors du paramétrage du modèle et était à ce moment assez important).

Notons aussi que dans les 10 paramètres jugés les plus sensibles, 4 portent sur des valeurs testées « qualitatives » (étendues, variables au cours du temps..., voir p. 27), alors que ces paramètres ne représentent que 20% de l'ensemble des paramètres testés. Là encore, on peut estimer que cette légère sur-représentation de ces paramètres tient à la difficulté de leur attribuer des étendues compa-

rables.

Un dernier constat, à l'échelle très agrégée qui caractérise ces analyses, nous paraît extrêmement rassurant en termes de validation interne du modèle : aucun des paramètres testés ne présente une sensibilité véritablement faible, qui plus est en considérant que la plus faible sensibilité mesurée (paramètre `croissance_demo`) correspond à un paramètre dont l'on sait qu'il a une influence réelle sur les sorties du modèle. Ce simple constat nous indique que dans la limite du jeu de paramètre issu du calibrage, aucun paramètre n'est inutile, et ne devrait par conséquent être supprimé du modèle si on recherchait une parcimonie plus importante à cette étape.

À la lecture du graphique, on remarque qu'autour des 10 premiers paramètres, on peut constater qu'un « saut » dans les valeurs de sensibilité se produit entre les paramètres `nb_min_fp_agregat` et `taux_prelevement_zp_chateau`. L'écart entre ce dernier paramètre et le suivant est en effet assez faible, et on décide de ne conserver que les 10 premières paramètres (le 10ème est `nb_min_fp_agregat`) pour l'analyse visuelle.

La sélection des trois paramètres ayant la plus forte sensibilité sur chaque indicateur complète cette liste avec 4 nouveaux paramètres, portant le total à 14 paramètres dont la sensibilité sera analysée visuellement. Le tableau 6.4 en présente la liste ainsi qu'une description complète.

Description	Intitulé	Type de paramètre	Agents caractérisés	Valeur par défaut	Valeurs testées	Origine de la sélection	Rang sensibilité
Étendue (variable au cours du temps) dans laquelle un foyer paysan calcule sa satisfaction religieuse	<b>dist_minmax_eglise</b>	Mécanisme	Foyers paysans	Avant 960 : de 5km à 25km Entre 960 et 1060 : de 3 à 10km Après 1060 : de 1.5 à 5km	- Valeur par défaut : base - Entre 5km et 25km tout au long : statique_large - Entre 1.5km et 5km tout au long : statique_reduit - Avant 960 : [1.5, 5], de 960 à 1060 : [1, 3], après : [0.5, 1.5] : dynamique_reduit - Avant 960 : [25, 50], de 960 à 1060 : [10, 25], après : [5, 10] : dynamique_large	Globale	1
Dimensions de l'espace du modèle en km de côté. (Une marge de 1km de large est ôtée dans les simulations afin d'éviter les effets de bord).	<b>taille_cote_monde</b>	Input	Monde	80	[50,75,100,125,150]	Globale	2
Distance maximale entre les foyers paysans et les attracteurs les plus proches pour qu'ils soient considérés comme faisant partie d'un même agrégat	<b>distance_detection_agregat</b>	Mécanisme	Agrégats	100 m.	[50, 100, 150, 200, 300]	Globale	3
Prélèvement de droits fonciers : montant des redevances perçues par le seigneur détenteur de la zone de prélèvement pour chaque foyer paysan assujéti	<b>droits_fonciers_zp</b>	Technique	Seigneurs	1	[0, 0.5, 1, 1.5, 2]	Globale	4
Rayon de distance dans lequel un foyer paysan effectue une migration locale	<b>rayon_migration_locale_fp</b>	Mécanisme	Foyers paysans	(800-2500) : 2500m tout au long	- 2500m tout au long : base - 1000m tout au long : statique_reduit - 2500m jusqu'à 1000, puis 5000m : dyn_croissant - 1000m jusqu'en 1000, puis 2500m : dyn_reduit - 5000m jusqu'en 1000, puis 10000m : dyn_large	Globale	5
Période au cours de laquelle les châteaux peuvent devenir des gros châteaux	<b>periode_promotion_chateaux</b>	Contexte	Châteaux	de 940 à 1040 inclus	940-1040 940-1200 1100-1200 1000-1100 940-1000	Globale	6
Pondération (en nombre de foyers paysans) de la probabilité qu'une nouvelle église paroissiale soit créée au sein d'un agrégat	<b>ponderation_creation_paroisse_agregat</b>	Mécanisme	Églises	2000	[500, 1000, 2000, 3000, 5000]	Globale	7
Probabilité pour un grand seigneur d'acquiescer des droits de haute justice à chaque pas de simulation.	<b>proba_gain_haute_justice_gs</b>	Contexte	Seigneurs	Avant 900 : 0 Entre 900 et 980 : 0.2 A partir de 1000 : 1	0 tout au long : statique_nul 1 tout au long : statique_fort Gradient défaut : base 0 + 0.1/an à partir de 900 : croissant_regulier 0 avant 1000, puis 0.5 : croissant_seuil	Globale	8
Date à partir de laquelle des châteaux peuvent être construits par les seigneurs	<b>debut_construction_chateaux</b>	Contexte	Châteaux	940	[820, 880, 940, 1000, 1060]	Globale	9
Nombre minimum de foyers paysans nécessaires pour constituer un agrégat	<b>nb_min_fp_agregat</b>	Mécanisme	Agrégats	5	[3, 5, 7, 10, 15]	Globale	10
Proportion de foyers paysans s'acquittant de droits dans chaque zone de prélèvement relevant d'un château (droits fonciers, droits de haute justice et autres droits)	<b>taux_prelevement_zp_chateau</b>	Mécanisme	Seigneurs	1 (100%)	[0.0, 0.25, 0.5, 0.75, 1]	ratio_charge_fiscale	11
Nombre de tirages successifs de la probabilité qu'un des petits seigneurs crée un château.	<b>nb_tirages_chateaux_ps</b>	Technique	Châteaux	1	[0, 1, 2, 3, 4]	nb_grands_chateaux	18
Probabilité, pour un petit seigneur, de créer une nouvelle zone de prélèvement d'autres droits dans son voisinage à chaque pas de simulation	<b>proba_creation_zp_autres_droits_ps</b>	Mécanisme	Seigneurs	0.15 (15%)	[0.0, 0.05, 0.15, 0.25, 0.35]	distance_eglises_paroissiales	36
Taux de croissance du nombre de foyers paysans à chaque pas de simulation	<b>croissance_demo</b>	Contexte	Foyers paysans	0	avec variation pop_init : pour pop 40k final constante : croissance_demo : 0.0 & init_nb_total_fp : 40000 croissance_demo : 0.1289 & init_nb_total_fp : 4000 croissance_demo : 0.0589 & init_nb_total_fp : 13500 croissance_demo : 0.0372 & init_nb_total_fp : 20000 croissance_demo : 0.0153 & init_nb_total_fp : 30000	ratio_charge_fiscale	57

TABLEAU 6.4 – Paramètres sélectionnés pour l'analyse visuelle.

### 6.2.3 Évaluation visuelle de la sensibilité

Les graphiques présentés dans la suite de ce chapitre ne concernent que les paramètres sélectionnés lors de l'analyse quantitative globale. Tous les paramètres peuvent toutefois être analysés individuellement, de manière interactive, dans la partie dédiée de la plate-forme SimEDB : [Mettre un lien direct](#).

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 de ces paramètres.

#### 6.2.3.1 Monde

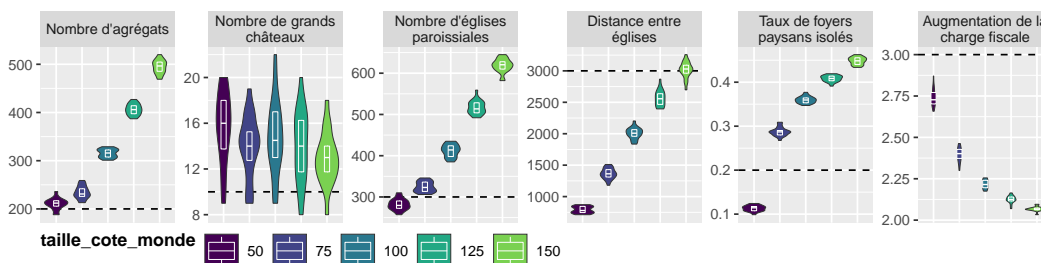


FIGURE 6.19 – Sensibilité à la taille du monde simulé.

Le paramètre régissant la taille du monde simulé est le seul des *inputs* présent dans la sélection. De manière peu surprenante, c'est un paramètre majeur qui affecte la totalité des indicateurs étudiés, selon une polarité assez intuitive (figure 6.19). 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 de superficie vraisemblablement supérieures par les effets des mécanismes de fusion des agrégats. 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 directement interprétable. On peut émettre l'hypothèse que cette augmentation est une conséquence de la dispersion des foyers paysans au sein de petits agrégats. De petites églises paroissiales seraient créées ou promues en plus grand nombre dans les zones 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.



### 6.2.3.2 Foyers paysans et agrégats

Les foyers paysans (et les agrégats de population qui résultent de leur concentration) sont 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.

Les trois paramètres, spécifiquement liés aux foyers paysans, qui montrent la plus forte sensibilité (figure 6.20) embrassent trois aspects bien différents des mécanismes des foyers paysans.

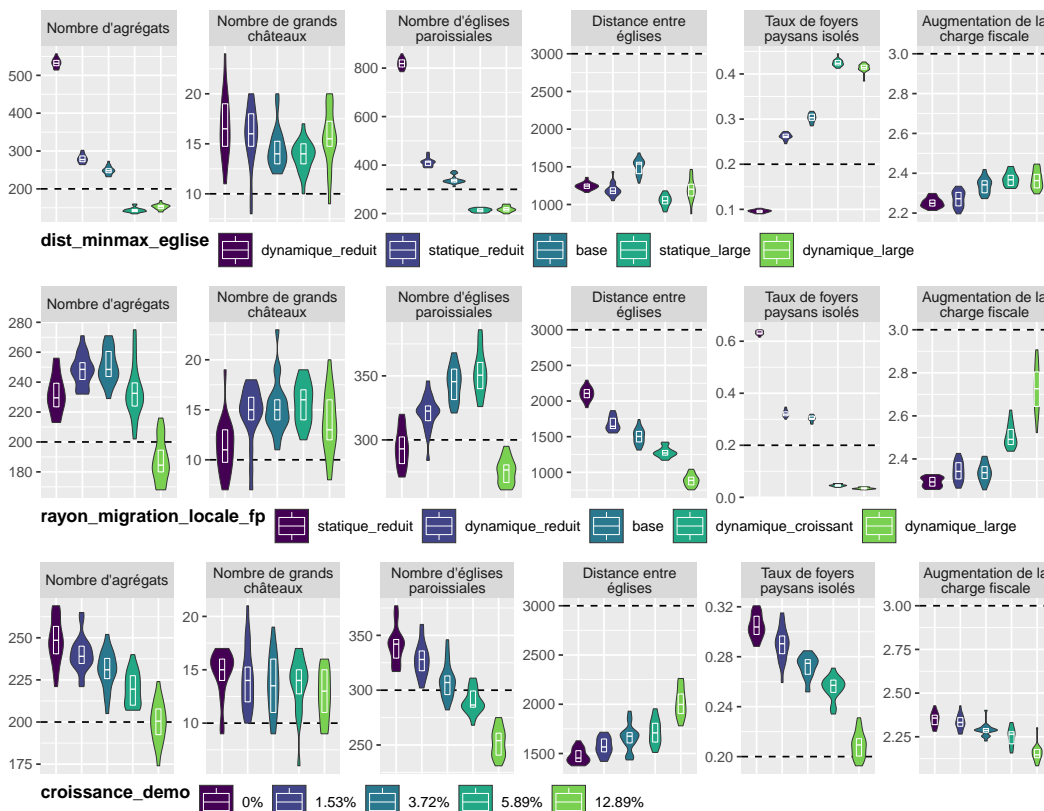


FIGURE 6.20 – Sensibilité des paramètres liés aux foyers paysans.

Le premier de ces paramètres, 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 des objectifs. En matière d'interprétation, le fait que cette satisfaction soit déterminante est inattendu : on a pu constater dans les résultats du modèle (??) que le facteur limitant de la satisfaction d'ensemble était la satisfaction protection. Les valeurs testées, qualitatives, ont une large influence sur les aspects liés à la concentration des foyers paysans, et leur action sur les autres indicateurs est moins linéaire et évidente. Plus les étendues testées sont restreintes, plus le nombre d'agrégats et de foyers paysans isolés est important : cette satisfaction joue sans doute à ce moment là le rôle limitant et « force » les foyers paysans à migrer à proximité d'églises paroissiales.

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. Celui-ci n'est pas une étendue, mais un seuil maximum qui peut (valeurs « dynamiques ») changer au cours du temps, la logique étant que plus le temps passe, plus ce rayon est susceptible d'augmenter en lien avec des aspects cognitifs et liés à l'intégration du système de peuplement qui voit ses distances-temps rétrécir. 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 63% à seulement 3%. Les effets sur le nombre d'agrégats, de grands châteaux et d'églises paroissiales ne sont pas directement linéaires et l'on peut penser que les variations de ce paramètre montre alors des tendances diverses selon l'ordre de grandeur des seuils empruntés.

La croissance démographique est particulière dans le modèle. On la considère comme un paramètre de contexte, mais elle aurait également entièrement sa place parmi les *inputs* tant son importance thématique est majeure. À la lecture des graphiques, on peut s'étonner de ce que ce paramètre soit le moins sensible (figure 6.18) tant son effet est direct sur tous indicateurs présentés (à l'exception du nombre de grands châteaux). On remarque qu'un taux de croissance plus élevé réduit le nombre d'agrégats tout en réduisant aussi le taux de foyers paysans isolés, deux effets qui sont souvent inverses dans les différents paramètres. C'est extrêmement intéressant pour le modèle tant ce paramètre permet d'approcher des objectifs de ces indicateurs qui sont tous deux dépassés alors qu'opposés. Ce paramètre fait l'objet d'un scénario thématique (section 6.3.1), et on le décrira plus avant à ce moment-là.

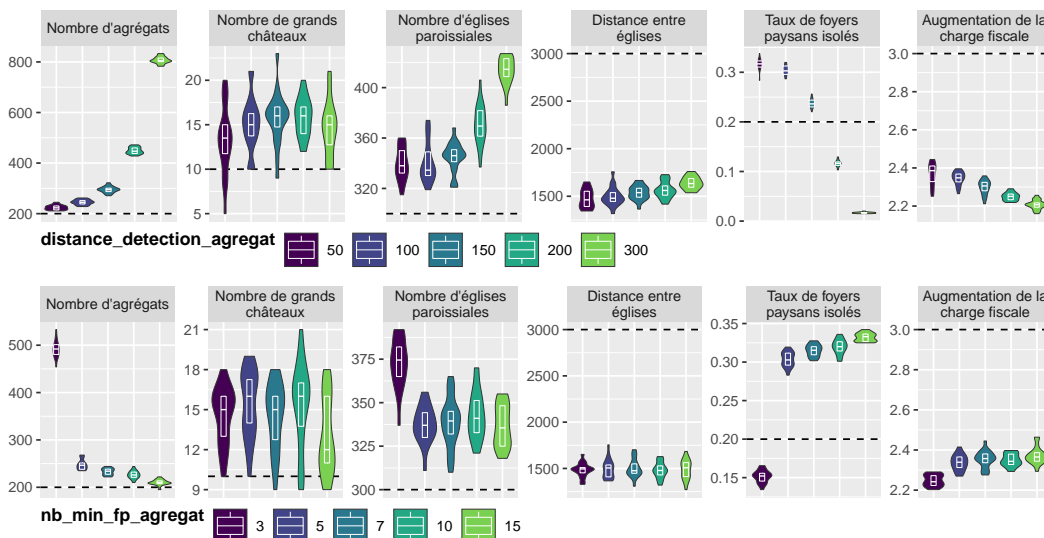


FIGURE 6.21 – Sensibilité des paramètres liés aux agrégats.

Les deux paramètres agissant sur la définition des agrégats (figure 6.21) sont très comparables et agissent de manière symétriquement opposée sur les indicateurs de sortie. Sans surprise, leur effet est largement circonscrit aux indicateurs relatifs aux foyers paysans (à l'exception surprenant des églises paroissiales, peut-être pour les mêmes raisons que le paramètre de taille du monde simulé), mais il est intéressant de noter leur extrême sensibilité, plus que linéaire, à des variations relativement fines dans les ordres de grandeur mobilisés (quelques foyers paysans de plus à l'échelle des 40 000, 50 ou 100

mètres à l'échelle d'un monde de 80 kilomètres de côté...). Les variations présentées dans la figure indiquent que les valeurs par défaut (100 m et 5 foyers paysans) sont au moins dans des intervalles assez sensées au regard des objectifs poursuivis.

### 6.2.3.3 Seigneurs et châteaux

Les paramètres liés aux seigneurs et aux châteaux (les premiers construisant les seconds) sont les plus nombreux du modèle et il est attendu qu'ils soient assez sensibles : ce sont les mécanismes associés qui façonnent le monde dans lequel les foyers paysans auront à évoluer et dans lequel ils essaieront de ne pas être trop insatisfaits.

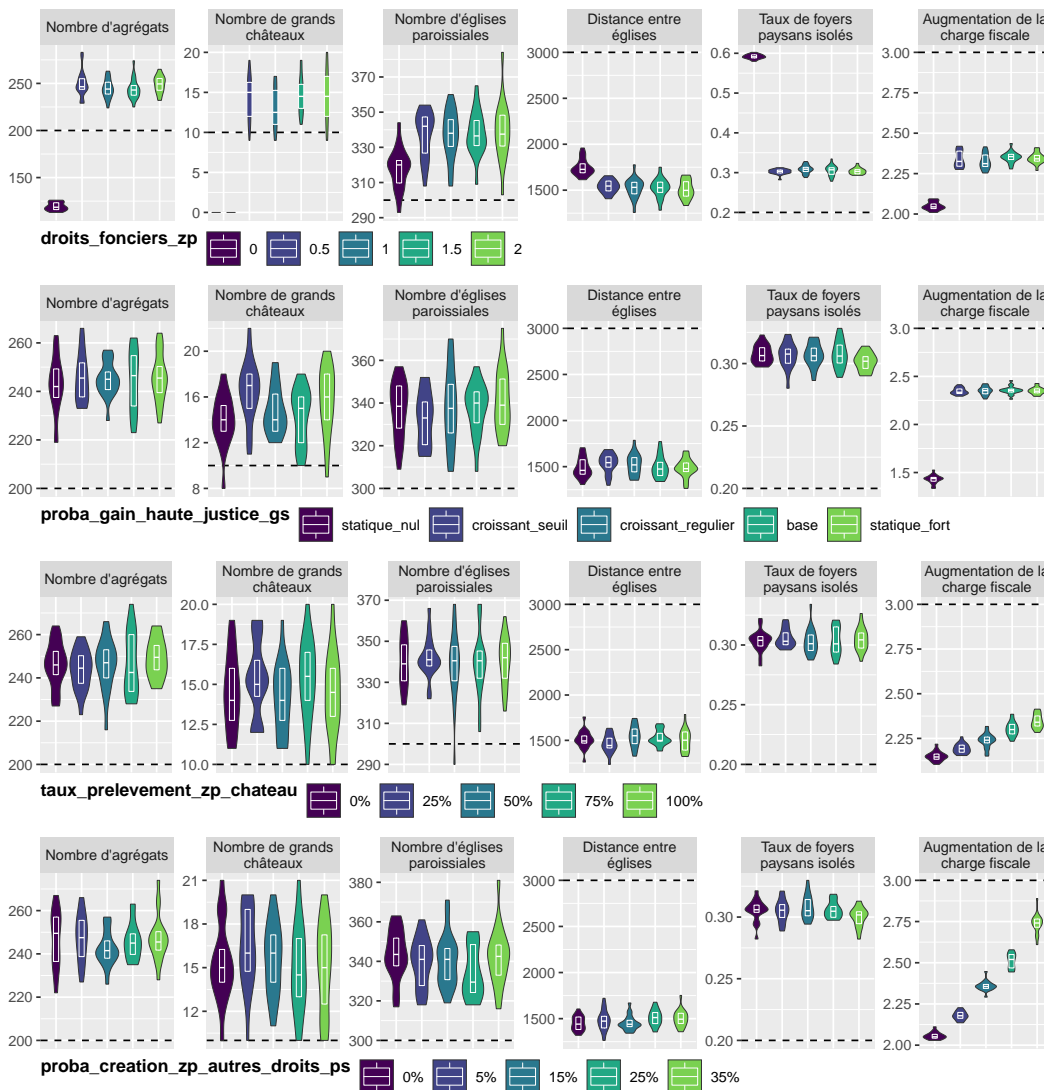


FIGURE 6.22 – Sensibilité des paramètres liés aux seigneurs.

Une caractéristique commune aux quatre paramètres isolés (figure 6.22) tient au test de valeurs nulles, ou autrement dit, à la désactivation des mécanismes associés. Dans les deux premiers paramètres, l'effet de rupture est net, par exemple sur l'augmentation de la charge fiscale. Le montant des droits fonciers collectés ne joue que par son activation ou non (les valeurs supérieures à 0 présentent des résultats très similaires), mais présente un effet clair sur le nombre d'agrégats et l'agrégation des foyers paysans. L'existence et la propension des droits de haute justice des grands seigneurs ne semble jouer que

sur la charge fiscale, mais y exerce une influence énorme : c'est le seul paramètre dont une valeur testée peut faire diminuer autant (1.5 alors que l'ordre de grandeur des simulations est plutôt entre 2 et 2.5)

Il est intéressant de remarquer que pour les deux paramètres suivant, où la valeur de 0% correspond aussi à une désactivation du mécanisme lié, cela n'a aucun effet de seuil notable : ces paramètres se comportent, dans l'étendue testée, comme des éléments linéaires sur l'augmentation de la charge fiscale. Ils semblent assez dépourvus d'influence sur les autres indicateurs, mais en particulier pour le dernier paramètre, leur action sur la charge fiscale est notable. Pour améliorer le calibrage du modèle sur cet indicateur, on aurait sans doute intérêt à augmenter la valeur par défaut du paramètre `proba_creation_zp_autres_droits_ps`, y compris au delà des valeurs ici testées. Ce serait d'autant plus adapté que la sensibilité globale de ce paramètre est relativement faible (36ème sur les 57 paramètres).

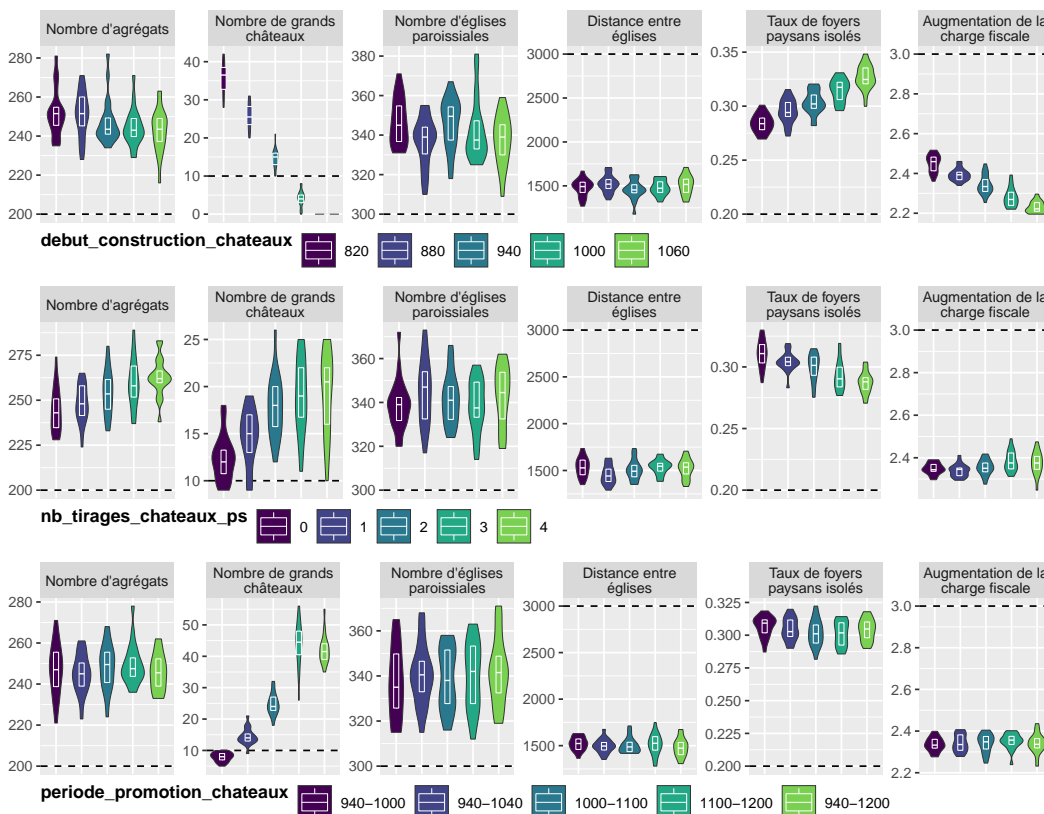


FIGURE 6.23 – Sensibilité des paramètres liés aux châteaux.

Les paramètres liés aux châteaux sont sensiblement sur-représentés parmi ceux qui ont la sensibilité la plus forte : il n'y en a que 6 sur les 57 paramètres (environ 11%), mais ils représentent 20% des paramètres les plus sensibles et 21% dans cette sélection de paramètres remarquables .

C'est assez remarquable, d'autant qu'on peut constater à la lecture des deux premiers paramètres de la (figure 6.23) que leur rôle ne se cantonne absolument pas à un simple raffinement du contexte où l'atteinte d'un certain nombre de grands châteaux serait à la fois un objectif et un élément déterministe. Sur ces deux premiers paramètres, on remarque certes une très forte variation de l'indicateur directement lié, le nombre de grands châteaux, mais aussi et surtout un lien net avec la concentration des foyers paysans (et un autre

lien moins significatif avec l'augmentation de la charge fiscale). La conception et le paramétrage des mécanismes liés aux châteaux ont demandé un travail conséquent (voir la partie dédiée leur calibrage – section 6.1.2 – p. 9), vraisemblablement trop important relativement à la complexité de leurs règles et à notre propre estimation subjective de leur apport concret au modèle. Pourtant, les résultats de cette analyse de sensibilité donnent tort à l'expertise du modélisateur 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.

Le troisième paramètre, la période durant laquelle les châteaux peuvent être promus en grands châteaux, apporte un léger contrepoint, ou au moins une précision à ce constat. Ce paramètre est déterminant dans le nombre de grands châteaux présents en fin de simulation, mais pourtant, il n'a aucune influence significative sur les autres indicateurs. Peut-on dès lors penser que la hiérarchie mise en place entre les châteaux, et la différence d'attractivité qui en découle, n'est pas indispensable au modèle ? Il faudra pour cela mener une étude plus approfondie des relations entre proportion de grands châteaux et les autres indicateurs. Rappelons en effet que l'analyse de sensibilité ici menée est grossière, et n'étudie aucunement les interactions entre valeurs de paramètres dont on peut imaginer, au sein d'un modèle aussi descriptif et complexe, qu'ils ont des effets locaux conséquents.

#### 6.2.3.4 Églises et paroisses

Le dernier paramètre de cette analyse visuelle est aussi le seul qui porte sur un élément aussi conséquent que les paroisses, lesquelles sont 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).

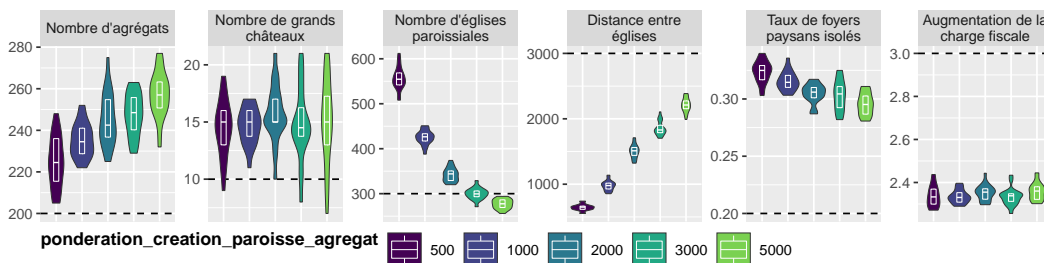


FIGURE 6.24 – Sensibilité du paramètre de pondération de création de nouvelles paroisses dans les agrégats.

La figure 6.24 présente les réponses des indicateurs à différentes valeurs du paramètre `pondération_creation_paroisse_agregat`. 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 cette trivialité, on notera avec intérêt la variation amenée par ce paramètre sur le nombre d'agréats : plus le seuil est élevé, plus les agréats sont nombreux, et cette corrélation apparaît visuellement significative et inverse à celle du nombre d'églises paroissiales, contrairement aux tendances que l'on a pu observer dans les paramètres liés aux foyers paysans et agréats (section 6.2.3.2). Avec ces paramètres, nombre d'agréats et d'églises paroissiales varient dans le même sens face aux valeurs de paramètres. Dans le cas du paramètre de pondération de la création de paroisses « urbaines », la relation est inverse, et il est difficile de l'expliquer, de même que le lien (plus ténu, mais lui aussi inverse) avec la concentration des foyers paysans. On peut émettre l'hypothèse que cette pondération influe largement sur la hiérarchie des paroisses et des pôles. En créant moins de paroisses en zone dense, la distribution de l'attractivité des pôles d'attraction – qui est mesurée en large partie sur le nombre d'églises paroissiales qui les composent – tend peut-être vers plus d'uniformité, et favorise ainsi relativement l'attraction locale vers des pôles de plus faible attractivité, et donc vers des agréats plus locaux et 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 tenait certainement à l'inégale amplitude des valeurs de paramètres testées, lesquelles influencent potentiellement plus directement ces indicateurs, mais cela ne nous semble pas être une explication suffisante.

De manière globale, on constate dans les résultats de la version calibrée de SimFeodal (tableau 6.2) que la variabilité des indicateurs émergents est assez forte en termes d'écart-type. L'écart-type se lisant dans l'unité de l'indicateur mesuré, 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	Écart-type	Coefficient de variation
<i>Agréats</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 lecture du tableau 6.5 nous montre que, rapportée à un paramètre de



dispersion relatif comme le coefficient de variation, la variabilité des indicateurs due à l'aléa est assez faible, et dans des ordres de grandeurs assez comparables entre les indicateurs (à l'exception du nombre de grands châteaux).

Pourtant, lors de l'analyse de sensibilité, on a pu constater des variations (étendue dans l'axe des ordonnées des *violin-plots*) bien plus importantes au sein des réplifications des paramètres testés.

Il nous paraît par conséquent utile de mener un bref complément d'analyse, dédié à l'étude de la variabilité au sein des réplifications d'une expérience. De telles analyses nous paraissent peu fréquentes dans la littérature liée aux modèles de simulation en géographie, mais on en trouve tout de même une définition chez GINOT et MONOD (2005), qui nomment ces approches des « 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

Pour mesurer cette incertitude, nous avons repris l'ensemble des réplifications correspondant à l'analyse de sensibilité de chaque paramètre, en avons mesuré la variabilité (écart-type) sur chaque indicateur. Afin à nouveau 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.

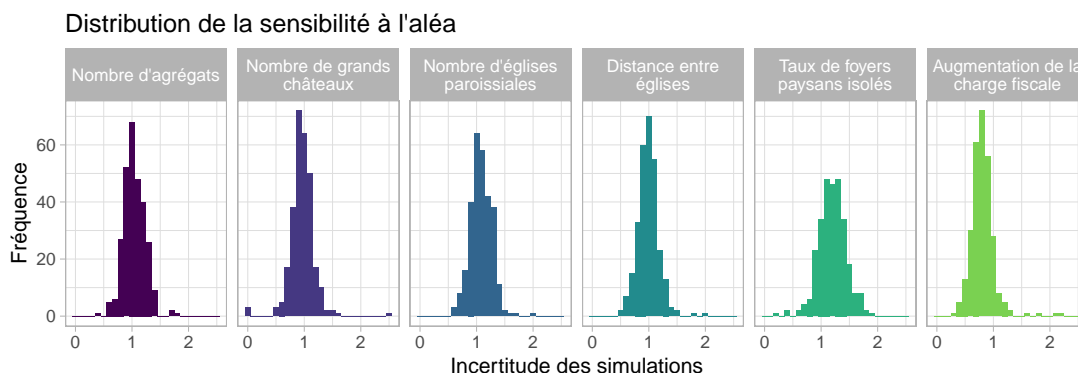


FIGURE 6.25 – Sensibilité à l'aléa selon les indicateurs.

La distribution de ces valeurs d'incertitude réduites est présenté dans la figure 6.25. Dans ces histogrammes, la valeur de 1 correspond à une incertitude moyenne identique à celle des réplifications de référence. 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 attendue. Autrement dit, cer-

taines 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 montre une absence presque totale de variabilité à l'aléa. Sans aller plus loin sur cet exemple, on peut penser qu'il s'agit 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 présentons une représentation graphique dans la figure 6.26.

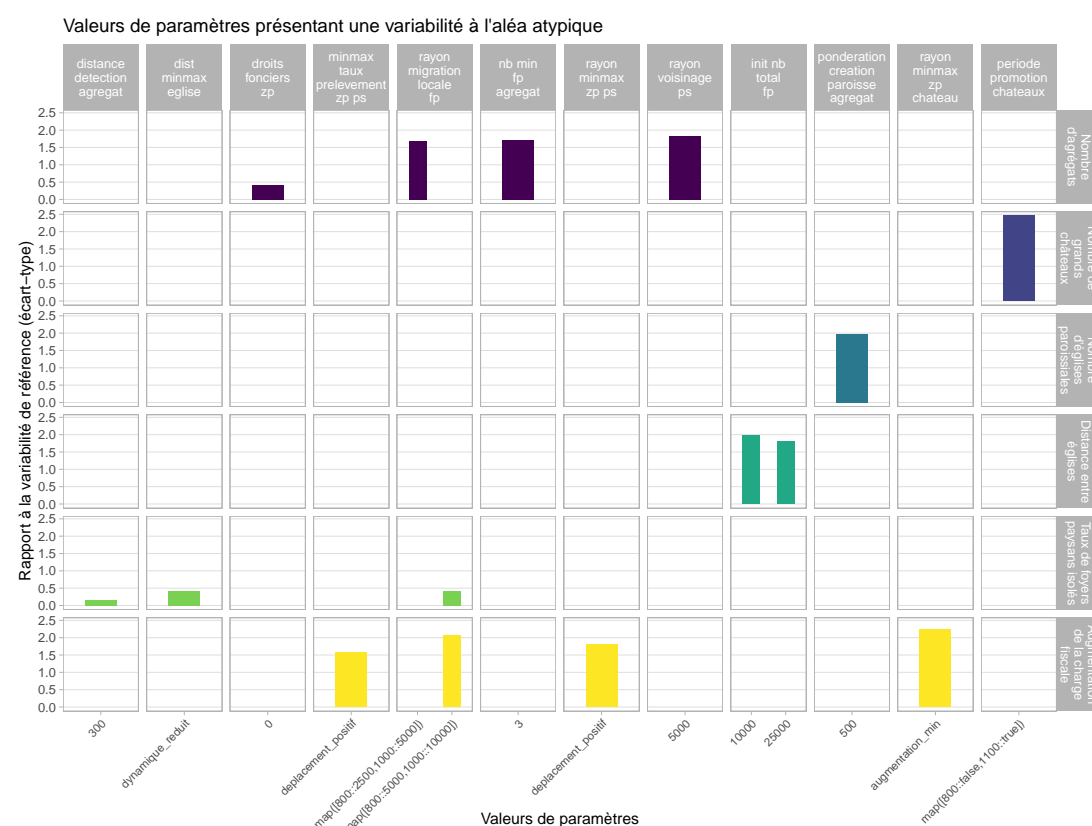


FIGURE 6.26 – Paramètres présentant des sensibilités atypiques à l'aléa.

De manière globale, on peut remarquer que les indicateurs relatifs aux foyers paysans présentent plus de valeurs extrêmes que les autres, c'est-à-dire sur une plus forte diversité de paramètres. Notons aussi que sur les 14 valeurs de paramètres isolées, la moitié concernent des paramètres qualitatifs. La difficile estimation de l'amplitude ceux-ci explique sans doute en partie la plus forte variabilité à l'aléa : si on a choisis des valeurs (étendues, évolutions temporelles etc.) trop éloignées des valeurs calibrées, il se peut que le modèle passe dans des régimes locaux différents du régime de base caractérisé par les réponses aux paramètres calibrés.

Plus spécifiquement, on peut constater le comportement très particulier du paramètre régissant le rayon maximum de migration locale des foyers paysans. Non seulement les valeurs de ce paramètre sont parmi les plus fortes sensibi-



lités à l'aléa (indicateurs nombre d'agréats et augmentation de la charge fiscale), mais de manière surprenante, la valeur correspondant à un rayon évolutif très étendue par rapport à la valeur de base (`map([800:5000, 1000:10000])`) se caractérise aussi par une très faible sensibilité à l'aléa sur l'indicateur de concentration des foyers paysans. Ce paramètre a donc la capacité de faire fortement varier les sorties (vu dans l'analyse de sensibilité visuelle), mais en plus de varier tout aussi fortement la part de l'aléa dans le modèle. Nous pensions exécuter un scénario thématique qui fasse varier ce paramètre afin d'étudier l'effet sur la hiérarchie des agrégats, et cette analyse confirme que ce serait tout à faire approprié.

### 6.2.5 Conclusion et apports de l'analyse visuelle de sensibilité

On peut tirer un bilan extrêmement positif de l'exécution de cette analyse de sensibilité, pourtant limitée et grossière.

Dans un premier temps, notons que la hiérarchie de la sensibilité des paramètres n'est pas véritablement proche de celle que l'on attendait, et qu'on a sans doute agit lors du paramétrage et du calibrage sur des paramètres qui n'étaient pas les plus déterminants. Cette étape d'évaluation qu'est l'analyse de sensibilité aurait sans doute gagné 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, peut-être que l'analyse de sensibilité n'aurait pas mis en avant les mêmes paramètres. Ceux-ci sont en interaction étroite, et il nous semble évident que les résultats de cette analyse de sensibilité sont eux même extrêmement sensibles au paramétrage de base.

Dans un second temps, l'analyse de sensibilité semble aller dans le sens d'une confirmation de la parcimonie du modèle. On l'a dit lors de l'analyse quantitative, mais il nous semble important de le répéter tant ce résultat est rassurant, mais aussi surprenant. 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 pense à minima comprendre à peu près son modèle et a des intuitions fortes sur les réactions de celui-ci aux différents mécanismes et paramètres, c'est une surprise très positive. 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.

Un autre point, classique, concerne les limites d'une telle analyse. L'approche entièrement quantitative, présentée dans l'introduction de cette partie, permet de s'affranchir des effets de mauvaise pondération 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. Dans un modèle comme SimFeodal, c'est-à-dire descriptif, exploratoire, et composé d'autant de paramètres hétérogènes, il semble toutefois illusoire de réussir à quantifier, pour une analyse

de sensibilité, tout ce qui n'a pas été quantifié dans le modèle en lui-même : pondération des objectifs, objectivation des attendus dans les indicateurs graphiques etc.

C'un point de vue subjectif assumé, la démarche mise en place, basée sur le visuel, nous semble tout à fait fructueuse. Elle s'inscrit, comme de nombreux aspects de ce travail de thèse, dans une approche d'analyse visuelle entièrement dédiée à l'exploration d'un modèle et paraît confirmer l'adéquation de ce type d'approches à la construction et évaluation commune et interdisciplinaire de modèles.

La dernière partie de ce chapitre, tournée vers l'usage du modèle pour tester des scénarios thématiques, trouve enfin une justification supplémentaire, dans le champ méthodologique et de la modélisation cette fois-ci. Les brèves analyses de sensibilités ont en effet renforcé le besoin criant d'études plus approfondies des variations de certains paramètres, qui plus est quand ceux-ci trouvent des correspondances dans les connaissances empiriques.

## 6.3 Comprendre le modèle par l'exécution de scénarios

Cette sous-partie est adaptée d'une partie d'un article collectif dont la première auteure est Cécile Tannier : **Mettre la ref quand ce sera soumis. Pour l'instant :**

Tannier C., Cura R., Leturcq S., Zadora-Rio E. (2020), « An agent-based modelling to explore the combined effects of social and demographic changes on the hierarchy of rural settlement patterns in North-Western Europe during the Middle Ages (800 CE to 1200 CE) - An application to the Tour's diocese, West of France ».

- On ne présente ici que 3 scénarios sur les 7 « familles » de scénarios testés dans l'article.
- Les simulations sont déjà effectuées et leurs résultats intégrés dans SimEDB.
- Cécile est en train de rédiger une analyse des résultats en anglais pour un article collectif.
- On présentera notamment les résultats de ces scénarios à l'ECTQG.
- = > je les intégrerai une fois que ce sera déjà rédigé et que la sélection de graphiques/indicateurs aura déjà été faite, autant ne pas perdre de temps là dessus pour l'instant

### 6.3.1 Tester l'hypothèse d'une croissance démographique

### 6.3.2 Modéliser la dépendance spatiale : le poids du servage

### 6.3.3 Quel rôle et importance des communautés paysannes dans la structuration du système de peuplement ?

## Conclusion

**A rédiger après les scénarios :**

- Le modèle à l'issue de la phase de calibrage est globalement satisfaisant.
- L'analyse de sensibilité a montré quelques pistes potentielles d'amélioration.
- Il serait pourtant difficile d'améliorer réellement le modèle : on entrerait dans de l'*overfitting*, aussi bien vis-à-vis de l'incertitude autour des *inputs* et paramètres non techniques que vis-à-vis de l'incertitude des données et connaissances expertes sur lesquelles on établit l'évaluation du modèle.
- Le modèle est toutefois déjà utile et utilisé, notamment avec l'exécution des scénarios qui nous semblent pouvoir donner de nouvelles hypothèses sur la période historique (croissance démo notamment).

- Dans l'ensemble, la calibration, l'analyse de sensibilité et les scénarios servent certes un rôle de validation interne, mais surtout permettent aux co-concepteurs du modèle de mieux en comprendre le fonctionnement et les biais.
- En cherchant à comprendre le fonctionnement du modèle, en cherchant à en raffiner le comportement, on effectue surtout un travail thématique poussé :
  - formalisation des hypothèses,
  - formalisation des attentes,
  - recherche approfondie de sources et de documentation sur les différents faits stylisés/mécanismes implémentés dans le modèle.

## Références

- CHAREILLE, Pascal, Xavier RODIER et Elisabeth ZADORA-RIO (2008). « Dynamiques Morphologiques et Échelles d'analyse Des Territoires ». In : *Des Paroisses de Touraine Aux Communes d'Indre-et-Loire. La Formation Des Territoires*. Sous la dir. d'Elisabeth ZADORA-RIO. FERACF, p. 260-263. URL : <https://halshs.archives-ouvertes.fr/halshs-00956186> (visité le 05/08/2019).
- CROOKS, Andrew et al. (2019). *Agent-Based Modelling and Geographical Information Systems : A Practical Primer*. 1 edition. Thousand Oaks, CA : SAGE Publications Ltd. 408 p. URL : <https://uk.sagepub.com/en-gb/eur/agent-based-modelling-and-geographical-information-systems/book250134>.
- GINOT, Vincent et Hervé MONOD (2005). « Explorer Les Modeles Par Simulation : Application Aux Analyses de Sensibilité ». In : AMBLARD, F et Denis PHAN. *Modélisation et Simulation Multi-Agents : Applications Pour Les Sciences de l'Homme et de La Société*, Chapitre. T. 3, p. 75-101.
- HAMBY, D. M. (1994). « A Review of Techniques for Parameter Sensitivity Analysis of Environmental Models ». In : *Environmental Monitoring and Assessment* 32.2, p. 135-154. DOI : 10.1007/BF00547132. URL : <https://doi.org/10.1007/BF00547132> (visité le 01/08/2019).
- HIRTZEL, Joanne (2015). « Exploration Prospective Des Mobilités Résidentielles Dans Une Agglomération Urbaine Au Moyen d'un Modèle de Simulation Multi-Agents (MOBISIM) ». Université de Franche-Comté.
- O'SULLIVAN, David et George L. W. PERRY (2013). *Spatial Simulation : Exploring Pattern and Process*. Chichester, West Sussex, UK : John Wiley & Sons Inc. 305 p.
- SCHMITT, Clara (2014). « Modélisation de la dynamique des systèmes de peuplement : de SimpopLocal à SimpopNet. » Thèse de Doctorat. paris : Université Paris I - Panthéon-Sorbonne. URL : <https://tel.archives-ouvertes.fr/tel-01077891/document>.
- ZADORA-RIO, Elisabeth, éd. (2008). *Des paroisses de Touraine aux communes d'Indre-et-Loire : la formation des territoires*. T. 34e Supplément. Revue archéologique du Centre de la France. 34e Supplément à la Revue Archéolo-

gique du Centre de la France, Tours. Tours : Ferac Editions. 302 p. URL : <https://halshs.archives-ouvertes.fr/halshs-00378136>.