

Chapitre 3

Évaluer un modèle de simulation complexe en situation d'inter-disciplinarité

Version 2018-04-30

3.1 Comment évaluer un modèle ?

3.1.1 Visual validation

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

3.1.2 Indicateurs

3.1.3 L'importance de la réplication

3.2 Indices et indicateurs

On attend d'un modèle, sans entrer encore dans le détail, qu'il reproduise au moins les grands traits de l'élément empirique qu'il cherche à simplifier. Ces grands traits peuvent s'entendre de multiples manières, et se formaliser avec encore plus d'approches. Ici, nous avons souhaité proposer une dychotomie simple entre le domaine de l'empirique et celui de la simulation, en forçant et systématisant l'usage d'un vocabulaire qui est souvent employé de manière plurielle. Pour être en mesure d'évaluer la vraisemblance du comportement empirique reproduit par le modèle, il est nécessaire de s'appuyer sur des éléments empiriques ayant une correspondance dans le modèle de simulation. Nous caractérisons ces éléments en deux grands ensembles : (1) **les indices empiriques**, éléments quantifiables ou au moins descriptibles du domaine empirique, et (2) **les indicateurs de sortie**, variables informatiques produites par le modèle de simulation et devant pouvoir être comparés à chacun des indices empiriques. La figure 3.1 reprend, sous forme de schéma ontologique synthétique, ces ensembles, explicitant le vocabulaire mobilisé dans cette partie.

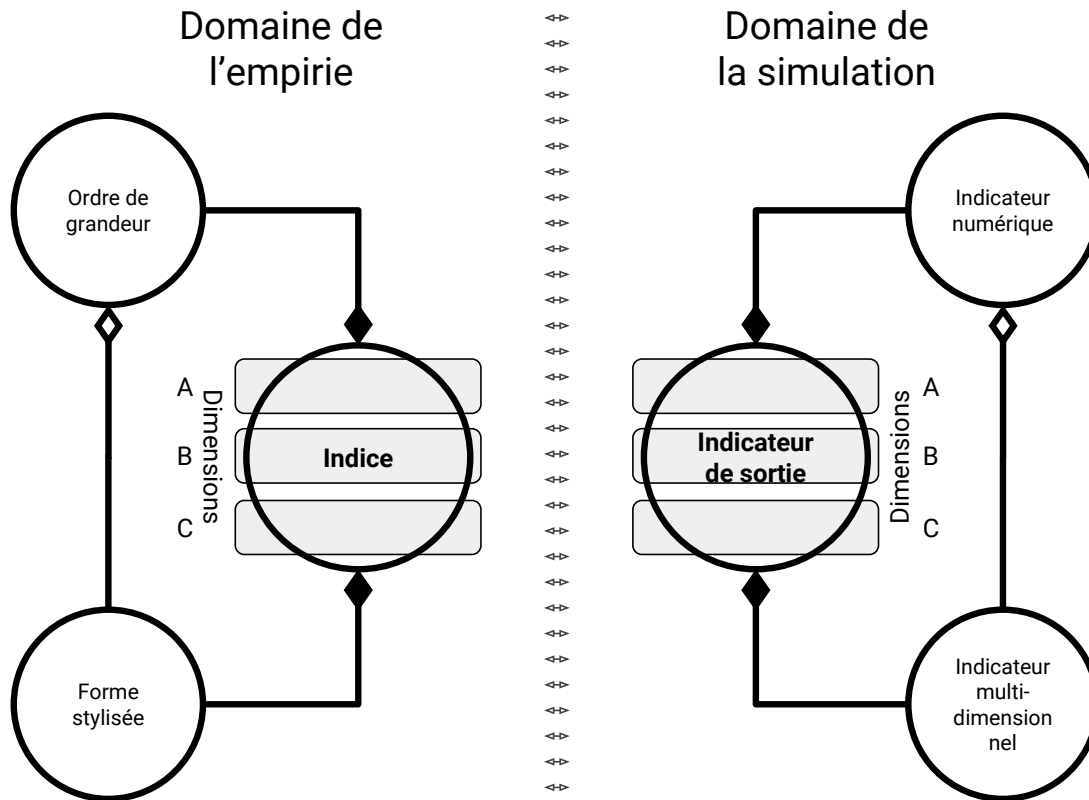


FIGURE 3.1 – Schéma de synthèse des correspondances entre empirie et simulation pour l'évaluation du modèle SimFeodal.

La correspondance des éléments est représentée par une symétrie axiale. Les losanges pleins désignent une relation de composition : un indice est soit un ordre de grandeur, soit une forme stylisée. Les losanges vides indiquent une relation d'agrégation : une forme stylisée est une agrégation d'ordres de grandeurs. Les dimensions (A à C) regroupent des indices (et les indicateurs qui leur correspondent) qui peuvent être de plusieurs types, et sont elles aussi comparables et en correspondance entre les domaines.

3.2.1 Les indices empiriques

Le domaine empirique, pour pouvoir être modélisé, doit être en mesure de proposer des « points de repère » sur lesquels le modèle s'appuiera. Ces éléments sont tirés de la thématique des dynamiques que l'on cherche à reproduire. Selon les modèles, ils peuvent revêtir de multiples formes et caractériser l'ensemble des échelles spatiales et temporelles que l'on choisit de mettre en scène dans le modèle. Leur point commun est qu'ils doivent pouvoir être mesurés, au sens le plus large, c'est-à-dire être en capacité d'être reproduits et comparables avec d'autres mesures. Dans cette étude, on a décidé de regrouper ces **indices empiriques** en deux catégories, basées la précision avec laquelle ils peuvent être décrits¹. La figure 3.1 montre cette catégorisation entre la première catégorie – les ordres de grandeur – et la seconde – les formes stylisées –.

3.2.1.1 Ordres de grandeur

La première catégorie est constituée d'**ordres de grandeurs** empiriques estimés – avec une précision plus ou moins importante (cf. tableau 3 p. 317 du chap. TMD, à reproduire dans chap 2). Certaines valeurs empiriques sont ainsi connues,

1. Et non sur la précision de leur connaissance, cf. section 3.3.1

que ce soit d'après des sources primaires ou secondaires, et peuvent ainsi constituer des indices. Par exemple, on connaît avec quasi-certitude le nombre d'églises paroissiales de la région Touraine en 1100. Certaines de ces valeurs sont toutefois plus proches de l'estimation, comme le taux de foyers paysans isolés en fin de période, que l'on ne peut qu'estimer à partir de sources secondaires et en menant des extrapolations. Ces différents indices observés ou estimés peuvent cependant tous trouver une correspondance presque exacte (cf. 3.2.2.2, Correspondance entre indicateurs et indices, p. 4) dans le modèle de simulation, et dès lors, mener à comparaison entre données observées/estimées et données simulées. Ils peuvent ainsi jouer le rôle d'éléments d'évaluation du comportement du modèle simulé.

3.2.1.2 Faits et formes stylisés

Le second type d'indice est moins précis, ne reposant pas sur une valeur estimable, mais plutôt sur la connaissance experte d'un phénomène. Il s'agit des « **faits stylisés** »², plus tendanciels que les indicateurs. On fait un large usage de ces faits stylisés en économie, mais aussi en géographie, par exemple quand observe que les systèmes de peuplement tendent à se hiérarchiser. On peut quantifier cette observation en observant que la pente formée par leur courbe rang-taille tend vers 1 à mesure que le système évolue (Trouver ref, sans doute Pumain/Saint-Julien). De la même manière, le modèle de transition démographique d'Adolphe Landry est un fait stylisé, énoncé à partir de l'observation de nombreuses récurrences de l'évolution des populations d'un pays en fonction des taux de natalité et de mortalité. Ces exemples montrent qu'au sein des faits stylisés, on retrouve aussi une diversité quant à la précision de leurs énoncés : on peut quantifier précisément la courbe de la rang-taille et l'allure de son évolution dans le temps, alors que la courbe logistique de la transition démographique revêt un caractère plus théorique. Dans notre cas d'étude, les faits stylisés sur lesquels on s'appuiera dans la modélisation peuvent être des allures de courbes (par exemple l'évolution dans le temps d'un indicateur) ou encore des formes de répartition spatiale. On pourra dès lors parler de « **formes stylisées** », aussi bien temporelles – courbe logistique estimée pour la polarisation des foyers paysans –, spatiales – différence dans l'occupation de l'espace par les agrégats entre le début et la fin de la période –, que correspondant à une agrégation à l'échelle du système, comme dans l'observation des hiérarchies grâce aux courbes rangs-tailles. Notons que ces formes stylisées relèvent le plus souvent d'une agrégation d'ordres de grandeurs (comme figuré dans la figure 3.1) : l'évolution dans le temps de la population, par exemple, correspond à un vecteur d'ordres de grandeur, c'est-à-dire à une succession de mesures pour chaque date étudiée. Dans le cas d'une agrégation spatiale, par exemple quand on observe la hiérarchie du système de peuplement, il s'agit aussi d'une agrégation d'ordres de grandeurs : cette forme stylisée est constituée d'un ensemble d'ordres de grandeurs, les populations de chaque agrégat de population.

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

3.2.2 Les indicateurs de sortie de simulation

Les ordres de grandeur et formes stylisées évoquées relèvent du domaine empirique. Afin de pouvoir les mobiliser, il est nécessaire de définir des **indicateurs en sortie** dans le modèle de simulation, c'est-à-dire des variables informatiques que l'on enregistrera durant l'exécution du modèle et que l'on pourra ensuite comparer aux indices empiriques définis.

3.2.2.1 Définition

Comme pour les indices empiriques qui sont leurs équivalents, on peut définir une typologie différenciant les indicateurs en sortie « numériques », en distinguant les formes numériques simples (des scalaires), et des indicateurs plus complexes, multidimensionnels, à même de permettre la confrontation du modèle de simulation avec les formes stylisées identifiées. Chaque indice empirique doit ainsi se voir correspondre, respectivement, un indicateur de sortie.

3.2.2.2 Correspondance entre indicateurs et indices

Cette correspondance n'est pour autant pas une équivalence exacte. En effet, si certains indices empiriques peuvent trouver un équivalent strict – le nombre de châteaux connus à chaque date est strictement équivalent au nombre de châteaux produits par le modèle –, d'autres indices trouveront une correspondance plus faible.

Il peut s'agir de correspondances ayant trait aux mêmes éléments de bases, donc « convertibles » entre elles. Par exemple, on connaît à peu près les populations de la région étudiée au début et à la fin de la période. Pourtant, dans SimFeodal, on ne modélise pas les individus en tant que tels, mais des foyers paysans. Le nombre de foyers paysans simulé n'est pas directement comparable à la population estimée, mais en estimant une moyenne de 4 ou 5 habitants par foyer paysan, ces quantités deviennent équivalentes.

D'autres indicateurs trouveront une correspondance plus lointaine. Certains indices empiriques ne sont ainsi pas observables directement, et il faut alors leur trouver des indicateurs analogues, que l'on nomme souvent « proxy » dans le champs statistique. Dans notre cas d'étude, il s'agit le plus souvent d'indices plus difficilement quantifiables. La puissance militaire des seigneurs, par exemple, est complexe à quantifier. On sait d'après connaissances expertes que la hiérarchie des puissances était forte, majoritairement dominée par deux seigneurs (les comtes de Tours et de Blois) et assortie d'une grande quantité de petits chevaliers. On sait de plus qu'avec les liens de vassalité, les grands seigneurs disposaient des forces militaires des seigneurs qui leur étaient assujettis. Faute à une capacité de quantification plus importante de ces indices, on leur associe, dans le modèle, un indicateur proxy relevant du nombre de foyers paysans s'acquittant de droits à chaque seigneur. De cette manière, on reproduit une hiérarchie implicite entre les seigneurs et leur puissance relative, qui est, elle, comparable aux connaissances empiriques de ces rapports de puissance.

La création d'indicateurs de sortie correspondant aux indices empirique permet donc de quantifier une information qui n'est pas forcément aisément quantifiable dans le domaine empirique.

3.2.2.3 Indicateur composite

La forme numérique (scalaire ou vectorielle) des indicateurs de sortie permet de trouver des manières plus simples d'évaluer le modèle que d'observer l'ensemble

des indicateurs. Chaque indicateur étant numérique, il devient en effet possible de les combiner au sein d'indicateurs composites, résultant en quelques indicateurs synthétiques permettant une évaluation plus rapide des résultats d'une simulation. Ces indicateurs composites sont très fréquemment utilisés en statistiques, permettant par exemple de résumer une information multidimensionnelle en un indicateur simple. L'Indice de Développement Humain (IDH), par exemple, est un indicateur composite dépendant de l'espérance de vie à la naissance, du niveau d'éducation et du niveau de revenu de chacun des pays caractérisés. On le trouve très souvent utilisé, parce qu'il permet de résumer le niveau de développement d'un pays en agrégeant trois dimensions majeures, l'aspect sanitaire, culturel et économique.

3.2.2.4 Indicateur synthétique

En renforçant cette logique de synthèse de plusieurs dimensions, on peut aller plus loin dans la définition d'un unique indicateur, synthétique, permettant d'évaluer la qualité de représentation d'un modèle. Là aussi, c'est une pratique très fréquente, qui plus est dans le domaine de la simulation informatique en particulier sur des modèles de type « KISS » (ref. à chap 2 là ou ce sera abordé). Il s'agit alors de définir une « fonction objectif », ou « fonction de *fitness* », composée d'une pondération des quelques indicateurs composites qui auront été identifiés. Être en mesure d'évaluer un modèle à l'aide d'un unique indicateur a des avantages majeurs en pratique, puisque cela permet par exemple d'explorer et de paramétrer un modèle de simulation de manière entièrement automatique (trouver refs dans JASSS, dans Rey ou Schmitt) puisqu'on peut alors générer une cartographie simple des résultats du modèle en fonction des valeurs de paramètres utilisés.

Ces indicateurs composites et synthétiques résultent d'une quantification des autres indicateurs (excluant donc les formes stylisées qui sont plus libres d'interprétation), et apportent un grand confort dans le paramétrage d'un modèle de simulation.

3.2.2.5 Quels types d'indicateurs pour SimFeodal ?

SimFeodal n'est pourtant pas adapté à de tels indicateurs : une large partie des faits stylisés et ordres de grandeur mobilisés proviennent de connaissances expertes, et les thématiciens qui les ont consolidées rechignent à créer de tels indicateurs composites, en ce que cela demande de pondérer précisément l'importance de chacun des indicateurs par rapport aux autres. Pour pouvoir pondérer cette importance, il faudrait de plus que les différents indicateurs mobilisés présentent le même niveau de certitude et de variabilité dans leurs résultats, ce qui est peu le cas des indices empiriques – et donc des indicateurs de sortie – choisis dans le cadre du modèle SimFeodal.

On aurait ainsi pu créer quelques indicateurs synthétiques, mais ceux-ci ne prendraient en compte qu'une faible proportion du comportement attendu du modèle, résultant en une forte perte du pouvoir explicatif attendu du modèle. Par exemple, pour caractériser la polarisation du système de peuplement, il pourrait suffire de définir un indicateur composite fonction du niveau de concentration – le taux de foyers paysans dispersés –, du nombre de pôles et de l'espacement moyen entre les agrégats. Les valeurs de l'indicateur généré renseigneraient certes efficacement sur le succès ou l'échec d'un ensemble de valeurs de paramètres, mais cette information serait grossière, agrégeant dans le groupe des « simulations réussies » des configurations extrêmement diverses, et donc potentiellement très

éloignés des connaissances empiriques : une information multivariée ne peut pas toujours être résumée de manière univariée.

On a donc fait le choix d'évaluer SimFeodal en conservant des indicateurs de sortie non composites. Ce choix implique toutefois un problème majeur dans l'analyse des sorties de simulation auquel la réduction de dimensionnalité est une réponse : il est plus simple d'analyser quelques indicateurs plutôt qu'un grand nombre d'entre eux.

3.3 Hiérarchiser et catégoriser les indicateurs

SimFeodal s'appuie sur une dizaine d'indicateurs numériques, ainsi que sur plus d'une trentaine d'indicateurs multidimensionnels. Tous ces indicateurs ne présentent pas le même degré de certitude, la même échelle d'observation, et surtout, le même niveau de précision sur les phénomènes modélisés. A chaque changement dans le modèle, pour une évaluation complète de la capacité de cette version à reproduire les indices empiriques, il faudrait donc observer et analyser chacun de ces nombreux et divers indicateurs. Dans le contexte du paramétrage d'un modèle s'appuyant sur une logique itérative et incrémentielle (voir encadré 3.1), on imagine bien que cela n'est pas possible : le nombre d'indicateurs est bien trop élevé pour avoir rapidement une vision globale de la qualité de représentation du modèle. Il faut dès lors, comme pour toute analyse synthétique, concevoir une hiérarchie d'observation et d'utilisation des indicateurs : il ne sera pas nécessaire d'analyser chacun des indicateurs dans la plupart des cas, seuls les indicateurs jugés plus importants pourront être analysés. Les indicateurs de moindre importance ne seront mobilisés que pour départager des situations dont la différence ne serait pas suffisamment explicitée par l'usage des indicateurs principaux.

3.3.1 Incertitude

Dans le modèle de simulation, les indicateurs de sortie sont à analyser en tenant compte de la précision des indices qu'ils représentent. Il ne faudra ainsi pas étudier la croissance du nombre d'agrégats au cours de la simulation de manière fine, par exemple en étudiant le coefficient directeur de la courbe, quand l'empirie ne donne quasiment aucune information à ce sujet si ce n'est qu'il y a bien plus d'agrégats en fin de période qu'au début. On peut vouloir quantifier la précision de ces données, par exemple à l'aide des méthodes développées dans le champ des observations floues et/ou incertaines (voir par exemple le travail de Cyril de Runz sur les données « imparfaites » (DE RUNZ 2008)). Cette quantification de l'incertitude pourrait alors servir de base à l'établissement d'une hiérarchie des indicateurs : on analyserait en premier lieu l'écart entre les ordres de grandeurs empiriques bien connus (cf. tableau du niveau de certitude des objectifs) et les indicateurs calculés sur les données simulées. Les ordres de grandeur plus incertains seraient analysés dans un second temps (augmentation de la charge fiscale entre 800 et 1100 par exemple), et les formes stylisées viendraient enfin clore cette hiérarchie d'indicateurs. Toutefois, SimFeodal se caractérise d'une part par une très forte hétérogénéité dans les niveaux de connaissance des ordres de grandeurs et faits stylisés modélisés, et d'autre part, se voulant un modèle théorique (A dire spécifiquement dans le chapitre 2 ; y faire une ref ici), il n'y a pas d'obligation de « coller aux données » à tout prix : la vraisemblance d'ensemble du modèle compte bien plus que la précision de chacune de ses composantes.

3.3.2 Catégoriser les indicateurs : définir des dimensions d'analyse

En présence de plus d'une quarantaine d'indicateurs, il est toutefois nécessaire, a minima, d'organiser leur analyse. On a vu qu'il n'était pas justifié de mener cet ordonnancement à partir des propriétés intrinsèques des indicateurs du modèle. Au contraire, et cela porte bien plus de sens vis-à-vis du rôle d'un modèle, la hiérarchisation des sortie du modèle doit suivre la hiérarchie implicite

qui structure les hypothèses et objectifs du modèle en lui-même. Ces hypothèses et objectifs sont multiples dans SimFeodal, et dès lors, une hiérarchie globale ne peut être définie. Il convient donc de catégoriser les indices empiriques – et les indicateurs de sortie de simulation leur correspondant –, avant d’organiser, au sein même de ces catégories, les indices les caractérisant. La hiérarchisation des indicateurs se fera donc relativement à chacune de ces catégories.

Dans le chapitre précédent (ref chap 2), nous présentions les principales dynamiques à reproduire avec le modèle SimFeodal : (1) polarisation, (2) hiérarchisation et (3) fixation des foyers paysans. En postulant que ces dynamiques sont caractéristiques du modèle, on peut s’appuyer sur cette triade pour caractériser les sorties du modèle, c’est-à-dire mener la confrontation entre indices empiriques et indicateurs de sortie. En reprenant ces catégories, que l’on nommera **dimensions** (voir figure 3.1), on va donc répartir chacun des indicateurs dans la dimension qu’il sera le mieux en mesure de décrire. Cette répartition n’a pas à être égale, chaque dimension pouvant s’appuyer sur un nombre différent d’indicateurs. De même, chaque dimension sera composée d’indicateurs dotés d’une qualité de représentation ou d’un niveau de certitude hétérogène. Le seul point commun des indicateurs de sortie de chaque dimension doit être thématique. Les trois dimensions choisies – polarisation, hiérarchisation et fixation –, et les indicateurs qui les caractérisent dans le modèle, sont dès lors considérés comme les trois dimensions d’analyse des sorties de SimFeodal.

3.3.3 Hiérarchiser les indicateurs dans chaque dimension

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

3.3.4 Une hiérarchie mouvante

Notons que l’analyse des indicateurs de sortie suit une hiérarchie parfois mouvante, et en tous les cas, assez peu quantifiable : si l’ordre d’observation est plutôt stable, l’importance que l’on portera à chacun des indicateurs peut varier. Les indicateurs principaux de chaque dynamique sont ainsi « incontournables », c’est-à-dire qu’un résultat trop loin de celui des indices empiriques est disqualifiant. Parmi les indicateurs secondaires, il n’est pas toujours possible, d’après les connaissances des experts sur le sujet, d’établir une priorité ou une pondération de chaque indicateur. L’évaluation de la polarisation par exemple (page 11), se définit principalement par rapport à un indicateur principal – le taux de foyers

paysans dispersés –, mais selon les résultats des autres indicateurs de sortie, on leur attribuera une importance variable. L'étude de la dispersion des agrégats et pôles peut ainsi se révéler plus importante que celle de l'évolution du nombre d'agrégats selon les paramètres que l'on souhaite ajuster, ou se montrer tout au moins plus différenciante selon l'état du paramétrage.

Encadré 3.1 : Incrémentalité des indicateurs

De la même manière que les paramètres et mécanismes d'un modèle de simulation tendent à évoluer³ au cours du temps de la construction, souvent afin d'affiner un comportement observé, les indicateurs de sortie sont amenés à évoluer aussi.

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

L'observation de la répartition spatiale requiert certes de nouvelles analyses, mais surtout, par exemple, d'enregistrer les positions des agrégats au cours du temps. Si cet indicateur de sortie n'était pas utile avant cela, il n'y avait aucun intérêt à l'enregistrer. Il faut donc adapter l'implémentation du modèle pour générer, faire évoluer et enregistrer une nouvelle variable informatique correspondant à cet indicateur. Dès lors, on pourra composer un nouvel indicateur synthétique, qui, dans cet exemple, pourrait prendre la forme d'un indice de concentration spatiale).

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

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

les indicateurs présentés.

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

3.4 Les indicateurs et dimensions de SimFeodal

3.4.1 Évaluer la polarisation des foyers paysans

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

Mettre un schéma pour rendre compréhensible cette contradiction apparente.

Pour analyser la polarisation du système de peuplement, il est nécessaire de définir des indices permettant de caractériser ce phénomène. Ces indices doivent d'une part avoir une logique thématique, c'est-à-dire être appropriés à la description et à l'étude de la polarisation, mais doivent pouvoir être produits et enregistrés dans le modèle de simulation, formant des indicateurs.

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

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

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

3.4.1.1 Taux de foyers paysans dispersés

Cet indicateur, et sa déclinaison temporelle, sont vraisemblablement les plus évidents : plus le taux de foyers paysans dispersés en fin de simulation est faible,

plus le système de peuplement est polarisé. On a vu⁴ qu’empiriquement, autour de 1160, on observe environ 20% de foyers paysans encore dispersés, alors qu’ils sont près de 95% au départ. Le modèle sera donc d’autant plus satisfaisant que cet indicateur s’approchera de 20% en fin de simulation.

La « déclinaison temporelle » mentionnée juste au dessus permet d’affiner légèrement la précision de l’information communiquée par cet indicateur : il faut certes atteindre un objectif quantifié (20%), mais les hypothèses empiriques permettent aussi de penser qu’il faut que l’évolution de cet indicateur au cours du temps présente une tendance stable à la diminution, diminuant ainsi plus ou moins, avec de faibles fluctuations à chaque pas de temps. Une configuration de paramètres présentant des valeurs d’indicateurs en sortie proches des objectifs en fin de simulation mais fluctuant fortement pendant son déroulement ne sera ainsi pas valide. Elle montrerait en effet un phénomène trop aléatoire au cours du temps, dont on ne peut penser qu’il s’est déroulé historiquement selon les dires d’experts qui considèrent que cette évolution s’est déroulée de manière continue dans le temps.

Dans la version 0 de SimFeodal, on atteint en moyenne (moyenne des répliques) **57%** de foyers paysans isolés en fin de simulation. C’est un taux bien plus important que les 20% attendus, illustrant dès lors une polarisation trop faible. Dans la première moitié de la période (jusqu’en 1000, (figure 3.2)), le taux est en diminution constante et linéaire, ce qui est une bonne tendance, et ne présente de plus presque pas de variabilité, ce qui l’est aussi. Pourtant, à partir de 1000, la variabilité augmente, et la diminution du taux au cours du temps s’interrompt, avant de prendre une allure croissante. L’allure de la courbe sur la seconde moitié de la simulation est donc très éloignée de ce qui est attendu au regard des connaissances empiriques, aussi bien en matière de tendance qu’à la vue du taux décevant atteint au final.

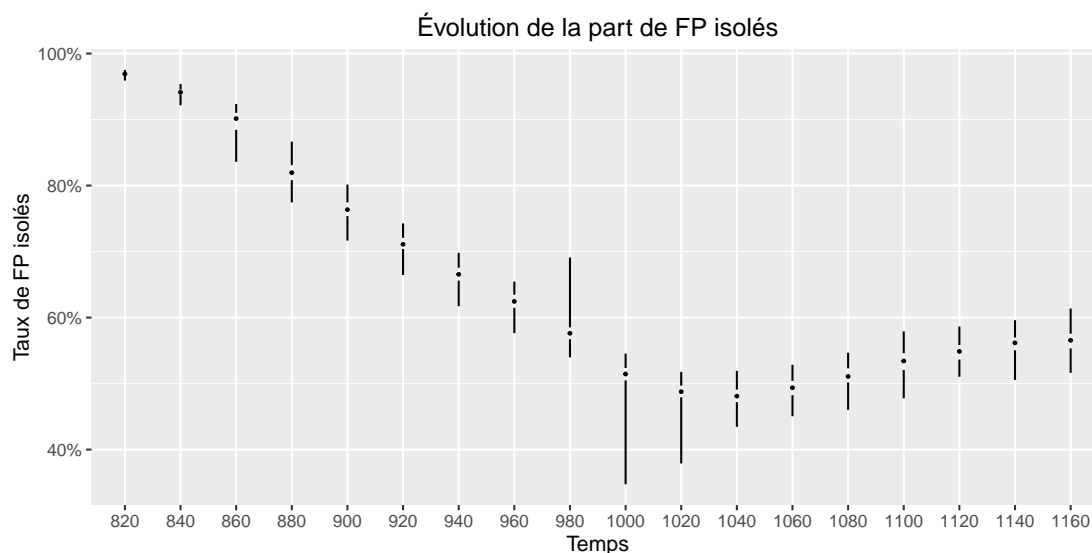


FIGURE 3.2 – Évolution de la part des foyers paysans isolés.

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

3.4.1.2 Nombre d'agrégats

Puisque les foyers paysans se concentrent au sein d'agrégats, il est logique d'observer l'évolution de ces derniers. Là aussi, (**objectif à définir**), on peut considérer qu'un nombre d'agrégats en fin de simulation proche de l'objectif, 200, permet de caractériser une polarisation réussie. Cet indicateur ne peut être lu seul, et c'est pour cela qu'il vient dans un second temps (cf. 3.3.3, Hiérarchiser les indicateurs dans chaque dimension, p.8) : en effet, un faible nombre d'agrégats peut aussi bien être révélateur d'une très faible polarisation des foyers paysans (ceux-ci restant dispersés) que d'une trop importante (un unique agrégat concentrant l'ensemble des foyers paysans par exemple). Une fois le taux de foyers paysans dispersés connu et ces potentiels biais pris en compte, le nombre d'agrégats et son évolution nous renseigne cependant sur les dynamiques de polarisation. D'après les connaissances expertes, on s'attend à ce que le nombre d'agrégats, très faible au départ (24 dans la version 0), suive trois phases : une première phase de croissante lente, le temps que les mécanismes agissent sur la polarisation, suivie d'une période de croissance plus rapide, une fois que tous les foyers paysans commenceront à être suffisamment attirés par les pôles pour y former des agrégats, et enfin, une nouvelle phase de croissante plus lente, une fois les foyers paysans répartis dans les agrégats existants et qui se déplaceront vers des agrégats plus importants, hiérarchisant le système de peuplement. Cette allure d'évolution rappelle les fonctions logistiques connues par exemple pour les cycles de diffusion/adoption des innovations **ref Pumain ou MN Comin**, et résulte des connaissances expertes des archéologues spécialistes de la période.

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

L'observation de cet indicateur au cours du temps (figure 3.3) est quant à elle assez satisfaisante : on retrouve bien les trois phases attendues, quand bien même le début de la croissance plus importante (vers 1000 ici) est un peu trop tardive. Ce moment coïncide de plus avec la stagnation puis l'inversion de la courbe d'évolution de la part de foyers paysans isolés.

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

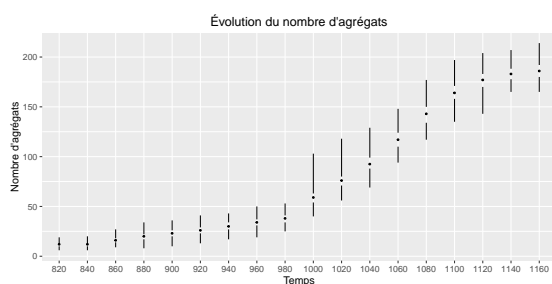


FIGURE 3.3 – Évolution du nombre d'agrégats.

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

3.4.1.3 Nombre de pôles

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

Par ailleurs, afin de renforcer la polarisation par l'action de l'attraction différenciée – selon le nombre de foyers paysans contenus dans chaque agrégat⁵ –, il faut que le taux de pôles contenant un agrégat soit important, et surtout croissant au cours du temps : comme dans l'empirie, cela est alors le marqueur que de petits pôles d'attractions, comme les églises paroissiales, parviennent à polariser suffisamment de foyers paysans de leur voisinage pour aboutir à la création d'un petit agrégat, un village par exemple.

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

L'évolution du nombre de pôles est assez représentative de l'empirie dans la version 0 de SimFeodal. On aboutit ainsi sur **190** pôles en moyenne en fin de simulation, ce qui est dans le bon ordre de grandeur, quoi qu'un peu trop faible. Ce nombre est en croissance constante à partir de 940 (figure 3.4). Ce départ quelque peu tardif peut expliquer le relatif manque de pôles en fin de simulation.

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

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

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

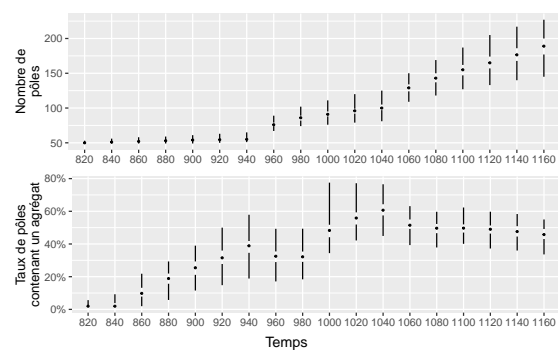


FIGURE 3.4 – Évolution du nombre de pôles et du taux de pôles contenant un agrégat.

3.4.1.4 Dispersion des agrégats et pôles

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

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

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

La lecture des cartes de répartition des agrégats et des pôles de deux répliques de la version 0 (figure 3.5) va, dans l'ensemble, dans le sens de l'empirie. On peut ainsi remarquer que ces entités ont bien tendance, au cours du temps, à se disperser dans l'espace modélisé. En fin de simulation, tout l'espace est occupé, et on remarque même que certaines zones voient une forte concentration en agrégats, reproduisant les faits stylisés connus. La répartition spatiale des agrégats – et des pôles – est donc bonne, et ne semble pas opposer d'obstacle à la polarisation attendue du système de peuplement.

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

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

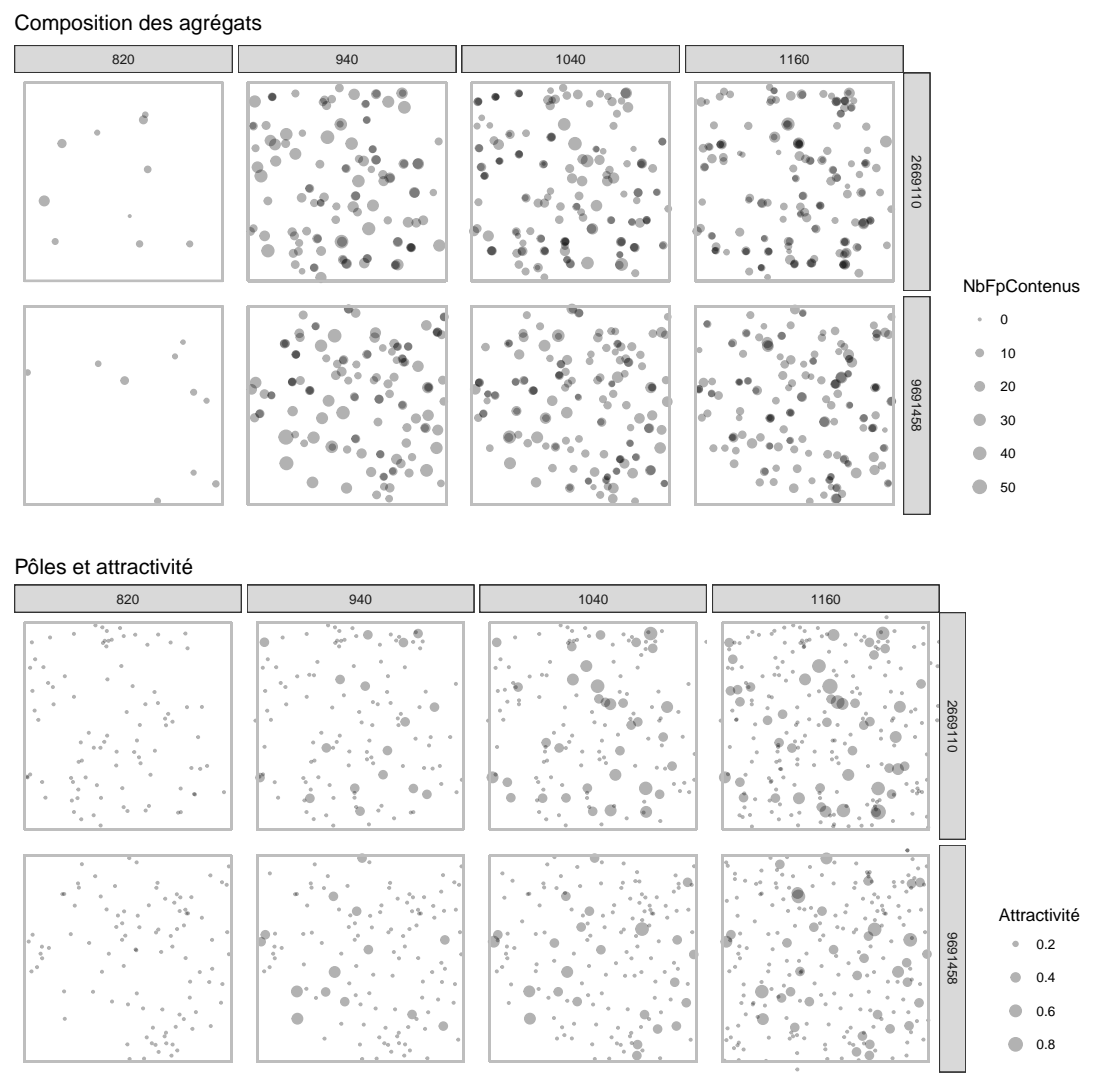


FIGURE 3.5 – Évolution de la répartition spatiale des agrégats et pôles, pour deux répliquations.

3.4.2 Évaluer la hiérarchisation du système de peuplement

La seconde dimension de l'évaluation du modèle SimFeodal correspond à l'étude de la hiérarchisation du système de peuplement. On déduit en effet des connaissances empiriques une forte hiérarchisation du système de peuplement sur la période, et plus généralement, des entités présentes. On passe ainsi, en 800, d'un habitat dispersé dans lequel coexistent quelques agrégats de taille uniforme, à un habitat concentré dans des agrégats de taille très hétérogènes à la fin du XII^{ème} siècle. La distribution des tailles des agrégats est estimée par les connaissances expertes, toutes proportions gardées, comme assez proche des distributions observées aujourd'hui dans les systèmes de peuplement. On souhaite ainsi que les agrégats modélisés suivent une distribution approchant la distribution log-normale.

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

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

De la même manière que pour la polarisation, les indicateurs secondaires ont aussi pour but de préciser cet indicateur principal, et en particulier d'analyser les moteurs de cette hiérarchisation du peuplement. On a en effet choisi d'observer plutôt la hiérarchisation des autres types d'entités – pôles, seigneurs, paroisses –, pour vérifier qu'elles accompagnent et/ou entraînent bien la hiérarchisation des agrégats. La hiérarchie des pôles, par exemple, a une influence directe sur l'attraction effectuée sur les foyers paysans (polarisation) et sur la hiérarchisation des agrégats : par effet d'attraction différenciée (voir la note de bas de page 5 page 15), des agrégats plus importants se constituent autour des pôles les plus importants.

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

3.4.2.1 Hiérarchie des agrégats

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

Dans le cas de SimFeodal, le faible nombre d'agrégats ainsi que la variabilité de leurs tailles rend difficile cette analyse quantifiée, le coefficient directeur,

par exemple, étant très sensible aux faibles effectifs. On utilise toutefois la représentation graphique décrite comme un indicateur majeur de la hiérarchie des agrégats.

Du point de vue des connaissances empiriques, la courbe doit ainsi voir sa pente augmenter avec le temps, tout en devenant plus convexe, ce qui représente la « longue traîne » des petits agrégats, empiriquement observée dans toutes les distributions de systèmes de peuplement.

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

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

La version 0 de SimFeodal présente une hiérarchisation des agrégats nettement trop faible. La courbe rang-taille (figure 3.6) est ainsi trop faiblement pentue, et présente une forme trop linéaire : la convexité due à la longue traîne n'est pas assez visible, en raison sans doute de trop faibles valeurs en haut de la hiérarchie, qui ne « tire » alors pas assez la distribution. Ce résultat de simulation est d'autant plus perturbant que son évolution est éloignée de l'empirie : on remarque en effet que la distribution se hiérarchise bien entre 800 et 1020, présentant à cette date une allure très satisfaisante. Pourtant, après cette période, les agrégats voient leur hiérarchie diminuer nettement et les agrégats les plus peuplés diminuer en taille.

On constate le même décrochage dans la discrétisation des agrégats (figure 3.7), où on remarque de plus que la hiérarchie la plus proche de l'attendu, où la position des classes suivrait une ligne droite, semble se dessiner entre 940 et 1040. En 1040, et plus encore en 1160, la proportion d'agrégats de taille moyenne et haute (de 30 à 100, et de plus de 100 foyers paysans) est trop faible, et pas assez hiérarchisée.

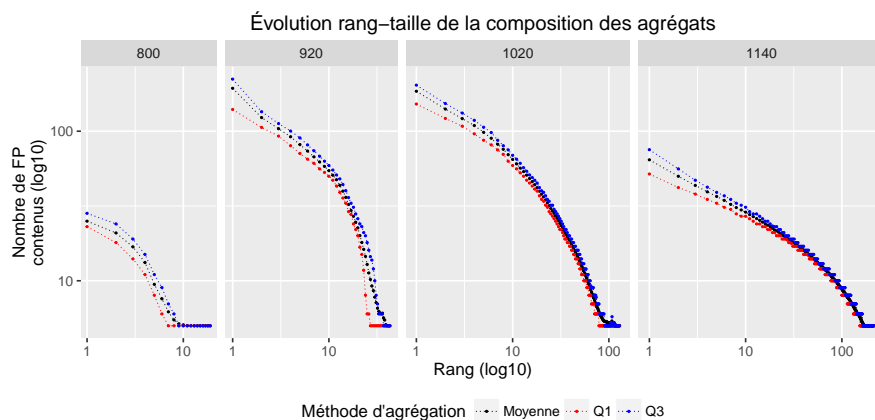


FIGURE 3.6 – Évolution de la courbe rang-taille des agrégats.

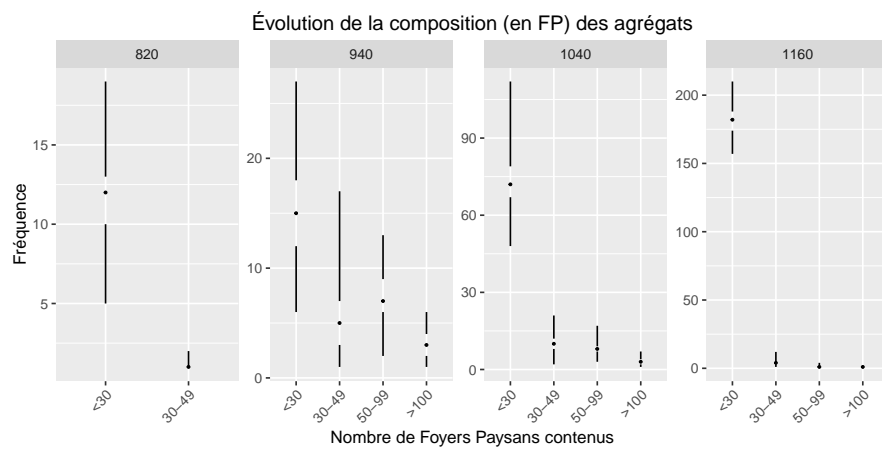


FIGURE 3.7 – Évolution de la composition des agrégats.

3.4.2.2 Hiérarchie des pôles

La hiérarchie des agrégats donne une bonne vision agrégée de la hiérarchisation du système de peuplement dans son ensemble. Pour autant, afin d'appréhender la dynamique de cette dimension, il est là encore nécessaire d'observer le comportement des composantes qui provoquent cette hiérarchisation. En effet, une forte hiérarchie des pôles entraînera une attractivité des foyers paysans très inégale⁶. De plus, comme indiqué plus haut, on cherche à obtenir une forte hiérarchie pour les différents types d'agents du modèle. L'observation de la hiérarchie des pôles est donc importante pour évaluer le modèle SimFeodal. Pour déterminer cette hiérarchie, on peut se fier à deux indicateurs complémentaires : le nombre d'attracteurs composant chaque pôle et l'attractivité de ces derniers. Ces indicateurs de sortie sont proches, mais apportent pourtant une vision légèrement différente : étant donné que chaque attracteur influe différemment, selon son type, sur l'attractivité globale d'un pôle, l'information sur l'attractivité et sur la composition ne sont pas redondantes, bien que fortement corrélées. On aurait pu présenter une information plus détaillée quant à cette composition, par exemple en différenciant le nombre de chacun des types d'attracteurs de chaque pôle, mais le nombre de combinaisons possible aurait rendu cette information confuse.

À partir des connaissances expertes qui guident l'évaluation de SimFeodal, on cherche à obtenir, pour ces deux indicateurs, une courbe d'allure similaire, c'est-à-dire une courbe décroissante, avec bien plus de pôles mineurs (faible attractivité ou nombre d'attracteurs) que de pôles plus importants. On cherche de plus à ce que cette courbe présente une allure log-normale, et donc que la proportion de pôles décroisse fortement à mesure que leur importance augmente.

Dans l'ensemble, on peut remarquer sur les deux indicateurs de la figure 3.8 que les pôles, dans cette version 0, sont très hiérarchisés. La tendance « évolutive va dans le bon sens : depuis une quasi-uniformité en 820, des pôles plus importants apparaissent et semblent se renforcer au cours du temps. On retrouve aussi, en observant les axes des ordonnées, la croissance du nombre de pôles identifiée plus haut (3.4.1.3, Nombre de pôles, p. 15) : de nouveaux pôles apparaissent tout au long de la simulation, et restent majoritairement peu importants, quand les pôles ayant commencé à se renforcer tôt continuent dans cette tendance.

En fin de période, les pôles sont très hiérarchisés, aussi bien en matière de composition que d'attractivité. Ils le sont même plus que ce que les connaissances expertes laissent entendre : au delà de 0.66 en attractivité ou de 4 attracteurs, on ne dénombre plus qu'un unique pôle de chaque importance, là où on attendrait que cette courbe continue plutôt à décroître. De plus, ces courbes montrent la variation des répliques de la version 0. Dès lors, on peut considérer que ces pôles majeurs ne sont en fait qu'un unique pôle (le plus souvent) d'importance variable selon les répliques, mais dans tous les cas d'importance bien supérieure aux pôles qui arrivent juste après dans le classement. Cette macrocéphalie ne correspond pas aux observations empiriques, d'autant que rappelons-le, la ville de Tours n'est pas modélisée dans SimFeodal.

6. En raison des logiques d'attachement préférentiel, voir la note de bas de page 5 page 15.

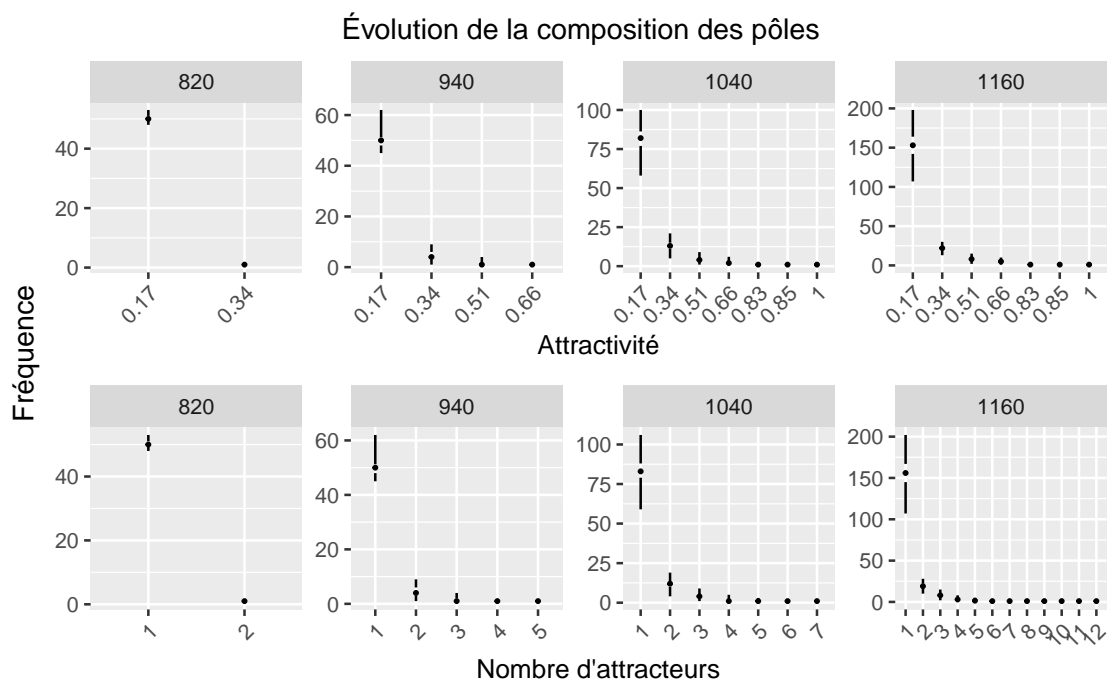


FIGURE 3.8 – Évolution de la composition et de l'attractivité des pôles⁷.

7. Les figures ne sont pas en log-log (ou log-lin) et on ne voit donc pas vraiment le haut de la hiérarchie. Peut-être faudrait-il changer ces axes, mais ça complexifiera la lecture pour les archéos/lecteurs non habitués aux axes log

3.4.2.3 Hiérarchie des paroisses

A l'instar des pôles, on attend aussi des ressorts paroissiaux d'être hiérarchisés. La période modélisée voit ainsi apparaître ces paroisses qui auront un rôle majeur dans la fixation du peuplement, et, pré-figurant le maillage communal, ont un double rôle de desserte efficace⁸ et équitable⁹. En effet, avec la volonté d'encadrement et de prélèvement de l'Église sur la population, de nouvelles paroisses apparaissent pour desservir au mieux leurs potentielles ouailles. Une structure double en résulte : dans les zones les moins denses, le maillage est régulier mais lâche, de manière à minimiser le nombre d'églises paroissiales tout en s'assurant que chacun puisse y accéder dans un temps raisonnable¹⁰. Il y a donc un certain nombre d'églises paroissiales desservant peu de paroissiens. Au contraire, dans les zones les plus denses, et en particulier au sein des petites villes naissantes, l'objectif est d'être au plus près des résidents tout en garantissant à chacun de pouvoir assister aux différents offices : il y donc une croissance du nombre d'églises paroissiales proches les unes des autres, visant à accompagner un encadrement maximum de la population, ainsi, avec une logique concurrentielle de ce clergé féodal, qu'à capter l'importante source de revenus qu'assure la collecte de la dîme. Cette logique concurrentielle doit aussi permettre de restreindre le nombre de foyers paysans desservis par une unique paroisse.

On s'attend donc à avoir une courbe hiérarchisée dans la lignée d'une courbe log-normale, mais avec toutefois un double seuil minimal (l'effet de « longue traîne »), autour des paroisses « rurales » peu peuplées (moins de 10 foyers paysans desservis) et des paroisses du bas de la hiérarchie classique, peuplées de 10 à 40 foyers paysans¹¹).

ici, impérativement, il faudra mettre un graphique schématique de ce qu'on attend.

En fin de simulation de cette version 0, la hiérarchie des paroisses est plutôt satisfaisante. On y retrouve en effet une forte hiérarchie, peut-être même légèrement trop importante. Il ne devrait ainsi pas y avoir de paroisse composée de plus de 100 foyers paysans en fin de simulation, et la courbe présente une pente trop importante (celles de 940 ou 1040 sont ainsi plus conformes aux attentes). Pour autant, on remarque bien un effet de hiérarchisation au cours du temps, ainsi qu'un accroissement du nombre de petites paroisses, montrant que la dynamique, bien que non ajustée, s'inscrit bien dans la dynamique observée empiriquement. (figure 3.9)

8. C'est-à-dire desservir de manière optimale la plupart de la population.

9. C'est-à-dire faire en sorte que même les populations les plus isolées aient un accès aussi rapide que possible à une église paroissiale.

10. Cette distance-temps évolue au cours du temps, en fonction de l'accroissement de la fréquence de l'obligation de fréquentation des églises paroissiales.

11. D'après les valeurs des paramètres `nb_min_paroissiens` et `nb_max_paroissiens`

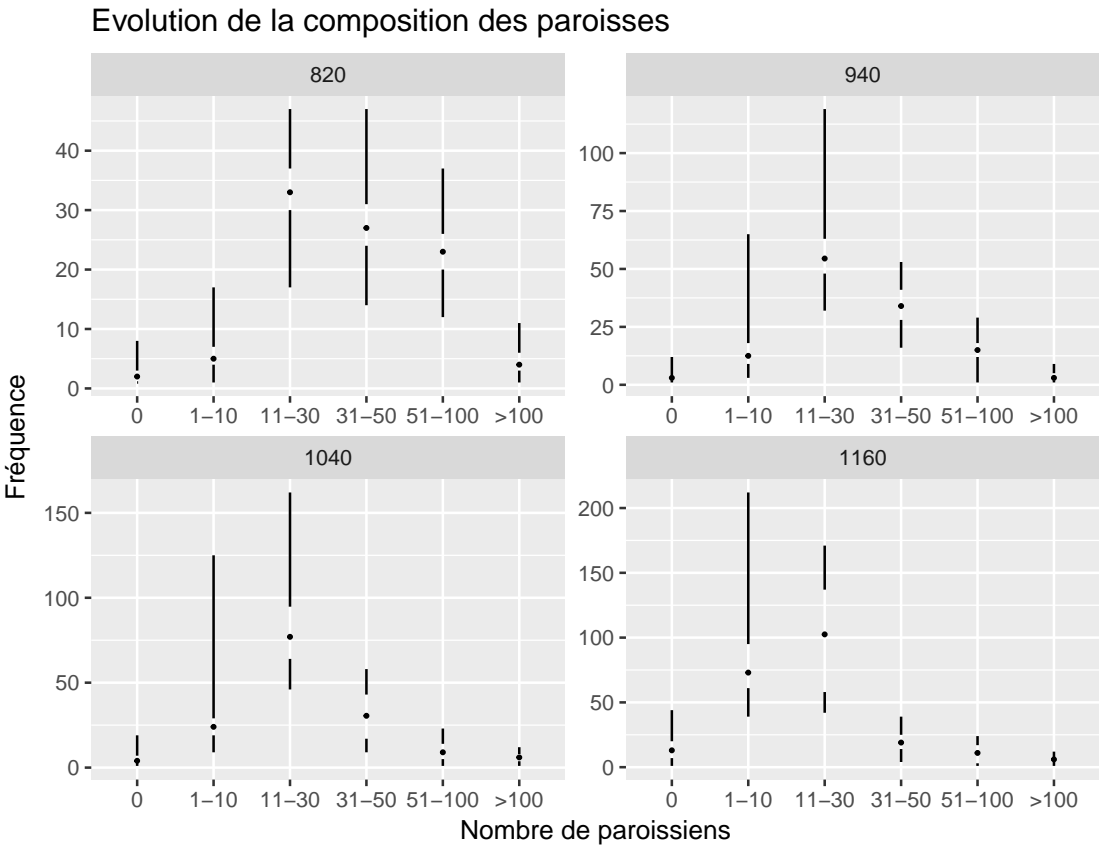


FIGURE 3.9 – Évolution de la composition des paroisses.

3.4.3 Évaluer la fixation et la dissémination du peuplement

La dernière dimension étudiée par SimFeodal est moins définie que les précédentes. On y observe ainsi la fixation du système de peuplement au sein du maillage naissant que constituent les paroisses. Cette notion de fixation pose problème par rapport à l'ensemble des indicateurs de sortie déployés jusqu'ici. En effet, SimFeodal est un modèle fondé sur un temps discret : on y observe, à chaque itération, le résultat des processus modélisés. On peut y observer une fixation d'un pas de temps par rapport à l'autre, par exemple en constatant qu'un agrégat constitué en 820 semble être toujours présent en 840. On peut de plus considérer que cet agrégat présente les mêmes caractéristiques au deux dates, en matière de localisation et de population. L'agrégat est donc stable dans le temps. L'échelle d'analyse mobilisée dans les indicateurs précédents – au niveau agrégé des agrégats, voir du système d'agrégats – pourrait donc s'appliquer ici aussi. Pourtant, quand on parle de fixation des foyers paysans, ce qui est observé empiriquement, il n'est pas question d'agrégats, mais des entités les composant, c'est-à-dire les foyers paysans. Ceux-ci ne sont pas observés directement quand on remarque que les agrégats semblent stable : on approche là à la différence entre un état stable et un état stationnaire. La relative stationnarité des agrégats dans le temps n'est ainsi pas garante de stabilité de leurs composantes : deux foyers paysans oscillant d'un agrégat à un autre, dans un mouvement opposé, produiraient ainsi une stationnarité de ces agrégats, mais, en se déplaçant, ils ne satisferaient pas au critère de fixation. Il y a donc un problème de changement d'échelle pour observer la fixation des foyers paysans : il n'est plus possible de raisonner à l'échelle des agrégats, et il faut se concentrer sur celle des foyers paysans.

Cette échelle pose un autre problème : les foyers paysans sont très nombreux (4000 dans la version 0) et se déplacent. Ils se déplacent de plus selon des modalités très différentes ([Faire ref aux mécanismes de déplacement du chapitre 2](#)), rendant complexe la caractérisation des mouvements de chacun et plus encore celle d'une agrégation de ces catégories.

Pour ces raisons, la production d'indicateurs synthétiques spatiaux – une ou plusieurs cartes – ne suffirait pas à communiquer une information intelligible sur l'éventuelle fixation des foyers paysans. Il nous faut donc faire appel à des « proxys », non spatiaux, pour évaluer la fixation des foyers paysans.

A cet effet, on a retenu des indicateurs relatifs aux déplacements des foyers paysans et à leur raison : combien de foyers paysans se déplacent à chaque pas de temps (moins il y a de déplacement, plus la fixation est importante) ? Quelles sont les modalités de ces déplacements (un déplacement entre deux agrégats lointains n'a pas les mêmes conséquences en terme de stabilité qu'un déplacement minime au sein d'un même agrégat) ? Ou encore, comment évolue la satisfaction des foyers paysans, et avec celle-ci, la probabilité de se déplacer ?

Ces indicateurs permettent d'évaluer la capacité du modèle à reproduire la fixation des foyers paysans. Pour autant, la contrainte est ici double : on recherche une fixation, mais celle-ci est, empiriquement, supposée se dérouler et se voir renforcer par la mise en place du maillage paroissiale qui doit servir de support à la nouvelle configuration spatiale émergente.

On s'appuiera donc aussi sur des indicateurs relatifs à cet espace support constitué par les paroisses : leur nombre, leur dispersion dans l'espace et l'efficacité de la desserte qu'elles assurent.

Notons que contrairement aux deux précédentes dimensions d'analyse, nous n'établissons ici pas de hiérarchie nette entre les indicateurs. Cette étude de la fixation est moins facilement appréhendable que celle de la polarisation ou de la hiérarchisation et les indicateurs qui la caractérisent apportent une complémentarité de points de vue plus qu'un affinement de l'évaluation de cette dynamique. Dès lors, les indicateurs présentés ci-après ne peuvent être catégorisés en indicateurs principaux et secondaires. On retrouvera cependant cette hiérarchie d'évaluation au sein des indicateurs, par exemple en suivant l'ordre des graphiques présentés. Par exemple, pour évaluer la fixation des foyers paysans, on observera d'abord le résultat produit (nombre de déplacements) avant d'entrer dans le détail de sa composition (types des déplacements).

3.4.3.1 Déplacement des foyers paysans

Le déplacement des foyers paysans est l'élément moteur de SimFeodal : c'est par le déplacement individuel de chacun des foyers paysans que la configuration spatiale évolue. Les déplacements affectent donc chacune des dimensions d'analyse – polarisation, hiérarchisation et fixation –, mais c'est au sein de cette dernière qu'il est le plus intéressant de les observer. Il est en effet attendu que de nombreux déplacements surviennent, afin que le système de peuplement puisse se structurer, mais pour autant, il est aussi nécessaire que ces déplacements tendent à diminuer au cours du temps, une fois le système en voie de stabilisation.

On pourrait donc attendre que les déplacements suivent une courbe négative (linéaire ou non) tendant vers 0, impliquant une absence de déplacements en fin de période. Pour autant, le mécanisme de déplacement est sans doute l'un des plus complexes du modèle, et on ne peut l'appréhender aussi simplement.

En premier lieu, les mécanismes de SimFeodal différencient deux types de déplacements ([Faire ref à chap2](#)) : les déplacements locaux (dans un rayon de 2500m dans la version 0) et les déplacements lointains.

- Les déplacements locaux visent à faire s'agréger des foyers paysans dispersés autour de pôles présents à proximités. Ce mécanisme peut être présent sur toute la durée de la simulation, sur un effectif faible. En effet, cet effet d'agrégation locale permet « d'optimiser » la répartition spatiale des agrégats, en renforçant leur hiérarchie et en faisant fusionner des agrégats qui seraient très proches les uns des autres. Pour autant, dans un objectif de fixation du peuplement, il est nécessaire de veiller à ce que les foyers paysans ne soient pas amenés à se déplacer localement de manière continue, par exemple en faisant des allers-retours entre des pôles ou agrégats spatialement proches.
- Les déplacements lointains servent un autre rôle : un foyer paysan qui ne serait pas en mesure d'augmenter sa satisfaction localement – faute de pôles suffisamment attractifs dans le voisinage – a une probabilité de se déplacer vers un agrégat situé n'importe où dans l'espace modélisé. Le modèle est très sensible à ce mécanisme qui agit comme une perturbation forte dans la structure spatiale du peuplement. Ce mécanisme « de dernier recours » ne doit être employé que rarement, en cas de situations où des foyers paysans seraient trop isolés pour pouvoir s'agréger localement. C'est notamment le cas pour les foyers paysans nouveaux arrivants, via le mécanisme de renouvellement ([ref dans chap 2](#)), qui peuvent se voir localisés n'importe où dans l'espace du modèle.

Pour ajouter à la complexité du mécanisme, et donc de l'évaluation de cet indicateur qu'est le nombre de déplacements au cours du temps, rappelons que dif-

férentes contraintes temporelles viennent bouleverser, à dessein, le comportement des foyers paysans. En particulier, entre 950 et 1050, les modalités d'évaluation de la satisfaction deviennent plus strictes (Voir dans chap2, frise). Cela engendre nécessairement une plus forte propension des foyers à se déplacer.

Si les connaissances empiriques d'un tel niveau de finesse ne sont pas disponibles, on peut tout de même avoir des attentes quant au comportement attendu du modèle. Au regard des éléments décrits plus haut, on peut ainsi chercher à ce que l'évolution des déplacements suive plusieurs rythmes au cours du temps :

1. Dans une première phase, du début de la simulation jusqu'aux perturbations débutant en 950 : quelques déplacements lointains marginaux ($\approx 5\%$ à 10%), stables au cours du temps ; et de plus nombreux (au moins $\approx 30\%$) déplacements locaux menant à la constitution de petits agrégats locaux. Les déplacements locaux doivent diminuer au cours du temps, une fois les agrégats constitués.
2. Pendant la deuxième phase, entre 950 et 1050, les nombreuses perturbations devraient voir une nette augmentation des déplacements locaux, et dans une moindre mesure lointains, prémices à la constitution d'agrégats plus hiérarchisés.
3. Après ces perturbations, on devrait retrouver le niveau de déplacement de la seconde période, et là aussi, tendre vers une diminution des déplacements locaux, le système se stabilisant à l'approche de la fin de la période.

Tout au long de cette période, on cherche de plus à ce que les foyers paysans soient polarisés, c'est-à-dire ici, qu'ils se regroupent dans des agrégats de population. Parmi les modalités de déplacement, un autre indicateur utile est ainsi l'observation des provenances et destinations des foyers paysans qui se déplacent : plus les foyers paysans originellement dispersés auront tendance à rejoindre des agrégats, plus la simulation sera satisfaisante au regard des hypothèses empiriquement émises.

Dans cette version 0 de SimFeodal, il apparaît en premier lieu que les déplacements sont trop peu nombreux (figure 3.10). Les déplacements lointains sont à peu près dans les proportions attendues, mais les déplacements locaux sont trop peu nombreux. Surtout, les trois phases attendues ne se retrouvent pas sur ces sorties. On y remarque bien l'impact des perturbations, mais celles-ci ne font qu'augmenter la part de déplacements, laquelle reste stable avant et après ces perturbations. La version 0 de SimFeodal n'est donc pas satisfaisante en termes de stabilisation et de fixation des foyers paysans. L'observation des modalités de déplacement (figure 3.11) montre certes une part croissante de déplacements « Isolé -> Agrégé », mais dans des proportions là aussi trop faibles passé le tout début de la simulation.

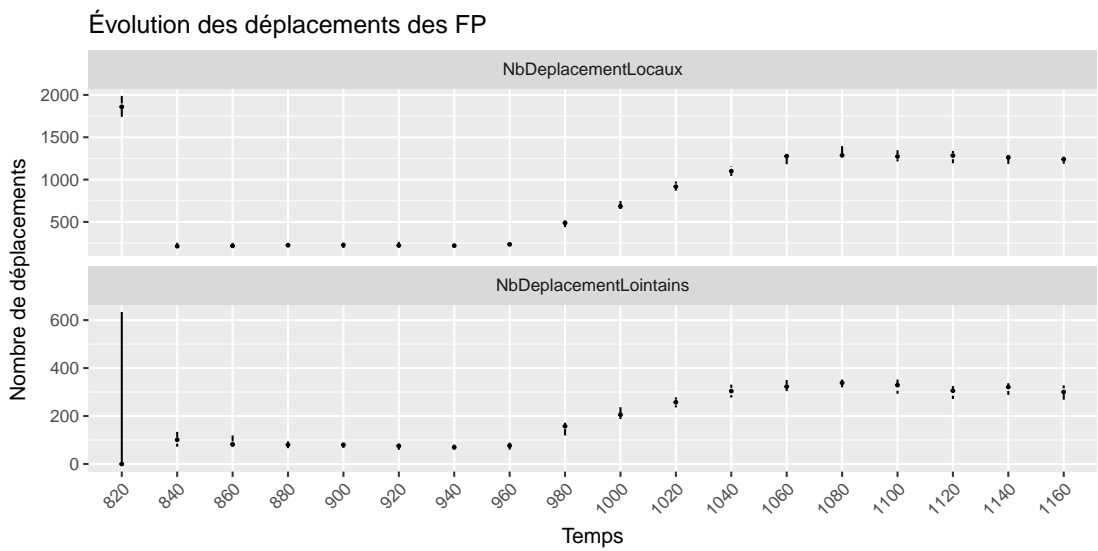


FIGURE 3.10 – Évolution du nombre de déplacement des foyers paysans.

Mettre le graphique en relatif

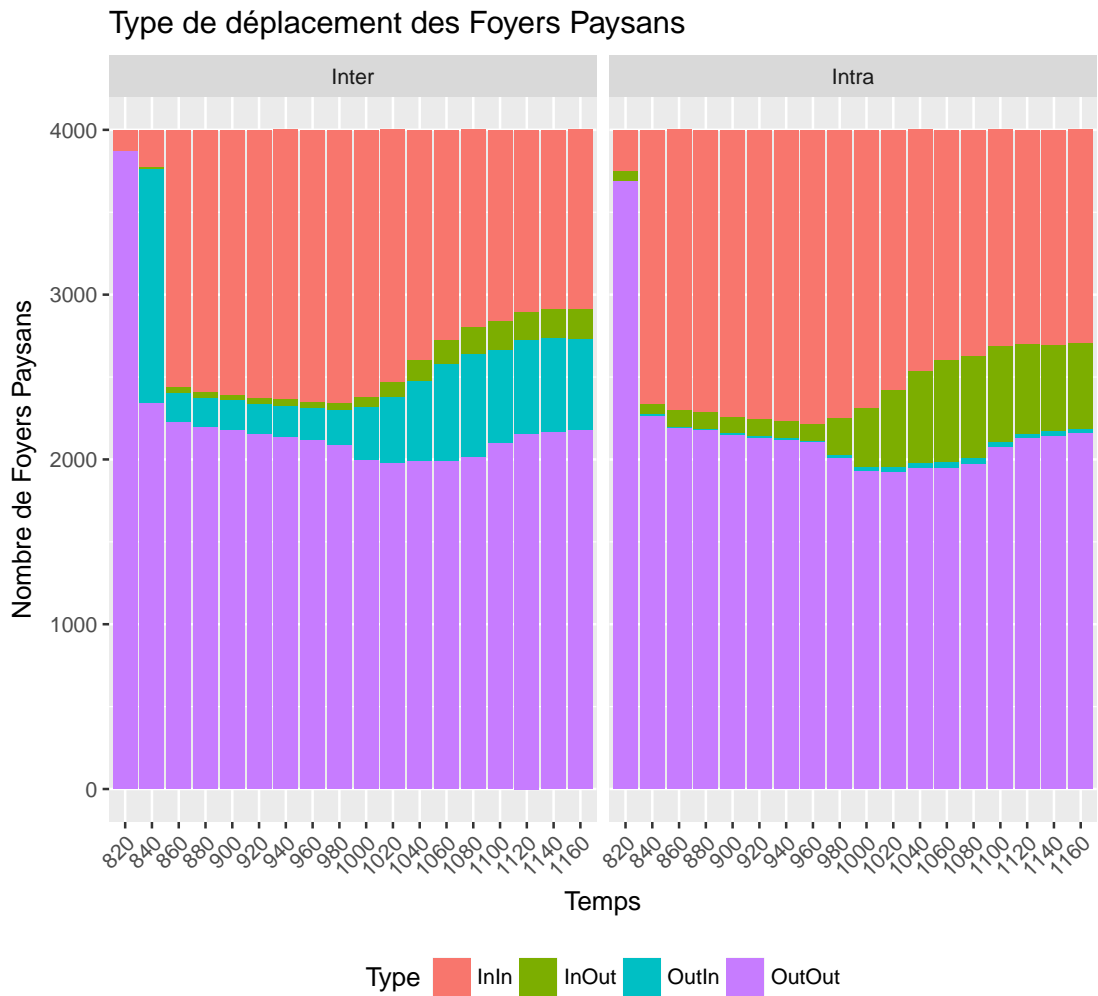


FIGURE 3.11 – Évolution des types de déplacement des foyers paysans.
Inter : Déplacement depuis le pas de temps précédent vs Intra : déplacement au sein d’un même pas de temps.

Ne garder que l’inter et mettre le graphique en relatif

3.4.3.2 Satisfaction des foyers paysans

On a vu dans les deux indicateurs précédents que les déplacements des foyers paysans sont très affectés par leur niveau de satisfaction. Pour comprendre ces déplacements, il est donc utile d'observer en détail l'évolution de la satisfaction qui les provoquent.

La satisfaction ne saurait être résumée en un simple indicateur de fixation, tant son rôle est prépondérant dans une large partie des mécanismes du modèle. Pourtant, mobilisé ici, cet indicateur apporte un éclairage différent. Il permet ainsi de préciser les indicateurs précédents en donnant une explication à leur éventuelle mauvaise réponse aux attentes.

Ainsi, une satisfaction globalement trop élevée ne serait pas assez motrice à des déplacements, résultant en une polarisation faible. Au contraire, une satisfaction globalement faible engendrerait une très forte mobilité, par exemple sous forme de mouvements pendulaires, d'où une absence de fixation du peuplement. Comme pour les déplacements, on attend, depuis les connaissances empiriques, qu'il y ait trois phases dans l'évolution de cette satisfaction : (1) une première phase, jusqu'en 950, où les foyers paysans sont globalement satisfaits, et le sont de plus en plus à mesure qu'ils s'agrègent ; (2) une seconde phase, entre 950 et 1050, où les restrictions fortes (distance à un château, à une église paroissiale) auront pour effet de violemment abaisser le niveau de satisfaction ; et enfin, (3) une dernière phase où, passées les perturbations, le niveau de satisfaction tend à remonter doucement, sous l'effet de l'agrégation et de la constitution généralisées de communautés paysannes et de la construction de châteaux et de nouvelles églises paroissiales.

Au vu des résultats de la version 0 de SimFeodal illustrés dans la figure 3.12, on constate que les trois phases attendues sont bien présentes. La perturbation en milieu de période est très forte, mais pourtant, le niveau de satisfaction général reste très élevé (à toute date, plus de 80% des foyers paysans ont une satisfaction supérieure à 0.5). Avec ce niveau de satisfaction et les mécanismes de déplacement de cette version 0, seuls les foyers paysans à proximité de très gros pôles (au moins un gros château et plusieurs églises paroissiales) pourront se déplacer localement. Et pour peu qu'il y ait plusieurs pôles importants à proximité, ils oscilleront entre les deux, gonflant artificiellement le nombre de déplacements locaux. L'observation de cet indicateur confirme que le niveau de déplacement est trop faible et ne présente pas l'allure attendue, tout en donnant une explication à la mauvaise fixation du peuplement observée.

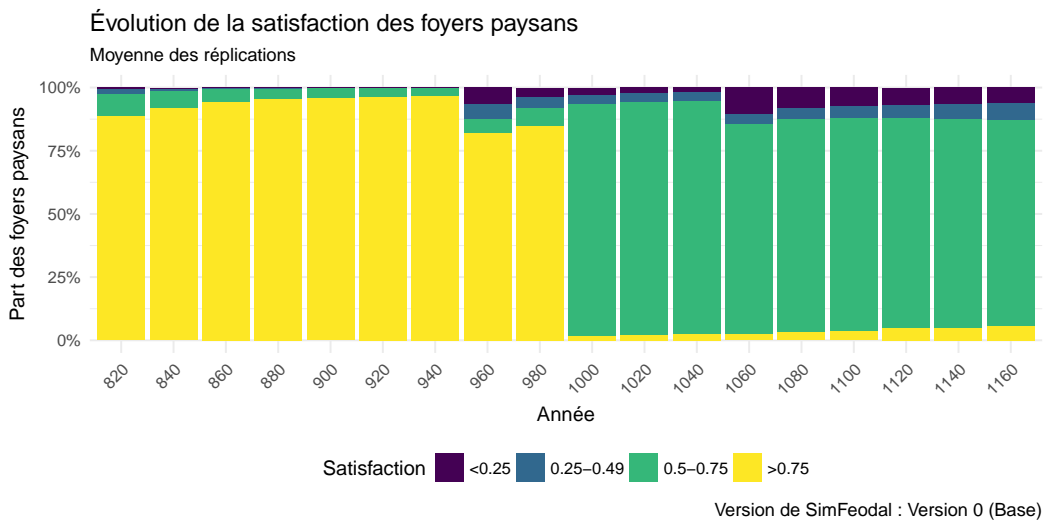


FIGURE 3.12 – Satisfaction des foyers paysans

3.4.3.3 Nombre et dispersion des paroisses

Tout au long de la période, de nouvelles églises paroissiales sont créées (cf. **mécanisme dans chap. 2**) et viennent renforcer l'encadrement des foyers paysans. L'évolution du maillage constitué par les ressorts paroissiaux représente donc l'évolution de la structure spatiale des foyers paysans. Pour que les foyers paysans soient satisfaits, il faut, à partir de 950, qu'ils soient suffisamment proches d'une église paroissiale. Celles-ci constituent donc, à mesure de l'avancement de la période, des pôles qui vont aider à la fixation des foyers paysans. Il est donc légitime d'observer la croissance du nombre et la répartition des paroisses telles que simulées dans le modèle SimFeodal.

Sur un plan purement numéraire, plus les paroisses seront nombreuses, mieux la population sera desservie, et moins les foyers paysans se déplaceront : la fixation sera donc plus forte à mesure que le nombre de paroisses augmente. Sur un plan spatial, l'accumulation de paroisses en zones denses (dans des agrégats de populations) doit renforcer la polarisation de ces zones, et avec elle, accroître les chances de fixation des foyers paysans.

D'après les mécanismes mis en places dans cette version 0 de SimFeodal, on s'attend donc à ce que le nombre de paroisses augmente régulièrement au cours du temps, depuis un nombre initial évalué à 50 en 800, jusqu'à atteindre un objectif numérique fixé à 200 d'après les connaissances empiriques de la région modélisée.

Concernant la répartition spatiale, on cherche à atteindre le double phénomène décrit dans la partie 3.4.2.3 (Hiérarchie des paroisses, p. 24). Spatialement, cela devrait mener à une diminution de la superficie des paroisses les plus larges dans les zones peu denses. Dans les zones plus denses, concentrant les agrégats, cela devrait aussi mener à une diminution de la superficie, bien plus drastique cependant : avec la création de nouvelles paroisses au sein des agrégats, on devrait voir apparaître de nombreuses paroisses se partageant un espace très réduit.

Notons que la dispersion des agrégats et des pôles, vue précédemment (3.4.1.4, Dispersion des agrégats et pôles, p. 17), constituerait ici aussi un bon indicateur de fixation, en observant non plus l'évolution de la couverture spatiale, mais plutôt la fixation et le renforcement des dynamiques locales de polarisation.

Dans la version 0 de SimFeodal, on peut en premier lieu constater que la tendance – à la croissance – de l'évolution du nombre de paroisses est bonne (figure 3.13). Ainsi, de nouvelles églises paroissiales sont bien créées ou promues régulièrement au cours du temps. La quantité atteinte (240 en moyenne) dépasse un peu trop fortement l'objectif empirique, mais surtout, on remarque dans la figure la très forte variabilité de ces résultats (l'intervalle interquartile est de 100). De plus, cette variabilité augmente fortement avec le temps.

La dynamique est donc plutôt satisfaisante, mais le nombre atteint autant que la variabilité sont améliorables et rendent clairement visible la nécessité d'un ajustement dans le modèle.

Sur le plan spatial, la figure 3.14 laisse bien apparaître la double évolution attendue : les paroisses les plus étendues, en périphérie, demeurent parmi les plus grandes mais voient leur superficie réduite à mesure que de nouvelles églises paroissiales y apparaissent. Dans le même temps, on remarque aussi les effets de fractionnement et de subdivision des paroisses les moins étendues. Cela résulte, dans la figure, en de nombreuses concentrations locales de paroisses que l'échelle des cartes nous permet plus de percevoir que de détailler précisément ¹².

12. La trop forte hétérogénéité dans les superficies des paroisses rend difficile la lecture de

Là aussi, courbe et nombre du nombre de paroisses est incohérent entre chapitre TMD (154) et résultats JIAP (240), + variabilité très forte dans JIAP et pas dans TMD.

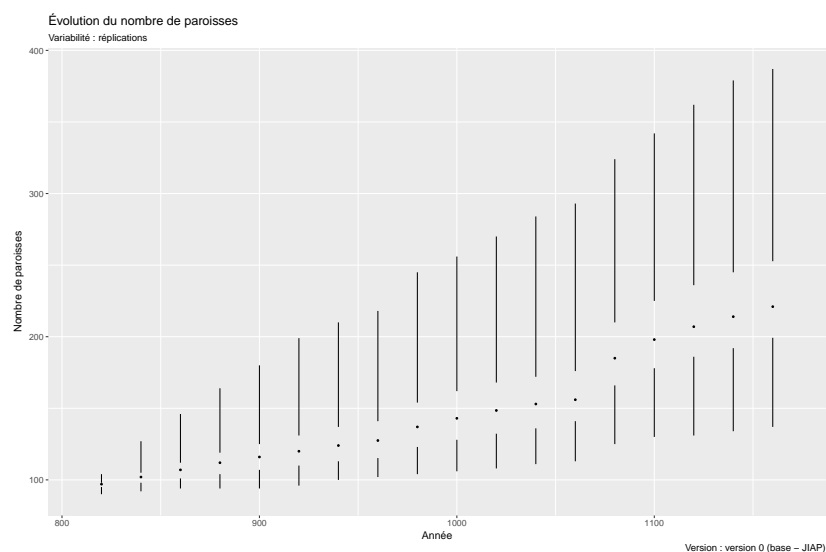


FIGURE 3.13 – Évolution du nombre de paroisses

l'espace. On pourrait y remédier par exemple avec des cartogrammes, mais alors, on ne verrait plus que ces zones de forte densité.

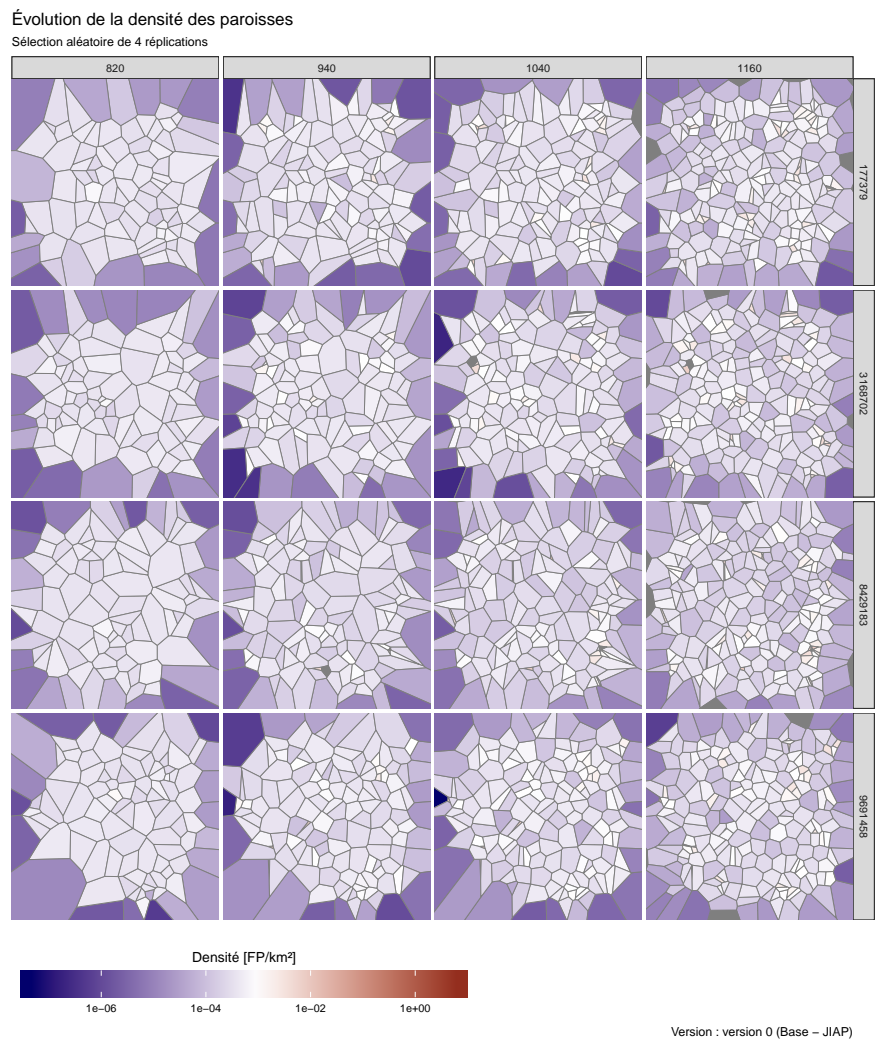


FIGURE 3.14 – Densité des paroisses

3.4.3.4 Efficacité et équité de la desserte paroissiale

On a vu précédemment (3.4.2.3 , Hiérarchie des paroisses, p. 24) que les paroisses devaient assurer conjointement une desserte efficace et équitable. Un indicateur possible pour évaluer l'efficacité du maillage simulé est d'en observer la couverture spatiale. Par définition, les paroisses couvrent ainsi l'ensemble du territoire simulé. Pour autant, certaines paroisses peuvent être très étendues, comme vu dans l'indicateur précédent. On peut toutefois quantifier la dispersion de ces paroisses en analysant leur répartition spatiale, ou plutôt, celle des églises paroissiales qui en constituent le cœur. Pour cela, on peut faire appel à une méthode d'analyse spatiale assez classique, similaire à la méthode des quadrats, en découpant l'espace en un maillage orthogonal régulier et en comptant le nombre d'églises paroissiales de chacune des mailles. On a vu que dans SimFeodal, les églises paroissiales sont réparties de manière à la fois dispersée (zones de faible densité) et concentrées (zones denses à proximité des agrégats importants). Plutôt que de faire un décompte des églises paroissiales et d'en tirer une hiérarchie, un indicateur plus simple est alors de faire un compte des mailles ne contenant pas de paroisse. Ces mailles étant régulières, l'indicateur résultant permettra alors d'appréhender simplement la part de l'espace couvert par des églises paroissiales. On a choisi ici de discrétiser l'espace modélisé (de dimensions 100×100 km) en mailles de 10×10 km, pour un total de 100 mailles de 100km^2 . Ces valeurs permettent d'approcher, à grands traits, la configuration théorique qui verrait chacune des 300 églises paroissiales être réparties dans une maille différente¹³.

Selon les mécanismes de SimFeodal, on s'attend à ce que l'indicateur ainsi produit diminue au cours du temps, à mesure que de nouvelles églises paroissiales viennent desservir le territoire, de manière régulière (comme l'évolution du nombre d'églises paroissiales). Cette diminution devrait toutefois être moins rapide que celle du nombre de paroisses, puisque les églises paroissiales créées au sein des agrégats devraient logiquement ne pas desservir de nouvelles mailles, mais plutôt se situer dans des mailles déjà occupées.

Du point de vue de l'équité de la desserte, on peut recourir à un indicateur plus partiel : il s'agit ici de décrire non pas l'évolution de la desserte des paroisses pour la plupart des foyers paysans, mais plutôt pour ceux d'entre eux qui seraient le plus isolés. Ainsi, on peut observer l'évolution de la proximité entre les foyers paysans et les églises paroissiales, et ce, pour les paysans les plus éloignés desdites églises paroissiales à chaque pas de temps. Une augmentation de l'équité devrait ainsi, par exemple pour les 10% de foyers paysans les moins bien desservis, montrer une diminution de la distance moyenne aux églises paroissiales.

On s'attend ici aussi à voir une augmentation de l'équité, et donc une diminution au cours du temps de cette distance. Empiriquement, on estime à 5km la distance maximale à laquelle un foyer paysan se situait d'une église paroissiale en fin de période (**Voir les seuils de distance dans chapitre 2**). Un paramétrage de SimFeodal sera donc considérée comme satisfaisant, sur ce plan de l'équité de desserte, s'il voit cette distance converger vers le seuil, voire passer en dessous, de 5km.

13. Il faudrait alors 300 mailles de ≈ 6 km de largeur, mais cette valeur ne constitue pas un multiple de 100. De plus, les églises paroissiales étant générées en partie au sein d'agrégats existants pour renforcer la desserte du plus grand nombre, cette valeur est très loin d'être accessible.

Il manque un indicateur d'équité. A compléter une fois qu'il sera mis en place. Par exemple :

Évolution de la distance des 10% de FP les plus éloignés d'une/à une église paroissiale ? ==> Sans doute très intéressant, mais pas « productible » pour l'instant. De toute façon, à voir les résultats JIAP, il faudra sans doute faire retourner toutes les itérations du modèle avec des sorties semblables à celles de la v8 (4_4_A)

==> Peut être sous forme de boxplot ou d'histogramme (ridgeplot) ou de moyenne (plus influencée par grandes distances que médiane, donc mieux ici)

En matière d'efficacité, la version 0 de SimFeodal est assez peu satisfaisante : malgré un nombre d'églises paroissiales trop élevé (voir l'indicateur précédent), la couverture du territoire est assez faible. En effet, la figure 3.15 montre qu'en fin de simulation, près de 90% de mailles restent inoccupées par des églises paroissiales. La tendance est toutefois à la diminution de ces mailles vides, et cette diminution est assez stable et régulière dans le temps.

Comme dans le cas de l'évolution du nombre de paroisses, la variabilité est ici préoccupante et très insatisfaisante : l'écart interquartile est modéré (environ 15%), mais les minimums et maximums présentent un écart énorme (de 95% à 55%).

Pas possible d'analyser les résultats de la version 0 en matière d'équité sans les mesures correspondant. A faire une fois que les graphiques auront été produits.

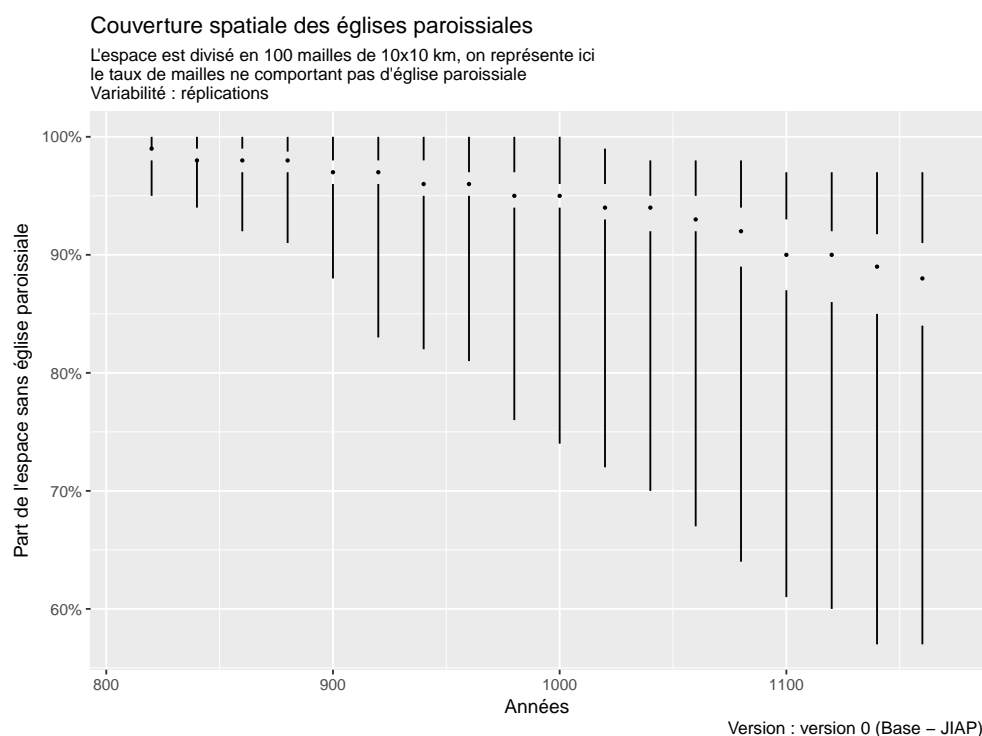


FIGURE 3.15 – Évolution de la couverture du territoire par les églises paroissiales.

Attention, résultats JIAP = pas fiables