

RECONSTITUER LA DEMOGRAPHIE DEPARTEMENTALE FRANCAISE DU XIX^e SIECLE¹

Pour un pays de tradition statistique ancienne, comme la France, les données démographiques couvrant les deux derniers siècles ne manquent pas. C'est plutôt leur fiabilité qui fait problème, et les incohérences apparaissent dès qu'on essaie de tirer parti des contraintes qui relient nécessairement ces données entre elles. Il faut alors corriger, ce qui suppose des hypothèses parfois fortes, et s'accompagne inévitablement du risque d'introduire d'autres incohérences.

Une ré-estimation des populations féminines des départements français (par âge) au cours du XIX^e siècle, à partir des données des divers recensements et de l'état civil, avait été publiée par Etienne Van de Walle en 1974 ; Noël Bonneuil vient d'en réaliser une autre, en utilisant des méthodes statistiques plus puissantes. Maks BANENS procède ici à un examen critique de ces deux reconstitutions, en insistant sur l'importance de pouvoir connaître les conditions locales de production des données démographiques ; quand ces conditions sont connues avec précision, il propose d'utiliser cette information pour guider la reconstitution, et en donne une application au cas d'un département, l'Hérault.*

Depuis le milieu du XIX^e siècle, de nombreuses statistiques démographiques départementales ont été publiées par le Bureau de la Statistique Générale de France². Elles permettent de suivre la structure par âge, sexe et état matrimonial de la population départementale à travers les

* Université de Picardie, Jules Verne.

¹ Une première version du texte se trouve dans Banens (1995), comme introduction méthodologique à la reconstitution démographique du département de l'Hérault et de ses communes. La méthode employée ici n'est pas exactement la même, car elle ne s'appuie que sur des données accessibles à tout chercheur.

² Pour un survol de l'évolution des statistiques, voir, parmi d'autres, Jacques Dupâquier et René Le Mée (1988).

recensements quinquennaux, ainsi que la mortalité, la nuptialité et le solde migratoire par sexe, par âge et par état matrimonial (précédent). Sauf pour les mesures de la fécondité (dont la publication détaillée date des années 1890), l'image démographique départementale semble donc complète dans ces traits fondamentaux.

Cependant, ces statistiques ont la réputation d'être peu fiables : par deux fois déjà, elles ont fait l'objet d'une « reconstitution »³. On devrait plutôt parler d'une correction, car les données qui sont reconstituées existent déjà. Mais le terme reconstitution (ou reconstruction) exprime bien la démarche de la correction : au lieu de corriger certaines pyramides ou statistiques annuelles, supposées erronées, la reconstitution remplace l'ensemble de la surface de Lexis (pyramides, naissances, décès et soldes migratoires par âge), par un nouvel ensemble qui répond aux critères de cohérence, fixés préalablement par l'auteur. Les méthodes s'apparentent à celles de la projection démographique⁴. Nous voulons discuter ici leur application aux populations départementales de la deuxième moitié du XIX^e siècle ; ce faisant, nous espérons apporter quelques réflexions méthodologiques au débat qui a cours autour des reconstitutions de populations.

Les méthodes de reconstitution se divisent en deux familles : les *rétro-projections* et les *projections inversées*. Les premières partent d'une pyramide par âge finale et reconstruisent les pyramides précédentes, une par une, en retranchant les naissances et en rajoutant les décès, répartis par âge à l'aide de tables types. Les deuxièmes partent d'une pyramide hypothétique au départ et reconstituent, une par une, les pyramides suivantes en ajoutant les naissances et en retranchant les décès répartis par âge⁵. Les deux méthodes doivent prendre en compte des migrations en cas de populations ouvertes.

Les partisans des *projections inversées* ont reproché à la *rétro-projection* l'incertitude du dernier groupe d'âges, que l'on fait revivre à partir de 0, et qui se trouve ensuite au départ de la cohorte reconstituée⁶. De leur côté, les défenseurs de la *rétro-projection* reprochent à la *projection inversée* l'incertitude liée à la pyramide choisie au départ⁷. Plus récemment, Noël Bonneuil a démontré qu'en *rétro-projection* l'effectif des derniers groupes d'âges est parfaitement déterminé, à condition de ne pas procéder pas à pas,

³ Etienne Van de Walle (1974), Noël Bonneuil (1997).

⁴ Les principales méthodes de reconstitution démographique sont celles d'Etienne Van de Walle (1974), Ronald D. Lee (1974), E. A. Wrigley et R. S. Schofield (1981), Jim Oeppen (1992, 1993) et Noël Bonneuil (1997).

⁵ L'appellation « projection inversée », donnée par Ronald D. Lee, est à prendre non pas dans le sens chronologique, mais logique ; elle fait référence à l'estimation des pyramides par âge à partir de totaux (naissances, décès et population totale), à l'inverse de la démarche de la projection habituelle qui consiste à estimer les totaux à partir des taux et pyramides par âge.

⁶ Voir notamment Ronald D. Lee (1985).

⁷ Kenneth W. Wachter (1986) a ensuite démontré l'influence décroissante de cette population de départ.

pyramide par pyramide, mais de résoudre l'ensemble d'équations déterminant la surface de Lexis de la reconstitution en même temps⁸ : dans ce cas, on ne peut plus parler d'une reconstitution en avant ou en arrière, on reconstitue les pyramides de l'ensemble de la période en même temps. C'est également la solution que propose Jim Oeppen en 1992⁹. On définit les hypothèses nécessaires à l'estimation de l'évolution de la population : naissances vivantes, décès et populations totales, réseau de tables types de mortalité, profil de migration, pyramides finale, intermédiaires ou de départ, etc. Puis on laisse l'ordinateur - par itérations - chercher la solution qui respecte toutes les contraintes.

Si on limite les hypothèses à une pyramide de départ, aux séries de décès, naissances et populations totales, à une famille de tables types et à un profil de migration, le calcul se réduit à une simple projection en avant (*projection inversée* selon la terminologie de Lee). La solution sera unique ; toute hypothèse supplémentaire concernant les pyramides (pyramide finale ou pyramides intermédiaires) conduirait à l'absence de solutions. Pourtant, dans le cas de la reconstitution du passé, on est le plus souvent amené à respecter des hypothèses supplémentaires concernant les pyramides intermédiaires ou finale : pour trouver la « meilleure » solution, on sera obligé d'introduire des zones de tolérance. Par exemple, autour des pyramides intermédiaires et finale, au lieu d'exiger la reconstitution exacte des pyramides, on tentera - par itérations - de s'en rapprocher le plus possible. Mais on peut aussi « fixer » les pyramides et laisser une certaine dispersion autour du profil de migration ; on cherchera alors à minimiser la dispersion autour du profil. Ou encore, on accepte la possibilité d'un sous-enregistrement des naissances ou des décès, d'incohérences dans la série des population totales, d'une dispersion systématique (mauvais choix de tables types) ou ponctuelle (crises de mortalité) autour de la mortalité, etc.

Théoriquement, cette méthode est certainement la plus satisfaisante. Dans la pratique, elle se heurte à la complexité des hypothèses : laisser l'ordinateur trouver la « meilleure » reconstitution suppose que l'on introduise au préalable l'ensemble des hypothèses de base, ainsi que l'ensemble de leurs dispersions possibles et la hiérarchie dans laquelle celles-ci doivent intervenir. Cela est peu réaliste. Dans la suite de cet article, nous défendons l'emploi d'une méthode semi-automatique qui offre l'avantage de pouvoir maîtriser les différents paramètres et de garder un contrôle visuel de l'itération en cours. On pourra intervenir en fonction de données objectives, mais aussi d'une certaine connaissance du terrain, qui permet de privilégier telle correction plutôt que telle autre, selon des critères qui ne seraient pas formalisables au préalable.

Auparavant, nous rappelons brièvement les méthodes de Van de Walle et de Bonneuil ; le département de l'Hérault servira d'exemple.

⁸ Noël Bonneuil (1992).

⁹ Jim Oeppen (1992 et 1993) revient sur le débat suscité par *The Population History of England 1541 - 1871 : a Reconstruction* de E. A. Wrigley et R. S. Schofield (1981), ouvrage pour lequel il avait conçu la méthode de reconstitution.

I.- Projection inversée et rétro-projection

La projection inversée : La méthode de Van de Walle s'appuie sur les séries de naissances vivantes, décès et populations totales de sexe féminin. Les

méthode de Van de Walle étapes de la reconstitution sont les suivantes :

a) Rapportant la population totale « naturelle » (calculée sur le seul mouvement naturel) à celle recensée, Van de Walle met en évidence l'apport migratoire net sur l'ensemble du XIX^e siècle. Il en dégage, par régression linéaire, une tendance régulière applicable à l'ensemble ou à une partie du siècle ;

b) Il choisit une répartition par âge en 1801, qui correspond à celle d'une population stable ayant connu, au cours de la période précédente, les taux de natalité et de mortalité constatés en 1801-1806. La pyramide est choisie parmi les populations stables des familles Ouest ou Nord de Coale et Demeny¹⁰ ;

c) S'appuyant sur la même famille de tables types de Coale et Demeny, il calcule les pyramides aux dates de recensement de telle sorte que le nombre de décès, produit par la projection, égale le nombre de décès observé. La migration intercensitaire est répartie par âge selon un calendrier observé à Paris en 1860 et 1890¹¹ ;

d) Pour s'approcher au plus près des recensements jugés les plus fiables (1861-1866 et 1891-1906), il ajuste la population stable choisie pour 1801, le nombre de naissances, la famille de tables de mortalité et le calendrier de migration. Comme il le démontre, le choix de la population en 1801 n'a plus guère d'incidence après 1830. Le calendrier de migration n'en a pas davantage là où la migration intercensitaire a été remplacée par une tendance régulière. L'ajustement se fait donc surtout par le nombre de naissances et par le choix de la famille de tables types pour la mortalité.

La méthode de Van de Walle est donc une méthode mixte. Son mode de calcul par projection en avant d'une pyramide choisie au départ en fait une *projection inversée* ; mais la reconstitution est surdéterminée, puisqu'il utilise

¹⁰ Ansley J. Coale et P. Demeny (1966).

¹¹ L'algorithme utilisé est le suivant ; avec $D, P', P, P_{-5}, L, PT, M, m, a$ et t représentant dans l'ordre : Décès, Population projetée, Population définitive, Naissances, Années vécues de la table, Population Totale, Migration nette, part dans la Migration nette, âge et temps :

$$\begin{aligned}
 PT_t &= \sum_5 P_{a,t} \\
 P'_{a+5,t+5} &= P_{a,t} (L_{a+5}/L_a) \\
 D_{t,t+5} &= \sum P_{a,t} (1-L_{a+5}/L_a) \quad (a = -5, 0, 5, \dots, 75) \\
 PT'_{t+5} &= \sum P'_{a,t+5} \\
 M_{t,t+5} &= PT'_{t+5} - PT_t \\
 P_{a,t} &= P'_{a,t} + m_a M_{t,t+5} \quad (\sum m_a = 1)
 \end{aligned}$$

des pyramides intermédiaires et finales. Pour obtenir la meilleure solution, il doit définir des marges de tolérance et procéder par itérations. Les marges de tolérance concernent le sens des écarts pour les naissances (seul le sous-enregistrement est accepté), la variation du profil de migration (le profil peut varier entre celui constaté en 1861-1866 et 1896-1901 à Paris, et une migration répartie de façon égale sur tous les groupes d'âges), la variation de la population stable en 1801, et, bien sûr, la dispersion autour des pyramides intermédiaires et finales. C'est la minimisation de cette dernière dispersion qui constitue l'objectif des itérations.

***La rétro-projection :
méthode de Bonneuil***

Par méthode de Bonneuil, on entend celle que Noël Bonneuil, dans son livre récent (1997), a appliquée à la période 1856-1906 ; pour la reconstitution de la première moitié du XIX^{ème} siècle, il a employé une méthode proche de celle de Van de Walle.

Noël Bonneuil s'appuie exclusivement sur la série annuelle des décès féminins par âge et celle, quinquennale, des femmes recensées par groupes d'âges. A la différence de Van de Walle, il n'utilise pas les totaux de population, de naissances vivantes et de décès. Les étapes sont les suivantes :

a) Il relève les erreurs de totalisation dans les données publiées. Il les corrige et en déduit un classement des recensements selon leur qualité supposée ;

b) Pour écarter les variations improbables d'un recensement à l'autre, il lisse les cohortes qui traversent les recensements. Aucun modèle démographique n'intervient de façon explicite¹² ; le modèle d'ajustement est mathématique, mais l'auteur intervient manuellement selon « l'évolution plausible » de l'ajustement obtenu. Le classement qualitatif des recensements, obtenu en (a), permet de pondérer les points à lisser. L'ajustement est ainsi fixé (en valeur et en tendance) par l'extrémité 1901-1906 (crédité d'un poids de 100000), puis guidé par les autres recensements, parmi lesquels 1861 et 1866 (aux poids de 10) jouent un rôle plus important que les autres (aux poids de 1) ;

c) Sur la base des cohortes lissées et de la série annuelle des décès par âge, préalablement répartis sur les cohortes, il calcule les tables annuelles de mortalité et de migration nette par âge. La migration nette est attribuée aux départements destinataires, dans les proportions déduites des résidents du département en question, recensés, en 1901, dans les autres départements français. Cette attribution s'effectue pour tous les départements en même temps, pour pouvoir satisfaire, pour chaque groupe d'âges et à chaque période inter-censitaire, aux trois contraintes imposées par : la migration nette de chaque département (estimée par les décès enregistrés et par les cohortes

¹² Noël Bonneuil (1997) p. 40. La méthode (BSVLS de l'IMSL Library) est un lissage B-spline dont l'utilisateur détermine l'ordre polynomial et le nombre de noeuds internes. Voir la figure 1, où nous montrons 2 ajustements BSVLS de la cohorte née en 1846-1850.

lissées, donc sans rapport avec les autres départements), les profils d'échanges migratoires interdépartementaux (empruntés au recensement de 1901), et la somme des soldes migratoires départementaux (correspondant au solde migratoire national, supposé nul en l'occurrence).

L'estimation simultanée des migrations nettes interdépartementales à chaque âge est certainement l'élément le plus innovant de la méthode de Bonneuil ;

d) Les tables de mortalité, calculées sous (c), sont ensuite remplacées par les tables types de Ledermann, qui, aux âges de 5 à 65 ans, sont les plus proches des tables observées¹³. Cela permet de ne pas utiliser la mortalité enregistrée avant l'âge de 5 ans, supposée moins fiable ;

e) La reconstitution des effectifs par âge peut s'effectuer maintenant par le rajeunissement de la population observée en 1906 (supposée correcte), à l'aide de la série de quotients de migration nette et de la série de quotients de mortalité Ledermann. L'effectif des derniers groupes d'âges (pour les cohortes éteintes avant 1906) est emprunté aux pyramides lissées ;

f) Les naissances sont ré-estimées par la *rétro-projection*, comme n'importe quel autre groupe d'âges.

La méthode de Noël Bonneuil est donc un exemple de reconstitution par *rétro-projection*. La reconstitution est déterminée par la pyramide finale, les derniers groupes d'âges, les quotients de mortalité et de migration retenus. Elle n'est pas surdéterminée, la solution est donc unique.

Quoiqu'en totale opposition, quant au mode de calcul et l'utilisation des sources, les méthodes de Van de Walle et de Bonneuil partagent l'objectif de rejoindre le recensement de 1906. Bonneuil fixe le lissage des cohortes en 1906 et fait partir la *rétro-projection* de là ; pour Van de Walle, le recensement de 1906 est l'objectif à atteindre par l'ajustement du nombre de naissances et du calendrier de migration. Si les données publiées ne sont pas perturbées par des crises de mortalité, par des migrations fluctuantes ou par des statistiques erronées - et c'est apparemment le cas pour de nombreux départements après 1856 - les résultats obtenus par les deux méthodes seront voisins. Par contre, là où les données sont perturbées, les deux méthodes donnent des résultats très divergents : c'est probablement le cas pour l'Hérault, car les statistiques de l'Hérault du XIX^e siècle sont connues pour être particulièrement perturbées. L'Hérault constitue, par conséquent, un cas particulièrement intéressant pour une étude comparative des deux méthodes¹⁴.

¹³ Bonneuil travaille avec des quotients prospectifs (1997, pp. 63-64). Les quotients considérés sont ceux de ${}_5q'_5$ à ${}_5q'_{60}$.

¹⁴ Une comparaison complète des résultats de Noël Bonneuil et d'Etienne Van de Walle est malheureusement impossible, car Noël Bonneuil n'a pas publié les mêmes résultats que Van de Walle, et n'a pas souhaité les communiquer pour les besoins de cet article.

Critique de la méthode de Van de Walle

Analysons les deux méthodes, étape par étape, en commençant par celle d'Etienne Van de Walle.

Soldes migratoires : tendance linéaire, fluctuations conjoncturelles

En remplaçant les soldes migratoires intercensitaires, tels qu'ils ressortent des données brutes, par une tendance séculaire, Van de Walle renonce à la conjoncture migratoire réelle. Il a de sérieux arguments pour le faire :

- le solde migratoire brut est soumis aux modifications de techniques de recensements, de définitions employées, etc ;
- ensuite, l'application d'un profil plus ou moins rigide de migration impose des flux plus ou moins constants, car la fluctuation du flux migratoire concentrerait celui-ci en quelques cohortes, laissant des cohortes voisines vides ou même de flux opposé.

Or, il est vraisemblable, nous le démontrerons plus loin, que les cohortes voisines ont connu des intensités migratoires finales voisines : les cohortes réalisent, à travers la conjoncture migratoire, par le jeu des retards et rattrapages, des intensités migratoires finales en évolution régulière, ce qu'un profil transversal rigide ne pourra jamais restituer sans lissage préalable des soldes migratoires.

Van de Walle a donc de bonnes raisons pour lisser le flux migratoire. Mais, ce faisant, sa reconstitution est également, en quelque sorte, lissée, et ne permet plus l'étude de la conjoncture migratoire.

Les recensements de 1851-1866 : utilisation et qualité

Van de Walle choisit la population stable de 1801 en se rapprochant des effectifs des plus de 50 ans, recensés en 1851-1866. Ces recensements occupent ainsi une place importante dans la reconstitution. Or, ce sont également les premiers recensements dont la récapitulation par âge a été effectuée et l'on pourrait douter de leur qualité. De fait, dans l'Hérault ce sont les recensements parmi les plus erronés : ils ont amené Van de Walle à proposer une reconstitution, qu'il juge lui-même « insatisfaisante »¹⁵. On aurait préféré un test préalable des pyramides de 1851 à 1866, indépendant de la reconstitution. Un tel test est facile à réaliser par l'analyse de l'attrait des âges ronds. Car les recensements de 1851-1866 sont également les seuls au XIX^e siècle, pour lesquels nous disposons de la répartition par année d'âge et non par groupe d'âges. Ce test sera développé plus loin.

¹⁵ Van de Walle (1974, p. 215). Ce même choix l'a conduit à corriger le nombre de naissances au début du XIX^e siècle de + 30 %, et à supposer une chute extrêmement rapide entre 1800 et 1820. Il répond ainsi à Le Roy Ladurie (1965), qui avait placé la chute de la natalité 20 ans plus tôt.

La fiabilité du nombre total de décès Le nombre de naissances est l'un des paramètres utilisés pour ajuster les pyramides calculées à celles jugées fiables ; l'éventuel sous-enregistrement des naissances peut ainsi être corrigé. Le nombre de décès, par contre, ne peut pas être corrigé de la même façon, puisque corriger les deux totaux ouvrirait l'éventail de solutions de façon inconsidérée ; il doit donc être sans erreurs.

Le choix des tables types de mortalité Dans l'ajustement final joue également le choix du modèle de mortalité : Nord, Sud, Est ou Ouest des tables types de Coale et Demeny. Le débat sur les reconstitutions ne fait pas toujours une grande place au choix des tables types : elles sont implicitement considérées comme proches les unes des autres. Or, les réseaux de tables types sont assez éloignés les uns des autres, notamment quand l'espérance de vie est faible et le choix peut avoir des conséquences importantes.

Critique de la méthode de Bonneuil

Le classement par qualité des recensements Noël Bonneuil établit un classement des recensements de 1851 à 1901, qui détermine la pondération du lissage des cohortes. Or, ce classement est réalisé par une vérification comptable au niveau national : il ne concerne pas les relevés réalisés aux niveaux de la commune, de la sous-préfecture et de la préfecture et ne prend pas non plus en compte les erreurs liées à la déclaration, à l'organisation du recensement, aux changements de définitions, etc. Comme Bonneuil le met en évidence, le classement des recensements reflète donc bien la vie interne du bureau national, mais ce classement national n'a pas de signification pour les relevés réalisés dans les départements.

Le lissage préalable des cohortes Le lissage des cohortes comporte une grande part d'aléatoire. L'exemple héraultais peut illustrer ce défaut : l'évolution des femmes nées en 1846-1850, ayant 0-4 ans au recensement de 1851 et 55-59 ans au recensement de 1906. La figure 1 montre les effectifs recensés, deux lissages selon la méthode mathématique BSVLS, utilisée par Bonneuil, et un lissage sur un modèle démographique construit sur la table type Sud de Coale et Demeny (espérance de vie de 45 ans à la naissance) et le profil migratoire selon Van de Walle. Le modèle est fixé en 1906 et minimise la distance avec les effectifs recensés, en variant l'intensité migratoire de la cohorte.

Figure 1. - Femmes héraultaises nées en 1846-1850 : effectifs publiés (SGF), lissages mathématiques (BSVLS 1 et 2) et démographique (voir texte). Pondération des recensements : 100 000 pour 1901 et 1906, 10 pour 1861 et 1866, 1 pour les autres.

Les deux lissages mathématiques correspondent à des B-splines de quatrième ordre, avec respectivement 0 et 2 noeuds internes. Il ne s'agit pas de juger lequel des trois lissages est le plus vraisemblable, mais de constater que les écarts entre les résultats peuvent atteindre 20 % de l'effectif initial.

Le lissage des cohortes et l'estimation des migrations

Cette part d'incertitude influence faiblement le calcul des quotients de mortalité, mais fortement celui des quotients de migration. Le tableau 1 compare les quotients quinquennaux perspectifs pour la cohorte née en 1846-1850, calculés sur la base des décès enregistrés et des effectifs lissés selon les méthodes de la figure 1¹⁶.

Le tableau 1 montre que le choix du modèle de lissage n'a qu'une incidence modérée sur l'estimation des quotients de mortalité, il est déterminant, au contraire, pour l'estimation des quotients de migration.

¹⁶ Le lissage des effectifs selon le modèle démographique alternatif comporte déjà un calendrier de migration : le calcul des quotients est donc biaisé. Le lissage mathématique peut sembler plus neutre, mais la différence n'est qu'apparente. Chaque modèle mathématique implique un calendrier démographique ; chaque intervention manuelle, au nom de la plausibilité démographique, fait référence, de façon plus ou moins consciente, à un calendrier démographique.

Age	Quotients de mortalité			Quotients de migration		
	lissage BSVLS 1 (= 100)	lissage BSVLS 2	lissage démogr.	Lissage BSVLS 1 (= 100)	lissage BSVLS 2	lissage démogr.
0-4/5-9	100	97	81	100	38	28
5-9/10-14	100	107	94	100	109	28
10-14/15-19	100	102	101	100	181	66
15-19/20-24	100	95	101	100	145	124
20-24/25-29	100	94	100	100	75	106
25-29/30-34	100	98	101	100	4	59
30-34/35-39	100	104	104	100	-25	43
35-39/40-44	100	108	106	100	60	32
40-44/45-49	100	106	107	100	-3165	-504
45-49/50-54	100	102	105	100	-87	-5
50-54/55-59	100	100	102	100	94	17

Tableau 1. - Quotients de mortalité et de migration des femmes héraultaises nées en 1846-1850, calculés sur la base des décès enregistrés et des effectifs recensés lissés, selon les différents modèles de lissage (voir figure 1).

Le problème est fondamental. Si le lissage des cohortes peut améliorer l'estimation des effectifs, s'il peut donc améliorer l'estimation de la mortalité et permettre une vérification approximative de celle-ci par la méthode des table types, il ne permet pas l'estimation de la migration. Et il n'y a aucune raison pour que les contraintes supplémentaires, résultant de l'attribution aux départements destinataires, arrivent à corriger cette première estimation fortement aléatoire.

Le lissage des quotients de mortalité

Le lissage (et l'extrapolation) des quotients de mortalité, sur le modèle des tables types de Ledermann, appelle la même critique que le lissage de la courbe de survie des cohortes : premièrement, il introduit un biais systématique si le type de mortalité du département n'est pas celui de Ledermann ; deuxièmement, le lissage n'est pas fait pour vérifier et corriger les quotients observés, mais pour les reproduire « ajustés », qu'ils soient exacts ou non. L'introduction du biais systématique n'est pas propre à la méthode de Bonneuil. Elle existe dans toute reconstitution recourant aux tables types¹⁷. La reproduction ajustée des quotients observés, par contre, est spécifique à la méthode de Bonneuil.

¹⁷ Méthodologiquement, le recours aux tables types relève d'un paradoxe : on tente de corriger des tables de mortalité avec des tables types qui ont été établies auparavant sur la base des mêmes tables non corrigées. En effet, Ledermann a construit ses tables types à partir de 154 tables, dont 3 proviennent de la France du XIXe siècle.

Prenons l'exemple héraultais de l'année 1859. Les quotients par âge, tels qu'ils résultent des décès publiés, semblent erronés (figure 2) : la mortalité de 0 à 5 ans est légèrement plus élevée que durant les périodes antérieures et postérieures, tandis que celle après 30 ans est beaucoup plus faible. S'agit-il d'un effet récupérateur après l'épidémie de choléra de 1854 ? Des décès aux âges plus élevés sont-ils par erreur comptés parmi les enfants ? Manque-t-il des décès ? Est-ce une variation normale de la mortalité ? A priori, on n'en sait rien. La figure 2 montre la solution de la méthode Bonneuil. L'ajustement Ledermann aux âges 5-69 ans reproduit, à ces âges, globalement le même nombre de décès (appliqué aux mêmes pyramides) : 1 836 au lieu des 1 914 enregistrés. Aux âges de 0 à 5 ans, par contre, il n'en reproduit que 931 au lieu des 2 470 enregistrés : le total des décès passe ainsi de 4 384 à 2 767.

Figure 2. - Mortalité des femmes héraultaises en 1859 ;
quotients observés et lissés sur le modèle Ledermann à 5-69 ans.

L'originalité de la méthode de Bonneuil est de s'appuyer sur une partie des décès (plus exactement sur une partie des quotients ; ceux entre 5 et 70 ans) et de laisser varier librement le total des décès. Ici, elle conduit à réduire le total des décès de 37 %, le total des décès avant 5 ans de 62 %, alors que le nombre de décès en 1859 est déjà faible par rapport à ceux des années environnantes. La solution paraît donc peu réaliste.

D'une façon générale, on peut dire que pour toute perturbation de la mortalité par âge, qui n'est pas répartie normalement autour des quotients réels, le lissage Ledermann conduit à une correction improbable. De surcroît, pour que la méthode n'introduise pas de biais systématique, les quotients « réels » doivent correspondre à la mortalité type de Ledermann.

Coale et Demeny ont construit les tables-types Ouest à partir de 129 tables, dont la France est le premier fournisseur avec 16 tables...

Le rôle capital du lissage préalable Les données utilisées dans la reconstitution sont les suivantes : la pyramide finale de 1906, qui est le point final de chaque cohorte lissée, et le point de départ de la *rétro-projection* ; les quotients de migration (étape 3) ; les quotients de mortalité Ledermann (étapes 3 et 4) ; et les effectifs des derniers groupes d'âge des cohortes éteintes avant 1906, estimés lors du lissage des cohortes (étape 1).

Toutes ces données s'appuient sur les cohortes lissées, avant de subir d'autres traitements (rééquilibrage des migrations interdépartementales, lissage des tables de mortalité) : on doit donc s'attendre à des cohortes reconstituées proches des cohortes lissées¹⁸.

La ré-estimation des naissances Dans la reconstitution de Bonneuil, le total des naissances vivantes peut varier aussi librement que le total des décès. Il résulte de la reproduction - ou non - du sous-enregistrement des 0-4 ans au lissage, du sens des possibles perturbations des décès enregistrés et du biais systématique introduit par le modèle Ledermann. Dans la mesure où le lissage des cohortes peut altérer l'estimation des 0-4 ans en augmentant comme en diminuant, l'estimation des naissances, fortement dépendante de l'effectif à 0-4 ans dans une rétro-projection, ne peut qu'en subir les conséquences. La méthode de Bonneuil va ainsi à l'encontre de l'idée générale, selon laquelle l'enregistrement des naissances à l'état civil, fin XIX^e siècle, fut plus fiable que le recensement des 0-4 ans.

¹⁸ Noël Bonneuil est conscient du rôle joué par le lissage des cohortes. Dans une section nommée « Discussion », il commente : « One could object, and rightly so, that the reconstruction depends on the smoothing procedure employed. But the point is to remain as close as possible to the original data, especially when censuses are known to be of good quality. Moreover, the reconstruction of censuses for the sake of having exact numbers of people is not really at stake. What matters is the reconstruction of demographic behaviour, with its full age, space, and period variability. » (Bonneuil 1997, p. 72) [« On pourrait faire remarquer, et avec raison, que la reconstitution est dépendante du procédé de lissage qu'on emploie. Mais l'objectif est de rester au plus près des données originales, surtout quand les recensements sont connus pour leur bonne qualité. En outre, l'enjeu n'est pas la reconstitution des recensements dans le but d'avoir les nombres exactes de personnes. Ce qui importe, c'est la reconstitution du comportement démographique, dans toute sa variabilité d'âge, d'espace et de périodicité. » Traduction M.B.)

Même si l'on suivait Noël Bonneuil dans son jugement des recensements, qui seraient « connus pour leur bonne qualité », le lissage des cohortes ne permettrait pas une estimation valable de la migration. On ne peut prétendre reconstituer le comportement migratoire par âge et par période, sans reconstituer en même temps, et avec beaucoup plus de précision, le « nombre exacte » des personnes recensées. Car toute incertitude du « nombre exacte » recensé se traduit mécaniquement par une incertitude multipliée du comportement migratoire.

II. - Proposition d'une méthode alternative

Les différences entre les méthodes de Van de Walle et de Bonneuil résultent, en dernier lieu, d'une appréciation différente des sources disponibles. Notre objectif est de concevoir une méthode flexible, qui utilise toutes les données disponibles, à la fois celles qui profilent normalement la reconstitution, et les informations supplémentaires permettant d'apprécier, voire de corriger, les données de base.

On rappelle que les effectifs totaux de naissances, de décès et de population, complétés d'un réseau de tables types, d'un profil de migration et d'une pyramide (de départ ou d'arrivée), suffisent à la reconstitution : toute information supplémentaire la sur-détermine. La méthode à concevoir consiste donc à analyser, en bon ordre, les données disponibles et à hiérarchiser leur utilisation. Ensuite, la reconstitution se fera pas à pas, du plus sûr au moins sûr¹⁹.

¹⁹ Comme il est dit dans l'introduction de cet article, on pourrait résoudre une reconstitution surdéterminée en une seule fois (comme Oeppen, 1991) en définissant autour de chaque déterminant (table type, recensement intermédiaire ou final, totaux de population, naissances et décès, profil de migration, etc.) une zone de tolérance, puis en cherchant la solution qui minimise la somme de tolérance concédée.

Quelques exemples suffisent pour montrer à quelles difficultés on se heurterait : 1) un des recensements contient une ou plusieurs erreurs importantes. A vouloir respecter la zone de tolérance, la reconstitution devra « corriger » les autres variables, diffusant ainsi l'erreur originale. Au lieu de prendre en compte le recensement déviant, on aurait dû l'exclure de l'observation. Bien sûr, on peut prévoir l'exclusion automatique du recensement le plus déviant, mais on ne sait pas, *a priori*, s'il y aura un, deux ou aucun recensement réellement déviant ; 2) comme on a vu pour l'année 1859, la reconstitution doit toujours choisir ou bien de respecter le nombre total de décès, ou bien de respecter l'enregistrement des décès par âge. Si nous définissons une zone de tolérance unilatérale autour du nombre total de décès, estimant qu'il ne peut être surestimé, que se passerait-il en 1859 ? La reconstitution ne corrigerait rien, car chaque correction des décès par âge entraînerait une diminution du nombre total de décès. Pour d'autres années, le problème serait inverse : une erreur sur l'enregistrement par âge obligerait la reconstitution à augmenter considérablement le nombre total de décès. Automatiquement, la reconstitution chercherait la solution intermédiaire, celle qui tire la table de mortalité un peu dans le sens de la table type, sans trop modifier le nombre total de décès. On se retrouve dans la situation du recensement déviant : la reconstitution est incapable de détecter s'il s'agit d'une erreur sur une des variables ; elle l'attribue toujours à toutes les variables.

Les deux exemples montrent que la simple définition d'une zone de tolérance autour des variables ne suffit pas, car l'incertitude des données ne provient pas d'une imprécision aléatoire des mesures ou d'une autre source aléatoire. Elle provient, d'une part, d'éventuels biais systématiques, comme le sous-enregistrement aux jeunes âges, le choix des tables types, celui du profil de migration, etc. ; d'autre part, elle trouve son origine dans des perturbations « naturelles », comme les crises de mortalité ou les fluctuations migratoires ; et enfin, on y trouve l'effet des erreurs administratives comme

Les erreurs principales

On distinguera deux types d'erreurs : erreurs de déclaration et erreurs administratives.

Pour les populations qui nous concernent, les premières sont de deux types : l'attrait des âges (ronds) et le sous-enregistrement aux jeunes âges.

Erreurs de déclaration

L'attrait des âges ronds a été mis en évidence par Tabah (1947) pour la population française au recensement de 1851. Tabah a proposé une correction dont nous ne commentons pas ici les étapes. Exprimons l'attrait A de l'âge i par la formule :

où p_i est l'effectif recensé à l'âge i et p_a l'effectif du groupe de 5 ans d'âge dont i est l'année centrale. La représentation graphique de l'attrait des âges ronds des femmes françaises au recensement de 1851 (figure 3) montre la grande régularité du phénomène : surenregistrement de 20 à 40 % aux âges finissant en 0 ; de 20 à 30 % aux âges finissant en 5 ; sous-enregistrement de 10 à 20 % aux âges finissant en 1 et en 9. Globalement, le surenregistrement aux âges finissant en 0 résulte pour moitié du vieillissement des âges finissant en 9 et pour moitié du rajeunissement des âges finissant en 1. Le surenregistrement aux âges finissant en 5 est plutôt dû au sous-enregistrement des âges finissant en 3 et en 7 ; mais là aussi, vieillissement et rajeunissement semblent s'équilibrer.

Le regroupement en classes d'âges quinquennales devrait annuler la quasi-totalité de l'erreur de déclaration, laissant un très léger avantage au premier groupe quinquennal de chaque décennie, car ce qu'il gagne sur les âges finissant en 9 excède ce qu'il perd sur les âges finissant en 4. La figure 4 montre que cet effet est invisible : la pyramide corrigée par Tabah - réalisée par année d'âge, puis regroupée en groupes d'âges - est quasiment identique à la pyramide originale. La correction moyenne par groupe d'âges après 20 ans n'atteint pas 0,5 %.

les erreurs d'écriture, de calcul etc., qui sont à exclure de l'observation plutôt que d'être pris en compte.

Figure 3. - Attrait des âges des femmes françaises en 1851 ; source : SGF.

Figure 4. - Effectifs recensés (SGF) et corrigés (Tabah) des femmes françaises en 1851 par groupe d'âges.

L'essentiel de la correction se situe à 45-54 ans et n'a pas de rapport avec l'attrait de l'âge à 50 ans ; cette correction vient de la ré-estimation par Tabah de l'histoire démographique des générations, nées durant la période perturbée de 1796-1805. On peut donc conclure que l'attrait des âges ronds, même lorsque celui-ci est important, comme au recensement de 1851, disparaît dans le regroupement des âges. Or, les sources sur lesquelles nous travaillons sont déjà regroupées : cette première erreur n'exige donc pas de correction.

Le sous-enregistrement des jeunes âges aux recensements a été constaté par Bourgeois-Pichat (1951), Tabah (1947), Van de Walle (1974) et d'autres. On doit donc s'y attendre, sans en connaître l'ampleur, ni le rythme de

sa disparition. Il est à rapprocher, d'ailleurs, du sous-enregistrement des naissances vivantes, des vrais et faux mort-nés et de la mortalité infantile (et donc du nombre total des décès). Ces différentes formes de sous-enregistrement ne sont pas sans rapport entre elles ; un sous-enregistrement des naissances vivantes ne peut se concevoir sans sous-enregistrement de la mortalité infantile, car déclarer le décès d'un enfant non déclaré à la naissance n'est possible que dans des circonstances exceptionnelles. Par conséquent, l'enregistrement complet de la mortalité infantile garantit l'enregistrement complet des naissances vivantes. On peut tenir le même raisonnement pour les couples mortalité infantile/mortinatalité ; naissances vivantes/mort-nés et décès/mort-nés. Dans la hiérarchie des sous-enregistrements, celui des 0-4 ans aux recensements et celui des mort-nés disparaissent en dernier ; celui des naissances vivantes, des décès et de la mortalité infantile disparaissent avant, généralement ensemble, car il s'agit des mêmes enfants.

Erreurs administratives

Les erreurs administratives sont indépendantes de la déclaration. Elles se produisent lors du traitement administratif des données déclarées à l'état civil ou à l'agent de recensement. Dans l'Hérault, on a rencontré trois types d'erreurs administratives : les erreurs de calcul et d'écriture, les reprises de données antérieures, et les données inventées.

- *Les erreurs de calcul et d'écriture* sont difficiles à déceler quand il s'agit d'unités ou de dizaines, mais elles deviennent évidentes quand il s'agit de centaines. Les premières ne gênent pas beaucoup, étant de faible ampleur et, a priori, aléatoires ; les dernières seront corrigées par estimation sur la base des années environnantes.

- *Les reprises de données antérieures* sont apparentes quand, pour des raisons qui restent inconnues, la commune ou la sous-préfecture ne fait pas le travail de récapitulation et reprend les mêmes données plusieurs années de suite, avec parfois des inversions, les garçons remplaçant les filles par exemple. Ces erreurs, évidentes au niveau où la manipulation s'effectue, n'apparaissent pas toujours clairement dans les chiffres départementaux. Dans le doute, on garde les données suspectes, considérant qu'elles constituent de toute façon une estimation non aberrante des vraies valeurs.

- *Les données peuvent aussi être inventées.* En général, l'auteur a pris soin de reproduire le total des décès, par sexe et par état matrimonial, puisque celui-ci a déjà fait l'objet d'une communication à la préfecture. Heureusement, les nombres inventés ont le plus souvent peu de vraisemblance démographique ; mais ici encore, ils peuvent être noyés parmi des valeurs réelles et ainsi disparaître.

Les erreurs administratives se distinguent des erreurs de déclaration par leur caractère discontinu : elles apparaissent et disparaissent brusquement, tandis que les erreurs de déclaration évoluent progressivement. Dans les deux cas, la correction par lissage n'est pas adéquate, puisque les erreurs ne sont pas distribuées normalement autour des vraies valeurs. Si les erreurs de déclaration

peuvent être ré-estimées, à condition de les avoir repérées avec certitude, les erreurs administratives sont à écarter de l'observation et à remplacer par des estimations *ad hoc*.

Le repérage des erreurs

Les erreurs administratives peuvent être présentes dans toutes les séries de données ; elles se repèrent par leur caractère discontinu et apparemment incohérent. On commence le repérage dans les séries les plus sûres et les plus régulières : les totaux de la population par état matrimonial, de naissances vivantes et de décès.

Figure 5. - Naissances vivantes et décès annuels, femmes de l'Hérault 1851 - 1910.

Dans la figure 5, la série des décès est dominée par un pic tout à fait exceptionnel en 1854, l'année du choléra : un excédent de 60 % semble élevé, mais ce n'est pas incohérent. Moins excessif, mais plus suspect, est le creux en 1862, pour les naissances et pour les décès féminins, tandis que, cette même année, les naissances et décès masculins sont plutôt élevés. Par contre, la chute des naissances en 1871 est naturellement liée aux événements politiques.

Figure 6. - Nombre de femmes de l'Hérault par état matrimonial aux recensements 1851 - 1911.

La série des totaux par état matrimonial (figure 6) montre un recensement incohérent en 1886 : le nombre de célibataires est anormalement élevé, celui des femmes mariées anormalement faible. Les statistiques des mariages des années précédentes (non représentées ici) ne montrent aucune irrégularité.

Analysons la distribution par âge des femmes aux différents recensements (figure 7) ; la figure n'est pas un modèle de clarté, mais elle montre bien la « fantaisie » des recensements de 1851 et de 1886. Il est vraisemblable que le recensement de 1851, tel qu'il a été publié par le SGF, fut entièrement inventé : les valeurs ne paraissent pas pouvoir se rattacher à une quelconque vraisemblance démographique. Même chose en ce qui concerne le recensement de 1886, du moins avant l'âge de 30 ans. L'incohérence, déjà constatée pour les nombre de mariées et de célibataires, est donc générale. Les autres recensements forment un faisceau cohérent qui ne laisse pas deviner des erreurs administratives importantes. En réalité, elles peuvent exister, notamment sous la forme de copies de valeurs antérieures.

Figure 7. - Distribution par âge des femmes de l'Hérault aux recensements 1851 - 1911.

Examinons de la même façon la distribution par âge des décès annuels (figure 8). Elle évolue de façon régulière entre 1878 et 1910. Avant 1878, elle est très perturbée à plusieurs reprises : en 1856-1863, une apparente sur-représentation des jeunes âges semblent compenser une faible représentation des âges moyens ; en 1862-1873 et 1876-1877, les âges élevés semblent sous-représentés au profit des âges moyens surtout.

On peut analyser les décès par âge de façon plus approfondie en les transformant en quotients, selon un procédé analogue à celui de Bonneuil. On réalise d'abord une première estimation des pyramides par lissage des cohortes, puis on calcule les quotients sur la base des décès enregistrés et les cohortes lissées. Deux points différencient ce lissage de celui de la méthode Bonneuil : premièrement, les recensements de 1851 et 1886 sont exclus du lissage en raison des anomalies relevées ; deuxièmement, les premiers groupes d'âges recensés ont été remplacés par les survivants à 0-4 ans des naissances et décès enregistrés, sous l'hypothèse d'une migration nulle²⁰.

²⁰ Le mode de calcul des quotients n'est pas celui de la méthode Bonneuil. J'ai estimé les pyramides d'âges en milieu d'année, puis calculé les quotients en passant par les taux de mortalité. Cela permet de comparer les quotients observés directement aux tables types de Ledermann, sans passer par les quotients perspectifs.

Je rappelle par ailleurs que cette première estimation des quotients de mortalité ne servira qu'à construire les figure 9 et 10, c'est-à-dire, à repérer les années où l'enregistrement des décès semblent contenir des erreurs. Les tables ainsi obtenues ne seront pas remplacées par celles de Ledermann ; et en aucun cas, les cohortes lissées pour obtenir ces tables ne serviront à estimer la migration.

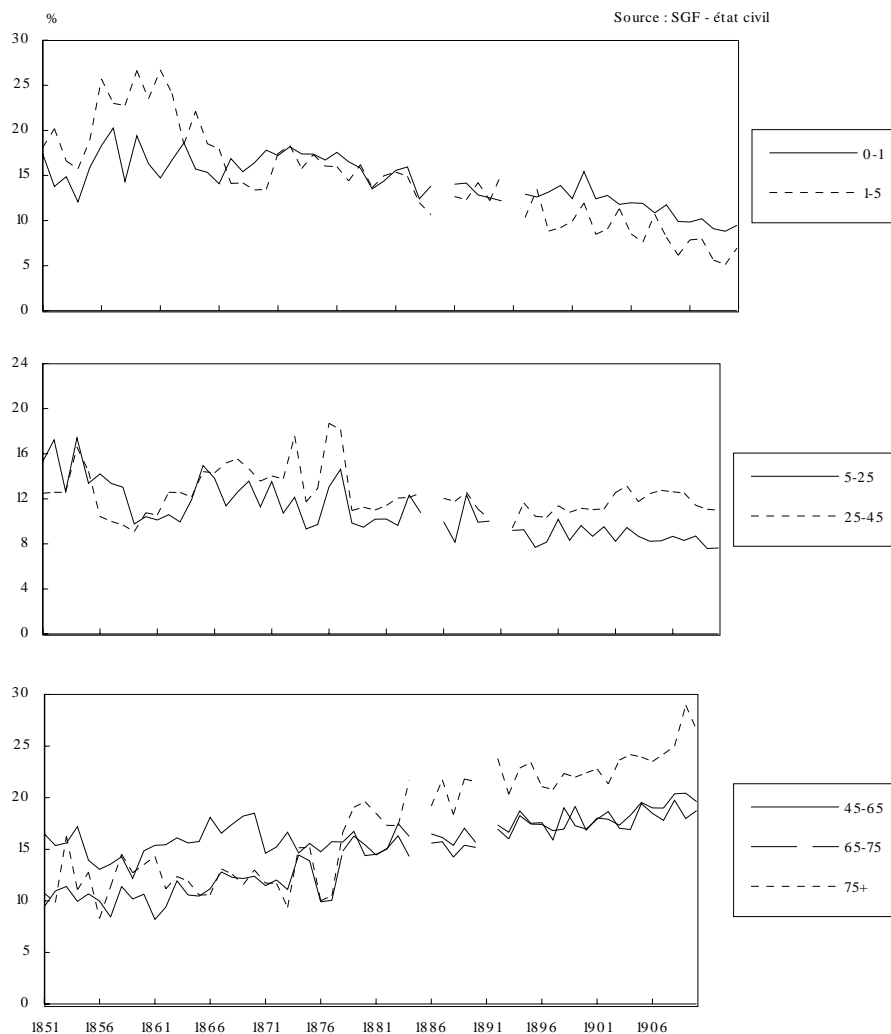


Figure 8. - Distribution des décès par âge, en % du total des décès ; femmes héraultaises.

Les quotients de mortalité ainsi obtenus forment des tables annuelles de mortalité. On a cherché ensuite, année par année, la table type Ledermann la plus proche de la table observée²¹. On s'est intéressé avant tout à l'écart qui sépare le quotient observé du quotient Ledermann d'ajustement : pour pallier la

²¹ Pour des raisons de calcul, j'ai cherché la solution de l'équation

$$\sum (q_{x,obs} - q_{x,Led}) = 0 \quad \text{pour } x = 0,1,5,10, \dots, 65$$

en variant l'espérance de vie à la naissance de la table Ledermann (réseau 100).

variation naturelle aux différents âges, on a standardisé les écarts obtenus par les écarts types que Ledermann a constatés aux différents âges pour une espérance de vie donnée. C'est l'écart standardisé, puis la moyenne des écarts standardisés une année donnée, qui donnent une indication de la vraisemblance des quotients de mortalité observés (figure 9). Notre objectif est d'étudier la structure par âge de la mortalité, de repérer les années incohérentes, où la discontinuité trahit les erreurs administratives. Une évolution progressive de l'écart standardisé, par contre, ne serait pas signe d'erreur : elle pourrait venir soit d'un réel rapprochement entre la mortalité locale et celle des tables de Ledermann, soit de l'estimation des pyramides, dont la qualité est de moins en moins assurée au fur et à mesure que l'on remonte le XIX^e siècle.

Figure 9. - Moyenne des écarts standardisés entre quotients observés et quotients Ledermann d'ajustement aux âges de 0 à 69 ans. Standardisation par les écarts-types de Ledermann, entrée 100.

Comme attendu, les années 1856-1866, 1871-1872 et 1876-1877 s'éloignent des tables types Ledermann.

Les années 1876-1877 sont un cas d'école : apparition et disparition soudaines d'une perturbation incompréhensible, sans changement notable du nombre total des décès (voir aussi figure 5). Les années 1871-1872 sont évidemment à mettre sur le compte de la guerre. La période 1856-1866 est plus ambiguë. S'agit-il de l'effet post-épidémique après la crise de 1854 ?

Rentrons dans le détail par âge.

En 1856-1866, l'ajustement semble surtout perturbé par une relative surmortalité aux jeunes âges. C'est conforme à l'hypothèse de l'effet post-épidémique : les adultes ont temporairement une mortalité plus faible après

une crise de mortalité. Mais l'effet semble trop long et trop constant ; il disparaît trop tardivement et pas assez progressivement²².

Figure 10. - Ecart standardisé par année et par âge entre les quotients observés et les quotients Ledermann d'ajustement ; standardisation par les écarts-types de Ledermann, entrée 100.

En 1876-1877, l'ajustement est perturbé par une apparente sous-mortalité des 60 ans et plus, tandis que le total des décès de ces années est relativement élevé. Aucune hypothèse plausible n'explique une telle perturbation.

Les périodes 1867-1875 et 1878-1905 semblent plus cohérentes. Les tables observées s'ajustent assez bien aux tables types Ledermann et les écarts sont systématiques : une mortalité plus élevée aux jeunes âges, moins élevée aux âges adultes. Le type de mortalité dans l'Hérault semble ainsi plus proche des tables types Sud de Coale et Demeny que de celles de Ledermann.

Revenons sur l'estimation de la mortalité de 0 à 5 ans. La figure 10 utilise l'estimation minimale de ${}_1q_0$ et ${}_4q_1$, basée sur les effectifs à 0-4 ans déduits de l'enregistrement de l'état civil, sous l'hypothèse d'une migration nulle. Toute mise en nourrice dans les départements ruraux environnants (l'inverse est très improbable par manque de grandes villes dans ces départements) entraînerait une diminution des jeunes enfants présents dans le département et donc une augmentation de la mortalité infantile constatée. Puisque l'estimation dépasse même la forte mortalité infantile du type méditerranéen (famille Sud de Coale et Demeny), on a préféré retenir celle-ci.

²² Les quotients observés à 1 an sont environ deux fois supérieurs à ceux de l'ajustement Ledermann, ceux à 50 ans et plus sont de 30 à 50 % plus faibles. La marge d'erreur des quotients observés, due à l'estimation des pyramides utilisées pour leur calcul, n'apporte pas d'explication suffisante.

Pour la même raison, on doit conclure à l'enregistrement complet de la mortalité infantile en 1856-1905 et, par voie de conséquence, à l'enregistrement complet des naissances vivantes.

Trois options de reconstitution Arrivé à ce stade de l'analyse, on est conduit à considérer deux scénarios de reconstitution :

I - Eliminer les années 1876-1877, mais accepter les années 1856-1866, considérant que l'effet post-épidémique a perturbé les tables de mortalité ;

II - Eliminer les années 1876-1877 et les années 1856-1866, considérant que l'hypothèse de l'effet post-épidémique n'est pas suffisamment plausible²³.

Le remplacement des années éliminées laisse également le choix entre deux méthodes :

Méthode A : elles peuvent être estimées utilisant la distribution des décès des années environnantes. Pour cela, le total des décès doit être accepté ou ré-estimé. Cette solution est indiquée pour les années 1876-1877 ; pour 1856-1866, elle prendrait la forme d'une extrapolation, ce qui serait plus discutable ;

Méthode B : on peut répartir le total de décès, qui à nouveau doit être correct, selon les âges à l'aide de tables types. Pour cela, on sera obligé de faire confiance soit aux pyramides estimées par lissage, soit aux pyramides reconstituées après estimation de la migration par âge.

La suite de la reconstitution dépend de l'option prise parmi les suivantes :

Méthode	Scénario	
	I	II
A	IA	IIA
B	-	IIB

Tableau 2. - Options de reconstitution suivant le scénario et la méthode choisies.

Les options IA et IIA permettent une estimation des décès par âge avant celle de la migration. L'option IIB nécessite une estimation de la migration avant celle des décès par âge. L'option IB est théoriquement possible, mais elle n'a pas de sens, car elle donnerait des résultats quasiment identiques à ceux de l'option IA ; nous l'avons donc éliminée.

²³ Les données non publiées, saisies aux Archives Départementales, permettent de trancher en faveur du deuxième scénario. Pour le besoin du développement de la méthode, nous nous en tenons aux seules données publiées.

Option IA

Après le remplacement des décès par âge des années 1876-1877, sur le modèle de la distribution observée en 1874-1875 et 1878-1879, on dispose pour la période 1856-1905 d'une pyramide finale supposée correcte, et d'une série de naissances et de décès par âge supposée fiable. On a également des pyramides pour 1856, 1861, ..., 1901, mais celles-ci sont jugées erronées ; leur correction est même l'objectif premier de la reconstitution.

La problématique de la reconstitution peut maintenant être ramenée à une estimation de la migration par âge. Le tableau 3 donne les valeurs obtenues.

	1856-60	1861-65	1866-70	1871-75	1876-80	1881-85	1886-90	1891-95	1896-00	1901-05	1906-10
N/0-4	-2622	-4431	-5626	-740	-3961	-5769	-660	-1816	-1120	-1599	-601
0-4/5-9	3177	2042	2442	3275	802	3784	5838	2002	1770	704	-92
5-9/10-4	-876	-1204	-1285	2216	-243	2023	-1481	522	408	368	-56
10-4/15-9	2595	2733	2766	6014	584	5224	-85	2604	2719	1528	1269
15-9/20-4	3682	3407	5411	5182	1142	11211	2835	4117	3496	1182	911
20-4/25-9	-2078	-1062	540	-1913	-2685	-3777	-10107	-4018	-127	-265	-276
25-9/30-4	1028	3019	2485	-1018	630	108	2072	1480	2034	-841	-530
30-4/35-9	1150	780	-169	-1426	1508	154	136	1565	1143	-165	116
35-9/40-4	660	1369	-1266	-1999	299	-2225	-125	748	120	-466	-375
40-4/45-9	-17	225	-1616	-276	302	-667	2260	1017	-228	-130	57
45-9/50-4	1710	1298	-271	414	-458	-524	883	342	-253	481	-258
50-4/55-9	78	876	220	327	-1795	-766	501	-168	-612	-912	-937
55-9/60-4	-1337	-225	-1709	-2414	-741	-509	1280	249	443	-166	518
60-4/65-9	-285	-1131	-2057	-388	314	-1105	1123	-44	-305	34	233
65-9/70-4	-1321	-843	-802	149	-188	376	229	-77	150	74	-112
70-4/75-9	-637	-173	13	472	37	-85	208	645	265	-71	220
75-9/80-4	87	315	309	358	794	-705	874	520	268	218	474
80-4/85-9	-28	-19	-377	84	336	-161	1493	270	40	187	343
85-9/90-4	72	23	-36	101	182	-25	224	145	106	169	202
90-4/+	-19	-15	-32	-57	-17	-89	26	-79	-94	-85	-67
SMiC	4883	4129	2941	5265	3647	4216	4189	1591	-472	-1691	-601
SMiP	5019	6984	-1060	8361	-3158	6473	7524	10024	10223	245	1039

Tableau 3. - Migration apparente par âge et par période intercensitaire, constatée sur la base des naissances, des décès corrigés et des pyramides brutes ; Solde Migratoire des Cohortes (SMiC) en 1911 ; Solde Migratoire des Périodes (SMiP).

A priori, les valeurs par âge sont erronées dans la mesure où les recensements le sont. Les Soldes Migratoires par Période (SMiP) sont également sujets à caution : ils représentent entre -1 et +2 % de la population totale recensée, qui, elle, connaît une marge d'erreur au moins du même

ordre²⁴. Ils accumulent également les effets des changements de méthode de recensement, de définitions employées, etc.

Les Soldes Migratoires par Cohorte (SMiC) sont, au contraire, des valeurs relativement fiables, car elles ne dépendent que des naissances vivantes et de la pyramide finale. Mais les SMiC ne sont connus que pour 12 cohortes, puisque la reconstitution ne porte que sur 55 ans : plus la période reconstituée est longue, plus les SMiC seront une aide à l'estimation de la migration par âge.

Malgré l'incertitude des valeurs par âge, celles-ci permettent quelques observations générales. D'abord, il y a une émigration apparente entre la naissance et le premier groupe d'âges, pour toutes les cohortes, suivie d'une immigration apparente entre le premier et le second groupe d'âges. Trois explications sont possibles : la mise en nourrice avec retour cinq ans plus tard (hypothèse peu probable en raison du retour à 100 % dans 5 cohortes sur 11, incompatible avec la forte mortalité infantile observée par ailleurs) ; le vieillissement à la déclaration (également peu probable, car il aurait dû être suivi d'une émigration apparente entre le deuxième et le troisième groupe d'âges) ; le sous-enregistrement des 0-4 ans. On retiendra donc cette dernière l'hypothèse.

On observe ensuite une émigration entre les groupes d'âges 20-24 et 25-29 ans (10 cohortes sur 11). Cette émigration intervient après 10 ans d'immigration (toutes les cohortes) et avant 10 ans d'immigration (8 cohortes sur 11). Les explications possibles sont les mêmes que pour l'émigration entre la naissance et le premier groupe d'âges, en remplaçant « mise en nourrice » par « départ en mariage » et « vieillissement » par « rajeunissement ». Mais ici, on n'arrive pas à départager facilement les hypothèses. Celle du sous-enregistrement est la moins probable. En effet, certains individus âgés de 23 et 24 ans déclarent avoir 25 ans, mais un peu plus nombreux sont ceux de 29 ans qui déclarent avoir 30 ans (cf. figure 3). L'hypothèse du rajeunissement des 25-29 ans, généralement retenue par les auteurs (par exemple, par Bonneuil et par Van de Walle) peut paraître crédible. Mais l'hypothèse d'une émigration réelle à l'âge du mariage l'est tout autant : l'émigration vers 25 ans, au milieu d'une immigration plus jeune et plus âgée, est une caractéristique de la démographie actuelle de l'Hérault et rien n'interdit de penser qu'il en était de même au XIX^e siècle.

Enfin, malgré la grande variation des valeurs calculées, on observe que la migration cumulée entre 30-34 ans et 100 ans, sur l'ensemble des 11 périodes intercensitaires, est de -745 femmes (-5 en moyenne par groupe d'âges et par période inter-censitaire), tandis que la migration cumulée entre la naissance et l'âge de 30-34 ans s'élève à 52 417 femmes (donnant une

²⁴ Vangrevelinghe (1963) et Coeffic (1993) concluent à une erreur de 1,5 à 2 % sur le dénombrement en 1962 et 1990. Il est probable que cette erreur fut plus importante aux recensements du XIX^e siècle.

moyenne de 681). La migration semble donc se concentrer avant l'âge de 30 ans, principalement entre 10 et 30 ans.

En somme, l'analyse du tableau 3 a donné quelques indications sans surprise : le sous-enregistrement des 0-4 ans est probable ; la migration se concentre vraisemblablement entre 10 et 30 ans ; et rien ne permet de départager un réel mouvement de retour à l'âge de 25 ans d'un rajeunissement à la déclaration.

Avec la correction des valeurs par âge, nous touchons aux limites des possibilités de reconstitution. Aucune table type, aucun profil type ne reconstitue la variété réelle des migrations par âge ; rien ne permet même d'affirmer avec certitude que les valeurs constatées sont erronées. Mais admettre que le nombre de migrants soit entièrement libre reviendrait à accepter les recensements publiés quelle que soit leur qualité ; il faut donc introduire des contraintes. Nous avons formulé les suivantes :

- le profil migratoire doit être constant sur la période 1856-1911²⁵ ;
- la migration se concentre entre les âges de 0 et 60 ans ;
- les Soldes Migratoires par Période (SMiP) sont considérés corrects, éventuellement après lissage.

Sous ces contraintes, on a cherché le profil de migration qui reconstitue le mieux les Soldes Migratoires des Cohortes (SMiC)²⁶. Le profil ne reconstitue pas exactement les SMiC ; on ne pourra donc respecter entièrement à la fois les naissances vivantes enregistrées en 1856-1910 et la pyramide finale de 1911 : nous avons privilégié les naissances. Les cohortes nées en 1856-1910 ont été vieillies jusqu'en 1911 ; les cohortes nées avant 1856 ont été rajeunies, à partir des effectifs recensés en 1911 (à 55 ans et plus). Le problème des derniers groupes d'âges n'existe pas, car nous disposons des décès et des migrations en effectifs réels, et non sous forme de quotients.

Option IIA

La démarche est la même que pour l'option IA : seuls les décès en 1856-1866 ne sont pas les mêmes. Le profil de migration - unique pour la période 1856-1910 - est identique à celui trouvé pour l'option IA, car les SMiP et les SMiC observés sont quasiment identiques. Les pyramides reconstituées sont assez proches de celles obtenues en IA, à l'exception de la cohorte née en 1856-1860, qui subit les plus importantes conséquences des différentes estimations concernant les décès en 1856-1866.

²⁵ Dans Banens (1995), nous avons introduit un deuxième profil, dit de "reflux", pour les périodes isolées de migration inverse. Nous avons toujours utilisé des profils transversaux, pour les commodités de calcul. Les reconstitutions qui s'étendent sur de longues périodes pourraient tirer profit d'un profil longitudinal : il faudrait alors résoudre le problème des deux triangles, en haut à gauche et en bas à droite, du diagramme de Lexis, où la carrière migratoire des cohortes est tronquée.

²⁶ Nous minimisons $\sum (SMiC_{i,obs} - SMiC_{i,reconst})^2$

Option IIB

Ici encore, de 1872 à 1906, les décès et les migrations par âge sont les mêmes : la particularité de la méthode IIB ne concerne que les décès de la période 1856-1866. Nous employons toujours le même profil de migration, car celui-ci s'accorde le mieux à la période 1872-1911. Les SMiP en 1856-1860, 1861-1865 et 1866-1870 acceptés, nous disposons donc d'une estimation de la migration par âge avant d'estimer les décès par âge à l'aide des équations suivantes :

$$D_{t,t+5} = P_t - P_{t+5} + SMiP_{t,t+5} + N_{t,t+5} \quad (1)$$

$$P_{t+5} = \Sigma(p_{i,t}(E_{i,t} + 0,5SM_{i,t})) + 0,5 SMiP_{t,t+5} \quad (i = -5, 0, 5, \dots, 85) \quad (2)$$

où $D_{t,t+5}$ est le nombre total des décès entre t et $t+5$, P_t la population totale en t , P_{t+5} la population totale en $t+5$, $SMiP_{t,t+5}$ le solde migratoire entre t et $t+5$, $N_{t,t+5}$ le nombre de naissances vivantes entre t et $t+5$, $p_{i,t}$ la probabilité de survie entre t et $t+5$ d'âge $i,i+5$ à $i+5,i+9$ déduite des tables types Sud de Coale et Demeny²⁷ ($p_{-,5}$ est la probabilité de survie entre la naissance et l'âge 0-4 ans), $E_{i,t}$ l'effectif à l'âge $i,i+4$ en t ($E_{-,5}$ est le nombre de Naissances vivantes), et $SM_{i,t}$ le solde migratoire entre l'âge $i,i+4$ en t et $i+5,i+9$ en $t+5$. Puisqu'il s'agit de trois périodes seulement, on a préféré la rétro-projection à partir de 1872, après estimation des derniers groupes d'âges en 1866 et 1861²⁸.

Les différentes techniques de correction des décès par âge n'introduisent que peu de variation dans la reconstitution finale (figure 11). La plus grande part d'incertitude inhérente à la reconstitution vient de l'estimation de la migration par âge. Or, ce serait une illusion de penser qu'une reconstitution, quelle que soit la méthode utilisée, puisse retrouver les véritables fluctuations de la migration par âge et par période.

Comparer les résultats

Toutes les méthodes de reconstitution décrites tentent de se rapprocher le plus possible des pyramides recensées en 1906.

Pour comparer les résultats, il faut donc s'éloigner de 1906. La figure 11 montre les différentes reconstitutions (celle de Van de Walle et nos trois options) pour 1856-1871, rapportées aux effectifs recensés en 1856-1872. Les pyramides reconstituées par Bonneuil n'étant pas disponibles pour comparaison, nous avons considéré le résultat de la première étape de sa méthode, le lissage des cohortes, que nous avons effectué selon la méthode

²⁷ Dans Banens (1995), nous avons utilisé un réseau de tables types local, construit sur la base de la mortalité héraultaise, observée de 1878 à 1990, et sur des observations antérieures de quelques arrondissements.

²⁸ Nous avons rétro-polé les 90-94 ans reconstituées en 1872-1911.

décrite par lui et commentée ci-devant, car celle-ci indique bien le sens de la reconstitution finale et en est probablement assez proche (cf. note 18)²⁹.

²⁹ La proximité entre les générations lissées selon la méthode et les paramètres BSVLS-1 (voir figure 1) et la reconstitution finale est confirmée aux âges de 20-24 à 25-29 ans, où toutes les cohortes montrent des taux de migration nette (publiés dans Bonneuil 1997, pp. 206-211) parfaitement en accord avec les cohortes lissées et les espérances de vie (ibid., pp. 188-196).

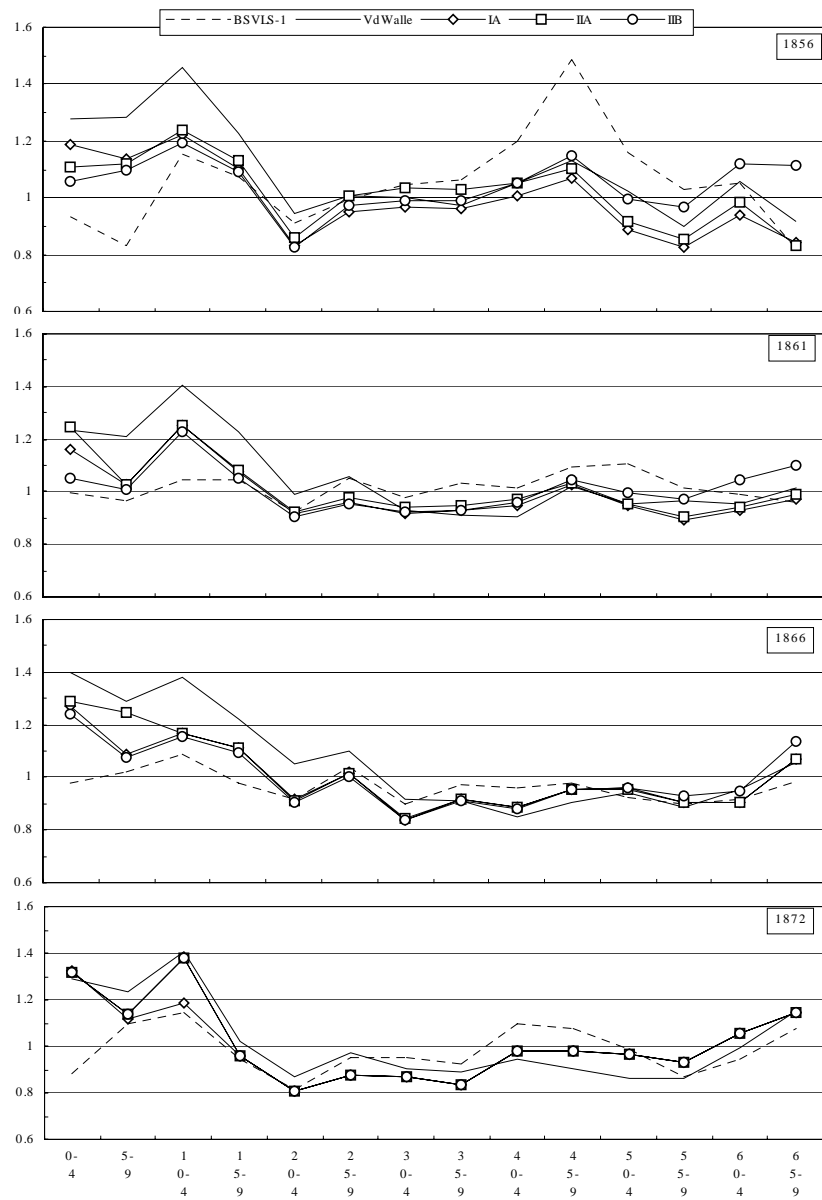


Figure 11. - Reconstitutions (Van de Walle, IA, IIA et IIB) et recensements lissés selon la méthode de Noël Bonneuil (BSVLS-1), rapportés aux recensements publiés (SGF)

Comme attendu, les trois options de notre reconstitution forment un faisceau serré de trois courbes, qui ne s'ouvrent qu'aux âges extrêmes. La reconstitution de Van de Walle se rapproche généralement des nôtres à partir

de l'âge de 20 ans. Le lissage mathématique, lui, en reste plus éloigné. Les plus grands écarts se situent avant l'âge de 15 ans. La correction y dépasse parfois 30, voire 40 % des effectifs recensés. Van de Walle a fortement corrigé pour sous-enregistrement, le lissage mathématique corrige pour sur-enregistrement. Nos options prennent une position intermédiaire.

L'exemple de l'Hérault en 1856-1906 a servi à exposer la logique des différentes méthodes de reconstitution, et à présenter une démarche alternative. La principale caractéristique de cette démarche est la place qu'occupe l'analyse critique des données disponibles et ce à toutes les étapes de la reconstitution. Celle-ci devient « opportuniste », la méthode étant choisie en fonction de la qualité des données disponibles. Le gain de fiabilité que l'on obtient ainsi, du moins à nos yeux, a bien sûr un prix : il n'est plus possible d'automatiser la reconstitution, ni de déterminer en une fois la matrice des migrations interdépartementales, comme le permet la méthode de Bonneuil. Le choix de la méthode de reconstitution ne devrait donc pas seulement prendre en compte la qualité et la quantité des données disponibles, mais aussi les objectifs de la reconstitution. A chaque reconstitution, ses objectifs et sa méthode.

Maks BANENS

BIBLIOGRAPHIE

- BANENS (Maks), (1995), *Transitions démographiques et peuplement dans l'Hérault, 1801-1990*, thèse, Université de Pau et des Pays de l'Adour, 3 tomes, 690 p.
- BLAYO (Yves), (1975), « La mortalité en France de 1740 à 1829 », *Population* numéro spécial « Démographie historique », pp. 123-142.
- BLAYO (Yves), (1975), « Mouvement naturel de la population française de 1740 à 1829 », *Population*, numéro spécial « Démographie historique », pp. 15-64.
- BLUM (Alain), BONNEUIL (Noël) et BLANCHET (Didier) éd., (1992), *Modèles de la démographie historique*, INED, 370 p.
- BONNEUIL (Noël), (1992), « Non-identifiabilité et cohérence démographique de la rétro-projection », in Blum A. éd., *Modèles de la démographie historique*, INED, pp. 99-108.
- BONNEUIL (Noël), (1997), *Transformation of the French Demographic Landscape 1806-1906*, Clarendon Press, Oxford, 217 p.
- BOURGEOIS-PICHAT (Jean), (1951), « Evolution générale de la population française depuis le XVIII^e siècle », *Population*, pp. 635-662.
- BOURGEOIS-PICHAT (Jean), (1952), « Note sur l'évolution générale de la population française depuis le XVIII^e siècle », *Population*, pp. 319-329.
- BOURGEOIS-PICHAT (Jean), (1966), *Le concept de population stable. Application à l'étude des populations des pays ne disposant pas de bonnes statistiques démographiques*, New York, Nations Unies, 235 p.

- BRASS (William), (1975), *Methods for estimating fertility and mortality from limited and defective data*, Chapel Hill.
- COALE (Ansley J.) et DEMENY (Paul), (1966), *Regional Model Life Tables and Stable Populations*, Princeton Princeton University Press.
- COALE (Ansley J.) et GUO (G.), (1989), « Revised Model Life Tables at Very Low Levels of Mortality », *Population Index*.
- COALE (Ansley J.) et GUO (G.), (1990), « New Regional Model Life Tables at High Expectation of Life », *Population Index*.
- COEFFIC (N.), (1993), « L'enquête post-censitaire de 1990. Une mesure de l'exhaustivité du recensement. », *Population*, pp. 1655-1682.
- DUPAQUIER (Jacques) et LE MEE (René), (1988) « La connaissance des faits démographiques, de 1789 à 1914 », in : Dupâquier Jacques (éd.) *Histoire de la population française*, Paris, tome III, pp. 15-62.
- HENRY (Louis) et BLAYO (Yves), (1975), « La population de la France de 1740 à 1829 », *Population* numéro spécial « Démographie historique », pp. 71-122.
- LEDERMANN (Sully), (1969), *Nouvelles tables-types de mortalité*, Travaux et Documents n° 53, INED/PUF, 262 p.
- LEE (Ronald D.), (1974), « Estimating Series of Vital Rates and Age Structures from Baptisms and Burials », *Population Studies*, pp. 495-512.
- LEE (Ronald D.), (1985), « Inverse projection and back projection: a critical appraisal, and comparative results for England », *Population Studies*, pp. 233-248.
- McCAA (R.) et VAUPEL (J.W.), (1992), « Comment la projection inverse se comporte-t-elle sur des données simulées ? », in Blum A. ed., *Modèles de la démographie historique*, INED, pp. 129-146.
- MESLE (France) et VALLIN (Jacques), (1989), « Reconstitution de tables annuelles de mortalité pour la France au XIXe siècle », *Population*, pp. 1121-1158.
- NATIONS UNIES, (1969), *Manuel IV. Méthodes permettant d'estimer les mesures démographiques fondamentales à partir de données incomplètes*, New York.
- NATIONS UNIES, (1984), *Manuel X. Techniques indirectes d'estimation démographique*, New York.
- OEPPEN (Jim), (1992), « La projection inverse généralisée et le problème des crises de mortalité », in Blum A. ed., *Modèles de la démographie historique*, INED, pp. 109-127.
- OEPPEN (Jim), (1993), « Back Projection and Inverse Projection : Members of a Wider Class of Constrained Projection Models », *Population Studies*, pp. 245-268.
- TABAH (Léon), (1947), « La répartition par âges de la population française en 1851 », *Population*, pp. 349-354.
- VAN DE WALLE (Etienne), (1972), « De l'emploi des modèles en démographie historique », *Annales de Démographie Historique*, pp. 153-177.

- VAN DE WALLE (Etienne), (1974), *The Female Population of France in the Nineteenth Century*, Princeton University Press, 482 p.
- VAN DE WALLE (Etienne), (1977), « La nuptialité des françaises avant 1851, d'après l'état civil des décédées », *Population* numéro spécial « La mesure des phénomènes démographiques », pp. 447-465.
- VANGREVELINGHE (G.), (1963), « Recensement démographique de 1962. Etude de l'exhaustivité du dénombrement », *Etudes Statistiques*, pp. 235-265.
- WACHTER (Kenneth W.), (1986), « Ergodicity and inverse projection », *Population Studies*.
- WEIR (David R.), (1993), « Family reconstitution and population reconstruction: two approaches to the fertility transition in France », in Reher ed., *Old and New Methods in Historical Demography*, Oxford.
- WEIR (David R.), (1994), « New Estimates of Nuptiality and Marital Fertility in France, 1740-1911 », *Population Studies*, pp. 307-331.
- WRIGLEY (E.A.) et SCHOFIELD (R.S.), (1981), *The Population History of England, 1541-1871: A Reconstruction*, Cambridge, Mass.