Evolution des Migrations dans le Temps en Fonction du Nombre Initial de Migrants Présents et de la Distance de Migration

M. Courgeau

- Essai d'application aux migrations internes françaises -

Le problème posé dans cette étude est celui de la prévision des mouvements migratoires en fonction des migrations antérieures. Une telle recherche est possible dans la mesure, où les migrants présents dans une zone gardent des contacts avec les individus demeurés dans leur zone d'origine. Cette hypothèse avait déjà été indiquée par Pinkney (1953) dans son étude sur les migrations vers la ville de Paris durant le second empire : selon lui, l'information sur les offres d'emploi à Paris était uniquement transmise oralement ou par lettres. A sa suite T. Hagerstrand élabora un modèle expérimental dans lequel il distinguait deux catégories de migrants : les migrants "actifs" qui choisissent leur lieu de destination indépendamment de leurs "connaissances,"les migrants "passifs" qui choisissent ce lieu en fonction de leurs"connaissances". L'application d'un tel modèle aux données Suédoises conduit à des résultats très satisfaisants. Partant de ces exemples nous avons cherché à voir si un modèle, basé sur des hypothèses voisines, permettrait de prévoir les migrations françaises.

- Modèle utilisé -

Supposons qu'à l'instant initial 0 le nombre de migrants nés dans la zone i (de population P_i^o) et présents en j (de population P_j^o et distante de r kms de la zone i) soit $X^o\ (r)$. Après l'intervalle de temps unitaire (1 an par exemple) l'accroissement des migrants originaires de i et présents en j sera :

$$\Delta X^o\ (r) = \underbrace{a\ (r)\ P_i^o\ P_j^o}_{\text{migrations actives}} + \underbrace{k\ (r)\ X^o\ (r)}_{\text{migrations passives}} - \underbrace{m(r)\ X^o\ (r)}_{\text{mortalité et nouvelles migrations}} \qquad (1)$$

où $a(r)$, $k(r)$, $m(r)$ sont des fonctions de la distance seule, correspondant aux trois types de variation du nombre de migrants (nous supposons que m n'est fonction de r, que par les nouvelles migrations). Les migrations actives, définies dans l'introduction, toucheront un effectif qui sera proportionnel aux populations de i et de j et fonction de la distance r. Les migrations passives porteront sur un effectif proportionnel au nombre de migrants originaires de i et présents en j et seront fonction de la distance r. Enfin, la mortalité des migrants, cumulée aux nouvelles migrations, introduira une correction nécessaire. L'hypothèse de base du modèle est que les trois fonctions $a(r)$, $k(r)$ et $m(r)$ restent constantes au cours du temps (1) ; cette hypothèse sera sans doute d'autant moins bien vérifiée que l'intervalle de temps considéré sera plus étendu.

Appliquant maintenant le modèle à l'évolution dans le temps du nombre de migrants considéré, il vient la récurrence facile à résoudre :

---

(1) par la suite ces trois quantités seront désignées plus simplement par a, k et m.

$$X_1 = a\,P_i^o P_j^o + X^o\,(1 + k - m)$$

$$X_2 = a\,P_i^1 P_j^1 + a\,P_i^o P_j^o\,(1 + k - m) + X^o\,(1 + k - m)^2$$

$$(2)\quad \boxed{X_n = a\,P_i^{n-1} P_j^{n-1} + a\,P_i^{n-2} P_j^{n-2}\,(1 + k-m) + \ldots\ldots + a\,P_i^o P_j^o\,(1+k-m)^{n-1} + X^o\,(1+k-m)^n}$$

Si l'on suppose que les populations des zones d'origine et de destination restent sensiblement constantes au cours de la période considérée, il vient la formule plus simple

$$X_n = a\,P_i^o P_j^o \left(1 + (1+k-m) + \ldots\ldots + (1+k-m)^{n-1}\right) + X_o\,(1+ k-m)^n$$

$$(3)\quad \boxed{X_n = a\,P_i^o P_j^o\left(\frac{(1+k-m)^n - 1}{k-m}\right) + X_o\,(1 + k-m)^n}$$

X_n est donc dans ce modèle lié linéairement à X_o ; la fonction trouvée sera croissante, si k est supérieur à m, en fonction du temps, X_o étant fixé.

Les principales critiques que l'on peut faire à ce modèle sont les suivantes :

- dans la définition des migrations actives, l'hypothèse de leur proportionnalité à la population de la zone j est valable lorsqu'on a affaire à des mouvements de remplacement. Elle devient moins satisfaisante lorsque des postes nouvellement créés entrent dans le circuit.

- Il suppose, que l'effet des autres variables (économiques, sociales) pouvant intervenir dans la détermination des migrations, est négligeable.

- Il s'agit d'un modèle déterministe : or en fait plus X_o sera faible moins la formule (1) sera applicable. Il serait dans ce cas nécessaire d'introduire la variance de la quantité ΔX dont la formule (1) donnerait la moyenne.

- Le modèle suppose que les migrants passifs nés dans la zone i ont été attirés dans la zone j uniquement par les migrants antérieurs nés en i. Or, des migrants passifs auront pu déjà effectuer d'autres migrations au cours de leur existence et être attirés dans la zone j par des migrants originaires d'une autre zone que i.

- Application aux migrations internes françaises -

Certains recensements français donnant le nombre de migrants présents dans un département en fonction de leur département de naissance, les recensements de 1911 et 1946 fourniront la base de cette étude.

Il convient en premier lieu de noter l'importance de l'intervalle de temps considéré. Il dépend bien sûr du fait que les recensements intermédiaires ne fournissaient pas les chiffres nécessaires. Mais son ampleur permettra, si le modèle est bien vérifié, de dire que l'effet des autres variables pouvant intervenir dans la détermination des migrations est négligeable. Un autre inconvénient se trouve dans le fait que les populations des zones d'origine et d'arrivée auront subi une modification non négligeable ; la formule (3) n'est donc pas utilisable directement. Or, cette formule avait l'avantage de ne faire intervenir qu'une variable $\frac{X}{P_i P_j}$ et l'ajustement des données pouvait se faire simplement par la méthode des moindres carrés. En vue de conserver cette simplification, la formule suivante a été finalement choisie :

(3) $$\frac{X_n}{P_i^n P_j^n} = a \frac{(1+k-m)^n - 1}{k-m} + \frac{X_o}{P_i^o P_j^o} (1+k-m)^n$$

Cette formule (3') permet une meilleure approche de (2). Introduisant la variable $y = \frac{X}{P_i P_j}$, elle s'écrit plus simplement :

$$y_n = a \frac{(1+k-m)^n - 1}{k-m} + yo (1+k-m)^n$$

En second lieu il n'était pas possible, étant donné leur nombre, de faire intervenir comme zones d'arrivées tous les départements français. Un choix a donc été effectué et il en a été pris 12 : dans cette première étude nous avons évité de faire intervenir les départements contenant un fort centre d'attraction (Seine, Seine-et-Oise, Rhône, Bouches-du-Rhône, Gironde). Les départements considérés ont été tirés arbitrairement de façon à avoir une répartition à peu près homogène sur toute la France : Allier, Ardennes, Cantal, Côte d'Or, Doubs, Finistère, Gers, Indre-et-Loire, Landes, Nord, Pas-de-Calais et Pyrénées Orientales.

En troisième lieu, pour chaque département d'arrivée considéré, il était nécessaire de définir des zones de départ fonctions de la distance parcourue. Pour ce faire il a été simplement tracé des cercles centrés au département d'arrivée et de rayon en progression arithmétique de raison 60 kms (largeur moyenne d'un département). Les départements dont plus du quart de la surface était présent dans une zone ont été comptés comme faisant partie de cette zone. De plus, dans cette première étude nous avons éliminés les départements de la Seine et Seine-et-Oise et nous avons choisi les zones 2 (60 - 120 kms), 3(120-180 kms), 5 (240-300 kms), 8 (420-480 kms) 12 (660-720 kms).

L'ajustement d'une droite par la méthode des moindres carrés donne comme estimation des coefficients de cette droite (si N est le nombre d'observation effectué)

$$A = (1+k-m)^n = \frac{\Sigma Y_n Yo - \frac{1}{N} \Sigma Y_o \Sigma Y_n}{\Sigma Yo^2 - \frac{1}{N} (\Sigma Yo)^2}$$

$$B = a \frac{(1+k-m)^n - 1}{k-m} = \frac{\Sigma Y_n}{N} - \frac{A \Sigma Yo}{N}$$

La précision R^2 de l'ajustement peut être mesurée par la réduction relative de variance obtenue en rapportant la valeur mesurée à la valeur prédite par la formule 3 '- Si 6^2_T est la variante totale et 6^2_r la variance résiduelle, alors $R^2 = 1 - \frac{6^2_r}{o^2_T}$

Les calculs effectués conduisent aux résultats suivants :

Zone	N	A	B x 10^{11}	R^2
2	76	1,282	121	0,89
3	105	1,296	168	0,79
5	186	1,291	89	0,55
8	174	0,735	90,9	0,16
12	85	1,677	48,3	0,07

L'examen de ces résultats amène un certain nombre de remarques :

1°) la précision de l'ajustement diminue avec la distance parcourue. Or, de nombreuses études ont montré que le nombre de migrants était une fonction fortement décroissante de la distance. Il s'ensuit que, du fait de la remarque indiquée ci-dessus, plus X sera faible plus la zone correspondante sera éloignée, donc moins la formule utilisée sera applicable. Nous éliminerons donc désormais les zones 8 et 12 des remarques suivantes.

2°) les variations du coefficient $A = (1 + k - m)^n$ en fonction de la distance parcourue sont très faibles. Il s'ensuivrait donc que les migrations passives et les nouvelles migrations ne seraient pas fonction de la distance sur un rayon de 300 kms : cette hypothèse est très vraisemblable. Elle revient en fait à dire que d'une part le nombre de contacts gardés avec la zone d'origine est indépendant de la distance de migration, d'autre part, la probabilité d'effectuer une nouvelle migration est indépendante de la distance parcourue. De plus, le nombre de migrants étant une fonction croissante du temps, k est supérieur à m.

3°) les variations du coefficient B, si A est constant, sont proportionnelles aux variations de a correspondant aux migrations actives. Bien que ces variations ne soient pas très élevées, elles ne permettent aucune conclusion. On aurait cependant pu penser que a serait de la forme $a = r^{-\alpha}$ introduisant ainsi la formule de Pareto, si souvent vérifiée dans l'étude des migrations, comme base du modèle.

Conclusion

L'application qui vient d'être effectuée est, en fait, une étude préliminaire des migrations en fonction du modèle proposé. Malgré l'intervalle de temps élevé séparant les deux recensements, elle nous a permis des conclusions très intéressantes.

En particulier la faible variation des coefficients, en fonction de r, est remarquable. Elle permet en regroupant l'ensemble des données d'obtenir une fonction indépendante de la distance parcourue :

$$\boxed{Y_{35} = 89{,}4 \times 10^{-11} + 1{,}29\ Y_0} \quad \text{avec} \quad \boxed{R^2 = 0{,}85} \quad \boxed{\begin{array}{l} a = 1{,}25 \times 10^{-11} \\ k - m = 7 \times 10^{-3} \end{array}}$$

Il s'agit donc d'une méthode pour la prévision des migrations, applicable dès que celles-ci ont atteint un certain niveau entre deux zones. Par contre, elle se trouve totalement inadaptée lorsque ces migrations sont au-dessous du niveau précédent : un modèle stochastique serait nécessaire pour l'étude de ces faibles migrations. De plus une étude sur des périodes de temps successives pourrait amener à considérer les termes a, k, m comme des fonctions du temps compliquant ainsi le modèle initial trop simple.

RESUME

Un modèle, faisant intervenir deux types de migrants de comportements différents, est appliqué aux migrations internes françaises. Il ne donne des résultats corrects que pour des distances inférieures à 300 kms ; de plus il ne permet pas de distinguer les migrations passives (correspondant au choix du lieu de migration en fonction des migrants antérieurs) de la mortalité et des nouvelles migrations. Compte tenu de ces restrictions, on obtient une valeur du coefficient annuel (k - m), correspondant à l'effet des migrations passives corrigé par la mortalité et les nouvelles migrations, de l'ordre de 7×10^{-3}, indépendant de la distance parcourue. Les valeurs du coefficient annuel (a), correspondant aux migrations actives (indépendantes des migrations antérieures) rapportées aux populations des zones de départ et d'arrivée, sont de l'ordre de $1,25 \times 10^{-11}$; ses variations en fonction de la distance sont faibles. Ce modèle est applicable à la prévision des migrations, sous certaines réserves.

SUMMARY

A model of French internal migration has been constructed on the assumption that there are two different types of migrants whose behaviour differs. The model only gives correct results for distances below 300 kilometres. The model does not distinguish between passive migration (choice of place depending on previous migrants) and mortality and new migration. Bearing these restrictions in mind, as value for an annual coefficient which gives an index of the effect of passive migration corrected for mortality and new migration is obtained, the value of this coefficient is 0.007 independently of migration distance. The value of the annual coefficient (a) (for active migration independent of previous migration) is of the order of 1.25×10^{-11}; it does not vary much with difference. The model may be applied to predicting migration movements.