

# Vers la représentation cartographique d'une extraction des motifs de flux mondiaux issue d'une régionalisation

Françoise Bahoken \*

\*Université Paris-Est, AME, SPLOTT,  
IFSTTAR, F-77447, Marne-la-Vallée, France  
UMR 8504 Géographie-Cités, Equipe PARIS.  
francoise.bahoken@ifsttar.fr,

**Résumé.** L'une des principales familles de solutions apportées à l'objectif de réduction de la complexité de la carte de flux consiste à réduire la résolution des données. Les solutions liées à l'agrégation spatiale conduisent à un changement d'échelle géographique, à la production d'un maillage territorial de niveau régional issu d'une classification fonctionnelle car fondée sur les valeurs des flux. Si l'approche est efficace, on souhaite s'interroger sur les raisons de l'absence d'analyse factorielle réalisées en amont de ces partitions de l'espace par des flux.

## 1 Introduction

L'analyse (carto)graphique<sup>1</sup> des matrices de flux ou des mouvements, c'est-à-dire des quantités mesurables transférées entre des couples de lieux d'origines et de destination, dans un espace assorti d'une métrique, par exemple géographique, comporte de nombreux défis liés à la complexité intrinsèque de l'objet. En présence de données de flux massives, la difficulté consiste, en effet, à produire une carte descriptive ou explicative des motifs des flux qui soit lisible et intelligible, autrement dit dotée de sens dans un registre thématique donné. Pour y remédier, trois familles d'approches sont traditionnellement déployées. La première approche consiste à filtrer l'information de flux disponible de manière à n'en représenter qu'une sélection significative. La seconde approche consiste à fusionner visuellement les tracés symbolisant les flux de manière à réduire leur complexité graphique. La troisième approche s'inscrit dans un objectif de réduction de la résolution des données qui conduit à mobiliser, voire à générer, un nouveau maillage territorial associé aux flux, à changer d'échelle géographique. Si ces approches descriptives s'inscrivent toutes dans le registre des analyses statistiques uni et bi variées appliquées aux données de flux, la troisième d'entre elles ouvre néanmoins la voie vers les analyses multivariées de type classification ascendante. En inscrivant notre communication dans cette troisième approche, nous souhaitons présenter un exemple classique de régionalisation fonctionnelle par les flux, i.e. la production d'un maillage territorial issu d'une procédure de classification ascendante hiérarchique (CAH), qui entraîne un changement d'échelle géographique des flux représentés, par conséquent une réduction de la complexité graphique de la

---

1. Ce terme signale un problème graphique que l'on appréhende au niveau cartographique, i.e. en tenant compte de la composante spatiale de localisation des lieux (et non des sommets ou objets, termes non spécifiquement géographiques).

cartographie associée. L'objectif de notre communication est d'ouvrir une discussion sur les conditions de l'application de méthodes d'analyses factorielles, en amont ou en complément d'une CAH, devant permettre d'extraire les principaux motifs de flux et de mouvements à cartographier. Nous rappelons en premier lieu l'objectif général de la cartographie des matrices de flux denses. Nous présentons ensuite la famille de solutions liées aux procédures d'agrégation, que nous illustrons par une méthode de régionalisation fonctionnelle. Le cas d'application est une matrice de flux commerciaux internationaux mondiaux, dont le niveau élémentaire d'expression (pays\*pays) va être réduit à un niveau interrégional.

## 2 L'objectif général de réduction de la complexité graphique

L'objet « matrice de flux », associé à un espace géographique, est formé d'objets vectoriels (le couple des  $(i, j)$  lieux d'origine et de destination) en sus des scalaires (la distribution des  $(i)$  lieux de la zone d'étude) à représenter. Cette particularité en fait un objet complexe à cartographier dans la mesure où la structure multidimensionnelle de la matrice nécessite l'articulation d'au moins quatre dimensions à représenter : les positions des lieux  $(i)$ ,  $(j)$  définies par un couple de coordonnées géographiques  $(x, y)$  dans un système de projection donné ; le tracé du lien caractéristique du couple  $L(i, j)$  ; la symbolisation de la valeur du flux  $F(i, j)$  comme composante attributaire ou thématique caractérisant le type de flux dont il est question (par exemple, migratoire ou commercial) ; ainsi que la temporalité  $(t)$ , soit :  $(x_i, y_i)$  ;  $(x_j, y_j)$  ;  $F_{ij}$  ;  $t$ .

Cette complexité intrinsèque de l'objet pose de nombreuses difficultés d'exploitation, d'une part, sur les aspects graphiques et sémiologiques de la représentation (la réalisation de tracés pertinents et signifiants au niveau thématique) et, d'autre part, sur les impératifs du traitement des données numériques : de l'information de flux (disponible dans la matrice de départ) et de la donnée cartographique (le maillage territorial). Toutes constituent autant de défis, souvent multiples, à résoudre qui sont renforcés dans le cas de données spatiales massives et/ou complexes ; la complexité étant entendue comme l'existence de matrices « empilées » sur un même espace géographique, autrement dit de sous-matrices qui correspondent à un découpage de la quantité de flux  $(Z)$  en catégories thématiques (par exemple : en types de produits transportés) et/ou en  $(T)$  périodes (par exemple, lors ces données sont disponibles à différentes dates).

La difficulté de cartographier ces matrices denses tient, en effet, dans la conjonction de plusieurs effets bien connus, parmi lesquels les effets graphiques suivants (Bahoken, 2015) :

- un « effet de surcharge graphique », caractéristique d'une densité de figurés par ailleurs hétérogènes trop importante ;
- un « effet de couverture spatiale », caractéristique d'une densité de teinte (surface visuelle) par rapport au fond trop élevée ;
- un « effet d'enchevêtrement » lié à la présence de figurés linéaires fortement enchevêtrés en raison d'orientations angulaires variées .

La combinaison de ces effets se répercute ainsi au niveau graphique par une complexité qui empêche, d'après D. Fairbairn (2006), la compréhension aisée de l'information. Cette situation explique que la clarté de ces représentations spatialisées de flux soit d'ailleurs considérée par (Tobler, n.-d.) comme une « propriété générale » de la carte de flux, comme l'objectif de la figure. La cartographie de flux est par conséquent fortement sensible aux choix réalisés en amont de la représentation proprement dite : lors des traitements numériques et cartographiques réa-

lisés pour produire une figure lisible et pertinente sur le plan thématique.

En matière de flux, indépendamment du caractère descriptif ou explicatif de l'approche mobilisée, plusieurs directions sont traditionnellement privilégiées pour réduire sa complexité analytique (voir Bahoken, 2016 : 205-206) : (i) agir sur les aspects graphiques ; (ii) filtrer les données affichées ; (iii) agréger les données, en réduisant leur résolution. Les procédures de type fusion graphique intervenant sur les liens (edge bundling : Holten, 2006, 2009 ; Lambert, Bourqui, Auber, 2010) et/ou sur les noeuds s'inscrivent dans la catégorie (i) et permettent de donner l'illusion d'un mouvement dans l'espace géographique. Suivant la catégorie (ii), la sélection des valeurs jugées significatives pour le propos et/ou par la sélection des objets spatiaux réduisent naturellement la quantité d'objets à représenter. Les procédures d'agrégation spatiale qui nous intéressent appartiennent à la (iii) catégorie, car elles regroupent les objets en communautés pertinentes suivant des critères qui portent sur les lieux géographiques et/ou sur les liaisons entre les lieux.

### 3 Les solutions liées aux procédures d'agrégation des flux

L'un des éléments d'explication de la complexité graphique de la carte de flux étant d'ordre géographique (l'augmentation du nombre de liens à représenter,  $N(N - 1)$ , est une fonction quadratique du nombre  $N$  de lieux de la zone d'étude), l'une des possibilités consiste à réduire le nombre théorique de liens, en regroupant les lieux.

Cette éventualité intéresse les questions liées à la partition des matrices de flux : par exemple, la création de sous-tableaux (des sommes marginales) ou l'identification de blocs internes (formé d'un ensemble de lieux), considérant une matrice externe des positions  $(x, y)$  de ces lieux. Ces blocs peuvent être appréhendés comme des classes, i.e. comme des régions, regroupant un certain nombre de couples de lieux qui partagent une caractéristique d'appartenance commune. La production de ces blocs régionaux intéresse la problématique du changement d'échelle géographique, car la production d'un maillage territorial peut être issue de données de flux. D'après C. Grasland (1997), un maillage territorial est une partition qui discrétise l'espace en deux catégories binaires : un couple de lieux appartient ainsi ou non à un ensemble régional, ce qui « [...] implique que l'on peut définir des maillages soit à travers l'étude des attributs d'un ensemble d'éléments [par exemple, ceux des lieux, d'origine ou de destination, des flux], soit à travers l'étude des relations entre les éléments d'un ensemble [...] le maillage est à la fois un réseau de relation et une partition ».

Cette solution, liée aux procédures d'agrégation, lorsqu'elle est apportée au problème de la complexité graphique des cartes de flux est appréhendée soit du point de vue des lieux, soit du point de vue des flux. Dans le premier cas, la partition est préexistante aux flux : la carte décrit alors les flux observés à l'échelle de pays voire de régions, échelons formés d'unités spatiales définies<sup>2</sup> indépendamment des flux. Dans le second cas, le niveau régional, qui sert de fond de carte, est issu d'une classification préalable des valeurs de flux observées à l'échelle des pays. C'est cette seconde possibilité qui nous intéresse ici.

---

2. Un exemple de cartes de flux mondiaux qui en résulte est proposé dans (Bahoken 2016 :317-322).

### 3.1 La génération de partitions territoriales issues des flux

Aborder la production de maillages du point de vue des flux est devenu un impératif, depuis que S. Openshaw (1977) a mis en évidence la manifestation d'un effet du problème des unités spatiales modifiables (Modifiable area unit problem MAUP) à partir de données de flux modélisées sur un maillage territorial inadéquat. Si la production de partitions territoriales par les flux permet de répondre aux prérequis des modèles (Combes et al. 1996), elle répond aussi à l'objectif de réduction de la complexité de la carte, par la réduction du nombre de lieux concernés par les flux.

Sur le plan méthodologique, la génération de maillages à partir d'unités spatiales élémentaires porte soit sur leur redécoupage, par modification de leur morphologie (forme, taille...), soit sur leur regroupement en régions fondé sur un ou plusieurs critères d'agrégation (contiguïté des zones, taille minimale ou maximale de la région, similitude ou non des zones intégrées, nombre de régions...). Dans le contexte de l'analyse géographique, ces procédures d'agrégation, qu'elles soient structurelles<sup>3</sup> ou fonctionnelles<sup>4</sup>, conduisent toutes à un changement d'échelle géographique, c'est-à-dire à la production d'un nouveau zonage de résolution spatiale supérieure.

L'agrégation entraîne donc un changement conceptuel fort, car elle modifie la résolution des données et leur thématique : l'analyse de flux intercommunaux différant de celle de flux inter régionaux. La procédure qualifiée de « régionalisation fonctionnelle » consiste donc à regrouper des unités spatiales élémentaires en régions qui sont fonction du type et de l'intensité des échanges entre les couples de lieux qui les composent ; ces régions correspondant à des classes définies selon différents critères d'homogénéité. L'un des intérêts de ces méthodes de régionalisation fonctionnelle réside donc dans la construction d'un espace multi scalaire fondé sur l'analyse des flux à différentes échelles géographiques, en incluant des critères spatiaux : de distance, de voisinage ou de contiguïté. La génération du maillage proprement dit intervenant après une opération de classification des flux qui conduit au regroupement des unités spatiales élémentaires en régions.

### 3.2 Une typologie des méthodes de partition territoriale par les flux

Plusieurs méthodes de partition territoriale par les flux sont envisageables, en fonction du modèle théorique sous-jacent, un présentation est disponible dans Grasland et al. (2009). Aussi pouvons-nous évoquer à minima trois grandes approches, sans compter les approches mixtes développées par Coombes et al. (1986). Il s'agit de l'approche nodale, de l'approche centre-périphérie et de la régionalisation fonctionnelle. L'approche nodale est inscrite dans le cadre d'analyse des flux dits majeurs et dominants. Elle est fondée sur la méthode initialement proposée par J. D. Nyusten et M. F. Dacey (1961), de définition de régions nodales autour des lieux les plus attractifs d'un système d'échanges. En révélant une organisation spatiale multipolaire, la méthode décrit l'arborescence des relations (et non des flux), de même que l'aire d'influence des pôles dominants. L'approche étant populaire, mais jugée restrictive, plusieurs variantes en

---

3. Les procédures structurelles sont fondées sur des caractéristiques morphologiques (géométrie, taille...) ou tributaires des lieux (appartenance à une même région administrative de niveau supérieur).

4. Les procédures fonctionnelles sont fondées sur les couples de lieux, dans notre cas sur l'analyse de l'intensité de leur relation (la valeur de flux), de la distance qui les sépare, etc. Elles pourraient également être fondées sur l'existence de relations, comme par exemple dans les approches de clustering de type Théorie des graphes.

ont été proposées (voir par exemple Cattan, 2004). L'approche centre-périphérie diffère de la précédente en ce sens qu'elle est fondée sur les valeurs de flux et mobilise en outre des données structurelles pour définir des zonages totalement indépendants des découpages administratifs. En France, le découpage en Aires urbaines, par exemple, est élaboré par l'INSEE à partir de tableaux de navettes domicile-travail. Sa cartographie décrit l'aire d'influence des territoires urbains, en les distinguant des territoires ruraux.

L'approche fonctionnelle est fondée sur le principe général de la CAH appliqué aux données de flux. Plusieurs versions en ont été proposées par l'INSEE, à la suite de la méthode Zonage élaborée dès 1965 pour générer un maillage fondé sur la hiérarchie des échanges. On peut citer les méthodes de Denimal et de C. Terrier. La première mobilise une matrice d'adjacence entre les lieux et non de flux, comme dans l'approche de J. D. Nyusten et M. F. Dacey. La méthode Mirabelle de C. Terrier (INSEE 1998) est, quant à elle, plus étendue que la méthode Zonage d'origine, car elle vise la production d'un « *zonage adéquat* » permettant l'analyse des configurations spatiales issues des volumes de flux (avec différentes variantes de pondération) ainsi que des liens hiérarchiques. Mirabelle propose ainsi « *des zonages de finesse différente plus ou moins emboîtés* » (Terrier, 1998).

L'une des méthodes anglo-saxonne très populaire<sup>5</sup> de régionalisation par les flux est la méthode Intramax (« *Intrazonal interaction maximisation* ») et sa variante Intramin développée par Masser et Brown (1975).

## 4 La méthode Intramax/min, principe et application

La méthode *Intramax/min* de régionalisation par les flux a ceci de particulier qu'elle n'est pas fondée sur un critère de contiguïté, elle ne mobilise que l'intensité des flux et une acception de la distance fonctionnelle.

### 4.1 Principe de l'Intramax/min

Le fondement de la méthode Intramax/min réside dans l'analogie du concept de « région fonctionnelle » avec l'hypothèse gravitaire de décroissance de l'intensité du flux avec l'augmentation de la distance parcourue. De ce fait, l'intensité des interactions entre deux zones sera d'autant plus forte que la « *distance fonctionnelle* » qui les sépare est faible. Ainsi, une zone quelconque sera agrégée avec une autre zone au sein d'une région si leur « *distance fonctionnelle* » est faible et leurs échanges intenses.

Appliquée pas à pas, la procédure permet de générer  $(n - 1)$  partitions territoriales par les flux dont le critère d'homogénéité maximise les flux intra régionaux cumulés à l'origine de la formation des régions, autrement dit le volume total des interactions entre les lieux regroupés dans les régions (la version Intramin étant une minimisation de ces flux intra régionaux). A la fin de la procédure, comme dans une CAH classique, l'ensemble des unités spatiales se retrouve dans la même zone.

Les résultats produits par la méthode sont : un dendrogramme, un rapport d'agrégation, sur les  $(n - 1)$  étapes précisant le pourcentage de flux intra régionaux cumulé à chacune des étapes ;

---

5. La popularité de cette méthode est probablement en partie liée à son implémentation dans le logiciel Flowmap© (Van der Zwan et al., 2003).

éventuellement des matrices de flux inter régionaux issues de l'état de l'agrégation par la fonction somme des valeurs de flux aux différentes étapes.

## 4.2 Application

Pour illustrer le fonctionnement de la méthode Intramax, nous l'appliquons à une matrice de flux internationaux mondiaux décrivant l'ensemble des échanges commerciaux, en valeur (US dollars), réalisés entre des pays ou des groupes de pays. Les données sont issues de la base CIN de CHELEM du CEPII (2004, 2005, 2006). Préparées dans le cadre du projet Eurobroadmap (2009-2011), elles ont été tri-annualisées entre 2004 et 2006 et décrivent ainsi la moyenne mobile de l'ensemble des flux commerciaux internationaux sur les trois années. La Fig. 1 présente le niveau géographique élémentaire de description de ces flux.

Au début de la procédure, les 220 unités spatiales (pays ou groupes de pays) sont indépendantes et l'interaction intra régionale est de 0%. La densité de la matrice est de 36% dont 32% de cellules différentes de zéro (20% de l'ensemble total). Le nombre maximal d'étapes est de 172.

Les Fig. 2 et 3 présentent deux des niveaux de partition de l'espace mondial formés par la classification de ces flux commerciaux. On observe que à 51,5% d'interactions intra régionale (Fig.2), l'espace commercial mondial est formé de sept grandes régions homogènes que sont une région Amérique, une région Europe du nord, une région formée de l'Europe occidentale, septentrionale incluant les pays du pourtour méditerranéen et une grande région englobant le proche orient, l'Asie et Australie ; une région fédération de Russie et une région Afrique sub-saharienne. A 77% d'interactions intra régionale (Fig.3), seules trois grandes régions subsistent à savoir : les amériques-Asie du sud et Australie ; l'Europe ? Afrique du nord et Russie ; l'Afrique sub-saharienne. Ces régions peuvent ainsi être interprétées comme autant d'aires d'influence commerciales.

Une des questions principales consiste donc à choisir la fonction de régionalisation qui conviendra le mieux aux données de flux disponibles, en particulier les éventuelles contraintes spatiales liées à l'échelle d'expression des flux (locale, régionale, mondiale) et la disponibilité des données. Plusieurs fonctions ont ainsi été définies dans les années 1970 de manière à maximiser ou minimiser le pourcentage cumulé de ces flux intra-zones (ou interzones), une revue est proposée par S. Alvanidès (2000).

La méthode Intramax, développé initialement à l'échelle locale, est également efficace pour générer des partitions territoriales à l'échelle mondiale (voir Fig. 2 et 3.). Elle conduit à une forme de segmentation des valeurs de flux qu'il est possible de cartographier de manière classique : en symbolisant les relations qui se produisent entre des lieux appartenant à une même région et celles qui se produisent entre des régions, jouant alors le rôle de *hub* permettant de passer d'une région à une autre.

## 5 Conclusion

Le fait est que cette approche descriptive soit directement mise en oeuvre sur les valeurs de flux et qu'elle conduise in fine à un filtrage des valeurs par l'espace (il est possible de ne représenter que les valeurs qui s'expriment entre ou au sein de certaines régions) peut être

préjudiciable pour l'analyse. De même que l'on peut s'interroger sur les raisons de l'absence d'analyse factorielle réalisée au préalable de cette partition de l'espace par des flux.

En effet, la régionalisation fonctionnelle n'intègre pas de module de traitement multivarié qui consisterait à extraire préalablement les axes factoriels, en l'occurrence les composantes principales de ces flux que l'on pourrait interpréter comme un « résumé » des motifs principaux des flux. Quelles en sont les raisons ? Est-il possible de trouver une (ou plusieurs) méthodes complémentaires à ces classifications ?

Pour la poursuite de ces travaux exploratoires sur les méthodes multivariées liées à l'agrégation des données de flux, nous faisons l'hypothèse que l'extraction des composantes principales d'une matrice de flux bilatérale, éventuellement catégorielle ou temporelle, permettrait de résumer efficacement l'information qu'elle contient et d'en proposer une représentation cartographique simplifiée, tout en mobilisant l'ensemble des valeurs disponibles. Cette analyse serait d'ailleurs plus robuste, car fondée sur l'analyse de la variance de données dont on sait qu'elles sont fortement dissymétriques.

## 6 Références mentionnées

Alvanides S., Openshaw S., Duke-Williams O., 2000, Designing zoning systems for flows datas, In : Atkinson P., Martin D. (eds.), 2000, Geocomputation, Part II : Zonation and Generalization, Taylor and Francis Group, Ed. CRC Press, pp. 115-134.

Bahoken, F., (2011), Evolution des limites de régions fonctionnelles Européennes liées aux flux internationaux, Proceedings of the EuroBroadMap International Conference, Session C : « Frontières, réseaux et régionalisations floues du monde », Rouen, France, 1-3 Décembre 2011, pp. 14-20.

Bahoken, F., (2015), Eléments pour une représentation (carto)graphique des matrices de flux, M@ppemonde, n° 115.

Bahoken, F., (2016), Contribution à la cartographie d'une matrice de flux, Thèse de doctorat en géographie, Sciences des territoires, Université Paris-Sorbonne (Paris 7), 520 p.

Cattan, N. (2004), Le monde au prisme des réseaux aériens, Flux, vol. 4, n°58, pp. 32-43.

Coombs M. G., Green, A. E., Openshaw, S. (1986) : An efficient algorithm to generate official statistical reporting areas : The case of the 1984 travel-to-work-areas revision in Britain. Journal of the Operational Research Society, 37(10), pp. 943-953.

Fairbairn, D. (2006), Measuring Map Complexity, Cartographic Journal, vol. 43, n°3, pp. 224 - 238.

Grasland, C. (1997), Contribution à l'analyse géographique des maillages territoriaux, Habilitation à Diriger des Recherches en géographie, Volume inédit, Université Paris I, 373 p.

Grasland, C., Bahoken, F., Beauguitte, L., Pion, G., Van Hamme, G. (2009), Toolbox for flows and network analysis (Methodological Paper). Deliverable D.5.1. EuroBroadMap.Vision of Europe in the World. Small or medium scale focused project FP7-SSH-2007-1, 100 p.

Holten, D. (2006), Hierarchical edge bundles : Visualization of adjacency relations in hierarchical data, Proceedings of the IEEE Transactions on Visualization and Computer Graphics, vol. 12, n°5, pp. 741-748.

Holten, D., Van Wijk, J. (2009), Force-Directed Edge Bundling for Graph Visualization, in : Hege, H.-C., Hotz, I., Munzner, T., (Guest editor), Journal compilation of The Eurographics Association /IEEE-VGTC Symposium on Visualization, Vol.28, n°3, pp. 983-990.

INSEE (1998) : « Les zonages : enjeux et méthodes » - INSEE Méthodes n°83 - décembre 1998

Lambert, A., Aubert, D., Mélançon, G. (2010), Living flows : enhanced exploration of edge-bundled graphs based on GPU-intensive edge rendering, Proceeding of the 14th Interna-



tional Conference on Information Visualization, pp. 523-530.

Masser, I., Brown, P.J.B. (1975), Hierarchical aggregation procedures for interaction data, *Environment and Planning*, A7, pp. 509-523.

Openshaw, S., (1977) A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling, *Revue Transactions of the institute of british geographers*, Institute of British Geographers, pp. 459-472.

Tobler, W. (1969), Geographical Filters and their Inverses, In : *Geographical Analysis*, Vol. 1, n° 3, pp. 234-253.

Tobler, W. R., (n.-d.) On viewing flow map, *Note de recherche non datée*, 6 p.

## **Summary**

One of the main families of solutions for reducing the complexity of flow mapping (the cluttering of a flowmap) is to reduce the spatial resolution of flow data. Solutions linked to spatial aggregation procedures lead to a change of geographical scale, they produce a territorial framework at a regional level, a functional regionalization. If the approach is efficient, we want to question the reasons for the absence of factorial analysis carried out before these partitions of space by flows.



FIG. 1 – *Le découpage du monde selon CHELEM.*

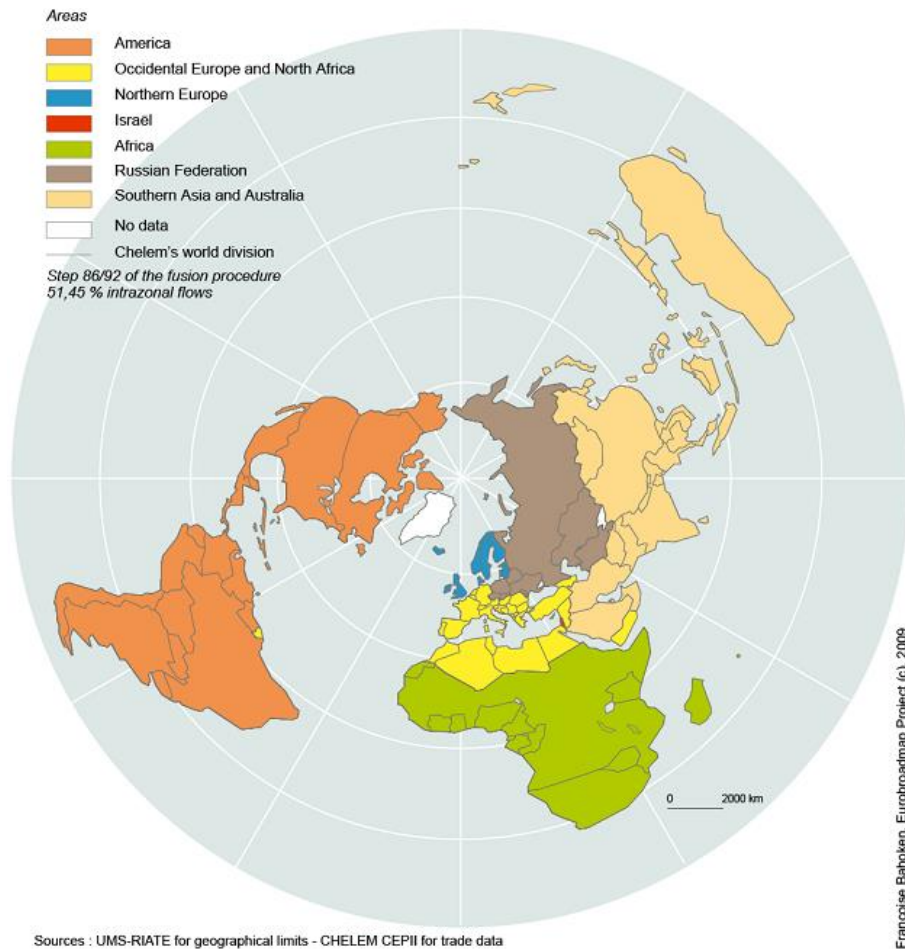


FIG. 2 – Exemples de partition de l'espace mondial issue d'une régionalisation fonctionnelle par des flux commerciaux mondiaux : 51,45% de flux intra zonaux cumulés

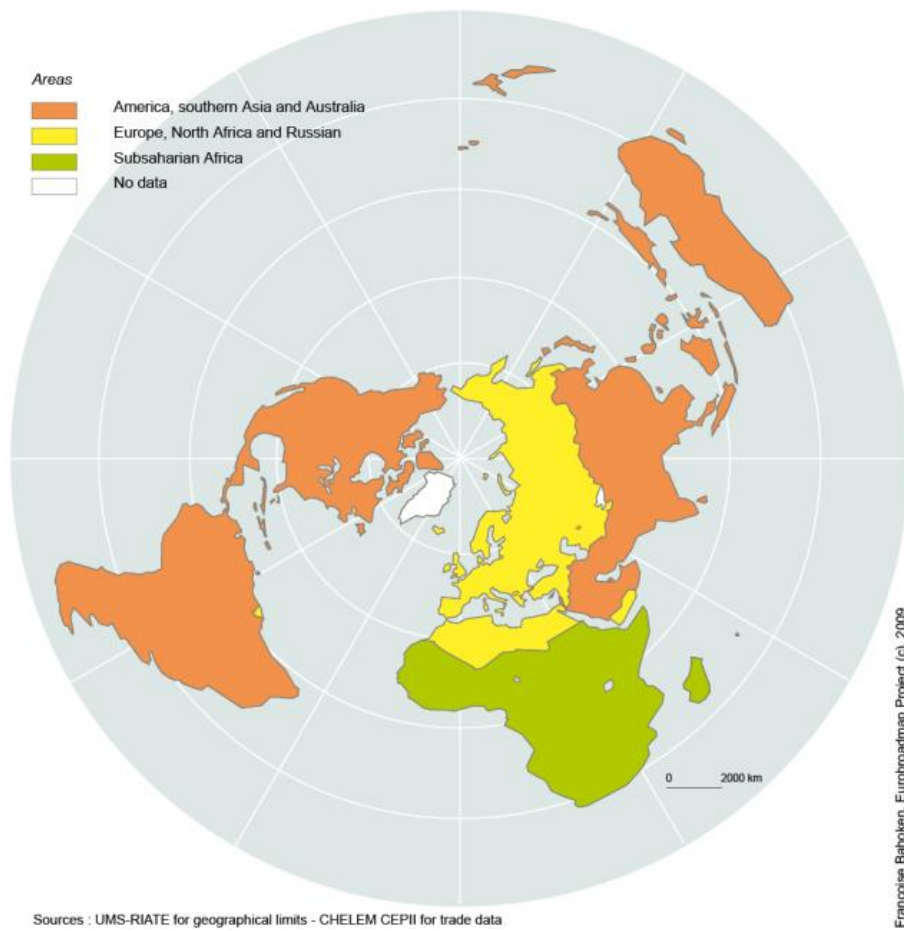


FIG. 3 – Exemples de partition de l'espace mondial issue d'une régionalisation fonctionnelle par des flux commerciaux mondiaux : 77,04% de flux intra zonaux cumulés