

LA PRISE EN COMPTE DE L'INCERTITUDE DANS UNE DEMARCHE DE MODELISATION PREDICTIVE

Reine Maria Basse

Doctorante, UMR 6012 ESPACE, CNRS - Equipe de Nice, basse@unice.fr
98 bd Edouard Herriot - BP 3209, 06204 NICE Cedex

Résumé

Les systèmes territoriaux sont composés d'éléments visibles et invisibles qui entrent en interaction et révèlent le dynamisme des territoires. Ces systèmes sont souvent caractérisés par des incertitudes du fait de leur fonctionnement complexe. Et puisque la modélisation aide à la connaissance et à la compréhension du fonctionnement des différents éléments qui entrent en interaction dans l'espace géographique, il est nécessaire d'utiliser des modèles ayant la capacité de prendre en compte ces incertitudes. La modélisation prédictive, à partir de la théorie des évidences de Dempster-Shafer et de la théorie de Dezert-Smarandache est un moyen de prendre en compte incertitudes, paradoxes et imprécisions en vue d'une prise de décision.

Mots clés : Modélisation prédictive, incertitude, territoire transfrontalier, interactions entre Territoire et Transports, LGV PACA, théorie des évidences de *Dempster-Shafer*, théorie de *Dezert-Smarandache*

A l'heure où le réseau ferroviaire européen à grande vitesse se densifie pour mieux rapprocher les territoires, la région Provence-Alpes Côte d'Azur, pour être connectée à toute l'Europe, a lancé une pré-étude fonctionnelle en 2000 sur le projet intitulé « ligne à grande vitesse LGV-PACA », qui doit relier la région PACA et la région Ligure en Italie. C'est dans ce contexte de « projet transfrontalier » que cette recherche doctorale¹⁶ intitulée « les enjeux transfrontaliers de la LGV PACA », a pour objectif final de mettre à la disposition des acteurs, des outils ainsi que des connaissances d'aide à la décision dans le cadre de l'aménagement du territoire transfrontalier.

Il s'agit, d'envisager le devenir du territoire transfrontalier à partir d'un objet encore inexistant, c'est-à-dire la LGV-PACA. On suppose au préalable de la recherche, que le TGV a des effets sur l'organisation d'un territoire. La question est : comment mettre en évidence ces effets, surtout si l'infrastructure n'existe pas encore.

Il existe des méthodes d'analyse en géographie qui permettent d'appréhender le devenir des territoires. Parmi celles-ci, on trouve la géoprospective et/ou la démarche prédictive. Ces deux approches ont pour finalité la production d'images du futur. Toutefois, elles n'ont pas les mêmes objectifs. L'approche

prospective cherche à anticiper l'évolution à moyen et long termes d'un phénomène à partir de la situation présente. Quant à l'approche prédictive, elle a pour vocation la prévision à court terme. Elle implique de sortir des approches mathématiques, tels les modèles cellulaires (automates cellulaires) et les modèles d'évolution (systèmes multi-agents) pour la concevoir. De plus, la modélisation prédictive inclut dans sa démarche l'incertitude souvent liée à l'objet d'étude lui-même. L'échelle temporelle et la prise en compte de l'incertitude en amont de la modélisation différencient donc ces deux approches qui peuvent être complémentaires.

Principalement utilisée en traitement d'images¹⁷ la modélisation prédictive est encore peu utilisée en sciences sociales. Du fait de sa capacité à prendre en compte l'incertitude qui entoure l'objet d'analyse, l'approche prédictive présente un intérêt certain dans l'évaluation des impacts à court terme de l'infrastructure ferroviaire. Il est donc important de se demander comment, tout en tenant compte de ces incertitudes « prédire » le devenir de cet espace ? Deux théories, qui procèdent de l'approche prédictive semblent apporter des éléments de réponse. Il s'agit de la théorie des évidences de *Dempster-Shafer* [Shafer, 1976] et de la théorie de *Dezert-Smarandache* [Dezert, 2003]. Elles ont toutes les deux l'avantage

¹⁶ La thèse intitulée « LGV PACA, les enjeux transfrontaliers : développement territorial et relations avec la Principauté de Monaco et l'Italie » est sous la direction de Christine Voiron, professeur à l'Université de Nice Sophia-Antipolis et directrice de l'UMR 6012 Espace.

¹⁷ Bloch. I., Maître H., 1994. *Fusion de données en traitement d'images : modèles d'information et décisions*. Traitement du signal, vol. 6.

de prendre en compte la notion d'incertitude dans le processus de modélisation.

C'est en trois sections principales que s'organise ce texte. Dans la première section, nous expliciterons la notion d'incertitude. Dans la seconde section, nous montrons la manière d'appréhender l'incertitude par rapport à notre espace d'étude. Enfin, la dernière section s'attache à présenter les applications possibles de la modélisation prédictive dans notre recherche.

I. L'incertitude dans la prospective territoriale franco-italienne

Pour comprendre le fonctionnement des territoires, il est important de tenir compte de tous les facteurs visibles et/ou invisibles qui entrent en interactions. Ces facteurs, souvent complexes sont généralement difficiles à analyser du fait des incertitudes qui peuvent les caractériser.

1.1. Définition de l'incertitude

L'incertitude est une notion généralement définie comme étant soit [1] *identifiable*, et dans ce cas le terme fait allusion aux données et aux mesures (échantillonnage, erreurs de mesures, nomenclature...) dont dispose le chercheur pour étayer son analyse; soit [2] *non déterminée*, ici, elle désigne l'objet étudié (dynamique temporelle, échelle d'analyse, modèle utilisé, connaissance, continuité et/ou hétérogénéité de l'objet étudié,...). [Hatem, 1993].

Pour bien distinguer les deux définitions, le terme « imprécision » est aussi utilisé pour faire référence à « l'incertitude identifiable ». On parle alors d'incertitude et d'imprécision en modélisation prédictive.

L'incertitude ne doit donc pas être comprise comme étant de « l'ignorance » ; elle est ce qui est difficilement quantifiable, non mesurable, et par conséquent, délicat à intégrer dans le processus de modélisation. Dans notre recherche, l'incertitude est présente à trois niveaux. Elle est liée aux données disponibles, à la connaissance du fonctionnement de la zone transfrontalière elle-même et enfin au projet de transport.

1.2 L'incertitude dans la connaissance de l'espace transfrontalier franco-italien et franco-monégasque.

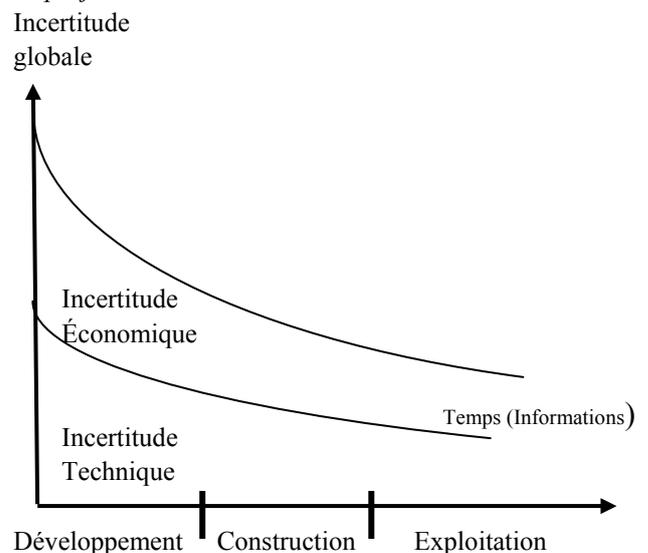
L'étape de la connaissance du fonctionnement de l'espace transfrontalier est complexe et nécessite la prise en compte

d'informations hétérogènes (données économiques, flux de mobilité...). Les limites des informations disponibles sur l'aire d'étude (informations non exhaustives, inexistantes, non homogènes...) constituent un handicap pour la connaissance du fonctionnement réel du territoire transfrontalier. C'est sous cette première forme que l'incertitude est présente dans l'aire d'étude.

1.3 Incertitudes dans le projet de transport LGV PACA

En France, l'infrastructure de transport grande vitesse est généralement débattue à des échelles de temps très importantes, entre quinze et vingt ans avant son exploitation. Cette nécessité de temps long conforte les incertitudes qui existent autour des projets à grande vitesse ferroviaire. En effet, comme le montre la figure 2, entre la phase du développement du projet et la phase de l'exploitation, il y a une grande place occupée par l'incertitude. On constate d'abord des incertitudes d'ordre technique, liées à l'insertion du projet dans un espace marqué par des enjeux environnementaux forts (nombreux massifs montagneux, urbanisation dense...). Les incertitudes sont aussi économiques : en effet, les coûts estimés (entre 5 et 8 milliards d'euros pour la LGV PACA), amènent très souvent les acteurs à « geler » le projet ou à le modifier ; il n'existe alors aucune certitude que le projet arrive au stade de la construction avec ses caractéristiques initiales. Mais avec le temps l'incertitude tant à diminuer.

Figure 2 : Les incertitudes autour du projet LGV PACA



Source : IMRI, Université Paris Dauphine, Prédit, 1996-2000

En géographie, les systèmes de transports sont souvent abordés à partir d'analyse des réseaux. L'approche que nous proposons est différente et innovante car elle apporte un élément d'analyse novateur : elle permet au géographe d'aborder la notion d'incertitude et de produire une évolution à court terme d'une dynamique.

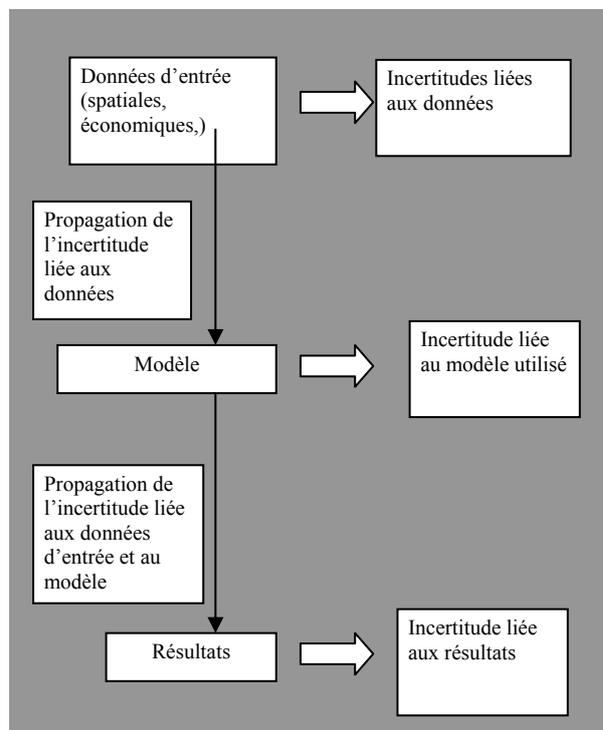
L'incertitude est la notion fondamentale de la modélisation prédictive, il est dès lors nécessaire de montrer la manière de le prendre en compte dans le processus de modélisation.

II. Place de l'incertitude dans la démarche prédictive

2.1 Incertitude, analyse spatiale et modélisation prédictive

L'analyse spatiale utilise des informations hétérogènes (qualitatives et quantitatives) et multisources, ainsi que des outils de plus en plus performants (télédétection, SIG, contrôle terrain...), en vue d'une prise de décision. Ces outils ont offert de nouvelles perspectives à l'analyse spatiale dans sa vocation à connaître le fonctionnement des territoires. Toutefois, des difficultés de gestion des bases de données apparaissent, et sont liées aux erreurs de mesures, aux contenus et à l'exactitude des informations. La gestion des imprécisions et/ou incertitudes est devenue un enjeu majeur pour l'analyse spatiale. La figure 3 indique la présence de l'incertitude dans les différentes phases de la modélisation en analyse spatiale. Trois niveaux d'incertitude doivent être pris en compte dans une démarche de modélisation prédictive. La première forme d'incertitude se situe au niveau des données spatiales. Face à cette forme d'imprécision, il existe deux méthodes : la première consiste à ne prendre en considération que les données les plus fiables ; la seconde pondère chaque donnée d'entrée en fonction de sa fiabilité à l'aide de méthodes géostatistiques ou statistiques. La deuxième incertitude est liée au modèle lui-même. Elle dépend beaucoup du paramétrage. Le modélisateur doit tenir compte de la complexité du modèle, mesurer sa fiabilité et minimiser ainsi la propagation des erreurs. La dernière incertitude se trouve au niveau des résultats, et dépend fortement de celles situées en amont du processus de modélisation (données d'entrée et modèle utilisé). La crédibilité des résultats est souvent vérifiée à partir d'une comparaison ou d'une simple vérification de terrain.

Figure 3 : Prise en compte de l'incertitude dans les phases de modélisation prédictive



Source : Corgne.S, 2004.

2.2 Apport de la prise en compte de l'incertitude en géographie

En géographie, la prise en compte de l'incertitude est intéressante pour minimiser la propagation des erreurs dans le cadre de l'analyse. Elle peut aussi contribuer à mieux mettre en évidence les interactions qui existent entre les différents éléments qui se situent dans l'espace géographique et dont la connaissance du fonctionnement est souvent incertain. Il est dès lors important de porter la réflexion sur des modèles dont la structure intègre le caractère incertain des phénomènes. C'est le cas des modèles experts utilisés en géographie (système expert, modèles probabiliste bayésiens, les ensembles flous...). Toutefois, seuls les modèles basés sur les théories de *Dempster-Shafer* et sur celle de *Dezert-Smarandache*, sont en mesure de gérer à la fois et de façon distincte l'incertitude liée au phénomène étudié et l'imprécision liée aux données d'entrée spatiale et au modèle. [Corgne S, 2004].

2.3. Présentations des théories des évidences de *Dempster-Shafer* et de celle de *Desert-Smarandache*

La théorie des évidences de *Dempster-Shafer* a fait l'objet d'applications dans des domaines comme la médecine ou l'aéronautique. Elle est récemment utilisée par les géographes, notamment en géographie physique où les applications ont été orientées sur le thème de l'occupation et de l'utilisation des sols.

L'utilisation de cette théorie dans le contexte agricole a permis de prédire la présence d'un couvert végétal de parcelles agricoles, affectées par un phénomène d'eutrophisation [Corgne, S, 2004]. Le modélisateur a utilisé des sources d'information différentes (images satellites, données économiques, expertises...) pour soutenir les hypothèses émises au départ : présence ou absence de couvert végétal. Ces hypothèses indiquent au départ de l'analyse, le caractère incertain du phénomène. Un niveau d'incertitude est attribué à chaque source d'information en fonction de son degré de pertinence.

Le modèle utilisé sur le bassin versant du Yar en Haute Bretagne basé sur la théorie des évidences, a permis de faire une distinction entre les prédictions erronées et celles qui sont correctes. La théorie de Dempster-Shafer est appréhendée à partir de fonctions de plausibilité¹⁸ et de croyance¹⁹. Elle présente l'inconvénient majeur de produire des résultats contrastés quand les données d'entrée ne sont pas que de type statistique. C'est pour gérer ce problème de fusion²⁰ de données hétérogènes, que la théorie de *Desert-Smarandache* a été développée en 2003. Elle présente l'avantage de pouvoir améliorer les scores de prédiction et d'optimiser ainsi les résultats.

Pour notre recherche, la modélisation prédictive à partir de la théorie de *Dempster-Shafer* et de la théorie de *Desert-Smarandache*, doit permettre de mettre en évidence les effets à court terme de la LGV PACA sur le territoire transfrontalier.

III Modélisation prédictive : quelles applications possibles aux enjeux transfrontaliers de la LGV PACA ?

Comme spécifié au début de la réflexion, « prédire » le devenir du territoire transfrontalier au lendemain de la mise en service de la LGV PACA,

exige au préalable de connaître son fonctionnement, c'est l'étape du diagnostic territorial. Les résultats du diagnostic ont permis d'acquérir des connaissances satisfaisantes sur l'organisation du territoire et justifient le choix des hypothèses suivantes.

3.1 Les hypothèses de la recherche

On suppose au départ de la modélisation prédictive que la future infrastructure grande vitesse aura des effets sur les formes de mobilité : mobilité résidentielle et mobilité récréative et sur l'accessibilité des territoires situés dans la zone transfrontalière. Concernant l'étape de la modélisation prédictive, pour chaque hypothèse émise, un niveau d'incertitude sera attribué afin de mesurer le degré de pertinence des sources à étayer les hypothèses.

Pour étayer ces hypothèses et conduire la modélisation prédictive, différentes données sont nécessaires.

3.2 Echelles spatiales

L'imbrication des territoires et son incidence sur la prospective territoriale amène à considérer différentes échelles spatiales. Au moins quatre échelles d'analyse ont été détectées : l'échelle locale qui concerne les territoires situés à proximité des frontières, c'est la zone transfrontalière elle-même. L'échelle régionale qui traite de la question des effets structurants du projet LGV PACA au niveau de la région. L'échelle nationale qui traite des relations entre la zone transfrontalière et les grands centres nationaux de décision et enfin l'échelle européenne qui aborde la question du transit marchandises se situe entre l'Italie la France et L'Espagne. Pour chaque échelle spatiale les enjeux sont différents ce qui crée des conflits d'enjeux territoriaux et rend complexe toute démarche prospective et ou prédictive.

La modélisation prédictive sera appliquée à l'échelle locale c'est-à-dire au niveau des communes situées à proximité de la frontière franco-italienne et franco-monégasque. Cette échelle est pertinente pour ce type de modélisation.

3.3 Données mobilisées

Les données mobilisées dans l'étape de la modélisation sont les suivantes :

- Données de type socio-économiques : population, emploi, flux et migrations, mobilité résidentielle, fréquentation touristique développement des entreprises, enquêtes ménages, enquêtes logements...

¹⁸ C'est-à-dire, la plausibilité que la vérité soit dans l'hypothèse A plutôt que dans l'hypothèse B)

¹⁹ C'est-à-dire la certitude que la vérité soit dans A plutôt que dans B.

²⁰ La fusion d'information consiste à combiner des informations issues de sources différentes afin d'améliorer la prise de décision [I.Bloch].

- Données de type spatiaux : les réseaux de transport, l'état du trafic, état de l'accessibilité des territoires, effets frontière...
- Données qualitatives : avis d'experts et estimations.

3.4 Traitements envisagés

Différents traitements sont envisagés dans l'étape de la modélisation prédictive. Parmi ceux-ci :

1. L'évolution de la population entre 1982 et 1999.
2. L'évolution possible de l'accessibilité des territoires frontaliers vers les grands pôles de décisions, régionaux, nationaux et européens à partir de la future gare TGV.
3. L'évolution possible de la mobilité résidentielle des populations.
4. Et enfin un dernier traitement sur la mobilité récréative. En effet la zone géographique étudiée est fortement touristique. L'amélioration de la fréquence touristique du fait de l'implantation de la gare TGV sera donc estimée.

3.5 Outils utilisés et formes de résultats et

L'outil SIG, la télédétection et les traitements géostatistiques restent les moyens les plus pertinents pour mettre en évidence les résultats attendus de la modélisation prédictive. Les Formes de résultats pour ces traitements seront de type cartographiques.

Conclusion

La prise en compte de l'incertitude à partir de la modélisation prédictive, est une piste de recherche intéressante pour l'analyse spatiale, qui a pour vocation, la compréhension des phénomènes qui entrent en interaction et structurent l'espace géographique.

L'analyse spatiale a aussi pour objectif de participer au devenir des territoires, en mettant à la disposition des acteurs et gestionnaires, des outils et des connaissances d'aides à la décision. C'est dans ce cadre que la thématique étudiée « les enjeux transfrontalier de la LGV PACA » se situe et cherche à offrir aux acteurs et gestionnaires, une vision globale des effets probables de la grande vitesse ferroviaire en vue de l'aide à la décision.

La prise en compte de l'incertitude à partir de la théorie des évidences de *Dempster-Shafer* et de *Dezert-Smarandache*, dans les étapes de la

modélisation est un moyen parmi d'autres de s'inscrire dans les objectifs du champ disciplinaire.

Bibliographie

- Bloch.I., Maître H., *Fusion de données en traitement d'images : modèles d'information et décisions*. Traitement du signal, vol. 6. 1994
- Corgne.S, *Modélisation prédictive de l'occupation des sol en contexte agricole intensif : application à la couverture hivernale des sols en Bretagne*, thèse 2004, 230 pages
- Dezert.J , *Fondation pour une nouvelle théorie du raisonnement plausible et paradoxale. Application à la fusion d'information incertaines et conflictuelles*, Rapport technique, ONERA, 2003, 76 pages.
- PREDIT, *Choix d'investissement dans les projets de rupture technologique et forme organisationnelle*. CD-1996-2000
- Hatem.F, *La prospective : pratiques et méthodes*, Paris, Economica, 1993, 385 pages
- Shafer.G., *A mathematical Theory of Evidence*. Princeton University press, 1976
- Smarandache.F, Dezert.J, *Advances and applications of DSMT for information Fusion*, Collected works, Américan Research Press, Rehoboth, 2004, 418 pages.