

# SADIC : SYSTÈME D'AIDE À LA DÉCISION EN INFORMATIQUE CARTOGRAPHIQUE

**Dominique LAFFLY**

**Pascal GILLON**

LEP-IRADES, URA 908 du CNRS  
Université de Franche-Comté  
32, rue Mégevand  
25030 BESANCON Cedex  
Tel 81 66 53 31  
Tel 81 66 54 02

## Résumé

SADIC est un logiciel qui gère et filtre l'information sur des bases de données thématiques de grande taille. Il intervient en amont du traitement de l'information et permet ainsi de valider ou d'infirmer rapidement des hypothèses de travail. Son principe repose sur une grande facilité d'utilisation quant aux choix des unités et des niveaux d'agrégation spatiale, des données à représenter et des modes d'analyses statistiques grâce à une interface graphique conviviale et à une mise en forme rapide des résultats cartographiques. SADIC est développé de manière à pouvoir s'adapter et évoluer facilement, notamment pour intégrer la dimension temporelle et la portabilité pour d'autres sources d'information.

## Mots clés

Agrégation spatiale - Banque de données - Cartographie automatique - Ergonomie - Typologie

## Introduction

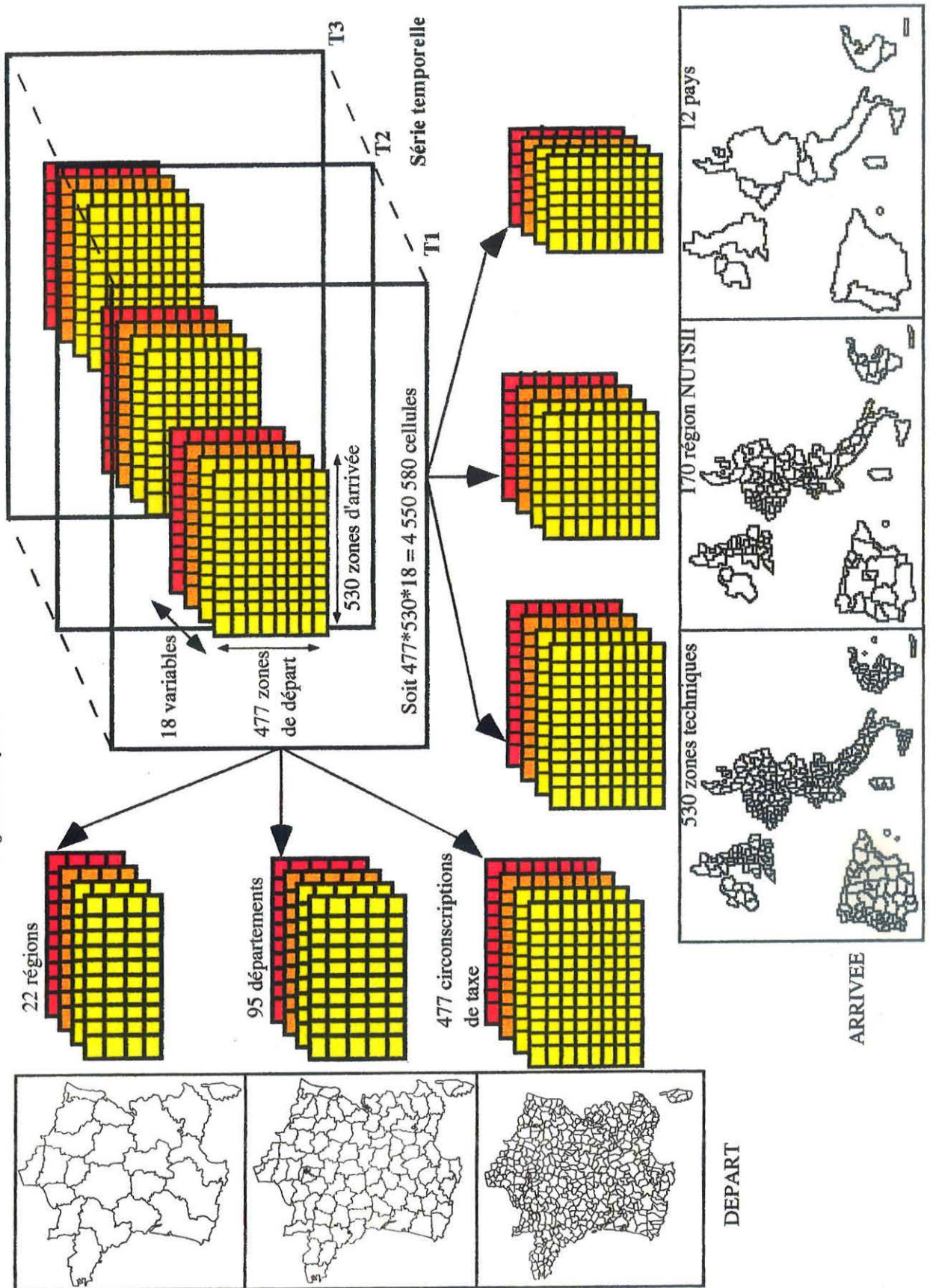
SADIC a été développé pour France Télécom, à l'occasion d'une étude sur les flux téléphoniques internationaux France - Union Européenne. Jusqu'à présent, FT ne possédait que peu d'éléments d'information sur les directions de ces trafics. La précision spatiale dans la destination des flux ne descendait pas en deçà de l'échelle du pays.

France Télécom a donc ressenti le besoin de constituer une banque de données qui recense et décrit les flux, à partir de 18 variables, entre les villes françaises (477) et européennes (530). Du fait de sa taille considérable (plus de 4,5 millions de cellules), l'exploitation de celle-ci ne pouvait se réaliser sur des logiciels courants. Ils sont en effet incapables de gérer de telles bases à cause des problèmes spécifiques que leur volume impose. En effet, la masse des données contenues dans la base est souvent un obstacle en soi à la compréhension du phénomène décrit.

SADIC a donc pour but de gérer et de faciliter le filtrage de l'information sur des bases de données volumineuses. Il se situe en amont dans le traitement de l'information en permettant une consultation rapide sous forme cartographique de l'ensemble des données. Ce procédé facilite l'exploration de la base et permet d'infirmer ou de confirmer des hypothèses parmi l'ensemble des possibles, sans perdre trop de temps en manipulations.

Afin de répondre aux besoins des utilisateurs, SADIC intègre aussi les principales caractéristiques des logiciels de cartographie automatique. Il gère donc l'agrégation spatiale, la cartographie des flux, mais il peut être aussi utilisé dans

Figure 1 : Représentation de la base de données



un but de consultation de la base de données, ou comme un atlas permanent qui peut être réalimenté chaque fois que l'application nationale sera lancée.

Après avoir exposé les principaux aspects qui ont présidé à la réalisation de SADIC, quelques exemples d'application seront utilisés pour démontrer les possibilités du logiciel.

### ***La base de données***

SADIC a été conçu pour gérer la base issue de l'application informatique de France Télécom. Elle est constituée de fichiers livrés par les Centres de Facturation Régionaux (CFR) sous la forme de fichiers ASCII avec une volumétrie de 10 000 articles maximum. Un article est composé de 20 variables :

- deux variables pour décrire l'origine et la destination des appels avec un emboîtement des échelles,
- 18 variables qui recensent le trafic à différentes heures (jour-nuit), périodes (semaine ouvrable, week-end, jour férié) et sous différentes formes (nombre d'appels, durée, nombre de taxes).

Les variables qui décrivent le lieu de départ et le lieu d'arrivée sont codées sur trois niveaux d'échelle (une lettre pour le pays suivie de deux chiffres pour les régions et deux chiffres pour la zone la plus fine) :

- en France, le niveau le plus fin correspond aux circonscriptions de taxe (477), l'échelle intermédiaire est représentée par les départements (95), et les régions (22) forment le troisième niveau ;
- en Europe, le découpage le plus précis a été défini au niveau des zones techniques (530) d'après le plan de numérotation de chaque pays, les régions NUTSII (170) déterminent la seconde échelle, les pays (12) constituant le dernier échelon.

Une fois la base organisée, on obtient plus de 4,5 millions ( $477 \times 530 \times 18 = 4,55$  M) de cellules à gérer au niveau le plus fin, sans tenir compte des autres combinaisons possibles grâce à l'agrégation spatiale, ce qui représente plus de 40 MégaOctets au format ASCII.

Sachant que le logiciel doit en plus être capable de gérer une série temporelle (plusieurs mesures à des époques différentes), la majeure partie des logiciels commercialisés à un prix raisonnable est inadaptée, d'où la création de SADIC.

### ***Les lignes directrices de SADIC***

Le soucis «d'ouverture du logiciel» a été pris en compte dans la programmation de SADIC. Il ne s'agit pas de refaire un logiciel complet de cartographie, mais d'adapter une partie des outils que nous possédons déjà (Line-Up,...). Il s'insère donc dans le traitement de l'information grâce à des passerelles entre l'amont (importation de tableau de données et de fonds de cartes) et l'aval (exportation des cartes réalisées et des statistiques traitées).

L'aspect convivial de SADIC a été particulièrement soigné afin qu'un utilisateur occasionnel puisse s'en servir sans que le temps d'apprentissage soit important. En fait, l'interface graphique fait partie intégrante de la performance du logiciel. Pour faire une carte, toutes les opérations se réalisent à partir de la souris : il suffit de sélectionner les options dans les menus proposés.

Enfin, SADIC a été conçu de façon modulaire afin de pouvoir adjoindre des développements futurs sans avoir à refaire toute l'ossature du logiciel. Chaque étape du travail est ainsi intégrée dans un module et l'ouverture du logiciel permettra l'adjonction de différents modules.

### ***L'organisation générale de SADIC***

SADIC présente trois parties bien distinctes qui reflètent les principales étapes de la cartographie.

La première partie est composée de deux modules d'acquisition des données, l'un pour les fonds de cartes, l'autre pour les valeurs à cartographier.

La seconde partie constitue le corps proprement dit du logiciel. Il est formé du module d'interrogation (choix de la variable à cartographier, des zones de départ et d'arrivée), du module de traitement statistique (plusieurs choix de découpage possibles) et du module de visualisation.

La troisième partie concerne l'exportation des différents résultats durant les phases de traitement. Nous n'avons donc pas créé un module spécifique d'exportation, mais la possibilité en est offerte dans la plupart des modules.

### *L'acquisition de données*

L'acquisition concerne deux types de données : d'une part les fichiers de fonds de cartes et d'autre part les fichiers de données.

Pour les fonds de cartes, SADIC gère uniquement le format vectoriel qui a été choisi pour sa facilité de gestion et sa rapidité d'affichage. Deux options sont alors possibles :

- si l'utilisateur de SADIC possède déjà des fonds vectoriels, ils peuvent être récupérés s'ils sont compatibles avec le format SAS (format classique et répandu) ;
- si l'utilisateur ne possède pas de fonds, il devra les créer. Pour notre part, nous nous sommes servi d'un logiciel de vectorisation (Line-Up) qui reprend un fond scanné (raster) et le transforme en vectoriel.

Pour les données numériques, le module convertit le format de départ (ASCII par exemple) au format propre à SADIC et permet l'extraction et l'organisation de la base. Chaque entité géographique, susceptible d'émettre un flux, est intégrée dans un fichier spécifique. Le nom du fichier est identifié par le numéro de code de l'entité émettrice.

### *Le corps du programme*

Comme son nom l'indique, le module de visualisation fonde toute la démarche d'interrogation de la base sur un mode visuel. Une session complète (choix du fond de carte jusqu'au résultat final) s'effectue par les choix suivants :

- La première étape consiste à déterminer la zone d'où sont émis les flux ainsi que le niveau d'agrégation (finesse du découpage) auquel on travaille. SADIC présente alors le fond de carte et l'utilisateur n'a plus qu'à sélectionner la ou les zones de départ en cliquant dessus. En effet, il peut être intéressé par l'émission de flux de la part d'une entité spécifique, mais aussi par un groupe d'entités qui ne recouvre aucun niveau d'agrégation effectif.

SADIC offre ainsi la possibilité de réaliser une collection de cartes (si l'on a sélectionné cinq zones de départ, on pourra obtenir une carte par entité) ou une carte qui représente la somme des flux en partance des zones sélectionnées.

Il sélectionne ensuite les directions d'arrivée des flux en procédant de la même façon (choix des pays et du niveau d'agrégation) avec le choix entre une et plusieurs entités.

Une fois la direction des flux déterminée, il suffit de sélectionner la variable à visualiser parmi toutes les variables du tableau de données.

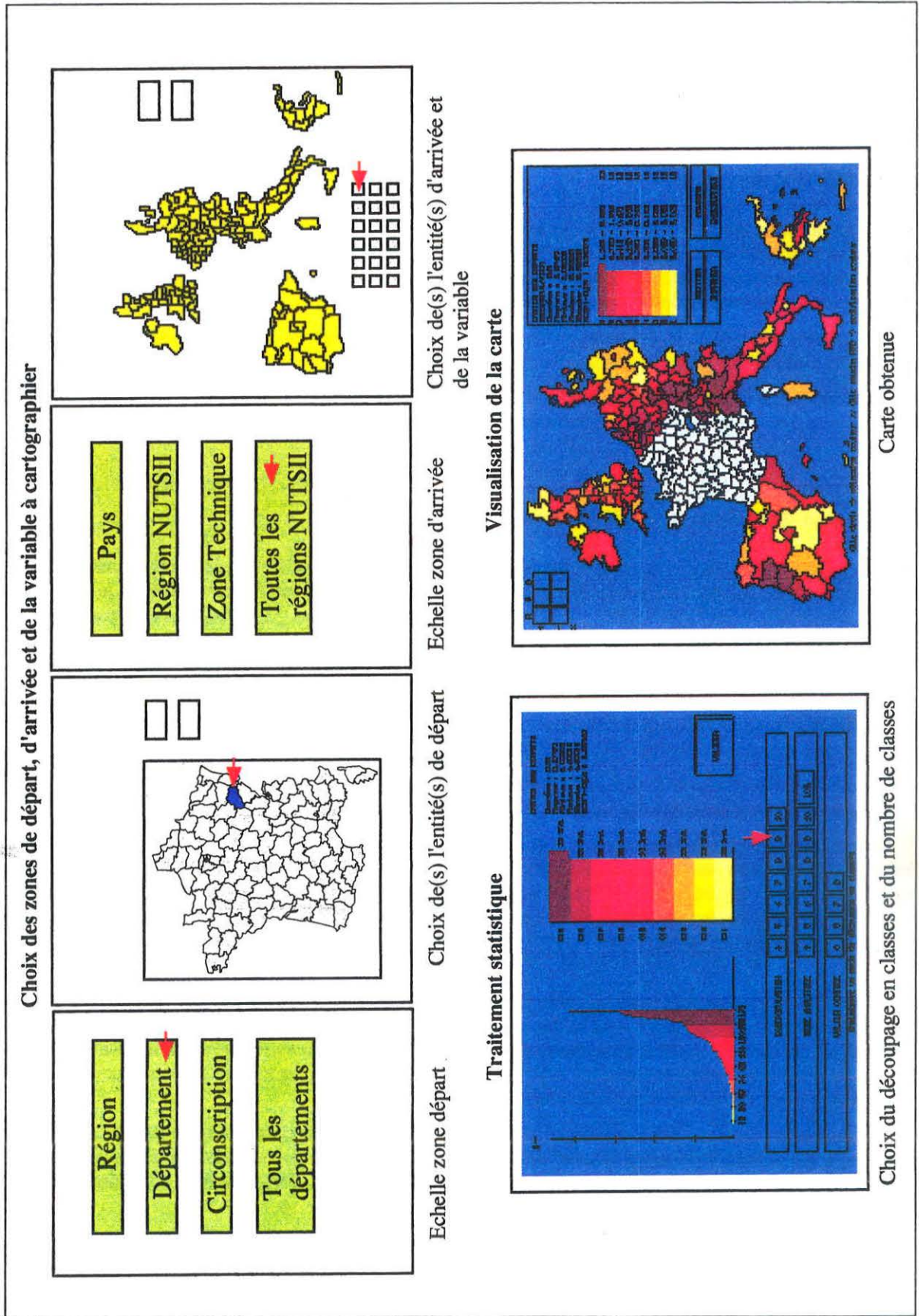
- L'analyse statistique représente la seconde étape dans le traitement de l'information. SADIC ne présente pas une batterie importante de tests statistiques, le but n'étant pas de produire une analyse fine mais de filtrer l'information. Le module permet ainsi de réaliser plusieurs types de discrétisations correspondant aux distributions statistiques rencontrées les plus fréquemment. Pour faciliter le découpage en classes, SADIC affiche deux histogrammes qui indiquent d'une part la distribution observée et d'autre part le nombre d'individus par classe lorsque le choix du nombre de classes a été défini. On peut tester autant de découpages que l'on veut avant de valider celui qui apparaît comme le plus pertinent.

- La troisième étape est constituée par l'affichage du résultat avec la carte et différents paramètres relatifs à la série statistique. Une palette de couleur standard a été définie, mais il est possible de modifier les teintes.

La présentation des données pose des problèmes de sémantique graphique. La représentation des flux s'opère souvent par un trait de taille ou de couleur variable qui relie la zone de départ à la zone d'arrivée.

Pour notre part, nous avons évité les traits pour des motifs techniques et parce que cette représentation ne correspond pas à toutes les situations de représentation cartographique envisageables. En effet, SADIC permet la cartographie de flux issus de plusieurs entités spatiales : le choix du barycentre des polygones (départ de la flèche) est à ce moment-là complètement subjectif. Nous avons donc préféré identifier la ou les zones de départ par une trame et

Figure 2 : Schématisation d'une session de traitement



représenter les flux reçus par les entités sous forme de cercles proportionnels pour les valeurs brutes, ou sous forme de trames lorsqu'il s'agit de valeurs pondérées.

Enfin une session de traitement ne serait pas complète sans l'aspect de transfert de l'information. Une fois que le découpage le plus pertinent a été défini, il s'agit de transférer les données et les résultats cartographiques afin de pouvoir cette fois-ci explorer plus à fond l'hypothèse de travail et améliorer l'illustration.

### ***Le module d'exportation***

L'exportation des données peut s'effectuer à trois étapes du processus de cartographie. On peut tout d'abord exporter le fichier brut de données sur lequel on a effectué des traitements statistiques simples, dans le but de le soumettre à des traitements plus complexes. L'exportation se réalise aux formats les plus courants afin de pouvoir réintroduire ces données dans des tableurs ou des logiciels de statistiques.

La deuxième possibilité d'exportation s'offre au moment du découpage en classes. On peut en effet récupérer les principaux paramètres de la série statistique (moyenne, écart-type ...).

Enfin, lorsque la carte est réalisée, il est possible d'exporter celle-ci par un procédé de copie d'écran. Le fichier ainsi créé peut être récupéré sous des logiciels de traitements d'images tel que Photoshop.

### ***Une session de traitement de l'information***

La constitution de la base de données a été réalisée afin d'améliorer à terme la prévision des flux entre les villes. Les recherches actuelles ont déjà souligné les relations qui existent entre les différents facteurs qui sont à la base des appels. En effet, quatre populations spécifiques engendrent les flux sur l'international : les entreprises qui tissent des relations avec des partenaires étrangers, les émigrants qui conservent des liens avec leur pays d'origine, les frontaliers qui ont tendance à communiquer avec leurs voisins, les résidents français ayant de la famille ou des amis à l'étranger, enfin les touristes.

Chacun de ces trafics utilise majoritairement des plages horaires spécifiques et des directions particulières, ce qui permet de dissocier assez nettement les flux d'entreprises des flux des émigrés et des touristes.

La période d'observation de la base de données, s'étendant entre la mi-juillet et la mi-août 1993, permet de focaliser notre attention sur le phénomène touristique au moment où il est le plus intense.

Une des hypothèses que nous allons tenter d'explorer concerne les lieux de séjour des touristes européens en France. Nous voudrions savoir si chaque nationalité exprime des préférences marquées pour une région touristique française ou s'il y a équirépartition entre les différentes régions. Pour chaque pays européen, l'attrait d'une zone touristique française s'effectue-t-il partout de façon indifférenciée ou bien assiste-t-on à l'apparition de préférences régionales pour un lieu de séjour ?

Dans un premier temps, on extrait les trafics de jour et de nuit au départ d'un département touristique afin de voir si la distinction des appels de jour (8 h-18 h) des appels de nuit (18 h-8 h) est pertinente. En effet, le trafic de jour est surtout le fait des entreprises et des administrations eu égard aux horaires d'ouverture des bureaux (seul le fax programmable peut être utilisé comme palliatif) ; le trafic de nuit correspond plutôt à un trafic de relations sociales émis par les particuliers, fortement incités par la tarification avantageuse.

La comparaison des cartes a et b de la figure 3 confirme l'hypothèse de la séparation horaire du trafic. Les relations de la Dordogne avec le reste de l'Europe montre l'évolution du trafic en fonction des horaires. Les directions principales se maintiennent (appels les plus nombreux vers les Pays-Bas et le Royaume-Uni) mais avec cependant des variations importantes.

Figure 3 : Trafic émis de Dordogne en direction du Nord de l'Europe de jour et de nuit

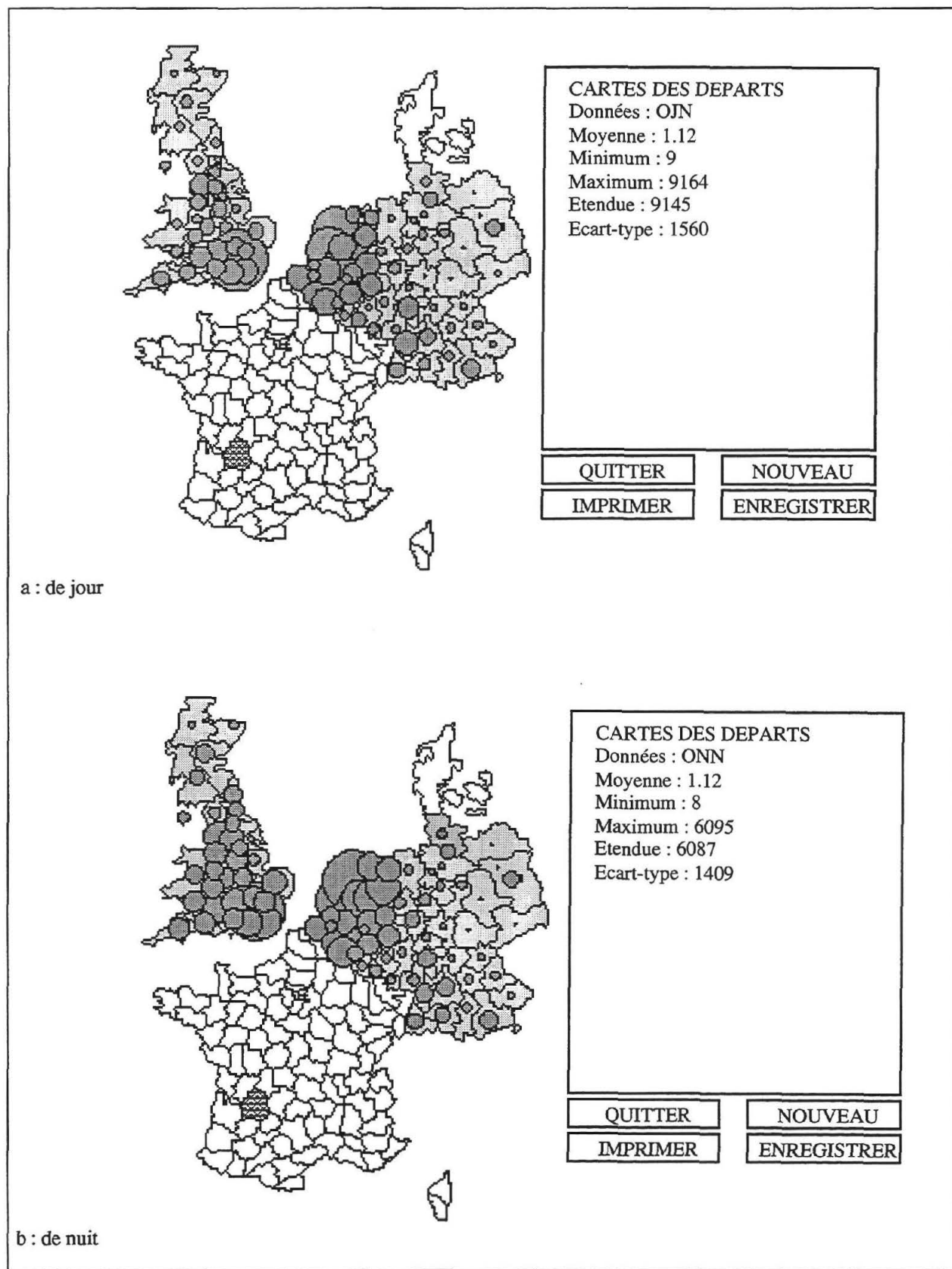
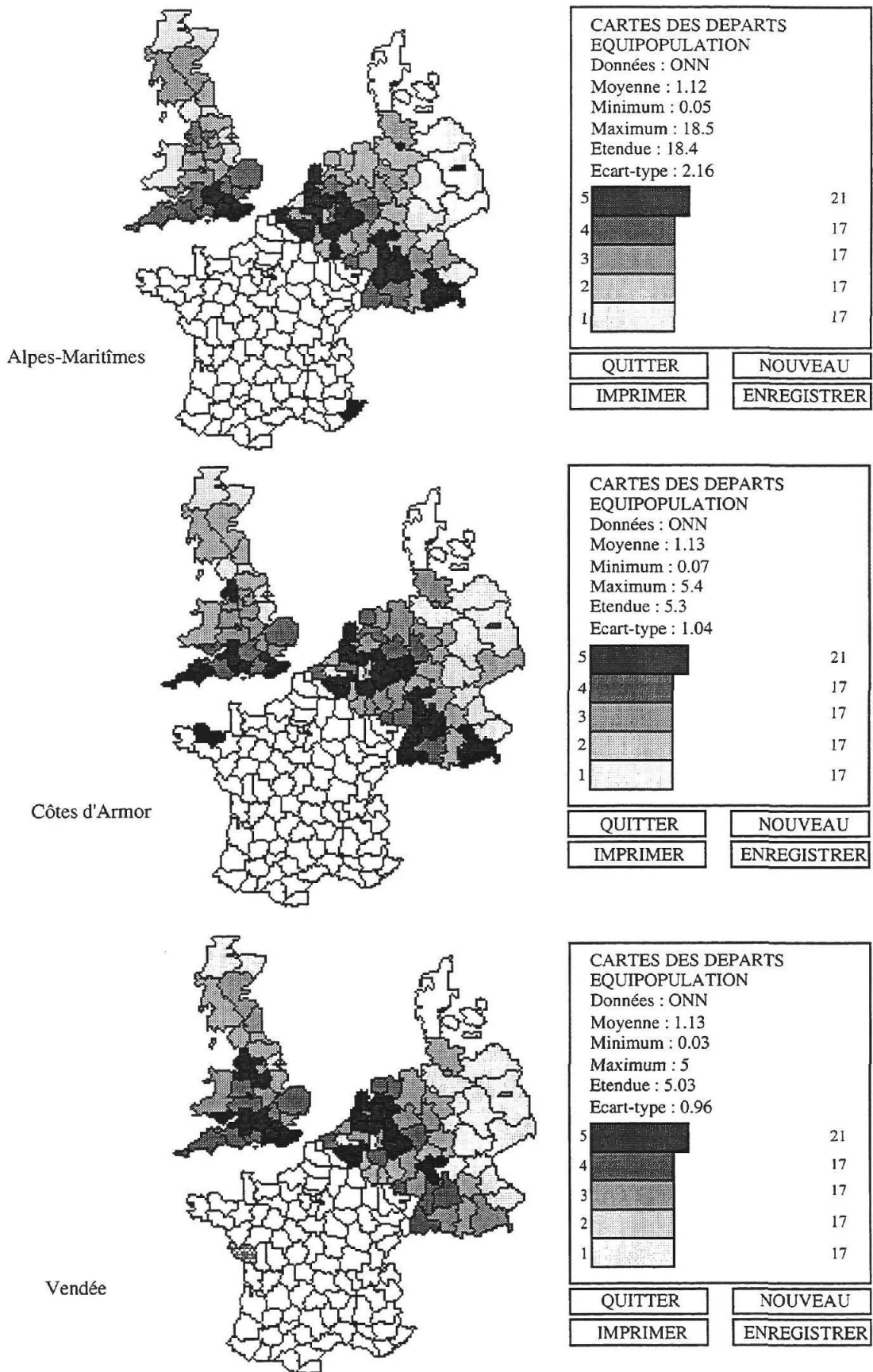


Figure 4 : Pourcentage du trafic émis par les départements en direction de l'Europe du nord





Tout d'abord, le trafic de jour privilégie toujours les pôles économiques ou politiques des pays européens. Or la nuit, le rôle de ces villes apparaît beaucoup plus atténué : Milan, Barcelone ou Francfort captent des flux plus faibles.

La répartition du trafic en direction des deux principaux partenaires de la Dordogne change sensiblement en fonction des horaires. Pour le Royaume-Uni la dissémination du trafic la nuit est particulièrement nette. Alors que le jour, Londres et son bassin attirent la majeure partie des appels, la nuit tous les comtés britanniques voient le nombre d'appels s'élever proportionnellement. Le phénomène est identique pour les Pays-Bas avec une concentration forte autour de Rotterdam et Amsterdam.

Ces flux sont émis de toute évidence par les touristes car la Dordogne n'a pas de relations économiques très suivies avec ces pays. La répartition du trafic confirme donc la validité de l'hypothèse sur la segmentation du trafic. Cette opération a bien sûr été réalisée non pas sur un département mais sur tous les départements touristiques français (notamment tous les départements côtiers et montagnards). La manipulation sur SADIC étant facilitée, le test réalisé sur tous les départements touristiques a permis rapidement (moins d'une heure) de vérifier cette hypothèse. Nous avons donc isolé ces fichiers afin de réaliser ensuite des traitements statistiques plus poussés destinés à mesurer plus précisément l'impact du tourisme sur ces régions.

La seconde hypothèse sur la répartition spatiale des touristes dans l'espace français a été testée en étudiant la part du trafic de chaque région NUTSII des pays du nord de l'Europe dans le trafic des départements français. Ces pays (Allemagne, Belgique, Pays-Bas et Royaume-Uni), ont été choisis car ils fournissent les principaux contingents de touristes, mais aussi pour leur proximité géographique et pour leur appartenance à une aire culturelle assez homogène.

L'analyse des trafics issus des départements côtiers de l'Atlantique et de la Méditerranée ainsi que des départements des Alpes et des Pyrénées montre que la distribution spatiale des nationalités ne se fait pas au hasard mais répond à des préférences qui varient en fonction du pays d'origine.

Le flux d'appels des allemands est dominant sur la majeure partie des côtes françaises. Ils sont implantés sur toute la façade atlantique (sauf les côtes vendéennes), entre Montpellier et Saint-Raphaël sur la côte méditerranéenne, partout en Corse et sur la façade bretonne de la Manche.

Les appels des Belges et des Hollandais émanent d'abord des zones de montagne. Les Alpes du Sud ainsi qu'une partie des Pyrénées et même la Haute-Savoie sont particulièrement attractives. La vallée du Rhône est aussi un lieu d'accueil privilégié pour ces deux nationalités.

La proportion des appels anglais est importante dans tout l'ouest français. Ils sont aussi présents sur la côte méditerranéenne mais jamais en forte position sauf sur une partie de la Côte d'Azur.

Le changement d'échelle et le passage au niveau des Circonscriptions de Taxe améliore encore la connaissance de la répartition spatiale des touristes. En effet, on s'aperçoit que, dans les départements côtiers, les Allemands ne sont présents que dans les circonscriptions côtières et qu'ils ne fréquentent que peu l'intérieur des terres. En revanche, les Britanniques se comportent de façon presque opposée puisqu'on les retrouve dans tout l'Arcoat breton et à l'intérieur de la Gironde.

## Conclusion

SADIC a été conçu comme un outil permettant de gérer les bases de flux de grande taille. Il doit permettre d'opérer un premier tri dans la masse de données, afin de confirmer ou d'infirmer des hypothèses. De plus sa facilité d'utilisation et sa rapidité sont un atout pour en faire un logiciel de consultation de la base de données.

La cartographie facilite la visualisation des données et permet d'apporter une valeur ajoutée par rapport au simple traitement statistique. Elle permet notamment de faire des rapprochements, des hypothèses à partir de configurations spatiales observées sur la carte.

Mais SADIC n'est pour le moment qu'un prototype. Sa puissance sera mise en évidence quand il s'agira de traiter d'évolution temporelle dans la base. En effet, les mesures réalisées au mois de juillet pourront être reconduites chaque mois. L'analyse dans le temps de l'évolution du trafic à différentes échelles apportera d'autres précisions, notamment pour le phénomène touristique.

Enfin il faut bien préciser que SADIC n'intervient que dans une étape du traitement de l'information, son rôle se limitant à celui de filtre. C'est pourquoi il nous a semblé très important qu'il puisse s'insérer et s'adapter à d'autres logiciels imbriqués dans l'ensemble du traitement, d'où le choix de la modularité et de son ouverture.