

POPULATION PRÉSENTE : méthodes de détermination à partir des enquêtes sur les touristes

par *Christophe TERRIER*^(*), *Marguerite SYLVANDER*^(*),
Abdel KHIATI^(*) et *Véronique MONCERE*^(**)

(*) DSPES, direction du Tourisme
(**) TNS-SOFRES

Ce texte est une version adaptée d'une communication présentée aux Journées de Méthodologie Statistique de l'INSEE en mars 2005

LA POPULATION « résidente » est en général la seule population connue, en particulier grâce au recensement de la population mené par l'INSEE. Mais avec le développement de la mobilité, la population « présente » en un lieu donné à un moment donné peut être notablement différente de cette population « résidente ». La préparation et la conduite d'une bonne gestion publique rendent nécessaire l'estimation de cette population présente. En effet certains équipements doivent être calibrés en fonction de la population maximale pouvant être présente sur les lieux.

L'écart entre population résidente et population présente est variable dans le temps et selon les lieux. Par exemple les bords de mer attirent les foules en été et la montagne voit converger les skieurs l'hiver et randonneurs et alpinistes l'été. Mais ces touristes ne viennent pas de nulle part : les villes se vident de leurs habitants pendant les vacances, tout en se remplissant parfois de touristes. On cherche donc à mesurer l'ensemble de ces mouvements et les variations de la population présente tout au long de l'année qui sont induites par ces mouvements.

En analyse de la mobilité, la définition du créneau spatio-temporel sur lequel on travaille a une importance considérable. Pour cet exercice nous avons adopté le département, territoire de base de la gestion publique, et la journée (24 heures), qui correspond au champ du tourisme. En effet, selon les normes internationales, est touriste toute personne qui voyage, quel qu'en soit le motif, avec au moins une nuit passée hors du domicile.

Nous avons donc cherché à estimer, jour par jour, la population présente dans chaque département de France métropolitaine. Cette population présente est formée d'habitants et de séjournants. Le nombre d'habitants est fourni par l'INSEE (recensement de population). Mais tous les habitants ne sont pas présents tous les jours de l'année dans leur département, certains d'entre eux pouvant être partis en voyage dans un autre département ou à l'étranger. Les séjournants peuvent être français ou étrangers ; ils ne seront comptabilisés ici que s'ils séjournent au moins une nuit dans le département. Par approximation, nous avons considéré qu'une personne était présente 24 heures dans le département où elle passait la nuit.

Nous avons combiné deux approches s'appuyant sur des sources de données différentes, l'une portant sur la population résidant en France, l'autre sur les touristes étrangers.

Pour la population résidant en France, nous avons utilisé l'enquête SDT (Suivi de la Demande Touristique), réalisée par TNS-Sofres pour le ministère du Tourisme, basée sur un panel de 20 000 Français interrogés mensuellement sur leurs déplacements touristiques du mois écoulé. On connaît, par l'INSEE (RP + actualisation annuelle) la population résidente d'un département. À partir du SDT, on détermine, dans un premier temps, la population résidente absente de chez elle pour cause de déplacement. Ces « absents » peuvent être en déplacement touristique (c'est-à-dire comportant au moins une nuitée hors du domicile) dans le même département, dans un autre département français métropolitain ou hors de France métropolitaine. S'ils séjournent hors de leur domicile mais toujours dans le même département, ils ne seront pas considérés ici comme absents de leur département, bien que comptabilisés comme touristes dans les statistiques habituelles. Toujours à partir de la même source, on estime la population des Français résidant dans un autre département et qui sont en voyage dans le département considéré. L'enquête ne portant que sur les plus de 15 ans, les chiffres sont extrapolés à la population totale. L'enquête ne porte que sur la France métropolitaine.

La problématique des touristes étrangers est théoriquement plus simple puisqu'ils ne génèrent pas d'absences mais seulement un surcroît de population. Mais pour estimer leur présence un jour donné nous disposons d'informations beaucoup moins précises. Il ne peut cependant pas être question d'ignorer cette population : on estime à environ 76 millions le nombre d'entrées de touristes étrangers en France chaque année. Nous avons donc été amenés à combiner un ensemble de sources, notamment l'Enquête aux frontières et les enquêtes mensuelles de fréquentation hôtels et campings (INSEE/Direction du Tourisme). Sur la base de ces données, nous avons estimé le volume et la répartition spatio-temporelle de ces touristes étrangers.

1 Estimation de la présence (et de l'absence) des Français.

1.1.

L'enquête « Suivi de la Demande Touristique »

La présence des Français en un lieu donné peut être estimée grâce aux résultats du Suivi de la Demande Touristique (SDT). Ce dispositif est basé sur l'interrogation tous les mois par voie postale d'un panel de 20 000 personnes par TNS-Sofres pour le ministère du Tourisme.

Le panel de la Sofres est composé d'individus représentatifs de la population de 15 ans et plus résidant en France métropolitaine. La représentativité du panel est assurée par une stratification. Les strates sont basées sur le croisement des critères suivants :

- La région de résidence : la Sofres utilise le découpage en régions UDA (Union Des Annonceurs ; les 22 régions administratives sont regroupées en 9 régions UDA) ;
- L'âge du chef de famille (5 modalités) ;
- La taille de l'agglomération de résidence (5 modalités) ;
- La catégorie socioprofessionnelle du chef de famille (7 modalités).

1.2.

Données de cadrage

1.2.1. Cadre spatio-temporel de l'étude

On raisonne dans le cadre de la définition internationale du tourisme, c'est-à-dire que l'on ne considère que les déplacements comportant au moins une nuitée hors du domicile quel qu'en soit le motif (loisirs, affaires...) et quelle que soit la distance parcourue. La maille territoriale retenue ici étant le département, tout déplacement à l'intérieur du département n'est donc pas pris en compte. Tout déplacement, même à longue distance, effectué dans la journée avec retour au domicile, n'est pas pris en compte. Par convention on comptabilise la durée du voyage en nuitées, en assimilant une journée de 24 heures à une nuitée.

Pour chaque individu enquêté, on peut donc en déduire, pour chaque jour de l'année, le département où il est présent. C'est l'information de base de cette étude. Cette information se décline de la façon suivante :

- **Soit l'individu est resté sur son lieu de résidence** (il peut avoir voyagé dans la journée, même en quittant son département mais il est rentré chez lui le soir),
- **Soit l'individu est en voyage**. Dans ce cas on s'intéresse au département de son lieu de séjour ce jour-là. Ce département peut être :
 - Le même que la résidence (il est touriste mais reste présent dans son département de résidence ; il n'est donc pris en compte ici dans les mouvements)
 - Un autre département métropolitain (il est touriste et absent de son département)
 - Un pays étranger ou un DOM-TOM (il est touriste et absent de métropole).

1.2.2. Présentation physique des données

L'enquête est réalisée tous les débuts de mois et porte sur les voyages terminés le mois précédent. En fin d'année, on cumule les 12 vagues mensuelles d'enquête pour former le fichier « cumul ». Les informations utilisées ici sont extraites du fichier de données du cumul annuel, auquel on ajoute le fichier de janvier de l'année suivante afin de comptabiliser les voyages débutés en décembre et terminés en janvier, notamment les voyages effectués lors des fêtes de fin d'année. Cela exclut en pratique les longs voyages de plus d'un mois commencés une année donnée et non encore terminés fin janvier de l'année suivante.

On constitue alors, à partir des répondants de chaque mois, un fichier comportant autant d'enregistrements que d'individus. Chaque enregistrement comporte, pour chaque date de l'année, un indicateur de son état touristique, en répondant aux questions suivantes par les règles :

1) Le panéliste est-il répondant ?

- Si le panéliste est répondant un mois donné, il l'est pour chaque nuit du mois. Dans ce cas, chaque nuit est marquée d'une valeur.

2) Pour chaque nuit, est-il en voyage ou non ?

3) S'il est en voyage, quelle est sa destination ?

- Si le panéliste est en voyage, sa destination (département ou pays) est indiquée pour chaque date du voyage.
- Si le panéliste n'a pas quitté son domicile à la date considérée, la date est marquée d'un 0.

Ainsi si le panéliste est répondant le mois m mais non partant, toutes les dates du mois m seront marquées d'un 0 pour ce panéliste.

La valeur manquante « . » signifie que le panéliste n'est pas répondant le mois considéré. Il peut toutefois avoir des nuits pour lesquels une destination est remplie, si cela concerne un voyage débuté le mois m , mais terminé le mois $m + 1$, mois pour lequel le panéliste a répondu.

Pour l'étude de la présence, qui est notre objet principal d'intérêt, cette information suffit. Cependant, si l'on veut étudier la relation entre cheminement et présence, on pourra s'intéresser également à des informations annexes :

1) Sur le trajet

- Est-ce le jour du départ en voyage ? (trajet aller)
- Est-ce un jour de milieu de séjour ? (pas de trajet ce jour-là)
- Est-ce le jour du retour ? (trajet retour)

On pourrait imaginer de compléter l'information par la distance parcourue (calculée à partir de l'origine et de la destination).

2) Sur la durée

- Quelle est la durée du voyage en cours ?

1.3.

Faiblesses connues de l'information

1.3.1. L'unité de temps : le jour ou la nuit ?

La notion de présence en un lieu un jour donné n'a son sens plein que si l'individu reste en ce lieu toute la journée. Or nous étudions des voyageurs. On a déjà établi la convention de ne considérer que les déplacements « touristiques », donc avec au moins une nuit hors du domicile. Il faut également convenir de n'affecter qu'un seul lieu de présence à un individu, y compris le jour où nous savons par l'enquête qu'il se déplace d'un lieu à un autre.

Travailler en **jour d'absence** impliquerait des doubles comptes lorsqu'un panéliste a effectué deux voyages successifs avec des dates contiguës. Par exemple, un individu parti du 10 au 12 dans le 75 et du 12 au 15 dans le 63 serait compté deux fois le 12 : une fois dans le département 75, une fois dans le 63.

Se baser sur la **nuitée d'absence** paraît plus robuste ; cela présente une meilleure cohérence par rapport au champ de l'étude et évite le problème des dates contiguës. Nous comptabilisons alors la journée de présence dans le département où la personne passe la nuit qui suit. Là encore on fait l'hypothèse – évidemment approximative – d'un comportement normal en considérant que tout individu dort la nuit et en un lieu fixe.

Cependant, cette notion de nuitée ne permet pas de distinguer des cas pourtant différents. Par exemple on ne tient pas compte du moment du voyage : selon que l'individu voyage le soir ou le matin, la durée est différente, et cela a d'autant plus d'impact que le voyage est court. On ne peut distinguer un week-end avec travail le vendredi toute la journée et retour tôt le lundi matin suivi d'une journée de travail, d'un week-end prolongé avec départ le vendredi matin et retour le lundi soir. Un séjour normal de week-end (samedi et dimanche) pourra être comptabilisé comme trois nuits si les trajets s'effectuent le vendredi soir et le lundi matin ou comme une nuit si les trajets s'effectuent samedi matin et dimanche soir.

Par convention, l'**unité de référence** utilisée ici est **la nuit qui suit** la date considérée : nuit 1 = nuit du 1^{er} au 2 janvier.

Présence ou risque de présence ?

Ces conventions ont été adoptées afin d'éviter les doubles comptes : un individu présent un jour donné en A ne pourra pas être considéré présent en B. Dans la réalité, quand un individu voyage un jour donné du lieu A vers le lieu B, il est présent alternativement en A et en B. Il est même – si l'on pouvait décomposer finement son itinéraire – présent successivement en chacun des lieux traversés au cours de son voyage entre A et B. On pourrait donc, à côté de la mesure de la « présence » à laquelle on attribue une valeur univoque, définir un « risque de présence » où le voyageur serait comptabilisé à la fois en A et en B. Cet indicateur génère bien sûr des doubles comptes mais peut offrir une réponse mieux adaptée à certaines problématiques. On peut cependant noter que, pour les mesures de précaution à adopter pour intervenir auprès des populations présentes (en cas d'alerte bactériologique par exemple), on doit mesurer le « risque maximal de présence » en un lieu donné un jour quelconque dans l'année. Pour un

lieu donné, la présence maximale un jour quelconque et le risque maximal de présence un jour quelconque ne seraient différents que si un important chassé-croisé entre A et B s'opérait le jour où la présence en A ou en B atteignait son maximum annuel.

1.3.2. La population non enquêtée résidant en France

• Les résidents non français

Jusqu'en 2003, la population de référence utilisée pour le SDT était la population de nationalité française, à l'exclusion des étrangers résidant en France. À partir de 2004, l'échantillon porte sur l'ensemble de la population résidant en France métropolitaine.

• Les enfants de moins de 15 ans

Pour les moins de 15 ans, la seule donnée disponible concerne les moins de 15 ans accompagnants. Il faudrait alors considérer que les moins de 15 ans résident dans le même département que l'individu accompagnant.

En outre, l'enquête SDT ne fournit aucune information sur les déplacements des enfants de moins de 15 ans voyageant seuls ou en groupe. On a donc fait l'hypothèse que les enfants de moins de 15 ans voyageant seuls ont des comportements en matière de déplacements touristiques semblables à ceux de leurs aînés. Pour valider cette hypothèse, nous avons eu recours aux résultats de l'enquête vacances de l'INSEE. La dernière vague de cette enquête porte sur les déplacements terminés entre le 1^{er} octobre 2003 et le 30 septembre 2004. Elle recense pour chaque individu les déplacements d'une durée de 2 nuits ou plus hors du domicile. Pour les déplacements de 4 nuits ou plus (vacances), la nature, la durée, la destination précise en France ou à l'étranger, le moyen de transport utilisé, le mode d'organisation et d'hébergement et les motifs de séjour sont précisés. Sur ce champ des vacances, on mesure de fait un comportement des enfants assez proche de celui des adultes.

Les très longs séjours

Les normes internationales englobent, dans la définition du tourisme, tous les voyages comprenant au moins une nuit et au plus un an. En pratique, la plupart des enquêtes sur les déplacements touristiques se limitent à une durée inférieure à un an, variable selon les enquêtes. Cette prise en compte ou non des longs séjours de plusieurs mois n'apporte pas de distorsion dans la mesure des flux touristiques. Par contre cette différence se répercute dès que l'on s'intéresse à la durée : un seul individu séjournant 4 mois génère plus de nuitées que 100 personnes séjournant chacune une seule nuit. La durée prise en compte dans le SDT n'est théoriquement pas limitée, mais des considérations pratiques de gestion de l'enquête peuvent intervenir (par exemple : une personne absente 6 mois de chez elle sera considérée comme non-répondante pendant 6 mois).

1.4.

Méthode de calcul des absences

1.4.1. Le calcul traditionnel des absences

Pour estimer un taux d'absence par date, nous exploitons jusqu'à présent le fichier annuel directement, avec les poids mensuels sur les 12 mois de l'année. Les poids mensuels sont calculés en « redressant » la structure des individus répondants selon les principaux critères sociodémographiques (par l'Enquête Emploi) afin de préserver la permanence de la représentativité statistique de l'échantillon. Ce redressement s'effectue selon la méthode de calage sur marges (algorithme "Iterative Proportional Fitting") pour chacun des critères sociodémographiques.

Parallèlement, le nombre de ces répondants est extrapolé à un effectif constant correspondant à la taille de l'échantillon de départ (soit 20 000 individus), évitant que les variations du taux de retour n'interfèrent sur les résultats. Un même individu, tout en appartenant à la même strate, peut avoir un poids différent selon le mois enquêté si ce ne sont pas les mêmes individus qui répondent et également parce que l'échantillon est renouvelé régulièrement.

Les critères de redressement appliqués au Suivi de la Demande Touristique sont les suivants :

- Région UDA x taille d'agglomération x sexe
- Catégorie socioprofessionnelle du chef de famille x sexe de l'individu
- Âge de l'individu x sexe de l'individu
- Sexe de l'individu x nombre de personnes dans le foyer
- Sexe de l'individu x activité de l'individu
- Région de résidence (22) x sexe de l'individu

Par ailleurs, faute de convention pour les voyages multiséjours, ce type de voyage était pris en compte dans le calcul des absences du lieu de résidence mais pas dans celui des présences dans un autre département.

1.4.2. Limites du traitement traditionnel

En traitant les données comme décrit ci-dessus, nous observons des distorsions date par date. En effet, si l'on calcule les absents par département de destination (+ étranger) d'une part, les partants par département de résidence d'autre part, et les non-partants par département de résidence, la population totale varie de date en date.

Deux points expliquent ces distorsions :

- 1) Les poids calculés mensuellement ne sont pas adaptés aux cas où un voyage a débuté le mois précédent
- 2) Dans le cas d'un voyage multiséjours, l'individu est considéré comme absent, mais n'est pas comptabilisé en destination.

1.4.3. Un nouveau traitement

Les poids par nuit

La périodicité mensuelle de redressement induit une source d'écarts de population nuit par nuit. En effet, le poids d'un individu à une date donnée est le poids du mois de fin du voyage. Ce mode de fonctionnement induit des erreurs dans le cas de voyages débutés en m-1 et terminés en m.

Voici un exemple : prenons le cas d'un panéliste parti du 29 janvier au 3 février. Pour chaque nuit d'absence, il a le poids du fichier de février :

Date d'absence	29 janvier	30 janvier	31 janvier	01 février	02 février
Poids affecté	Poids de février	Poids de février	Poids de février	Poids de février	Poids de février

1^{er} cas de figure : le panéliste n'a pas répondu au questionnaire de janvier, mais a répondu en février et décrit son voyage. On rajoute alors à 3 dates du mois de janvier un individu redressé selon les répondants du mois de février. Or, la strate de redressement de cet individu a été pondérée en janvier de façon à le représenter. Cela revient à surreprésenter cette strate et aboutit à une somme des poids supérieure à 20 000.

2^e cas de figure : le panéliste a répondu en janvier et février. Pour les 3 dates de janvier, le poids de l'individu est celui calculé en février et remplace celui de janvier. Cela distord la population totale. Pour les 3 derniers jours de janvier, la somme des poids est différente de 20 000.

Pour éviter ces distorsions, on pourrait appliquer pour le 2^e cas le poids de janvier aux dates de janvier. Cependant, cela ne résout pas le problème du 1^{er} cas : l'individu « n'existe pas » en janvier.

La seule solution consiste à calculer un poids pour chaque nuit de l'année. Les individus pris en compte pour le calcul du redressement sont tous ceux dont on sait où ils se trouvent la nuit considérée, chez eux ou en séjour. De cette façon, nous pouvons calculer des flux pour chaque date, vers chaque destination, et pour chacune des dates, la somme des individus pour lesquels nous savons où ils se trouvent est égale à 20 000.

Le redressement pour chacune des dates est réalisé selon le même processus qu'au niveau mensuel.

Le fichier final contient donc 365 poids pour une année et chaque date doit être traitée avec son propre poids.

Les voyages multidestinations

Les voyages itinérants ou comportant plusieurs lieux de séjour successifs ne sont décrits que de façon partielle dans l'enquête. Pour les prendre en compte, il a fallu adopter des conventions. On fait ici correspondre un voyage à une seule et unique destination : toutes les nuits sont attribuées au département ou pays dans lequel l'individu a séjourné le plus longtemps. En cas d'égalité de durées entre plusieurs destinations, les absences sont par convention attribuées à la destination du premier séjour.

Par exemple : 1 voyage du 1^{er} au 15, en 3 séjours :

- 1 séjour de 4 nuits dans le Cantal
- 1 séjour de 5 nuits dans le Cantal
- 1 séjour de 6 nuits dans le Puy de Dôme

Les nuits du 1^{er} au 14 seront attribuées au département du Cantal (15).

1.5.

Résultats obtenus

En additionnant les informations individuelles on obtient, pour chaque jour de l'année, la population présente par département.

Pour un département donné, un jour donné, la population présente est égale à la population résidente à laquelle on retranche les personnes absentes (effectuant un voyage hors du département) et à laquelle on ajoute la population touristique (en séjour dans le département et dont le lieu de résidence est hors du département).

Chaque jour la population résidente du département, considérée comme fixe tout au long de l'année, se répartit entre ceux qui sont restés chez eux, ceux qui voyagent (avec nuitée hors du domicile) dans le département, ceux qui voyagent en France métropolitaine dans un autre département et ceux qui voyagent hors France métropolitaine (DOM-TOM ou étranger). Les trois dernières catégories sont des touristes mais seules les deux dernières forment la population absente du département. Les deux premières représentent la population stable à laquelle il faut ajouter les touristes venus d'autres départements pour former la population présente.

Pop-présente (D) = pop-stable (D) + touristes (D)
Pop-stable (D) = pop-résidente (D) – pop-absente (D)

On notera bien que les « touristes infradépartementaux » (séjournant hors de leur domicile mais dans le même département) ne sont pas comptabilisés ici dans les « touristes ».

1.6.

Qualité de l'information obtenue

Il reste la question : l'échantillon est-il suffisant pour donner une information fiable par jour et par département ?

1.6.1. Calcul d'intervalles de confiance

Pour évaluer la fiabilité des résultats obtenus, on peut établir deux types d'intervalles de confiance par jour et par département : un intervalle de confiance pour la proportion de panélistes ayant quitté leur département de résidence un jour donné d'une part, et un autre pour la proportion de panélistes présents dans un certain département (autre que le leur) d'autre part.

Si l'on était dans le cas d'un échantillonnage aléatoire simple, ces intervalles de confiance auraient la forme suivante :
 $[p - u \sqrt{(p(1-p)/n)} ; p + u \sqrt{(p(1-p)/n)}]$

Où :

« p » est la proportion de panélistes pour laquelle on veut calculer un intervalle de confiance

« n » est la taille de l'échantillon

« u » est le fractile de la loi normale (par exemple $u = 1,96$ pour un intervalle de confiance bilatéral avec un risque de 5 %).

Cette méthode ne doit toutefois pas être appliquée dans notre cas. En effet, deux des hypothèses requises ne sont pas vérifiées :

« n » désigne le nombre de personnes de l'échantillon du SDT résidant dans le département pour lequel on cherche à calculer un intervalle de confiance. Pour les petits départements, le nombre de panélistes « n » peut-être assez petit (de 17 panélistes résidant en Lozère à 912 panélistes résidant à Paris ; 220 panélistes en moyenne par département). Or pour pouvoir appliquer le théorème central limite et donc supposer que la proportion suit bien une loi normale, il faut que « n » soit au moins égal à 30 individus.

Le problème de taille d'échantillon n'est pas le problème principal nous empêchant d'utiliser cette méthode. Le SDT n'est pas issu d'un sondage aléatoire simple. Les panélistes sont sélectionnés après diverses stratifications et les données sont ensuite redressées suivant des variables différentes des variables de stratification.

Une estimation de l'intervalle de confiance pour la proportion de personnes quittant un département un jour donné à l'aide d'une méthode de bootstrap semble donc plus appropriée. Cette méthode permet d'estimer les intervalles de confiance sans faire d'hypothèses préalables sur l'échantillon. La seule contrainte est d'avoir suffisamment d'individus pour pouvoir faire suffisamment d'itérations successives. On estime que pour avoir une convergence, un minimum de 30 observations est nécessaire. Par conséquent, cette méthode nous permet de calculer les intervalles de confiance, uniquement pour les grandes destinations (Paris par exemple) ou pour la haute saison.

1.6.2. Améliorations envisageables

Comment consolider l'information disponible ? Plusieurs pistes peuvent être envisagées :

- **Augmenter la maille géographique** (par exemple en travaillant par région) réduirait l'incertitude mais réduirait l'intérêt. D'une part, plus la maille est grande et plus les mouvements sont faibles (les déplacements internes à la maille ne sont pas pris en compte). D'autre part l'intérêt pour les acteurs publics se situerait plutôt à un niveau de proximité, parfois même au niveau de la commune quand il s'agit d'adapter des équipements publics. On s'en tiendra donc au département.

- **Cumuler les informations collectées sur plusieurs années** est une possibilité que l'on a utilisée pour de nombreuses études portant sur les destinations fréquentées. Ici cette voie ne semble pas pouvoir être utilisée car la répartition journalière des séjours touristiques est assez variable d'une année à l'autre, selon le positionnement des week-ends, des ponts et des fêtes mobiles, en particulier au printemps. Les valeurs lissées sur 3 ans sont données ici à titre de cadrage.

- **Extrapoler l'information à un niveau territorial plus fin.** Pouvoir estimer la population présente jour par jour dans chaque département est une avancée importante qui permet de prendre réellement en compte certains problèmes de gestion de l'espace et des services publics. Mais travailler dans un maillage départemental lisse évidemment l'amplitude des variations mesurées : à l'intérieur d'un département qui voit doubler sa population présente, il peut y avoir des communes qui multiplient la leur par des facteurs nettement plus élevés, de l'ordre de 10 ou 100, ce qui pose des problèmes locaux

très aigus, par exemple dans le calibrage des réseaux d'assainissement. Sauf enquête complète menée sur place, il n'est pas possible de mesurer ces variations de façon fiable.

Une piste, actuellement tentée par une équipe universitaire de Créteil-Paris-12 autour de Laurent Davezies, consiste, partant des estimations journalières par département, à en extrapoler les composantes par commune. Les « émissions » de touristes du département sont réparties entre les communes de résidence (et donc de départ des touristes) selon les caractéristiques de la population communale. L'utilisation de l'indicateur « revenu moyen des ménages » donne d'assez bons résultats, le taux de départ en déplacement touristique lui étant assez fortement lié. Pour les « arrivées », on répartit les touristes reçus par le département selon la capacité d'hébergement touristique de la commune. Pour tous ces calculs, il convient de prendre en compte tous les touristes, y compris ceux qui, bien que partant de chez eux, ne quittent pas leur département, alors que ce tourisme intradépartemental est bien évidemment exclu des calculs de variation de la population présente dans le département.

Cette méthode d'extrapolation donne en général de bons résultats ; elle présuppose cependant une saisonnalité cohérente sur l'ensemble du département, ce qui n'est pas toujours le cas, en particulier lorsqu'un même département comprend du littoral, de la montagne et de la ville.

1.7.

Quelques voies d'exploration

1.7.1. Représentativité de l'échantillon au lieu de séjour

On évoquera ici, sans la résoudre, la question récurrente : la représentativité de l'échantillon au lieu de résidence garantit-elle une représentativité au lieu de séjour ? Cette question pourrait être formulée autrement : y a-t-il neutralité de l'échantillonnage (au lieu de résidence) sur la représentativité des lieux de séjours ? On livre ici quelques éléments de réflexion sur cette question : on peut supposer que la propension d'un individu à voyager est en partie fonction de facteurs forts mais dont la base statistique n'est pas connue.

Par exemple, un Parisien qui dispose d'une résidence secondaire en Bourgogne va, en général, faire un nombre de voyages vers cette région supérieur à la moyenne de ses voisins. Or on ne dispose hélas pas d'une base statistique croisant lieu de résidence, possession d'une résidence secondaire et localisation de cette résidence secondaire.

Autre exemple, une personne originaire de Marseille et en poste dans « le Nord » va également générer des déplacements multiples entre sa famille et lui (vacances ou week-end au pays dans un sens, visite à la famille dans l'autre). Là non plus on ne dispose pas de base statistique sur l'historique territorial des individus. On se retrouve donc avec des facteurs dont on sait qu'ils peuvent être générateurs de mobilité mais sur lesquels on ne peut pas fonder un échantillonnage, faute de base statistique.

L'enquête SDT offre dans ce cadre l'avantage certain d'être un panel : si l'on peut supposer que tout échantillon pourrait être entaché d'un biais, on lisse largement les effets de ce biais en conservant le même échantillon d'un mois sur l'autre. Une étude sur les territoires fréquentés par les touristes habitant dans chacune des régions françaises a été menée sur 5 ans d'enquête SDT : elle montre une grande stabilité de la relation origine-destination.

1.7.2. D'autres méthodes de mesure journalière de la fréquentation touristique

Les enquêtes cordons

Les enquêtes cordons – appelées enquêtes aux frontières quand les limites du territoire étudié forment des frontières – offrent une voie intéressante pour déterminer la présence un jour donné. Le principe consiste à établir un cordon – non pas sanitaire mais statistique – autour du territoire concerné. En comptant les passages et en effectuant des enquêtes auprès d'un échantillon représentatif des personnes qui franchissent ce cordon pour entrer ou sortir du territoire on peut déterminer la population présente dans le territoire. Ce type d'enquête est particulièrement adapté aux îles où l'aéroport constitue le point de passage obligé des entrées et des sorties.

L'opération est beaucoup plus complexe sur le territoire métropolitain de la France où les modes de transports et les points d'entrée-sortie sont multiples. Une difficulté est venue récemment se rajouter avec la suppression de fait des frontières routières entre pays d'Europe (accords de Schengen) et l'impossibilité pratique de recours aux forces de l'ordre pour arrêter les véhicules dont on aurait souhaité enquêter les passagers.

Actuellement ce type d'enquêtes aux frontières n'existe plus au plan national, remplacé par une autre enquête (EVE) que l'on ne détaillera pas ici. Au plan local (départemental ou régional) ce type d'enquête cordon intégré à un ensemble méthodologique (méthode dite des « flux », cf. *biblio*) donne actuellement encore des résultats opérationnels sur quelques régions ou départements.

2 Estimation de la présence (et de l'absence) des touristes non-résidents.

La construction de l'Union Européenne a rendu plus difficile la mesure du tourisme international. L'ouverture des frontières a compliqué l'estimation de la fréquentation des touristes étrangers en France ; le passage à l'euro a perturbé l'évaluation de leurs dépenses sur le territoire français.

Pour avoir une idée du nombre de touristes étrangers présents par département et par jour, il a donc fallu procéder à tout un ensemble d'estimations successives. D'abord évaluer les dépenses des touristes étrangers en France pour estimer leur fréquentation globale en nuitées ; puis répartir ces nuitées par régions, puis par mois, puis par département, et enfin par jour.

2.1.

Estimation des recettes de la ligne « voyages » de la Balance des Paiements

Jusqu'en 1996, la direction du Tourisme menait régulièrement des enquêtes aux frontières, ce qui permettait d'avoir une bonne estimation de la fréquentation en arrivées et en nuitées des touristes non-résidents en France métropolitaine. Avec les accords de Schengen qui suppriment les contrôles – et donc les possibilités d'enquête – aux frontières terrestres, la direction du Tourisme a dû revoir sa méthode et a lancé une nouvelle enquête auprès des visiteurs étrangers. En attendant que les résultats de cette enquête soient disponibles, l'évolution de la fréquentation des touristes non-résidents est chaque année estimée en fonction de plusieurs indicateurs économiques dont le principal est le niveau des recettes du poste « voyages » de la Balance des Paiements.

Le passage à l'euro a perturbé la mesure des dépenses des touristes non-résidents sur le territoire français. Jusqu'au 1^{er} janvier 2002, la Banque de France établissait les flux de la ligne « voyages » via les échanges de billets, les paiements transfrontières par cartes bancaires ou les transactions des agences de voyages. L'exhaustivité des flux était ainsi en théorie retracée de façon satisfaisante pour l'établissement des chiffres définitifs.

Depuis l'arrivée de l'euro, on n'a qu'une idée partielle du montant des dépenses des étrangers en France. En effet, non seulement les résidents de la zone Euro n'ont plus besoin de changer de monnaie avant leur visite en France, mais les touristes non européens peuvent se procurer des euros partout et ne les dépensent pas nécessairement sur le territoire français. Pour pallier ce manque, la direction du Tourisme a développé une méthodologie d'estimation des dépenses des étrangers en France.

Un modèle économétrique a été mis en place pour chacune des principales clientèles de la France. Pour chaque pays, on établit une relation entre les dépenses des touristes étrangers en France et un certain nombre de variables explicatives, comme les nuitées hôtelières du pays, les nuitées en camping, éventuellement le taux de change entre la monnaie du pays et l'euro ou encore des données de trafic aérien. Cette relation est calculée sur les années pour lesquelles on dispose des données Banque de France (à partir de 1995 et jusqu'aux dernières don-

nées publiées). L'estimation des recettes du poste « voyages » de la Balance des Paiements pour l'année est cours est ensuite obtenue en prolongeant de douze mois la relation établie sur le passé de la série.

Le type de modèle choisi est un modèle à retards échelonnés (Régression linéaire sur les séries corrigées des variations saisonnières ; parmi les variables explicatives figurent des variables endogènes retardées).

2.2.

Estimation du volume annuel de nuitées

Le volume de nuitées passées par les touristes étrangers en France est estimé à partir des recettes de la ligne « voyages » de la Balance des Paiements. Pour chacun des principaux pays émetteurs de touristes, on fait l'hypothèse suivante :

L'évolution annuelle de la dépense moyenne par nuitée (à prix constant) est proportionnelle à l'évolution de la consommation privée des ménages en volume dans le pays d'origine.

Cela revient à résoudre une équation à une inconnue : on dispose pour chaque pays de l'estimation de l'évolution de la consommation privée des ménages, d'une estimation des dépenses des touristes étrangers en France et du volume de nuitées touristiques de l'année précédente. On obtient ainsi une estimation du volume annuel de nuitées pour chaque pays.

Il reste toutefois un problème à traiter pour les pays dont la monnaie n'est pas l'euro : faut-il considérer que c'est la dépense moyenne en euros ou la dépense moyenne exprimée dans la devise du pays émetteur qui évolue comme la consommation privée des ménages ? Dans le premier cas, les touristes ont, en vacances en France, un train de vie équivalent d'une année sur l'autre, quel que soit le taux de change entre l'euro et leur monnaie ; dans le second cas, les touristes adaptent leur train de vie sur place en fonction du taux de change, de telle sorte que leurs dépenses, exprimées dans leur monnaie locale, ne varient que de façon marginale.

Il est probable que la vérité se situe entre ces deux extrêmes ; on a donc fixé un paramètre, pour chaque pays hors zone euro, qui détermine le niveau de prise en compte du taux de change dans l'évolution de la dépense moyenne par nuitée. La valeur de ce paramètre a été choisie différente suivant le pays d'origine des touristes.

Prenons l'exemple des Britanniques et des Japonais. De nombreux touristes britanniques se rendent régulièrement en France : en 1996, l'enquête aux frontières montre que seuls 8 % d'entre eux venaient pour la première fois sur le territoire français (contre 44 % des touristes japonais). En outre, le Royaume-Uni est entouré de pays de la zone Euro : les destinations hors zone euro sont plus éloignées et rarement accessibles en voiture. Plutôt que de décourager les touristes anglais de venir en France, une dévaluation de la livre face à l'euro les incite vraisemblablement à surveiller leurs dépenses lors de leur séjour.

La situation est un peu différente pour les touristes japonais. Pour eux, un voyage en Europe revêt un caractère exceptionnel ; il est probable que même lorsque l'euro est fort, une fois qu'ils ont décidé de partir en Europe, les touristes japonais choisissent de profiter pleinement de leur voyage sur place. Une hausse de l'euro encouragerait plus les touristes japonais à différer leur voyage en Europe qu'à chercher à limiter leurs dépenses une fois en France.

2.3.

Estimation du volume annuel de nuitées par région

En 1996, les résultats de l'enquête aux frontières ont permis de connaître la fréquentation en nuitées des touristes non-résidents dans les régions françaises. On dispose également, pour chaque année depuis 1997, de l'estimation du volume de nuitées passées par les étrangers sur l'ensemble du territoire métropolitain, pour les principales clientèles étrangères. On n'a en revanche plus de données à l'échelle régionale. Il faut donc répartir le total des nuitées entre les différentes régions françaises pour les années postérieures à 1996. L'idée consiste à faire évoluer la structure régionale des nuitées de 1996 en fonction de l'évolution régionale observée dans les hébergements enquêtés : les hôtels classés et les campings.

Les enquêtes de fréquentation en hôtellerie et hôtellerie de plein air, menées conjointement par l'INSEE et la direction du Tourisme, nous permettent de connaître, pour chaque année depuis 1997, la répartition régionale des nuitées étrangères passées dans les hôtels et campings classés. En revanche, la fréquentation des touristes étrangers dans les autres modes d'hébergement n'est pas connue. Il reste donc à répartir par région le total des nuitées passées dans les hébergements non enquêtés, et ceci pour chaque pays d'origine des touristes non-résidents.

Les nuitées passées dans les hébergements non-enquêtés peuvent être réparties par région de deux façons différentes, selon que l'on considère que leur évolution régionale d'une année sur l'autre est corrélée ou non à l'évolution régionale des nuitées en hôtel et en camping.

Première méthode. On fait l'hypothèse que l'évolution régionale des nuitées dans les hébergements non-enquêtés est indépendante de l'évolution régionale des nuitées observée dans les hôtels et campings. Pour un pays d'origine donné, l'évolution des nuitées hors hôtels et campings est la même dans chaque région, et est égale à l'évolution estimée à l'échelle de la France entière.

Deuxième méthode. On fait l'hypothèse que l'évolution régionale des nuitées dans les modes d'hébergement non-enquêtés est parfaitement corrélée à l'évolution des nuitées dans les hôtels et campings. Pour chaque pays d'origine, les parts de marché régionales des nuitées dans les autres hébergements connaissent la même évolution que les parts de marché régionales des nuitées passées dans les hôtels et campings.

C'est en fait une méthode hybride qui a été retenue. Elle consiste à fixer, pour chaque pays, un coefficient déterminant le niveau d'entraînement de l'évolution des nuitées en hôtels et campings sur l'évolution des nuitées dans les autres modes

d'hébergements. La valeur de ce coefficient est choisie entre 0 et 1. En fixant un coefficient d'entraînement inférieur à 1, les variations des nuitées régionales dans les hébergements non enquêtés sont moins accentuées. Comme le degré d'incertitude est assez élevé, il paraît plus prudent de lisser les variations hôtelières qui peuvent parfois être assez radicales.

Le niveau du coefficient diffère suivant les pays. Le choix du coefficient a été fait en fonction de la part des nuitées en hôtels et campings parmi l'ensemble des nuitées. En effet, plus la part des nuitées dans les hébergements enquêtés est forte, plus les évolutions des nuitées de ces hébergements peuvent être considérées comme représentatives de celles de l'ensemble des nuitées.

2.4.

Estimation du volume de nuitées étrangères par région et par mois

Il s'agit désormais, pour chaque région, de répartir par mois le volume annuel de nuitées.

Les enquêtes de fréquentation dans l'hôtellerie et l'hôtellerie de plein air permettent de connaître, pour chaque région, deux types de saisonnalités : la saisonnalité mensuelle régionale des nuitées étrangères dans les campings (pour les mois de mai à septembre), et la saisonnalité mensuelle régionale des nuitées étrangères dans les hôtels classés. Ces deux types de saisonnalités vont être utilisés pour répartir les nuitées régionales par mois.

Il s'agit de classer les modes d'hébergement selon que leur fréquentation connaît une saisonnalité plutôt proche de celle des hôtels ou plutôt proche de celle des campings. Les résultats trimestriels de l'Enquête aux Frontières de 1996 suggèrent d'appareiller la saisonnalité des nuitées dans les villages et clubs de vacances à celle des nuitées en campings et d'appareiller la saisonnalité de l'ensemble des autres modes d'hébergement (location, résidence secondaire, hébergement chez parents ou amis...) à celle des hôtels.

Pour chaque région, on calcule la part des nuitées qui, en 1996, avaient été passées dans des campings ou clubs et villages de vacances, et la part des nuitées qui avaient été passées dans tous les autres types d'hébergement. On utilise les mêmes proportions pour séparer les nuitées de l'année étudiée qui doivent suivre la saisonnalité mensuelle des campings de celles qui doivent suivre la saisonnalité hôtelière. Pour chaque région, on applique à chacun des deux agrégats de nuitées la saisonnalité qui lui correspond ; on obtient ainsi une estimation du nombre de nuitées de touristes non-résidents par région et par mois.

2.5.

Estimation des nuitées étrangères par département et par mois

Pour procéder à l'estimation des nuitées des non-résidents par département et par mois, on part de la structure par région et par mois estimée lors de l'étape précédente, et on répartit, mois par mois les nuitées de chaque région entre les différents départements qui la composent. On ne peut pas pour ce faire avoir

recours aux résultats de l'enquête aux frontières de 1996 : en effet, si cette enquête renseigne sur la fréquentation étrangère par région administrative, elle n'apporte aucune information sur la fréquentation touristique au niveau départemental.

Les enquêtes de fréquentation INSEE-Direction du Tourisme permettent quant à elles de connaître les volumes de nuitées des touristes étrangers dans les hôtels et campings par département. Cependant, comme rien ne garantit que la répartition départementale des nuitées des non-résidents dans les hébergements autres que les hôtels et campings soit identique à la répartition départementale des nuitées dans ces deux structures d'hébergement, il paraît délicat de n'utiliser que les séries issues des enquêtes de fréquentation dans l'hôtellerie et l'hôtellerie de plein air pour répartir l'ensemble des nuitées régionales des non-résidents par département et par mois.

On va donc faire appel à une autre source pour estimer les volumes mensuels et départementaux des nuitées passées dans les hébergements non-enquêtés.

On va faire l'hypothèse que pour un mois donné, la répartition par département des nuitées hors hôtels et campings des touristes étrangers au sein d'une région est proche de celle des nuitées des touristes français.

Il faut donc pour chaque région et pour chaque mois, calculer la répartition des nuitées passées par les Français dans les départements composant la région. On utilise pour cela les résultats de l'enquête « Suivi de la Demande Touristique ». Seules sont prises en compte les nuitées des Français incluses dans des séjours d'agrément (séjours personnels non motivés par la visite de famille ou d'amis), afin de ne pas surestimer la fréquentation des étrangers dans les départements peu touristiques. On a par ailleurs choisi de prendre cinq ans d'observation pour calculer ces parts de nuitées par département (par exemple les années 1999 à 2003 plutôt que seulement l'année 2003). Cela risque d'estomper des phénomènes propres à l'année étudiée (comme par exemple l'affluence accrue en juillet 2003 en Seine-Maritime pour l'Armada de Rouen ou la commémoration du débarquement en 2004), mais ce risque est faible par rapport au gain de fiabilité qu'apportent cinq années d'observation plutôt qu'une seule.

Pour récapituler, le passage du niveau régional au niveau départemental se fait en deux étapes :

- Les nuitées des non-résidents passées dans les hôtels et campings par département et par mois sont connues grâce aux enquêtes de fréquentation en hôtellerie et hôtellerie de plein air.
- Les nuitées passées dans les hébergements non-enquêtés sont, pour chaque mois et au sein de chaque région, réparties par département suivant le même partage que les nuitées d'agrément des touristes français.

2.6.

Estimation des nuitées étrangères par département et par jour

À ce stade on dispose des nuitées par département et par mois. Il s'agit désormais de répartir ces données mensuelles par jour. On va encore une fois utiliser les résultats des enquêtes de fré-

quentation dans l'hôtellerie et l'hôtellerie de plein air. Ces enquêtes permettent en effet d'avoir une idée de la fréquentation journalière des touristes par le biais du taux de chambres occupées pour les hôtels et du taux d'emplacements occupés pour les campings. Ces données journalières couvrent malheureusement un champ assez large, puisqu'elles ne sont calculées qu'au niveau régional et qu'elles regroupent à la fois les nuitées françaises et étrangères, d'affaires et de loisir.

Il faut donc construire pour chaque mois et chaque région une saisonnalité journalière des nuitées. On appliquera ensuite à chaque total mensuel et départemental de nuitées la saisonnalité journalière du mois correspondant et de la région à laquelle le département appartient.

Pour chaque région et pour chaque mois, on répartit les nuitées en deux groupes, selon qu'elles ont été passées dans des hébergements qui suivent une saisonnalité plutôt proche de celle des campings ou plutôt proche de celle des hôtels. On réutilise pour ce faire la répartition des nuitées « assimilées hôtels » et « assimilées campings » qui a servi à l'estimation des nuitées par région et par mois. Pour les nuitées « assimilées campings », on se contente, pour chaque région, de répartir par jour les nuitées mensuelles en fonction des données journalières de taux d'emplacements occupés. Pour les nuitées « assimilées hôtels », les choses se compliquent légèrement. En effet, les données journalières dont on dispose révèlent que les taux d'occupation sont plus forts les jours de semaine que les jours de week-end. Ceci s'explique par le fait que les nuitées d'affaires représentent certains mois plus de la moitié de l'ensemble des nuitées. Or les nuitées d'affaires sont pour l'essentiel effectuées par des touristes français. Ces pics de fréquentation en milieu de semaine ne concernent pas les touristes étrangers. On a donc choisi de lisser les taux d'occupation journaliers de leurs variations hebdomadaires à l'aide d'une moyenne mobile d'ordre 7. Pour chaque région et pour chaque mois, on estime ainsi la fréquentation journalière des touristes étrangers dans les hôtels et hébergements assimilés à l'aide de ces taux d'occupation journaliers lissés.

On fait ensuite la somme de ces deux séries journalières de nuitées et on obtient, pour chaque région et pour chaque mois, une estimation de la répartition journalière des nuitées. Ces saisonnalités journalières sont ensuite appliquées aux totaux de nuitées par département et par jour : les 567 millions de nuitées de touristes non-résidents en France en 2005 sont ainsi distribués par jour et par département.

La méthode d'estimation que l'on a choisie implique qu'il y a parfois des ruptures dans les courbes journalières lors du passage d'un mois à l'autre. En effet, faute d'autres données, on a appliqué les répartitions journalières des nuitées dans les hôtels et campings, toutes nationalités et tous motifs (affaires ou loisir) confondus, aux totaux mensuels et départementaux des nuitées tous modes d'hébergement confondus. Il n'y a donc aucune raison qu'il y ait une parfaite continuité entre les données journalières de la fin d'un mois et celles du début du mois suivant. Pour estomper ces ruptures artificielles, on a lissé par moyenne mobile les données des quelques jours de transition correspondant au passage d'un nouveau mois. Cela améliore la représentation graphique des courbes journalières sans perte d'informations pertinentes.

Limite.

La saisonnalité que l'on calcule à partir des taux d'occupation journaliers dans les hôtels et campings n'est pas toujours très satisfaisante. C'est notamment le cas pour la région Île de France, où la part des nuitées d'affaires dans l'hôtellerie est très importante et les nuitées de camping sont quasiment inexistantes. La saisonnalité journalière semble donc refléter essentiellement les mouvements du tourisme d'affaire (creux entre le 14 juillet et le 15 août, creux autour du 11 novembre, du premier et du 8 mai) et est donc peu adaptée aux nuitées des touristes non-résidents. Les saisonnalités journalières des autres régions ne rencontrent en revanche pas cet écueil.

Évolution à venir.

L'enquête EVE (Enquête sur les Visiteurs de l'Étranger) actuellement en cours fournira à partir de fin 2006 et de façon continue des informations permettant d'ajuster le cadre de ces estimations. Les premiers résultats de cette enquête laissent apparaître – par rapport à l'enquête aux frontières de 1996 – une notable diminution des nuitées touristiques étrangères, conforme au mouvement général d'accélération des voyages et de raccourcissement des séjours.