

Bilan qualité de l'enquête sur l'utilisation des véhicules utilitaires légers 2011



Fiche descriptive de l'enquête

« carte d'identité »	
nom	Enquête sur l'utilisation des véhicules utilitaires légers 2011
n° visa	2011 X069 TR
année : 2011	périodicité : quinquennale
service enquêteur responsable	SOeS/SDST/BSRV (bureau de la statistique de la route et des véhicules)
mode de collecte	collecte postale et électronique

caractéristiques
<p>Définition du champ :</p> <p>L'enquête sur les véhicules utilitaires légers (VUL) porte sur l'usage des véhicules classés selon leur certificat d'immatriculation comme véhicule de catégorie N1, selon l'article R311-1 du code de la route. Elle concerne les véhicules utilitaires de poids total autorisé en charge inférieur ou égal à 3,5 tonnes, immatriculés en France (y compris DOM). Les fourgonnettes dérivées de voitures particulières ayant certaines caractéristiques et homologuées pour des raisons fiscales en tant que véhicules utilitaires, font partie du champ de l'enquête. C'est un complément à l'enquête permanente sur l'utilisation des véhicules de transport routier de marchandises (TRM). En effet, celle ci couvre annuellement le champ des transports routiers de marchandises effectués par les véhicules utilitaires lourds immatriculés en France : camions et tracteurs routiers supérieur à 3,5 tonnes de poids total autorisé en charge et de moins de quinze ans.</p> <p>Les enquêtes précédentes sur les véhicules utilitaires légers ont couvert les années 1981, 1986, 1991, 1996, 2000 et 2005. Elles excluaient les véhicules âgés de plus de 20 ans. La nouvelle enquête a permis de les inclure grâce à une amélioration de la qualité de l'échantillon par l'usage des données relatives au contrôle technique des véhicules de 20 ans ou moins, soit +5,2% (+1,0% par an). Le changement de méthode n'a pas affecté la comparabilité des deux enquêtes, il permet d'améliorer la qualité de l'échantillon et donc réduire le coût de l'enquête pour un volume identique de questionnaires recueillis.</p> <p>Collecte par internet : Le dispositif VUL 2011 a été enrichi par rapport aux dispositifs antérieurs par la possibilité de répondre en ligne.</p>
<p>unité statistique : véhicule. Le propriétaire (ou l'utilisateur) est interrogé sur l'utilisation globale du véhicule pendant l'année 2010 ainsi que pendant deux jours précis d'une semaine du mois de mars 2011</p>
<p>taille du parc en service : 5 785 000 taille de l'échantillon : 25 000</p>
<p>temps de réponse médian (estimé) entre ½ heure et 1 heure</p>
<p>Caractère obligatoire : oui</p>

Les dates-clés

	Prévu	Réalisé
réunion de la formation CNIS		16/04/2010 (avis d'opportunité)
réunion du comité du label	Automne 2011	16 novembre 2011
Déclaration à la CNIL (pour les entrepreneurs individuels et les particuliers selon la norme n°18 pour les EI et la norme n°19 pour les particuliers)	Début d'année 2011	23 février 2011
lancement de la collecte	Fin mars 2011	21 Mars 2011
fin du travail des gestionnaires	Début octobre 2011	04 octobre 2011
base de données finale	Fin 2011	25 février 2012
retour d'information aux entreprises	Non	Non
mise à disposition du fichier	été 2012	Date inconnue à ce jour
Diffusion des premiers résultats	30 mars 2012	6 avril 2012
Diffusion de tableaux détaillés	été 2012	Date inconnue à ce jour

La préparation de l'enquête

existe-t-il un comité d'utilisateurs des résultats de l'enquête?	avec utilisateurs externes <input checked="" type="checkbox"/>	sans externe <input type="checkbox"/>	util <input type="checkbox"/>	Oui
nombre d'entreprises ayant fait l'objet d'un test de questionnaire				10
l'échantillon a-t-il fait l'objet d'une démarche de coordination ?				oui non <input checked="" type="checkbox"/>

L'échantillon enquêté

Type de sondage :

Sondage aléatoire simple stratifié

Plan de sondage :

Le tirage stratifié simple repose sur les croisements de cinq variables : le PTAC, l'année de mise en circulation, l'activité des entreprises, le statut des utilisateurs (professionnels, particuliers), le type de carburant utilisé. La population mère (composée de 5 792 000 véhicules en service au 1er janvier 2011) a été découpée en 38 strates

(cf. description de la stratification en annexe 1).

Ce tirage a été effectué dans la partie du fichier central des automobiles correspondant au champ de l'enquête. Les données du fichier central des automobiles sont issues des cartes grises et ont été appareillées avec les données du fichier des contrôles techniques afin d'écarter les véhicules détruits ou vendus à l'étranger.

Taux de sondage :

Les taux de sondage résultent :

- à l'origine d'un calcul d'optimisation au sens de Neyman réalisé en 2000.
- ensuite de la volonté de sur-représenter le transport de marchandises par rapport aux autres activités pour une meilleure description du TRM et d'échantillonner de façon progressive avec la classe de PTAC pour une meilleure précision des tonnes et des tonnes-kilomètres.
- volonté de sur-représenter également les autres énergies et les véhicules récents (les doubler car ils roulent plus que les autres)
- enfin la nécessité de se limiter à une taille d'échantillon d'environ 25000 véhicules comme en 2005 étant donné nos contraintes de budget et en moyens humains pour traiter l'information.

Le taux de sondage global est de 4,3 ‰ (3,5 ‰ en 2006) : la taille de l'échantillon de l'enquête est de 25 000 véhicules et la population mère est de 5 792 291.

Redressement de la non réponse totale :

Les poids de tirage ont été ensuite améliorés par un calage sur marges (programme CALMAR) précédé de redressement par la méthode des groupes de réponses homogènes (cf. note sur le calage en annexe 3). Les marges ont été calculées pour chacune des 38 strates.

Redressement de la non réponse partielle :

Le redressement de la non réponse partielle a porté sur une vingtaine de variables dont le kilométrage effectué en 2010 (cf. note en annexe 2). Hormis pour cette variable, le redressement de la non réponse partielle a été réalisé suivant deux logiques :

- imputation de la réponse la modalité la plus répandue dans la strate de diffusion pour les variables qualitatives (ex. CSP, statut (actif inactif)) ;
- imputation du taux moyen de la strate de diffusion pour les variables quantitatives (ex. : répartition du kilométrage en % en agglomération, sur route, autoroute...etc.).

Dans le cadre des précédentes enquêtes VUL, le redressement de la non réponse partielle n'avait concerné que le kilométrage parcouru durant l'année qui précédait le lancement de l'enquête.

Taux de réponse :

Années	Taux de questionnaires exploitables(*)
2001	62 %
2006	63 %
2011	62 %

(*) les questionnaires sont jugés exploitables lorsque la situation des véhicules associés est connue avec certitude, qu'ils soient en services, détruits, vendus à l'étranger ou hors champs

Taux de NPAI (retour à l'envoyeur) :

Enquête VUL	Taux de NPAI
2001	8 %
2006	7 %
2011	5 %

Taux de véhicules hors services (nombre de véhicules détruits, vendu à l'étranger ou hors champ rap l'échantillon) :

Du fait de l'appariement du parc des VUL issu de SIDIV avec le fichier des contrôles techniques en 201 le nombre de véhicules hors service contenu dans l'échantillon a été divisé par 2,5 entre 2006 et 2011.

Enquête VUL	Taux de véhicules hors service
2001	5 %
2006	5 %
2011	2 %

variables objectif pour la diffusion (ou pour les études)

▶ nombre : 6

▶ liste : tonnes transportées, tonnes-kilomètres réalisées, véhicules-kilomètres parcourus, kilométrage moyen réalisé par les véhicules, parc en service, parc utilisé

sous-populations de diffusion

▶ nombre 6

▶ liste :

1) variables relatives au véhicule :

Origine du véhicule (propriété, location), charge utile, PTAC, âge, carrosserie, consommation, climatisation,

2) variables relatives à l'utilisateur :

Situation juridique, catégorie socioprofessionnelle, statut (actif, inactif), APE.

3) variables relatives au kilométrage :

Kilométrage 2010, kilométrage journalier, kilométrage au compteur, kilométrage à l'achat.

Répartition du kilométrage par nature de trajet (agglomération, route, autoroute, site fermé ou autre), par zone de trajet (en local ou régional, national ou international).

Répartition du kilométrage en fonction des usages professionnels.

4) variables relatives aux usages :

fréquences d'utilisation (par classe de distance, par tranche de poids de la marchandise transportée, par tranche de poids de la marchandise tractée), usages exclusivement non professionnels, usages professionnels

5) variables relatives aux parcours durant deux jours de 2011 :

nature du transport (national/international) compte d'autrui /compte propre, lieux de chargement et de déchargement, nature de la marchandise, durée du parcours, distances parcourues (en charge et à vide), poids de la marchandise, nature du déplacement (transport de bien, marchandises, déménagement...).

Distribution des taux de sondage par strate de sondage

ENERGIE	supercarburants et gazole									autres énergies	
ACTIVITES	TRM et Poste : 4941A, B, C, 4942Z, 5229A, B, 5310Z, 5320Z			Autres activités (autres activités d'entreprise, administration, associations, particuliers)						Toutes activités	
				professionnels			particuliers			professionnels	particuliers
année de mise en circulation	avant 1991	1991 à 2000	2001 à 2010	avant 1991	1991 à 2000	2001 à 2010	avant 1991	1991 à 2000	2001 à 2010	Tous âges	
Classes de PTAC											
moins de 1,5 t	0,067	0,100	0,100	0,002	0,002	0,003	0,001	0,001	0,003	0,0667	0,0667
1,5 A 2,599 t	0,067	0,100	0,100	0,002	0,002	0,002	0,001	0,001	0,003		
2,6 A 3,499 t	0,067	0,100	0,100	0,002	0,002	0,002	0,001	0,001	0,003		
juste 3,5 t	0,067	0,100	0,100	0,002	0,002	0,002	0,001	0,001	0,003		

Indicateurs liés à la production

y a-t-il des relances ? **Oui**
 si oui, de quel type et combien ? **15 888 1ers rappels (dont 12 313 pour les professionnels et 3 575 les entreprises)**
9 272 seconds rappels pour les professionnels uniquement
 Pour l'enquête précédente (VUL 2006), deux rappels avaient été envoyés à l'ensemble des enquêtés mise en demeure pour les professionnels du transport routier de marchandise et des postes.

Nombre d'unités de l'échantillon initial **25 000**

Nombre de retours **17 878**

Nombre de « hors champ » (y compris unités disparues) dans les retours **706**

Nombre de questionnaires exploitables **15 606**

Taux de non-réponses selon les regroupements de strates de départ :

Définition : nombre d'entreprises ni répondantes, ni hors champ rapporté au nombre d'entreprises de l'échantillon diminué des hors champ (y compris les disparues) et des NPAI. Soit un taux de non-réponses global de 32,4 %.

=1-[15 606 / (25 000 - 706 - 1 196*)]

*** le chiffre de 1 196 désigne les Npai**

Pourcentage de non réponses-partielles, pour les variables qui ont été redressées :

Variable	% non réponse partielle
Kilométrage journalier habituel	22,3 %
Nombre de km parcourus en 2010	20,1 %
Fréquence d'utilisation du véhicule	19,4 %
CSP des utilisateurs de VUL à usage exclusivement non professionnel	20,2 %
statut des utilisateurs de VUL à usage exclusivement non professionnel (actifs, inactif)	6,7 %
Charge utile	3,4 %
Détention du véhicule (propriété, location)	2,7 %
Situation juridique	3,3 %
Usage du véhicule en agglomération, sur route,	

autoroute, site fermé ou autre	1,1 %	
Usage du véhicule en agglomération, sur route, autoroute, site fermé ou autre	1,2 %	
Usages exclusivement non professionnels	< 1 %	
Usages professionnels	< 1 %	
Dont		
transport de biens	< 1 %	
transport de marchandises	< 1 %	
Pourcentages de kilométrage correspondant aux professionnels	1,3 %	

Liste des variables (pas forcément variables-cibles) pour lesquelles le pourcentage global de non-réponses partielles est le plus élevé :

Kilométrage journalier habituel	22,3 %
Nombre de km parcourus en 2010	20,1 %
Fréquence d'utilisation du véhicule	19,4 %
CSP des utilisateurs de VUL à usage exclusivement non professionnel	20,2 %

Indicateurs de précision

Coefficient de variation pour les estimations des km parcourus en 2010	
Nombre de km parcourus en 2010	1,42 %
Nombre de km parcourus en 2010 par les professionnels	1,22 %
Nombre de km parcourus en 2010 par les particuliers	2,08 %

La diffusion

l'enquête a-t-elle fait l'objet d'une procédure d'archivage ?	Elle est prévue
existe-t-il une documentation pour la diffusion ?	Oui
aucune documentation <input type="checkbox"/>	une documentation simple <input checked="" type="checkbox"/>
un dictionnaire de données complet <input checked="" type="checkbox"/>	une documentation à usage externe <input checked="" type="checkbox"/>
nombre de publications, par type	
↪ type « Infos Rapides »	1
↪ type « 4 Pages » (Insee-Première, Le 4 pages des statistiques industrielles, ou équivalent)	Au moins 1
↪ type CD-Rom, « Insee-Résultats », web	Une mise en ligne de résultats détaillés
↪ articles, études	Selon l'actualité du moment, des analyses seront réalisées soit à la demande notamment pour évaluer des politiques publiques, soit diffusées largement.

Moyens relatifs à l'opération

Taille de l'équipe (en équivalent temps plein)

- équipe « statistique » : 1
- gestionnaires : 5

Moyens informatiques (nombre de jours de développement, maintenance dans le cas d'une opération régulière) : 10

2 mois de maintenance

Annexe 1 :

POPULATION MERE DU PLAN DE SONDAGE VUL (PARCS au 1-1-2011) :

ENERGIE	supercarburants et gazole									autres énergies	
ACTIVITES	TRM et Poste : 4941A, B, C, 4942Z, 5229A, B, 5310Z, 5320Z			Autres activités (autres activités d'entreprise, administration, associations, particuliers)						Toutes activités	
				professionnels			particuliers			professionnels	particuliers
année de mise en circulation	avant 1991	1991 à 2000	2001 à 2010	avant 1991	1991 à 2000	2001 à 2010	avant 1991	1991 à 2000	2001 à 2010	Tous âges	
Classes de PTAC											
moins de 1,5 t	349	2 215	4 182	33 687	108 752	121 867	243 954	415 934	123 026	15399	24378
1,5 A 2,599 t	199	3 758	64 230	16 462	156 033	1 126 238	147 384	525 909	703 020		
2,6 A 3,499 t	102	1 263	24 914	9 831	70 885	536 230	65 464	192 160	314 037		
juste 3,5 t	316	2 326	22 570	16 294	71 778	326 813	68 776	114 458	117 098		

Total 5 792 291

Echantillon

ENERGIE	supercarburants et gazole									autres énergies	
ACTIVITES	TRM et Poste : 4941A, B, C, 4942Z, 5229A, B, 5310Z, 5320Z			Autres activités (autres activités d'entreprise, administration, associations, particuliers)						Toutes activités	
				professionnels			particuliers			professionnels	particuliers
année de mise en circulation	avant 1991	1991 à 2000	2001 à 2010	avant 1991	1991 à 2000	2001 à 2010	avant 1991	1991 à 2000	2001 à 2010	Tous âges	
Classes de PTAC											
moins de 1,5 t	21	177	188	47	129	153	95	328	67	964	1431
1,5 A 2,599 t	14	390	6 481	27	335	2 708	74	598	1 629		
2,6 A 3,499 t	6	141	2 551	16	152	1 296	35	198	743		
juste 3,5 t	22	248	2 370	26	141	799	31	112	257		

Total 25 000

Annexe 2 : Note sur l'appariement du parc VUL avec le fichier des contrôles techniques

MINISTÈRE DE L'ÉCOLOGIE, DU DÉVELOPPEMENT DURABLE, DES TRANSPORTS ET DU LOGEMENT

Commissariat général au développement durable

Paris, le 19 mars 2012

Service de l'observation et des statistiques
Sous-direction des statistiques des transports
Bureau des statistiques de la route et des véhicules

Note sur l'appariement du parc VUL avec le fichier des contrôles techniques

Nos réf. : CGDD/SOeS/OST/BSRV/XXXX

Affaire suivie par : Véronique Coutant

veronique.coutant@developpement-durable.gouv.fr

Tél. 01 40 81 73 55 – Fax : 01 40 81 17 72

Courriel : SOeS.CGDD@developpement-durable.gouv.fr

Objet : actualisation du parc VUL au 1er janvier 2011 avec le fichier des contrôles techniques
PJ :

Le parc des VUL au 1er janvier **2011** a été estimé à partir du fichier SIV en retenant :

- les VUL récents du SIV (n'ayant pas passé de 1er contrôle technique) : 1ère immatriculation en 2010, 2009, 2008, 2007 ou 2006; (1)
- les VUL de plus de 4 ans issus du fichier des contrôles techniques pour lesquels le dernier contrôle techniques avait eu lieu en 2008 ou après. (2)

Le fichier des contrôles techniques utilisé était celui en date du 1er avril 2010.

Le parc VUL au 1er janvier 2011 est ainsi estimé à 5 281 000 véhicules ou plutôt à 5 280 950 véhicules si l'on élimine les doublons (104).

Cette estimation semble néanmoins plutôt basse par rapport à l'estimation du parc au 1er janvier 2006 (6 236 000 VUL de 20 ans et moins). Effet, on observe de fortes immatriculations de véhicules neufs sur la période même si aucune information n'est disponible par ailleurs sur les « sorties de parc »¹ pour tempérer l'effet de ces « entrées de parc ».

Un minorant possible à l'estimation du parc ainsi calculé est le retard entre deux contrôles techniques pour un nombre de véhicules non négligeable. Le délai maximal entre deux contrôles techniques est en pratique de 5 ans. En théorie, ce délai maximal est de 1 an (après la 4ème année du véhicule) puisque un contrôle complémentaire pollution a lieu 12 mois après un contrôle technique. Mais, 9,7% des véhicules ont un délai supérieur à 2 ans et 2 mois.

Pour estimer le nombre de véhicules exclus peut être « à tort » du parc des VUL au 1er janvier 2011, un appariement entre le parc des 5 281 000 VUL et le fichier des contrôles techniques actualisé au 1er avril 2011 a été entrepris. Les premières immatriculations de 2006 devraient se retrouver toutes puisque faisant partie du parc récent (1).

Le fichier des contrôles techniques du 1er avril 2010 au 1er avril 2011 comprend 3 693 903 immatriculations de VUL. Ce nombre ne correspondant pas exactement aux nombre de véhicules ayant passé un contrôle technique entre le 1er avril 2010 et le 1er avril 2011 car des erreurs de saisie ne sont pas rares et certains VUL

1 Si la carte grise n'est pas restituée à la préfecture.

peuvent se retrouver sur deux lignes s'ils ont passé deux contrôles techniques (si le premier s'avère négatif ou s'il y a vente).

Un traitement préalable a été nécessaire pour comparer les deux fichiers. En effet, le numéro d'immatriculation dans le parc est sous la forme 10MS2815 pour les anciennes immatriculations alors que dans le fichier des contrôles techniques, le numéro d'immatriculation est sous la forme 2815MS10. Cette distinction n'existe pas pour les véhicules immatriculés sous le nouveau système d'immatriculation (AB032FR par exemple). C'est pourquoi les deux types de véhicules ont été traités l'un après l'autre.

Le tableau ci-dessous résume les différents appariements réalisés :

	méthode	nombre d'appariés	restants à appairer
1	numéro d'immatriculation (ancien système)	1 616 979	2 076 854
2	numéro d'immatriculation nouveau système)	266 081	1 810 773
3	vin_date = code VIN + date de 1ère mise en circulation	390 863	1 419 910
4	vin_date2 = code VIN + année et mois de 1ère mise en circulation	649 995	769 915
5	vin_date4 = code VIN + année et jour de 1ère mise en circulation	1 366	768 549
6	numéro d'immatriculation sans 0	113 603	654 946
7	vin avec contrainte de distance sur date 1ère circulation et immatriculation*	3 960	650 985
8	vin_date2 à partir du parc récent puis appariement avec l'identifiant SIDIV récupéré	27 669	623 316
9	immat à partir du parc récent puis appariement avec l'identifiant SIDIV récupéré	58	623 258
10	immat précédente à partir du parc récent puis appariement avec l'identifiant SIDIV récupéré	226	623 032
11	vin_date2 à partir du parc CT puis appariement avec l'identifiant SIDIV récupéré	6	623 026
12	immat précédente à partir du parc CT puis appariement avec l'identifiant SIDIV récupéré	72	622 954
13	immat à partir du parc CT puis appariement avec l'identifiant SIDIV récupéré	107	622 847

* une observation manquante.

Les appariements 8, 9 et 10 ont été réalisés non pas avec le fichier du parc final mais avec le parc récent. Le parc final est censé être la fusion du parc récent (1) et du parc issu des contrôles techniques (2) mais il s'avère que certains véhicules ont été « perdus » lors de cette fusion.

L'appariement avec le parc récent en fonction du vin_date2 a permis d'apparier 38 282 véhicules mais seulement 27 669 véhicules ont pu être ensuite appariés avec le parc final. Cet appariement ayant été réalisé à partir de l'identifiant SIDIV (T_SIDIV_VEHICULE_ID), les 38 282 véhicules auraient dû tous être appariés mais 10 613 véhicules ne l'ont pas été ⇒ **Le parc final ne comprend pas l'intégralité du parc récent.**

De la même manière, l'appariement avec le parc récent en fonction de l'immatriculation a permis d'apparier 9 118 véhicules mais seulement 58 véhicules ont pu être ensuite appariés avec le parc final (à partir de l'identifiant SIDIV). Il reste alors 9 060 véhicules non appariés.

Un autre appariement similaire a été réalisé en fonction de l'immatriculation précédente répertoriée dans le fichier parc récent. 10 016 véhicules ont pu être ainsi appariés mais seulement 226 véhicules ont été appariés avec le fichier final (à partir de l'identifiant SIDIV). Il reste alors 9 790 véhicules non appariés.

En tout, 29 463 véhicules récents avec un identifiant SIDIV ne se retrouvent pas dans le parc final et il ne peut pas s'agir d'erreur de saisie puisque l'appariement s'est fait à partir de l'identifiant SIDIV.

Les appariements 11,12 et 13 ont été réalisés à partir du fichier du parc issu des contrôles techniques (2) :

- l'appariement en fonction du vin_date2 a permis d'apparier 19 716 véhicules mais seulement 6 véhicules ont été ensuite appariés avec le parc final (à partir de l'identifiant SIDIV). Il reste alors 19 706 véhicules non appariés.
- l'appariement en fonction de l'immatriculation précédente a permis d'apparier 429 véhicules mais seulement 72 véhicules ont été ensuite appariés avec le parc final (à partir de l'identifiant SIDIV). Il reste alors 357 véhicules non appariés.
- l'appariement en fonction de l'immatriculation a permis d'apparier 18 608 véhicules mais seulement 107 véhicules ont été ensuite appariés avec le parc final (à partir de l'identifiant SIDIV). Il reste alors 18 505 véhicules non appariés.

En tout, 38 568 véhicules supplémentaires avec un identifiant SIDIV ne se retrouvent pas dans le parc final et il ne peut pas s'agir d'erreur de saisie puisque l'appariement s'est fait à partir de l'identifiant SIDIV.

Il reste alors 622 847 véhicules issus des contrôles techniques qui n'ont pas pu être appariés avec le parc final dont 38 568 +29 463 = 68 031 véhicules qui ont un identifiant SIDIV.

Une nouvelle extraction d'un parc VUL SIDIV permet de retrouver un grand nombre de ces véhicules non appariés. Les appariements suivants ont permis d'ajouter 416 695 véhicules.

14	num éro d'im m atriculation (ancien systèm e)	186 518	436 329
15	num éro d'im m atriculation (nouveau systèm e)	30 897	405 432
16	im m atriculation précédente	54 558	350 874
17	vin_date avec longueur du VIN sur 17 caractères	60 593	290 281
18	num éro d'im m atriculation sans 0 avec vin sim ilaire**	14 965	275 322
19	vin_date2	85 912	189 410
	- doublons du fichier des contrôles techniques	16 748	172 662
	TOTAL	416 695	

** 6 doublons dans le fichier des restants à apparié

Par ailleurs, un nombre non négligeable de véhicules récents (1) ont pu être récupérés à partir de ce nouveau parc : 96 595 VUL récents viennent s'ajouter au parc initial (5 280 950 VUL).

Au final, le parc VUL est estimé à 5 794 240 véhicules. = 5 280 950 véhicules du fichier de départ + 416 695 véhicules appariés avec le parc nouvellement extrait + 96 995 véhicules récents.

Pour mémoire :

D'autres essais d'appariement entre le fichier des contrôles techniques et le parc initial au 1er janvier 2011 n'ont pas abouti à des résultats cohérents (véhicules appariés à tort). Il s'agit de l'appariement réalisé à partir du « vin_date3 » des véhicules représentant la concaténation des 15 premiers caractères du numéro VIN et de la date de première immatriculation : 22 786 véhicules pourraient être ainsi appariés mais un test sur quelques VUL montre que certains véhicules appariés le sont visiblement à tort. → En utilisant la fonction compged (comparateur de distance) entre les vin_date pour ne garder que les petites distances (distance=20 soit 157 véhicules...).

Un autre appariement à partir des 6 premiers caractères du numéro d'immatriculation a été jugé inexact.

Un autre appariement est effectué à partir du vin_date en supprimant les O (c'est-à-dire les zéros contenus dans le vin). Il n'est pas retenu car les lignes appariées ne sont pas des véhicules identiques (numéro d'immatriculation très différents). Il y a peut être eu un changement d'immatriculation pour certains d'entre eux, ce qui explique leur doublonnage dans le fichier des contrôles techniques.

Parmi les 622 847 véhicules non appariés, on observe une proportion anormalement élevée de véhicules avec une 1ère immatriculation égale à 2006 : 17% des non appariés contre 9% avec l'année 2005 et 1% avec l'année 2008 en 1ère immatriculation. Cette observation semble surprenante puisque le parc VUL initial comprenait tous les véhicules récents (1ère immatriculation de 2006 à 2010).

Pour information, lors de l'appariement des bus et cars, un problème était apparu : les bus et cars présents dans le fichier des contrôles techniques n'ont pas tous pu être appariés aux bus et cars du parc SIDIV. Il a fallu reprendre le parc de l'année précédente qui avait été constitué à partir de l'ancien système FCA pour constituer un parc roulant cohérent.

Enquête VUL : Traitement de la non réponse totale

(version modifiée suite à la révision à la hausse de la population mère en février 2012)

Sur les 25 000 unités enquêtées, 15 093 ont répondu, 9 872 sont non répondantes et 35 sont hors-champ.

Pour corriger la non réponse totale nous avons deux options :

- Le calage sur marges direct : en une étape, sans corriger la non-réponse au préalable, on modifie les poids d'échantillonnage en les calant sur les marges de variables auxiliaires connues.
- Le redressement en deux étapes : modèle de réponse suivi d'un calage sur marge
Après avoir corrigé les poids par les GRH, on effectue un calage sur marges à partir de variables dont on connaît l'estimation sur la population totale.

Nous avons opté pour la deuxième option dans la mesure où l'efficacité du calage sur marge direct est conditionnée par la présence, dans la définition des marges, de l'ensemble des variables expliquant la non réponse. Hors nous n'avons pas la certitude que les variables de stratification expliquaient à elles seules la non réponses.

- **Etape 1 : Constitution des GRH**

Un modèle logistique a été mis en place pour expliquer les comportements de non réponse.

34 % des entreprises n'ont pas répondu.

Les « hors-champ » ont été exclus de la modélisation, la population considérée s'élève donc à 24 465 unités.

Plusieurs variables ont été testées :

- la localisation géographique (ZEAT2 : IDF_DOM, est_BP et reste)
- PART (véhicule particulier ou pas)
- ANNEE (avant 91, entre 1991 et 2000, après 2000)
- TRANCHE_PTR (4 tranches de PTAC)
- CARB (gazole ou autres)
- Secteur d'activité REGROUP (4 regroupement de secteurs d'activité)

Le khi-deux et le V de Cramer permettent de repérer les variables les plus explicatives de la non-réponse (variable ETAT)

D'après le V de Cramer, la zone géographique, le secteur d'activité et le statut de l'utilisateur (particulier/professionnel) sont les variables les plus liées à la non réponse (leurs statistiques sont plus proches de 1).

V de Cramer

ZEAT2	Part	Année	Tranche_ptr	carb	regroup
0,0897	0,1839	0,0651	0,0345	0,0397	0,0769

Cependant, d'après le test du Khi-2 les six variables sont significatives et ont été incluses dans le modèle.

La construction du modèle implique de choisir une modalité de référence pour chaque variable explicative. La modalité de référence est celle dont le taux de réponse est le plus proche du taux de réponse global (environ 60 %).

Construction du modèle :

```
PROC LOGISTIC DATA =cal.exploit_mars_n (where = (etat in ("RE" "NR")))
outest=est;
CLASS part(REF = '0') / PARAM = reference; /*la modalité dont le tx de
réponse est le plus proche du tx de réponse global 60%*/
CLASS annee (REF = 'après 2000      ') / PARAM = reference;
CLASS tranche_ptr(REF = '          3,5t ') / PARAM = reference;
CLASS carb (REF = 'gazole' ) / PARAM = reference;
*CLASS zeat (REF = 'centre-est' ) / PARAM = reference;
CLASS regroup (REF = 'trm_large ' ) / PARAM = reference;
CLASS zeat2 (REF = 'reste' ) / PARAM = reference;
MODEL etat = part annee tranche_ptr carb regroup zeat2/ RSQUARE;
output out=cal.estim predicted=probest;
TITLE 'Modélisation Logistique Forward';
RUN;
```

Sortie SAS :

Estimations par l'analyse du maximum de vraisemblance						
Paramètre		DDL	Valeur estimée	Erreur type	Khi-2 de Wald	Pr > Khi-2
Intercept		1	-0.6028	0.0385	245.5811	<.0001
part	1	1	0.8307	0.0401	430.0639	<.0001
annee	avant 91	1	0.0916	0.0796	1.3240	0.2499
annee	entre 91 et 2000	1	0.2823	0.0407	48.1759	<.0001
tranche_ptr	entre 1,5t et 2,599t	1	-0.2036	0.0398	26.2455	<.0001
tranche_ptr	entre 2,600t et 3,5t	1	-0.0152	0.0438	0.1211	0.7279
tranche_ptr	moins de 1,5t	1	-0.1206	0.0645	3.4957	0.0615
carb	autres	1	-0.2725	0.0487	31.2824	<.0001
regroup	autr	1	0.1897	0.0360	27.7597	<.0001
regroup	sect_1	1	-0.1637	0.0433	14.2958	0.0002
zeat2	IDF_DOM	1	0.5698	0.0358	252.7582	<.0001
zeat2	est_BP	1	-0.2875	0.0324	78.8222	<.0001

Il ressort que les modalités significatives sont :

Véhicule particulier : part= « 1 » (professionnel),

Annee « entre 91 et 2000 »,

tranche_ptr « entre 1,5t et 2,599t »,

carb= »autres »

localisation géographique ZEAT2 : Ile-de-France et DOM (IDF_DOM) et Est et Bassin Parisien (est_BP)

et secteur d'activité « sect_1 » (agriculture, commerce et construction) « et « austr » (hors TRM).

Les modalités non significatives sont reclassées dans la modalité « xxxx_autre » de la variable associée.

La variable « Année » se retrouve avec 2 modalités :

- la modalité significative : « entre 91 et 2000 » ;
- les autres modalités : « année_autre » (ces dernières comprennent également la modalité de référence)

La variable « secteur d'activité » se retrouve avec 3 modalités :

- la modalité significative : « sect_1 »

- la modalité significative : « autre »
- les autres modalités : « année_autre » (ces dernières comprennent également la modalité de référence)

...etc.

En croisant toutes les ensembles composés de 6 modalités (six variables), et chacune des variables ayant on obtient $2 \times 2 \times 2 \times 2 \times 3 \times 3 = 144$ croisements. Comme certains croisements ne comportent aucune observation, seuls les GRH renseignés sont listés sous SAS :

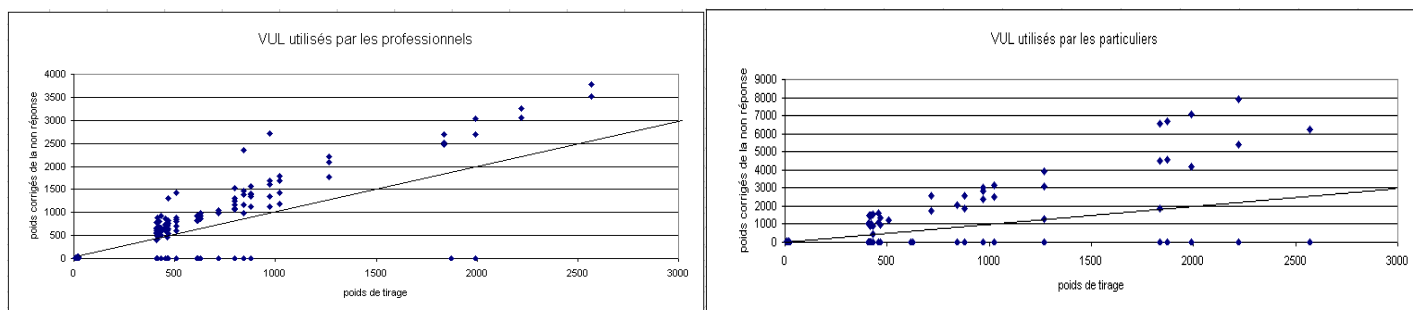
Les poids initiaux (d'échantillonnage) sont modifiés :

- pour les véhicules répondants (ETAT= « RE »), on les augmente en divisant par le taux de réponse dans le GRH
- pour les hors champ (ETAT= « HC »), on conserve les poids initiaux
- pour les non répondants (ETAT= « NR »), les poids de non réponse sont nuls

A l'issue de cette repondération, la somme des poids corrigés de la non réponse est égale à la somme des poids initiaux (5 792 291) mais les effectifs de strate ont subi des distorsions comprises entre 0,62 et 1,26.

Numéro de Strate	marges après repondération	Pourcentage	population mère	Pourcentage population	Rapport de distorsion des marges (marges après repondération / population mère)
1	328	0,01	349	0,01	0,94
2	1 583	0,03	2 215	0,04	0,71
3	4 308	0,07	4 182	0,07	1,03
4	170	0,00	199	0,00	0,85
5	2 808	0,05	3 758	0,06	0,75
6	60 791	1,05	64 230	1,11	0,95
7	46	0,00	102	0,00	0,45
8	860	0,01	1 263	0,02	0,68
9	21 164	0,37	24 914	0,43	0,85
10	270	0,00	316	0,01	0,85
11	1 748	0,03	2 326	0,04	0,75
12	19 278	0,33	22 570	0,39	0,85
13	22 092	0,38	33 687	0,58	0,66
14	82 666	1,43	108 752	1,88	0,76
15	125 604	2,17	121 867	2,10	1,03
16	261 708	4,52	243 954	4,21	1,07
17	477 555	8,24	415 934	7,18	1,15
18	155 611	2,69	123 026	2,12	1,26
19	12 770	0,22	16 462	0,28	0,78
20	134 665	2,32	156 033	2,69	0,86
21	1 092 810	18,87	1 126 238	19,44	0,97
22	132 572	2,29	147 384	2,54	0,90
23	547 482	9,45	525 909	9,08	1,04
24	758 143	13,09	703 020	12,14	1,08
25	6 122	0,11	9 831	0,17	0,62
26	58 026	1,00	70 885	1,22	0,82
27	520 863	8,99	536 230	9,26	0,97
28	75 134	1,30	65 464	1,13	1,15
29	202 718	3,50	192 160	3,32	1,05
30	310 440	5,36	314 037	5,42	0,99
31	14 192	0,25	16 294	0,28	0,87
32	70 227	1,21	71 778	1,24	0,98
33	311 523	5,38	326 813	5,64	0,95
34	66 881	1,15	68 776	1,19	0,97
35	85 654	1,48	114 458	1,98	0,75
36	110 232	1,90	117 098	2,02	0,94
37	15 027	0,26	15 399	0,27	0,98
38	28 220	0,49	24 378	0,42	1,16
	5 792 291	100,0	5 792 291	100,0	

Les pondérations initiales ont subi des distorsions comprises entre 1,0 et 2,8 pour les VUL utilisés par les professionnels et entre 1 et 3,6 pour les VUL de particuliers (graphiques ci-dessous). L'amplitude plus importante pour les particuliers qui s'explique par des taux de non réponses plus élevés que pour les professionnels.



Le calage sur marge permet de rétablir les populations initiales par strates de diffusion.

- **Etape 2 : Calage sur marges**

Les populations mères sont connues sur les 38 strates d'échantillonnage. Ces dernières vont servir de marges de calage.

GENRE	CAMIONNETTES : 5										
ENERGIE	supercarburants et gazole								autres énergies		
ACTIVITES	TRM et Poste : 4941A, B, C, 4942Z, 5229A, B, 5310Z, 5320Z			Autres activités (autres activités d'entreprise, administration, associations, particuliers)						entreprises	particuliers
				Autres activités (autres activités d'entreprise, administration, associations)			particuliers				
modèle de questionnaire	A			B			C			D	E
année de mise en circulation	avant 1991	1991 à 2000	2001 à 2010	avant 1991	1991 à 2000	2001 à 2010	avant 1991	1991 à 2000	2001 à 2010	Tous âges	
moins de 1,5 t et N.D.	strate 01	strate 02	strate 03	strate 13	strate 14	strate 15	strate 16	strate 17	strate 18	strate 37	strate 38
1,5 A 2,5 t	strate 04	strate 05	strate 06	strate 19	strate 20	strate 21	strate 22	strate 23	strate 24		
2,6 A 3,4 t	strate 07	strate 08	strate 09	strate 25	strate 26	strate 27	strate 28	strate 29	strate 30		
juste 3,5 t	strate 10	strate 11	strate 12	strate 31	strate 32	strate 33	strate 34	strate 35	strate 36		

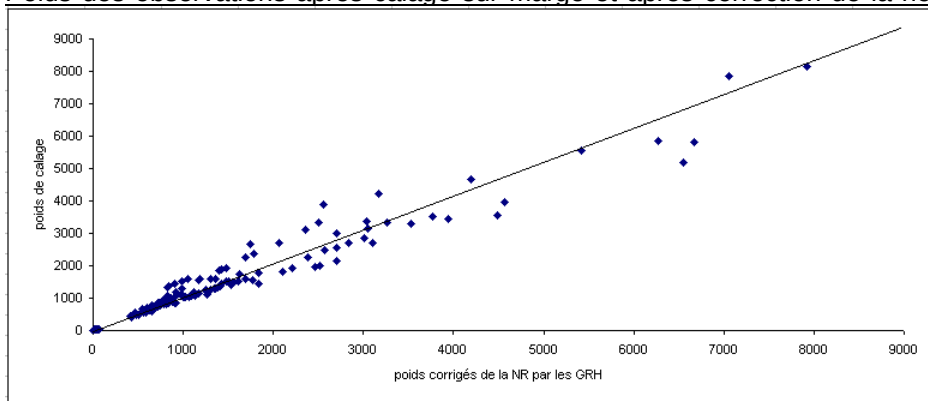
Par la méthode du raking ratio, l'algorithme converge en 5 itérations. Après cette étape, les marges de l'échantillon coïncident avec celles de la population. Les rapports de poids (final/corrigés de la non réponse) varient entre 0,8 et 2,2.

numéro de strate	Effectif combinaison (*)	Rapport de poids
1	15	1,06
2	103	1,40
3	135	0,97
4	10	1,17
5	220	1,34
6	4288	1,06
7	2	2,23
8	77	1,47
9	1509	1,18
10	14	1,17
11	145	1,33
12	1415	1,17
13	20	1,52
14	63	1,32
15	105	0,97
16	43	0,93
17	160	0,87
18	35	0,79
19	15	1,29
20	185	1,16
21	1780	1,03
22	33	1,11
23	303	0,96
24	858	0,93
25	7	1,61
26	82	1,22
27	869	1,03
28	16	0,87
29	94	0,95
30	341	1,01
31	16	1,15
32	92	1,02
33	525	1,05
34	14	1,03
35	40	1,34
36	123	1,06
37	589	1,02
38	787	0,86

(*) l'effectif de combinaison est l'effectif de la strate associé à une pondération finale non nulle

Le calage a modifié légèrement les pondérations corrigées de la non réponse totale (graphique ci-dessous). Les distorsions de pondérations sont beaucoup plus faibles qu'à l'issue de la phase intermédiaire de correction de la non réponse totale décrite plus haut : les rapports de poids sont compris dans l'intervalle [0,8 ; 2,2] pour les VUL utilisés par les professionnels et [0,8 ; 1,5] pour les VUL utilisés par les particuliers.

Poids des observations après calage sur marge et après correction de la non réponse par les GRH.



Annexe 4 : statistiques sur la charge statistique générée par la réponse à l'enquête VUL 2011.

(liste des entreprises ayant rempli au moins 10 questionnaires).

Obs	Siren	Raison sociale	Nombre de questionnaires VUL remplis	temps de réponse total en mn (*)
1	356000828	LA POSTE	281	12 645
2	356000117	LA POSTE	279	12 555
3	356000729	LA POSTE	268	12 060
4	356000737	LA POSTE	172	7 740
5	356000521	LA POSTE	168	7 560
6	356000414	LA POSTE	162	7 290
7	356000547	LA POSTE	141	6 345
8	356000240	LA POSTE	139	6 255
9	356000257	LA POSTE	137	6 165
10	356000935	LA POSTE	129	5 805
11	356000000	LA POSTE	125	5 625
12	444608442	ERDF	114	5 130
13	356000232	LA POSTE	113	5 085
14	356000265	LA POSTE	113	5 085
15	356000539	LA POSTE	113	5 085
16	356000315	LA POSTE	106	4 770
17	356000919	LA POSTE	104	4 680
18	356000224	LA POSTE	102	4 590
19	356000422	LA POSTE	95	4 275
20	356000745	LA POSTE	95	4 275
21	444786511	GRDF	84	3 780
22	356000216	LA POSTE	82	3 690
23	356000430	LA POSTE	72	3 240
24	356000836	LA POSTE	61	2 745
25	342630340	TSE EXPRESS MEDICAL	45	2 025
26	552049447	SOC NAT DES CHEMINS DE FER FRANCAIS	29	1 305
27	217500016	COMMUNE DE PARIS	26	1 170
28	380129866	FRANCE TELECOM	26	1 170
29	494956774	DHL INTERNATIONAL EXPRESS (FRANCE)	25	1 125
30	334175221	UNITED PARCEL SERVICE FRANCE SNC	21	945
31	419448683	OCCITAN TRANSPORTS ENTRETIEN	20	900
32	785792433	SOCIETE GENERALE DE PRESTATIONS DE SERVICES	20	900
33	323658062	LOCA POIDS LOURDS	17	765
34	417552460	AXEREAL SERVICES	16	720
35	546650334	DACHSER FRANCE	16	720
36	356000976	LA POSTE	15	675
37	377895404	DHL EXPRESS (FRANCE) SAS	14	630
38	217603513	COMMUNE DE LE HAVRE	13	585
39	447895954	FRAIKIN ASSETS	12	540
40	552046955	GDF SUEZ ENERGIE SERVICES	12	540
41	344680855	ALTEAD LOCATION	11	495
42	381061027	A TOUTE VITESSE	11	495
43	309543288	QUEGUINER TRANSPORTS	10	450
44	332747690	AYALA	10	450
45	339379984	SAUR	10	450
46	423262609	MOUSSET LOCATION	10	450
47	552081317	ELECTRICITE DE FRANCE	10	450
48	662043116	OFFICE NATIONAL DES FORETS	10	450

(*) le temps de réponse total est estimé sur la base d'un temps de réponse moyen par questionnaire de 45 mn.