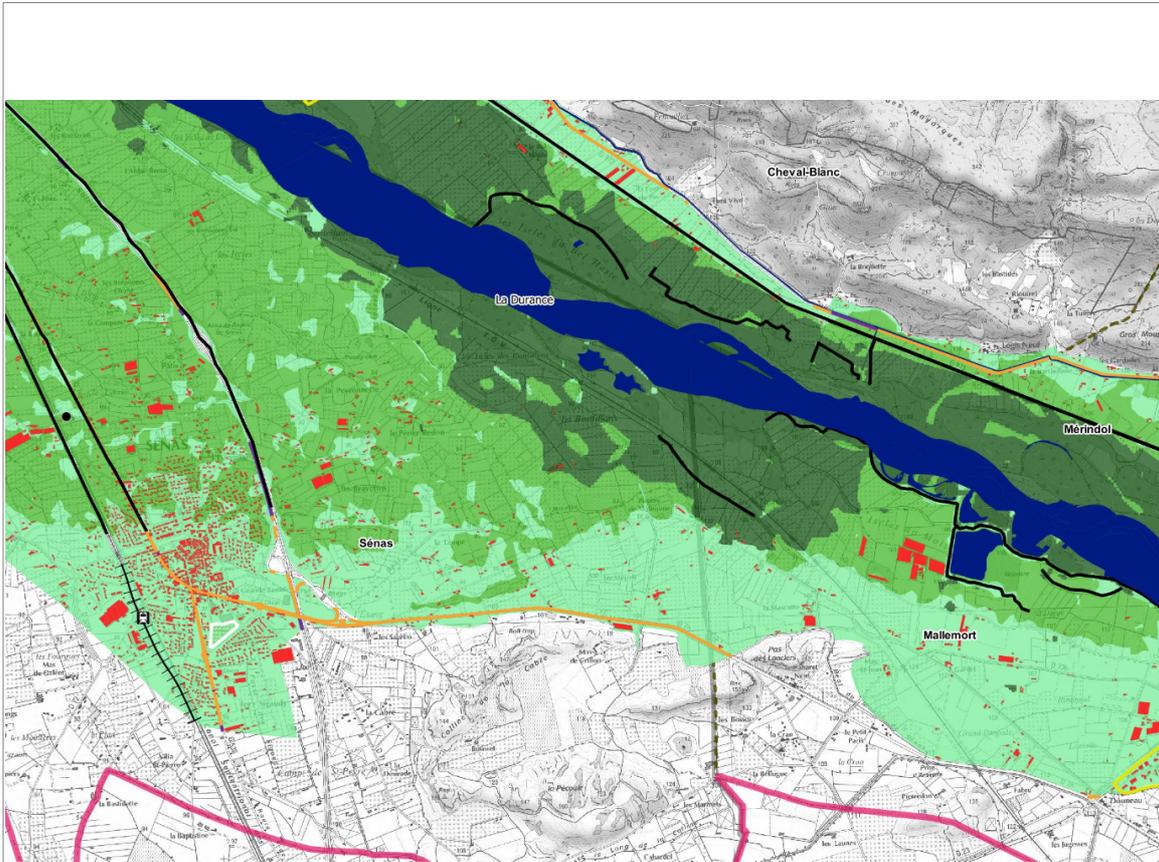


# Directive inondation – cycle 2

Fourniture de données pour le calcul de la population et des emplois

Directive inondation – Cartographie des risques du TRI d'Avignon



# Directive inondation – cycle 2

## Fourniture de données pour le calcul de la population et des emplois

### Historique des versions du document

Version	Date	Commentaire
1	29 juin 2018	

### Affaire suivie par

<b>Christophe MOULIN</b> - Département Risques Eau Construction – Service vulnérabilité et gestion de crise
Tél. : 04 42 24 71 56
Courriel : <a href="mailto:christophe.moulin@cerema.fr">christophe.moulin@cerema.fr</a>
<b>Site d'Aix-en-Provence</b> : Cerema Méditerranée – CS 70499 – 13593 AIX-EN-PROVENCE

### Références

n° d'affaire : C17MR0050

maître d'ouvrage : DGPR (M. Emmanuel Vullierme)

Rapport	Nom	Date	Visa
Établi par	Christophe Moulin, Sylvain Charaud	29/06/18	
Avec la participation de	Bernard Rongione		

---

### Résumé de l'étude :

L'application de la directive inondation par les services déconcentrés nécessite, dans la phase dite « cartographie », un traitement de données dont les populations résidentes et les emplois.

Le Cerema a procédé à un traitement national de données pour faciliter et homogénéiser la mise en œuvre. Le présent rapport présente les méthodes utilisées pour produire les fichiers déposés sur un site FTP.

# SOMMAIRE

<b>1 CONTEXTE.....</b>	<b>4</b>
<b>2 CALCUL DE LA POPULATION PERMANENTE.....</b>	<b>5</b>
2.1 Données sources.....	5
2.1.1 Populations à l'IRIS 2014, INSEE.....	5
2.1.2 Fichiers fonciers 2016 (« MAJIC 2016 »), DGIFP.....	5
2.2 Méthode de traitement.....	5
2.2.1 Répartition des données de population sur les parcelles, IRIS par IRIS.....	5
2.3 Données livrées.....	6
2.4 Calcul de la population en zone inondée.....	6
<b>3 VARIATIONS SAISONNIÈRES DE LA POPULATION COMMUNALE.....</b>	<b>7</b>
3.1 Introduction.....	7
3.2 Capacité d'accueil et populations touristiques.....	7
3.2.1 Capacité d'accueil.....	7
3.2.2 Définitions.....	8
3.2.3 Ratios.....	9
3.3 Données disponibles.....	9
3.3.1 Données INSEE publiques.....	9
3.4 Méthode proposée.....	10
3.4.1 Structure des fichiers fournis.....	11
<b>4 ESTIMATION DU NOMBRE D'EMPLOIS DANS LES SURFACES INONDABLES.....</b>	<b>11</b>
4.1 Objectif.....	11
4.2 Les bases de données de l'INSEE.....	12
4.3 Constitution du semis de points « emplois ».....	14
4.3.1 Extraction de la base SIRENE.....	14
4.3.2 Le géocodage de SIRENE Base de données.....	15
4.3.3 Exclusions préalables.....	16
4.3.4 Calcul de la fourchette communale et des ratios correctifs.....	17
4.3.5 Calcul des fourchettes pour les établissements géocodés.....	18
4.3.6 Fichiers obtenus.....	19
4.3.7 Sources d'erreur et biais.....	20
4.4 Évaluation du nombre d'emplois dans les surfaces inondables.....	20
4.4.1 Croisement avec les surfaces inondables.....	20
4.4.2 Représentation.....	21

# 1 Contexte

La directive européenne « inondation » du 23 octobre 2007 a été transposée en droit français par l'article 221 de la LENE du 12 juillet 2010, et par le décret 2011-227 du 2 mars 2011, dans les articles L566 et R566 du code de l'environnement.

La mise en œuvre comporte plusieurs étapes :

- une évaluation préliminaire des risques d'inondation (EPRI),
- une identification des territoires à risque important d'inondation (TRI),
- la cartographie des surfaces inondables et des risques d'inondation dans ces TRI,
- l'élaboration de plans de gestion (PGRI).

Cette mise en œuvre suit un rythme cyclique de 6 ans. Le deuxième cycle a débuté et il a été demandé au Cerema de fournir les données concernant la population et l'emploi servant de base à l'établissement des cartographies de risque.

Ces données avaient été fournies par le Cerema (alors CETE Méditerranée) lors du premier cycle, sur la base d'un travail méthodologique permettant :

- de dénombrer la population permanente dans les surfaces inondables,
- d'estimer le nombre d'emplois dans ces surfaces,
- de déterminer un taux de variation saisonnière de la population communale.

Un rapport datant de 2012 rend compte de ces méthodologies.

Le présent rapport rend compte des calculs effectués pour le deuxième cycle et des méthodologies utilisées, adaptées, à la marge, de celles du 1er cycle.

## 2 Calcul de la population permanente

### 2.1 Données sources

#### 2.1.1 Populations à l'IRIS 2014, INSEE

Il s'agit de données de population infracommunales, issues du recensement de la population INSEE. Les données de population (tableaux) sont associées aux contours des IRIS (polygones). Les données utilisées sont :

- Contour IRIS de 2016, disponibles sur le site [data.gouv.fr](http://data.gouv.fr)
- Population en 2014 à l'IRIS, disponibles sur le site de l'INSEE

Les communes d'au moins 10 000 habitants et la plupart des communes de 5 000 à 10 000 habitants sont découpées en IRIS (Îlots Regroupés pour l'information Statistique). Ce découpage, maille de base de la diffusion de statistiques infracommunales, constitue une partition du territoire de ces communes en "quartiers" dont la population est de l'ordre de 2 000 habitants. La France compte environ 16 000 IRIS dont 650 pour les DOM.

Par extension, afin de couvrir l'ensemble du territoire, on assimile à un IRIS chacune des communes non découpées en IRIS.

#### Remarque :

Au moment de lancer les traitements, les données carroyées de population à 200 m n'étaient plus mises à jour depuis 2010. C'est pourquoi ce sont les données de population à l'IRIS, plus récentes, qui ont été utilisées.

#### 2.1.2 Fichiers fonciers 2016 (« MAJIC 2016 »), DGIFP

Ces données ont été préalablement traitées (géoréférencement, anonymisation, remontée de données utiles à l'échelle de la parcelle...) et livrées par le Cerema Nord Picardie à tous les services du ministère.

### 2.2 Méthode de traitement

Les traitements ont été développés et réalisés au Cerema Méditerranée sur PostGIS / Postgresql, grâce à des scripts permettant d'automatiser ces calculs.

#### 2.2.1 Répartition des données de population sur les parcelles, IRIS par IRIS

Dans chaque IRIS la formule suivante est appliquée sur chacune des parcelles :

$$pop_{Parcelle} = pop_{IRIS} \cdot \frac{Stoth}{\sum_{IRIS} Stoth}$$

$pop_{Parcelle}$  = estimation de la population dans la parcelle

$pop_{IRIS}$  = population dans l'IRIS INSEE

$Stoth$  = Surface totale dédiée à l'habitation dans la parcelle

IRIS = « Pour toutes les parcelles qui intersectent l'IRIS » (requête géographique)

Les « parcelles » sont matérialisées par un point (le « localisant parcellaire ») dont les coordonnées sont fournies par le Cerema Nord Picardie dans sa livraison des fichiers fonciers.

## 2.3 Données livrées

Les semis de points de population sont mis à disposition des services sur le site ftp du Cerema méditerranée, au format shape (.shp).

Chaque point correspond au localisant parcellaire d'une parcelle.

Les données attributaires des semis de points sont composés des colonnes suivantes :

IDPAR	Numéro de parcelle : code INSEE+numéro de section+numéro de parcelle
CODE_INSEE	Code INSEE de la commune
SURFACE_HA	Surface d'habitation dans la parcelle, issue des fichiers fonciers 2016 (dits « MAJIC »), utilisée pour la répartition de la population sur la parcelle
POP_PARCEL	Population à la parcelle issue de la répartition de la population à l'IRIS 2014
NLOCHABIT	Nombre de locaux d'habitation dans la parcelle, issue des fichiers fonciers 2016

Tableau 1: Attributs des données livrées

## 2.4 Calcul de la population en zone inondée

C'est le champ POP\_PARCEL qui doit être utilisé pour calculer la population en zone inondable pour chaque scénario : HAB\_PERM dans la table N\_[prefixTri]\_ENJEU\_RAPPORT\_ddd.TAB .

Remarque : Le champ HAB\_PERM de la table N\_[prefixTri]\_ENJEU\_RAPPORT\_ddd est différent de celui de la table communale N\_[prefixTri]\_COMMUNE\_S\_ddd.TAB

Pour la table « ENJEU\_RAPPORT », il s'agit du « Nombre d'habitants permanents du TRI impactés par les inondations survenant dans le scénario étudié. »

Pour la table « COMMUNE », il s'agit de la population permanente totale dans la commune.

## 3 Variations saisonnières de la population communale

### 3.1 Introduction

Dans le cadre des travaux méthodologiques pour la mise en œuvre de la phase « cartographie » de la directive inondation, il a été mis en évidence lors du 1<sup>er</sup> cycle le besoin de disposer de données non seulement sur les populations permanentes, mais également sur les populations saisonnières, au travers de la variation de population pour une commune donnée.

L'évaluation de cette variation a été jugée indispensable pour les communes littorales. Il serait en effet difficilement compréhensible pour des territoires touristiques d'évaluer et de cartographier les risques sans tenir compte des populations présentes dans la pleine saison.

Le but de l'étude est de donner un **taux par commune, pour l'ensemble des communes françaises**.

Un territoire est susceptible d'accueillir :

- la population permanente,
- la population touristique saisonnière, qui séjourne sur ce territoire,
- des travailleurs saisonniers,
- des visiteurs à la journée.

Si on ne dispose généralement pas de chiffres sur les travailleurs saisonniers et sur les visiteurs à la journée, des données existent en revanche sur les populations touristiques.

En théorie, il conviendrait de confronter la saisonnalité de l'aléa (lorsqu'elle existe et c'est souvent le cas pour les inondations) aux variations saisonnières réellement constatées.

En pratique, seules des études très ponctuelles s'intéressent aux variations réellement constatées de population et il reste difficile de « saisonnaliser » l'aléa.

Pour ces raisons, il est ici choisi de s'intéresser à la capacité d'accueil des communes, plutôt qu'aux fréquentations. La notion a le mérite d'exister dans les méthodes du ministère du tourisme, et permet de s'exonérer d'une baisse conjoncturelle éventuelle de la fréquentation pour revenir vers des notions de bâtis, propres aux territoires étudiés.

### 3.2 Capacité d'accueil et populations touristiques

#### 3.2.1 Capacité d'accueil

La capacité d'accueil touristique d'un territoire est un indicateur qui peut être suivi tant au niveau quantitatif que qualitatif. Du point de vue quantitatif, cette capacité se définit toujours par un nombre de « lits » et représente le potentiel du territoire à accueillir dans le cadre d'une « nuitée » un certain nombre de touristes.

L'accueil des touristes peut être réalisé au travers de plusieurs types d'hébergements :

- les hébergements collectifs :
  - villages de vacances ;
  - maison familiale de vacances ;

- résidences de tourisme ;
- auberge de jeunesse ;
- centres internationaux de séjour ;
- centres sportifs ;
- l'hôtellerie ;
- l'hôtellerie de plein air (campings) ;
- les locations saisonnières :
  - meublés classés de tourisme ;
  - chambres d'hôtes ;
- les résidences secondaires.

### 3.2.2 Définitions

Des définitions sont proposées par le mémento du tourisme 2017.

Les **hôtels de tourisme** sont des établissements commerciaux d'hébergement, qui offrent des chambres ou des appartements meublés en location à une clientèle de passage ou à une clientèle qui effectue un séjour à la journée, à la semaine ou au mois mais qui, sauf exception, n'y élit pas domicile. On distingue les hôtels indépendants et les hôtels de chaînes. Les hôtels de chaînes sont des enseignes appartenant à des groupes hôteliers. Au 1er janvier 2017, l'hôtellerie de tourisme compte 18 172 hôtels en France métropolitaine (14 688 hôtels indépendants et 3 484 hôtels de chaînes) et 236 hôtels dans les départements d'outre-mer (225 hôtels indépendants et 11 hôtels de chaînes).

Les **résidences de tourisme** sont des établissements commerciaux d'hébergement, faisant l'objet d'une exploitation permanente ou saisonnière. Elles sont constituées d'un ou plusieurs bâtiments d'habitation individuels ou collectifs regroupant, en un ensemble homogène, des locaux d'habitation meublés ou des locaux à usage collectif. Les locaux d'habitation meublés sont proposés à une clientèle touristique qui n'y élit pas domicile, pour une occupation à la journée, à la semaine ou au mois. Elles sont dotées d'un minimum d'équipements et de services communs.

Les **campings** sont destinés à l'accueil de tentes, de caravanes, de résidences mobiles de loisirs et d'habitations légères de loisirs. Ils sont constitués d'emplacements nus ou équipés de l'une de ces installations ainsi que d'équipements communs. Les données de l'Insee intègrent désormais les établissements n'offrant que des emplacements loués à l'année. Est comptabilisé l'ensemble des emplacements offerts, qu'il s'agisse d'emplacements de passage ou d'emplacements résidentiels. Un emplacement de passage est un emplacement destiné à une clientèle touristique qui n'y élit pas domicile. Dans les emplacements de passage, on distingue les emplacements nus des emplacements locatifs, c'est-à-dire équipés d'un hébergement léger de type chalet, bungalow ou mobile home. Un emplacement résidentiel (ou loué à l'année) est un emplacement réservé à la location résidentielle, c'est-à-dire à un seul client pour l'ensemble de la période d'ouverture du camping.

Les **villages de vacances** sont des établissements touristiques constitués d'hébergements individuels ou collectifs proposant des séjours de vacances et de loisirs sous forme de forfait, comprenant la restauration ou des moyens individuels pour la préparation des repas, ainsi que l'usage des équipements collectifs de loisirs sportifs et culturels. La plupart des villages de vacances sont gérés par des associations de tourisme social (établissements affiliés à l'UNAT) ; les autres sont exploités par des sociétés commerciales (par exemple le Club Méditerranée).

Les **maisons familiales de vacances** sont des établissements sans but lucratif, à caractère social, familial et culturel qui ont pour principale vocation l'accueil des familles pendant leurs vacances et leurs loisirs. Elles sont, en priorité, ouvertes aux familles ayant des revenus

modestes.

Les **auberges de jeunesse** sont des établissements généralement régis par une association à but non lucratif. Elles offrent aux usagers un hébergement et un service de restauration limité et/ou une cuisine individuelle de même que d'autres prestations, programmes et activités. Ces derniers sont principalement destinés aux jeunes dans un objectif éducatif et récréatif.

Les **centres internationaux de séjour** (CIS) ont une double vocation : lieu d'hébergement et de restauration, c'est aussi un lieu culturel. Le concept du CIS repose sur la convivialité, la rencontre et les échanges internationaux.

Les **centres sportifs** sont des centres d'accueil et d'hébergement proposant des installations et des activités sportives. Ce sont des établissements du type UCPA ou « les Glénans ».

Les **meublés classés de tourisme** sont des villas, appartements, ou studios meublés, à l'usage exclusif du locataire, offerts en location à une clientèle de passage qui y effectue un séjour caractérisé par une location à la journée, à la semaine ou au mois, et qui n'y élit pas domicile. Ils doivent être déclarés en mairie, où leur liste est consultable. Les meublés classés de tourisme comprennent les meublés labellisés et les meublés non-labellisés.

Les **chambres d'hôtes** sont des chambres meublées situées chez l'habitant en vue d'accueillir des touristes, à titre onéreux, pour une ou plusieurs nuitées, assorties de prestations.

Les **résidences secondaires** sont des logements utilisés pour les week-ends, les loisirs ou les vacances. Les données présentées ici sont comptabilisées lors du recensement de la population ; elles comprennent une partie des villages de vacances, des résidences de tourisme et des meublés qui sont présentés par ailleurs.

### 3.2.3 Ratios

Le code du tourisme a défini dans son article R133-33, en vue de qualifier une commune de « touristique », des ratios pour estimer la capacité d'hébergement d'une population non permanente.

Cette capacité est estimée par le cumul suivant :

- nombre de chambres en hôtellerie classée et non classée multiplié par deux ;
- nombre de lits en résidence de tourisme répondant à des critères déterminés par décret ;
- nombre de logements meublés multiplié par quatre ;
- nombre d'emplacements situés en terrain de camping multiplié par trois ;
- nombre de lits en village de vacances et maisons familiales de vacances ;
- nombre de résidences secondaires multiplié par cinq ;
- nombre de chambres d'hôtes multiplié par deux ;
- nombre d'anneaux de plaisance dans les ports de plaisance multiplié par quatre.

Nous proposons de retenir ces ratios pour le calcul de la capacité d'accueil.

## 3.3 Données disponibles

### 3.3.1 Données INSEE publiques

Dans ses « chiffres détaillés », l'INSEE recense la « Capacité des communes en hébergement

touristique », par commune (géographie du 1<sup>er</sup> janvier 2017) en 2018.

L'hébergement touristique des communes de la France métropolitaine comprend ici :

- les hôtels et le nombre de chambres de l'hôtellerie non classée ou classée (de 0 à 5 étoiles) ;
- les campings non classés et classés (de 1 à 4 étoiles) et leurs emplacements ;
- le nombre de lits en villages vacances ;
- le nombre de lits en résidence de tourisme ;
- le nombre de lits en centres sportifs et auberges de jeunesse.

Par ailleurs, des données dans le thème « Logement » sont également disponibles, pour l'année 2014. Les logements recensés par l'INSEE sont répartis en 4 catégories :

- résidences principales ;
- les résidences secondaires et logements occasionnels ;
- logements vacants.

Un logement occasionnel est un logement ou une pièce indépendante utilisée occasionnellement pour des raisons professionnelles (par exemple, un pied-à-terre professionnel d'une personne qui ne rentre qu'en fin de semaine auprès de sa famille).

La distinction entre résidences secondaires et logements occasionnels est parfois difficile, et les deux catégories sont souvent regroupées (c'est le cas dans les données relatives à l'hébergement touristique citées ci-dessus). Les logements meublés loués (ou à louer) pour des séjours touristiques sont également classés en résidences secondaires.

**Ainsi les données publiques de l'INSEE du niveau communal occultent essentiellement les chambres d'hôtes. Au sein des résidences secondaires, il convient de noter que la proportion de logements dévolus à la location saisonnière est une inconnue.**

Les fichiers et attributs disponibles sont les suivants (téléchargeables sur

<https://www.insee.fr/fr/statistiques/2021703>

<https://www.insee.fr/fr/statistiques/3137421>

Les données utilisées sont listées dans le tableau suivant.

Variable	Description	Source
P14_RSECOCC	Nombre de résidences secondaires et logements occasionnels	RP2014
HTCH17	Nombre de chambres dans les hôtels en 2017	Insee, DGE et partenaires territoriaux
CPGEO17	Nombre total d'emplacements offerts clientèle de passage dans campings en 2017	Insee, DGE et partenaires territoriaux
VVLIT17	Nombre total de places lit dans les Villages vacances - Maisons familiales en 2017	Insee, DGE et partenaires territoriaux
RTLIT17	Nombre total de places lit dans les résidences de tourisme et hébergements assimilés en 2017	Insee, DGE et partenaires territoriaux
AJCSLIT17	Nombre total de places lit dans les auberges de jeunesse - Centres internationaux de séjour et centres sportifs en 2017	Insee, DGE et partenaires territoriaux

Tableau 2: Variables INSEE utilisées

### 3.4 Méthode proposée

Il est proposé, sur la base des fichiers INSEE et des ratios du code du tourisme, de procéder aux

calculs sur les communes des TRI métropolitains.

La formule retenue est donc :

$$\mathbf{HAB\_SAISON = 2 * HTCH17 + VVLIT17 + RTLIT17 + AJCSLIT17 + 5 * P14\_RSECOCC + 3 * CPGEO17}$$

Sur les TRI des DOM, les données sur les hébergements collectifs et les campings sont absentes : seules sont prises en compte les places en hôtels et les résidences secondaires.

L'analyse des variations saisonnières par les services seront différentes selon qu'ils s'intéressent à des territoires littoraux, ou de montagne.

**Dans les territoires de montagne, les chiffres importants correspondent parfois à une variation hivernale (stations de ski par exemple), en dehors des périodes à risque d'inondation. On pourra alors revenir aux différentes variables d'origine, permettant un interprétation nuancée.**

### 3.4.1 Structure des fichiers fournis

2 fichiers sont fournis : pop\_saiso.csv et pop\_saiso.ods.

Pour l'ensemble des communes françaises (hors Mayotte), sont fournis les variables suivantes :

- Numéro INSEE de la commune : **insee** (dans ce fichier Paris, Lyon et Marseille sont représentés par leurs arrondissements),
- Numéro INSEE de l'arrondissement : **insee2** (dans les communes à arrondissement),
- Nom de la commune : **nom**
- Population permanente : **p14\_pop**
- Ratio entre la population permanente de l'arrondissement et la population communale (communes à arrondissement) : **ratio**
- Nombre de résidences secondaires et de logements occasionnels : **p14\_rsecocc**
- Nombre total de chambres dans les hôtels classés et non classés : **htch17**
- Nombre total d'emplacements offerts à la clientèle de passage dans les campings classés : **cpgeo17**
- Nombre de lits en village de vacances et maisons familiales : **vvlit17**
- Nombre de lits en résidences hôtelières et résidences de tourisme : **rtlit17**
- Nombre de lits en auberges de jeunesse et centres internationaux de séjour : **ajcslit17**
- Surplus de population saisonnière théorique : **hab\_saiso**
- Taux de variation saisonnière théorique : **taux\_saiso**.

## 4 Estimation du nombre d'emplois dans les surfaces inondables

### 4.1 Objectif

La circulaire du 16 juillet 2012 relative à la mise en œuvre de la phase « cartographie » de la

directive inondation prévoyait que les cartographies des risques comportent une indication du nombre d'emplois impactés par commune et par scénario.

La présente étude rappelle la méthode permettant une évaluation de ce nombre d'emplois.

## 4.2 Les bases de données de l'INSEE

L'INSEE produit de nombreux rapports et de nombreuses données, homogènes au niveau national, concernant la question des emplois.

Seules les données suffisamment précises, à l'échelle communale ou infra-communale sont susceptibles de contribuer aux calculs d'emplois sur des zones infra-communales. On trouve celles-ci dans la rubrique « Bases de données » du site<sup>1</sup> de l'INSEE, il s'agit, concernant l'emploi :

- des données locales, à l'IRIS, à la commune, au département, à la région, à l'EPCI, à la zone d'emploi,
- des fichiers détail : enquêtes sur des fractions de population, la zone géographique étant a minima le département,
- du répertoire SIRENE recensant les entreprises et leurs établissements à l'adresse, avec une évaluation du nombre d'employés et une classification de l'activité (par code NAF).

Le répertoire SIRENE, qui comporte les adresses, est le plus adapté à une analyse géographique (infra-communale) des emplois : c'est la base de donnée qui a été retenue lors du 1er cycle. La nouveauté est qu'elle a été ouverte et qu'elle est disponible sur le site [data.gouv.fr](http://data.gouv.fr) et qu'elle a été géoréférencée.

### Base de données Sirene

« La base de données Sirene reprend, pour les entreprises et les établissements administrativement actifs, les informations contenues dans le répertoire Sirene en les restructurant et en les complétant. Elle rassemble ainsi des informations économiques et juridiques sur environ 10 millions d'établissements appartenant à tous les secteurs d'activité.

De nombreux organismes ont pour mission de déclarer à l'Insee les immatriculations, radiations et modifications au répertoire. Ils contribuent quotidiennement à l'actualisation du répertoire Sirene. En moyenne, 10 000 modifications par jour sont enregistrées au répertoire.

Sirene enregistre toutes les entreprises et leurs établissements, quelle que soit leur forme juridique et quel que soit leur secteur d'activité, situés en métropole ou dans les départements d'outre-mer (Guadeloupe, Guyane, Martinique, Mayotte et Réunion) et à Saint-Pierre et Miquelon. Les entreprises étrangères qui ont une représentation ou une activité en France y sont également répertoriées.

En outre, lors de la constitution de la base de diffusion, des données supplémentaires sont ajoutées pour faire de Sirene® base de données un outil opérationnel. »<sup>2</sup>

Les données figurant dans la base sont de trois ordres, pour les entreprises et leurs établissements :

- des données d'identification (numéro SIRET, nom, raison sociale, forme juridique, adresse) ;
- des données économiques (code d'activité principale APE, importance de l'effectif salarié par tranche, chiffre d'affaires) ;
- d'autres données (dates de création, de fermeture, ...).

Les adresses de cette base de données sont soumises à un traitement « RNVP » (Restructuration, Normalisation et Validation Postale) visant à les rendre conformes à la norme NF Z10-011, décrivant notamment comment doivent être stockées les données d'adresse dans une base de données.

Ceci les rend particulièrement propices à un géocodage à l'adresse : A l'aide d'un géocodeur et

1 <http://www.insee.fr/>

2 Extrait du site de l'INSEE : <http://www.sirene.fr/sirene/public/static/contenu-base-sirene>

d'une ou plusieurs bases de données adresses, ce procédé permet de rechercher les coordonnées géographiques (généralement la longitude et la latitude) correspondant à l'adresse la plus proche.

Les établissements présents dans les zones d'activités certaines zones rurales ne disposent cependant pas toujours d'une adresse très explicite (ex : « ZI n°1 », « CENTRE COMMER REG CRETEIL » ou « ROUTE DE PELVOUX »), et il peut s'avérer difficile de les localiser précisément à l'aide d'outils informatiques.

La notion d'entreprise renvoie à un ensemble d'établissements (le siège et des antennes par exemple) ; celle d'établissement est ici la plus adaptée.

La variable EFETCENT correspond à l'effectif salarié à la centaine près de l'établissement. C'est une variable statistique, millésimée au 31/12 d'une année donnée, majoritairement l'année n-2. Elle est à rapprocher de sa date de validité, c'est-à-dire de la date de mise à jour de l'effectif salarié de l'établissement.

EFETCENT	Correspondance
NN	Unités non employeuses (pas de salarié au cours de l'année de référence et pas d'effectif au 31/12). Cette tranche peut contenir quelques effectifs inconnus
0	0 salarié (n'ayant pas d'effectif au 31/12 mais ayant employé des salariés au cours de l'année de référence)
1	1 à 2 salariés
3	3 à 5 salariés
6	6 à 9 salariés
10	10 à 19 salariés
20	20 à 49 salariés
50	50 à 99 salariés
100	100 à 199 salariés
200	200 à 299 salariés
.../...	de 100 en 100 jusqu'à 999 999 salariés

Tableau 3: Valeurs possibles de TEFET (tranche d'effectif salarié de l'établissement)

Des biais existent :

- certains établissements importants comportent la valeur NN ;
- les gérants ne sont pas comptés (ex : autoentrepreneurs pour lesquels **EFETCENT** = 0) ;
- les fourchettes sont larges (jusqu'à 200 la borne supérieure vaut environ 2 fois la borne inférieure) ;
- les effectifs des établissements du secteur public sont souvent mal répartis : la gestion du personnel centralise souvent les effectifs de plusieurs établissements dans un établissement donné, qui n'est pas toujours l'établissement siège.

La variable **LIEUACT** permet de connaître le lieu de l'activité. Les codes 91 (sur chantier), 92 (sur marché) et 93 (en clientèle) correspondent à des activités où les personnes ne sont pas présentes la majeure partie du temps sur le site du siège. Leur travail peut cependant être impacté par une inondation, et il est proposé de ne pas utiliser cette variable pour sélectionner les emplois réellement présents sur le secteur (à noter la présence sur le secteur de professionnels issus d'autres sièges).

La variable **APET700** renvoie l'activité principale de l'établissement selon la nomenclature NAF (3 chiffres 2 lettres). Les codes commençant par « 84 » correspondent aux établissements du secteur public.

D'autres variables pourraient s'avérer intéressantes dans le cadre d'analyses sur la vulnérabilité

des territoires, mais n'ont pas été utilisées ici :

- **ACTISURF** : renvoie la surface du type de magasin (sans objet / 1 =moins de 300 m<sup>2</sup> / 2=de 300 à 400 m<sup>2</sup> / 3=de 400 à 2500 m<sup>2</sup> / 4=plus de 2500 m<sup>2</sup>) ;
- **ACTIVNAT** : renvoie la nature de l'activité de l'établissement selon une nomenclature simplifiée (sans objet / extraction / fabrication, production / montage, installation / réparation / transport / import, export / commerce de gros ou intermédiaire de commerce / commerces de détail / profession libérale / services / location de meublés / bâtiments, travaux publics / services aux entreprises / donneur d'ordre / autre / non renseigné).

## 4.3 Constitution du semis de points « emplois »

### 4.3.1 Extraction de la base SIRENE

Depuis janvier 2017, des extractions régulières de la base de données SIRENE sont mises à disposition du public, sous licence ouverte (« open data ») sur la plateforme ouverte des données publiques [data.gouv.fr](http://data.gouv.fr).

En particulier, le « fichier stock », mis à jour mensuellement, permet de récupérer les informations sur l'ensemble des établissements enregistrés dans la base SIRENE.

Le fichier utilisé pour la production du semis de points « emplois » est le fichier stock de mars 2018.

Les variables utilisées par la suite sont les suivantes :

- **SIREN** : numéro unique d'identification de l'entreprise (« numéro SIREN ») ;
- **NIC** : numéro interne de classement, qui permet de distinguer les établissements d'une même entreprise. Par association au SIREN, il forme le SIRET ;
- **L1\_NORMALISEE**: Raison sociale ou dénomination commerciale – Civilité pour les entreprises individuelles ;
- **L2\_NORMALISEE** : Complément d'adresse (destinataire, bâtiment...) ;
- **L3\_NORMALISEE** : Complément d'adresse (point géographique) ;
- **L4\_NORMALISEE** : Numéro et libellé de la voie ;
- **L5\_NORMALISEE** : Lieu-dit ou service particulier de distribution ;
- **L6\_NORMALISEE** : Code postal et localité ;
- **L7\_NORMALISEE** : Pays ;
- **COMET** : Code INSEE de la commune de localisation de l'établissement (3 derniers chiffres, sans le numéro de département) ;
- **DEPET** : Code INSEE du département (2 premiers caractères) ;
- **APET700** : Code d'activité principale exercée (APE) de l'établissement, codifié selon la nomenclature d'activités françaises (NAF rév. 2, 2008) ;
- **TEFET** : Code correspondant à la tranche d'effectif salarié de l'établissement ;

TEFET	Catégorie
NN	Unités non employeuses (pas de salarié au cours de l'année de référence et pas d'effectif au 31/12) ou unités sans mise à jour d'effectif
00	0 salarié (unités ayant eu des salariés au cours de l'année de référence mais plus d'effectif au 31/12)
01	1 ou 2 salariés
02	3 à 5 salariés
03	6 à 9 salariés
11	10 à 19 salariés
12	20 à 49 salariés
21	50 à 99 salariés
22	100 à 199 salariés
31	200 à 249 salariés
32	250 à 499 salariés
41	500 à 999 salariés
42	1 000 à 1 999 salariés
51	2 000 à 4 999 salariés
52	5 000 à 9 999 salariés
53	10 000 salariés et plus

Tableau 4: Valeurs possibles de EFETCENT (fourchettes d'effectifs de l'établissement)

- **EFETCENT** : Effectif salarié de l'établissement à la centaine près (cf Tableau 3).

### 4.3.2 Le géocodage de SIRENE Base de données

Ce fichier stock a été géocodé à l'aide de plusieurs bases de données d'adresse (Base adresse nationale - « BAN », base adresse nationale ouverte - « BANO », points d'intérêt d'openstreetmap). Les scripts utilisés sont disponibles sur <https://github.com/cquest/geocodage-sirene>.

Il est possible de télécharger (voir [ici](#)) un fichier géocodé par département et pour chaque arrondissement parisien. Sur le site [data.gouv.fr](http://data.gouv.fr), ces données peuvent être retrouvées dans la rubrique « données communautaires » de la page dédiée aux données SIRENE.

Les fichiers utilisés pour la préparation des données décrites ci-dessous datent du 13 mars 2018.

Les champs ajoutés par ce procédé automatique sont :

- **longitude** (en degrés décimaux, WGS84) ;
- **latitude** (en degrés décimaux, WGS84) ;
- **geo\_score** : indice de similarité fourni par le moteur de géocodage ;
- **geo\_type** : type de géocodage ;

geo_type	Catégorie
houenumber	Numéro trouvé
interpolation	Numéro interpolé
street	Voie trouvée
locality	Lieut-dit – ou position de la mairie pour les adresses indiquées « MAIRIE » ou « HOTEL DE VILLE »
municipality	Position de la commune, lorsque l'adresse n'a pas été trouvée
poi.[xxx]	Point d'intérêt, [xxx] précisant la nature du point d'intérêt, généralement en anglais (ex : « hotel » pour un hôtel, « mall » pour un centre commercial, ...)

Tableau 5: Valeurs possibles de geo\_type (type de géocodage)

- **geo\_adresse** : libellé de l'adresse trouvée ;
- **geo\_id** : Identifiant dans le référentiel BAN ou BANO (si commence par "BANO\_"), ou lien openstreetmaps (si commence par « http://osm.org/ ») ;
- **geo\_ligne** : Ligne d'adresse géocodée ;

geo_ligne	Catégorie
G	géographique
N	normalisée
D	déclarée

Tableau 6: Valeurs possibles de geo\_ligne (ligne d'adresse géocodée)

#### Limites :

Comme à l'issue de tout procédé de géocodage automatisé, un certain nombre de points ne sont pas positionnés au bon endroit. Cela peut venir d'une adresse de départ imprécise (ex : localisation à la commune car l'adresse n'a pas pu être retrouvée), mais aussi du procédé de géocodage en lui-même (ex : mauvais rapprochement d'adresse). Il est possible de signaler les problèmes liés au géocodage sur <https://github.com/cquest/geocodage-sirene/issues>

### 4.3.3 Exclusions préalables

Les collectivités territoriales représentent des effectifs importants mal répartis quelle que soit la méthode utilisée. Elles ont donc été exclues au préalable en supprimant les lignes qui comportaient à la fois une activité **APET700** valant 8411Z (« Administration publique générale ») et dont l'attribut **L1\_NORMALISEE** contenait les mots « COMM », « AGGLO », « DEPARTEMENT », « REGION » (avec un espace), « COM » (idem), « METROPOLE », « COLLECTIVITE », sans contenir les mots « PREFECTURE », « DIR », « LABO », « COMMUNI ».

Requête SQL<sup>1</sup> permettant de sélectionner les lignes utilisées :

```
SELECT * FROM [nom de la table départementale]
WHERE apet700 <> '8411Z'
OR (
  I1_normalisee not like '%COMM%'
  AND I1_normalisee not like '%AGGLO%'
```

1 Sous postGIS, sensible à la casse, il est plus sûr de remplacer « I1\_normalisee » par « UPPER(I1\_normalisee) »

```

AND I1_normalisee not like '% DEPARTEMENT%'
AND I1_normalisee not like '%REGION %'
AND I1_normalisee not like '%COM %'
AND I1_normalisee not like '%METROPOLE%'
AND I1_normalisee not like '%COLLECTIVITE%'
)
OR (
I1_normalisee like '%PREFECTURE%'
OR I1_normalisee like '%DIR%'
OR I1_normalisee like '%LABO%'
OR I1_normalisee like '%COMMUNI%'
);

```

Exemples d'exclusions : « COMMUNE DE NEUF BERQUIN », « DEPARTEMENT DE PARIS », « COMMUNAUTE COMMUNES DE MEUSE ET SEMOY », « METROPOLE NICE COTE D AZUR », « ALES AGGLOMERATION »

Exemples d'administration publique générale non exclues : « PREFECTURE DES ARDENNES », « CHAMBRE REGIONALE DES COMPTES D ALSACE », « COMITE REG INFORMATION COMMUNICATION », « SYND ETUDE MASSIF NOVES CHATEAURENARD », « DIR REGIONALE DES DOUANES DE PROVENCE »

#### 4.3.4 Calcul de la fourchette communale et des ratios correctifs

Nombre de petites entreprises ne comprennent qu'une personne, non comprises dans le chiffre donné par l'INSEE. Le gérant d'une société n'est pas non plus compté. Pour rétablir un certain équilibre, les fourchettes suivantes ont été choisies :

EFETCENT	Minimum	Maximum
NN	1	1
0	1	1
1	2	3
3	4	6
6	7	10
10	11	20
20	21	50
50	51	100
100	101	200
200	201	300
N*100	N*100 +1	(N+1)*100

Tableau 7: Valeurs retenues pour le calcul des effectifs

Tous les établissements de la commune ont été sélectionnés sur la base du code INSEE de la commune, obtenu par concaténation de **dep**et (département) et de **com**et (commune)

Pour chaque commune, les minima et maxima de la tranche d'effectif calculés avec les valeurs du Tableau 7 ont été sommés pour tous les établissements : les valeurs obtenues sont **min\_com** et **max\_com**.

Les minima et maxima des établissements bien géocodés (c'est-à-dire disposant d'un géocodage à l'adresse<sup>1</sup>) calculés avec le même tableau ont également été sommés : les valeurs obtenues sont **min\_geocod** et **max\_geocod**.

Les ratios correctifs communaux ont ainsi pu être initialement calculés par :

$$r\_mincom = min\_com / min\_geocod$$

$$r\_maxcom = max\_com / max\_geocod$$

**Attention** : ces ratios sont ceux livrés avec les données de base ; ils doivent être revus dans le cas où des établissements seraient ajoutés / retirés après vérification par les services (après calcul des nouveaux **min\_geocod** et **max\_geocod**)

### 4.3.5 Calcul des fourchettes pour les établissements géocodés

Pour chaque établissement bien géocodé (c'est-à-dire disposant d'un géocodage à l'adresse), la fourchette **emploi\_min** et **emploi\_max** a été calculée avec les valeurs du Tableau 7, puis corrigée en les multipliant par les ratios correctifs **r\_mincom** et **r\_maxcom**.

**Attention** : dans le cas où des établissements seraient ajoutés / retirés après vérification par les services, chaque **nb\_emploi\_min** et **nb\_emploi\_max** doit être recalculé par application des nouveaux ratios sur **nb\_emp\_min\_nc** et **nb\_emp\_max\_nc**, valeurs des fourchettes non corrigées fournies dans le fichier des communes

La variable **TEFET** classe les entreprises en fonction de leur taille :

Les entreprises peuvent être classées en 4 catégories en fonction du nombre d'employés et de leur chiffre d'affaires (**Décret n° 2008-1354 du 18 décembre 2008**), pour les analyses statistiques.

Un attribut **taille** sur la taille de l'entreprise a été créé, valant de 1 à 4 :

- **taille** : Code calculé par rapport à la tranche d'effectif salarié de l'établissement

taille	Correspondance
1	Pas d'effectif salarié : auto-entrepreneurs, professions libérales, ...
2	0 < Effectif salarié < 250 : Petites et moyennes entreprises <sup>2</sup>
3	250 ≤ Effectif salarié < 5000 : Entreprise de taille intermédiaire <sup>3</sup>
4	5000 ≤ Effectif salarié : Grandes entreprises <sup>4</sup>

Tableau 8: Valeurs possibles de « taille »

Cet attribut a vocation à permettre aux services de cibler rapidement les plus grands établissements afin de vérifier leur géocodage. En effet, la présence d'établissements de Taille 3 et 4 (voire de taille 2) est de nature à influencer fortement les résultats du nombre calculé sur les différentes surfaces inondables (à noter : on applique ici les fourchettes de taille d'entreprise aux établissements ; or un petit établissement peut faire partie d'une grande entreprise).

Dans le cas où cette vérification est opérée, il est nécessaire de recalculer les ratios correctifs communaux, qui s'appuient sur la somme des minima et maxima des établissements bien géocodés.

1 geo\_score > 0 AND geo\_type NOT LIKE 'municipality'

2 Sans considération du chiffre d'affaires

3 idem

4 Idem

### 4.3.6 Fichiers obtenus

Un fichier zip est fourni par département.

Contenu :

Les différents composants d'un fichier shape, contenant un semis de point calculé suivant la méthode décrite précédemment, sur tout le département.

Attributs :

- **siren** : numéro unique d'identification de l'entreprise (« numéro SIREN ») ;
- **nic** : numéro interne de classement, qui permet de distinguer les établissements d'une même entreprise. Par association au SIREN, il forme le SIRET ;
- **I1\_normali** : Raison sociale ou dénomination commerciale – Civilité pour les entreprises individuelles ;
- **I2\_normali** : Complément d'adresse (destinataire, bâtiment...) ;
- **I3\_normali** : Complément d'adresse (point géographique) ;
- **I4\_normali** : Numéro et libellé de la voie ;
- **I5\_normali** : Lieu-dit ou service particulier de distribution ;
- **I6\_normali** : Code postal et localité ;
- **I7\_normali** : Pays ;
- **comet** : Code INSEE de la commune de localisation de l'établissement (3 derniers chiffres, sans le numéro de département) ;
- **code\_insee** : Code INSEE complet de la commune (5 caractères) ;
- **apet700** : Code d'activité principale exercée (APE) de l'établissement, codifié selon la nomenclature d'activités françaises (NAF rév. 2, 2008) ;
- **tefet** : Code correspondant à la tranche d'effectif salarié de l'établissement ;
- **taille** : Code calculé par rapport à la tranche d'effectif salarié de l'établissement ;
- **emp\_min\_nc** : Seuil inférieur du nombre d'employés de l'établissement, non corrigé ;
- **emp\_max\_nc** : Seuil supérieur du nombre d'employés de l'établissement, non corrigé ;
- **emploi\_min** : Seuil inférieur du nombre d'employés de l'établissement, ajusté après répartition du seuil inférieur de nombre d'employés mal géolocalisés dans la commune ;
- **emploi\_max** : Seuil supérieur du nombre d'employés de l'établissement ajusté après répartition du seuil supérieur de nombre d'employés mal géolocalisés dans la commune ;
- **longitude** (en degrés décimaux, WGS84) ;
- **latitude** (en degrés décimaux, WGS84) ;
- **geo\_score** : indice de similarité fourni par le moteur de géocodage ;
- **geo\_type** : type de géocodage ;
- **geo\_adresse** : libellé de l'adresse trouvée ;
- **geo\_id** : Identifiant dans le référentiel BAN ou BANO (si commence par "BANO\_"), ou lien openstreetmap (si commence par « <http://osm.org/> ») ;
- **geo\_ligne** : Ligne d'adresse géocodée.

### 4.3.7 Sources d'erreur et biais

Les erreurs possibles sont :

#### Erreurs liées au fichier SIRENE :

- informations non renseignées : **EFETCENT** = NN ; cela concerne 301 091 établissements sur 4 429 279, dans des secteurs d'activité variés : on ne peut isoler de typologie majoritaire mais cela peut conduire à sous-estimer notablement l'effectif ;
- effectif de l'établissement reparti sur une seule adresse, alors qu'en réalité il l'est sur plusieurs (exemple : Aix-en-Provence, 2900 personnes situées place de l'Hôtel de Ville !) ; cela concerne principalement les collectivités territoriales. Dans la méthode proposée, ces effectifs ont été exclus.

#### Erreurs liées au géocodage :

- géocodage non pas à l'adresse mais à la rue, à la commune ou même absence de géocodage : pour y pallier en partie, la méthode proposée répartit les effectifs non localisés sur les établissements bien localisés ;
- mauvaise localisation de l'établissement, que le géocodage situe généralement au niveau du portail d'entrée, pouvant être éloigné des bâtiments
- erreur de rapprochement entre l'adresse du fichier SIRENE et une adresse des bases de données utilisées par le géocodeur. Hors des zones très peu denses (ex : dans la commune du Dévoluy, rapprochement entre une adresse contenant « résidence Aurouze » et un lieu dit « aurouze agnieres », non construit), seule une expertise de terrain permettrait de détecter et corriger ce genre d'erreurs.

#### Concernant la méthode :

Des établissements dont la surface est parfois importante (industries automobiles, chimiques ...) sont représentés par un point (essentiellement des entreprises de taille intermédiaire et des grandes entreprises) appartenant ou pas à une surface inondable alors qu'une partie des installations seulement peut être concernée.

Il est choisi de sommer les effectifs quel que soit le type d'établissement ; or certains emploient des personnes travaillant sur d'autres sites (ex : entreprises de nettoyage, sociétés de services en ingénierie informatique...).

Étant donné les fourchettes individuelles de la variable **EFETCENT** (ou la borne supérieure vaut souvent deux fois la borne inférieure), les fourchettes totales restent imprécises, mais permettent d'approcher l'ordre de grandeur.

## 4.4 Évaluation du nombre d'emplois dans les surfaces inondables

### 4.4.1 Croisement avec les surfaces inondables

Les semis de points complétés et vérifiés devront être croisés avec les différents scénarios d'inondation, par commune.

La somme des minima et la somme des maxima sera été réalisée pour chaque scénario (somme des valeurs **emploi\_min** et **emploi\_max** des points situés dans une surface inondable et une commune donnée), et permettra de remplir les champs **EMPLOI\_MIN** et **EMPLOI\_MAX** de la table N\_[prefix\_Tri]\_ENJEU\_RAPPORT\_ddd, définie dans le géostandard COVADIS de la Directive Inondation.

#### **4.4.2 Représentation**

Les différents chiffres obtenus figureront dans le rapport d'accompagnement.

Les moyennes arithmétiques des minima et maxima pour chaque scénario figureront dans les cartouches type, au même titre que les chiffres de population permanente.

Cependant, pour tenir compte de l'imprécision de la méthode, le chiffre ne sera pas indiqué sous un seuil minimal de 50 emplois, : on écrira alors « - de 50 ».



[Cerema Méditerranée](#) - Pôle d'activité – 30 rue Albert Einstein – CS 70499 – 13593 Aix-en-Provence CEDEX 3  
Tel : 04 42 24 76 76 – mel : [DTerMed@cerema.fr](mailto:DTerMed@cerema.fr)

Siège social : Cité des Mobilités - 25, avenue François Mitterrand - CS 92 803 - F-69674 Bron Cedex - Tél : +33 (0)4 72 14 30 30

Établissement public – Siret : 130 018 310 00313 - TVA Intracommunautaire : FR 94 130018310 - [www.cerema.fr](http://www.cerema.fr)