

Document technique

**Méthodologie d'estimation des enjeux
exposés aux inondations**

L'Ifen met à disposition des documents techniques en dehors de ses collections.



Méthodologie d'estimation des enjeux exposés aux inondations

Document de travail
Philippe Dorelon, Bernard Poupat

Sommaire

1	PRESENTATION GENERALE DE LA DEMARCHE	3
2	DONNEES SOURCES.....	3
2.1	CONTOURS IRIS	3
2.2	STATISTIQUES DU RECENSEMENT DE POPULATION 1999	3
2.3	CORINE LAND COVER (CLC) 2000.....	3
2.4	BDCARTO	4
2.5	CONTOURS DES ZONES INONDABLES.....	4
3	TRAITEMENT DES DONNEES SIG.....	5
3.1	DONNEES POUR CONSTRUIRE LE MODELE	5
3.2	CONTOURS DES ZONES INONDABLES.....	5
3.3	LIMITES DE PRECISION DES DONNEES CARTOGRAPHIQUES	6
4	ESTIMATIONS PAR UN MODELE LINEAIRE	6
4.1	MODELE	6
4.2	APPLICATION DU MODELE	7
4.3	COMPARAISON DES MODELES	7
4.4	TABLEAU DE SYNTHESE, ENSEMBLE DES IRIS, TOUS MODELES.....	9
4.5	ESTIMATIONS EN ZONES INONDABLES, COMPARAISON AVEC LES DONNEES DISPONIBLES	9
4.5.1	<i>Languedoc-Roussillon.....</i>	<i>9</i>
4.5.2	<i>Val d'Authion, tous modèles.....</i>	<i>11</i>
4.5.3	<i>Centre (communes > 10000 habitants), tous modèles</i>	<i>11</i>
4.6	BILAN DES MODELISATIONS.....	11

1 Présentation générale de la démarche

*La Direction de la prévention des pollutions et des risques et la Direction de l'eau ont confié à l'Institut français de l'environnement (IFEN) le soin de développer **un outil d'évaluation des enjeux présents dans les zones exposées aux risques** d'inondation, en commençant par l'estimation de la population et du nombre de logements. Cet outil doit être simple d'application pour permettre une mise à jour régulière et utilisable sur tout le territoire français, pour une cohérence entre les données.*

Ce document résume les travaux méthodologiques qui ont été élaborés par l'Ifen pour répondre à ce besoin de systématiser une démarche d'observation des enjeux au niveau national. Différentes méthodes ont déjà été appliquées dans les DIREN mais elles sont soit lourdes et coûteuses (estimations à partir de cartes et de bordereaux du recensement), soit trop imprécises (estimations par proratas surfaciques).

- La méthode proposée est automatisable et applicable aux contours des zones inondables de tout le territoire métropolitain. Elle fournit des résultats sur la population et les logements issus du recensement de 1999.
- Les modèles d'estimations s'appuient sur des régressions linéaires.
- Les données fournies par les DIREN ainsi que le test réalisé par la direction régionale de l'Insee du Centre ont permis de comparer les résultats obtenus.

2 Données sources

Les données utilisées sont nationales issues du MEDAD, de l'IGN et de l'INSEE :

- Corine Land Cover (IFEN/AEE) 2000
- BDCarto (IGN) 2000
- Contours Iris (IGN / INSEE) 1999
- Base de Données Communales du recensement 1999 (INSEE)

2.1 Contours Iris

Le zonage administratif utilisé est le contour des Iris de la Métropole, référence 1999. Les iris sont un zonage de l'INSEE, constitués de sommes d'îlots et regroupant environ 2000 personnes. Il y a 15375 iris provenant de 1852 communes. Pour les 34755 communes sans Iris, on considère que le contour de la commune est un iris. Au total, le fond est donc constitué de 50130 objets.

2.2 Statistiques du recensement de population 1999

La base de données communale (BDCOM) est issue des résultats du recensement de la population effectué par l'INSEE en 1999. Elle fournit les statistiques de bases (population, logement) pour chaque iris.

2.3 Corine Land Cover (CLC) 2000

Corine Land Cover est une base de données géographique environnementale européenne. Elle constitue un inventaire de l'occupation des terres (pour plus d'information voir sur le site web de l'Ifen <http://www.ifen.fr/donIndic/Donnees/corine/presentation.htm>).

Parmi les 44 catégories de la base, seuls les tissus urbains continus et discontinus sont conservés, le reste du territoire est regroupé dans une catégorie « autres ».

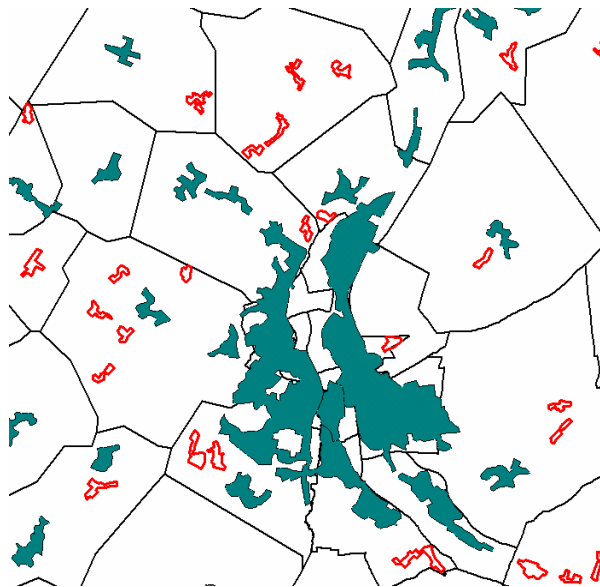
2.4 BDCarto

La superficie minimale des polygones dans CLC est de 25 ha. C'est pourquoi de très nombreuses petites communes n'ont aucun tissu urbain identifié dans CLC. Le thème occupation du sol de la BDCarto de l'IGN fournit une information plus fine (seuil de 8 ha) pour certaines zones urbaines, en revanche, il ne distingue pas les tissus urbains continus et discontinus qui serviront lors de la modélisation, c'est pourquoi les deux produits sont utilisés conjointement.

L'identification par la BDCarto de surfaces urbaines, notamment dans la partie inondable des communes est donc un gain de précision supplémentaire : la BDCarto est utilisée en complémentarité de CLC, elle impacte essentiellement les petites communes.

A partir du fichier CLC a été élaboré un fichier CLC corrigé avec la BDCarto. Il comprend toutes les communes de France métropolitaine et répartit celles-ci selon les surfaces urbaines et non urbaines, le tissu urbain supplémentaire, issu de la BDCarto étant considéré comme un tissu urbain discontinu de CLC, codifié S112.

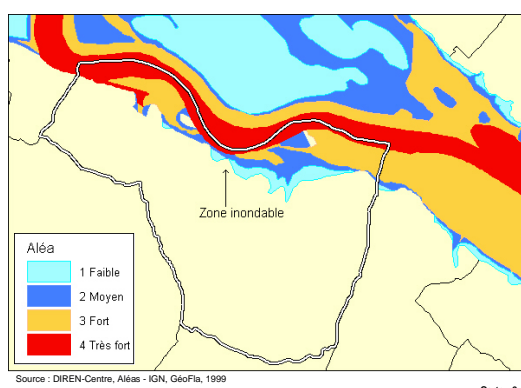
Exemple de superposition des trois fonds (iris+clc zones urbaines +bdcarto)



2.5 Contours des zones inondables

La constitution de contours de zones inondables au format SIG est en cours d'achèvement. Elle fait intervenir différents acteurs (DIREN, directions de l'équipement, agences de l'eau, CETE, MEDAD,...). L'état d'avancement varie en fonction des régions.

Exemple de cartographie d'une zone inondable de la Loire



Une cartographie des zones inondables encore perfectible

Quelques problèmes ont été mis en évidence lors des tests :

- Différents formats de fichiers (shape, mapinfo, arcinfo,...)
- Différents concepts (aléas, plus hautes eaux connues, crues centennales,...)
- Cartorisque :

La base de données Cartorisque devrait fournir dans quelques mois des contours de zones inondables pour tous les départements de métropole. L'objectif est de mettre en place une chaîne de production utilisant ces données normalisées même si quelques défauts ont été mis en évidence dans les fichiers déjà disponibles :

- utilisation de libellés au lieu d'identifiants stables (exemple : nom du département au lieu du code)
- problèmes éventuels de topologie
- données disponibles variables selon le département

3 Traitement des données SIG

3.1 Données pour construire le modèle

Le traitement consiste à « croiser » géométriquement les trois bases CLC, BDCarto et Contours IRIS afin de déterminer pour chaque iris les types de surfaces CLC associées.

Une table ayant la forme suivante est ainsi obtenue :

IRIS	Superficie tissu urbain continu (S_111)	Superficie tissu urbain discontinu (S_112)	Superficie autres (S_autres)
452340101	1.4	5.6	10.42
450010000	0	2.5	8.1
...			

3.2 Contours des zones inondables

Le traitement à réaliser ressemble au précédent, il faut ajouter le contour des zones inondables et ainsi déterminer pour chaque iris les superficies respectives en zones inondables. Lorsque plusieurs contours sont proposés, les calculs ont été effectués sur l'enveloppe de ceux-ci.

3.3 Limites de précision des données cartographiques

La méthode fait intervenir différentes données cartographiques avec des précisions différentes. Le contour des zones inondables est précis à 10 ou 20 mètres, Corine Land Cover à 20 mètres, améliorée en partie par la BD Carto. Le contour des Iris a quant à lui une précision variable de quelques mètres à cent mètres. L'analyse des résultats obtenus devra tenir compte de toutes ces imprécisions auxquelles s'ajoutent des problèmes de calage des différentes cartographies.

4 Estimations par un modèle linéaire

4.1 Modèle

Un modèle de référence a été conçu puis décliné en sous-modèles adaptés aux spécificités locales. L'enjeu (population, nombre de logements,...) est supposé suivre ce modèle pour chaque iris :

$$X_i = d_{S_{111}} \times S_{111_i} + d_{S_{112}} \times S_{112_i} + d_{S_{autres}} \times S_{autres_i} + \varepsilon_i$$

Avec

X_i : enjeu de l'iris i

d_{S_*} : densité associée au poste

S_{*_i} : superficie du poste pour l'iris i

ε_i : erreur pour l'iris i

Et avec les hypothèses :

les erreurs sont centrées
de même variance (homoscédasticité)
et non corrélées entre elles.

Ce modèle correspond à un modèle linéaire sans constante. Il s'agit d'estimer les densités associées à chaque type de surface connaissant la valeur X pour chaque iris. Seules les catégories tissu urbain continu (S_{111}) et discontinu (S_{112}) ont été conservées, les autres surfaces ont été regroupées dans une catégorie « autres ». Des tests ont été effectués en conservant également les zones industrielles ou commerciales mais le gain obtenu était faible, beaucoup de densités associées n'étaient pas significativement différentes de zéro ou parfois négatives.

$$\hat{X}_i = \hat{d}_{S_{111}} \times S_{111_i} + \hat{d}_{S_{112}} \times S_{112_i} + \hat{d}_{S_{autres}} \times S_{autres_i}$$

Avec

\hat{X}_i : estimation pour l'iris i

\hat{d}_{S_*} : densité estimée associée au poste S_*

S_{*_i} : superficie du poste pour l'iris i

Afin d'améliorer l'estimation, le modèle est découpé en plusieurs sous-modèles permettant de regrouper les communes ayant des densités voisines. Plusieurs découpages ont été testés. Ils sont présentés dans ce document.

Estimation des densités : une régression linéaire par la méthode des moindres carrés est effectuée sur chaque groupe d'Iris (Tous les iris de métropole pour le modèle global), afin de déterminer les densités

associées aux catégories de surfaces, avec la contrainte d'obtenir des densités positives (parfois nécessaire pour la catégorie « autres »).

Les résultats obtenus permettent de calculer une estimation de X pour chaque iris (\hat{X}_i). Un coefficient correctif (Coeff) est calculé afin de caler l'estimation des densités sur la valeur observée X:

$$\hat{X}_i = \hat{d}_{S_{111}} \times S_{111_i} + \hat{d}_{S_{112}} \times S_{112_i} + \hat{d}_{S_{autres}} \times S_{autres_i}$$

$$Coeff_i = \frac{X_i}{\hat{X}_i}$$

4.2 Application du modèle

A partir d'un zonage quelconque défini par un contour, un croisement entre Corine Land Cover, Bd Carto et les contours Iris déterminent, l'occupation du sol par catégorie pour chaque iris ayant une intersection avec le zonage.

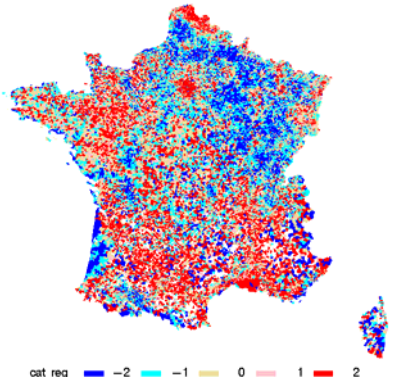
L'estimation dans chaque iris est alors égale à :

$$X_{esti} = Coeff_i \times (\hat{d}_{S_{111}} \times S_{111_{zonage_i}} + \hat{d}_{S_{112}} \times S_{112_{zonage_i}} + \hat{d}_{S_{autres}} \times S_{autres_{zonage_i}})$$

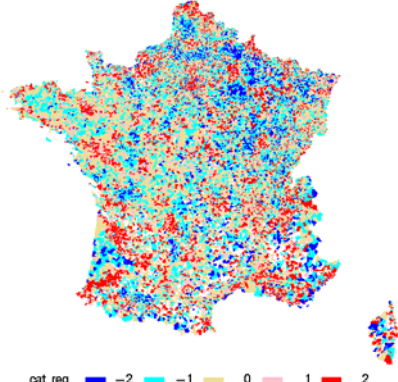
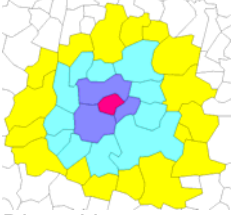
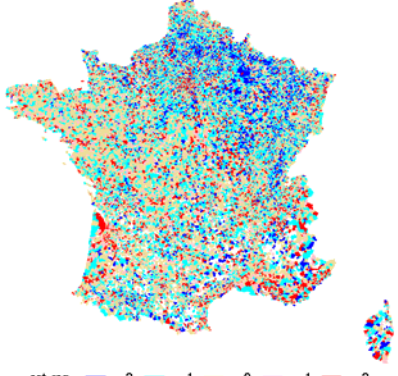
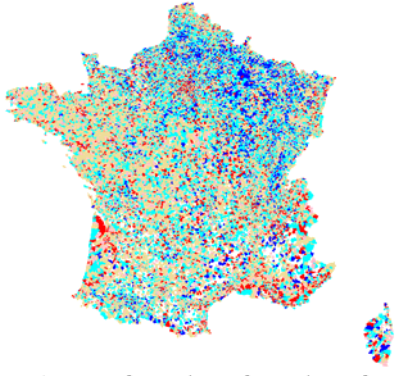
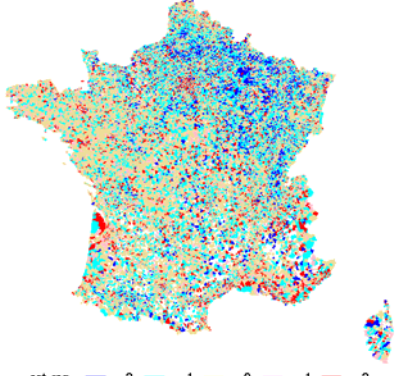
On suppose ici implicitement que le modèle déterminé pour des iris complets s'applique sur des fractions d'iris, ce qui constitue une hypothèse forte et doit conduire à analyser avec prudence les estimations obtenues au niveau de l'iris ou de la commune.

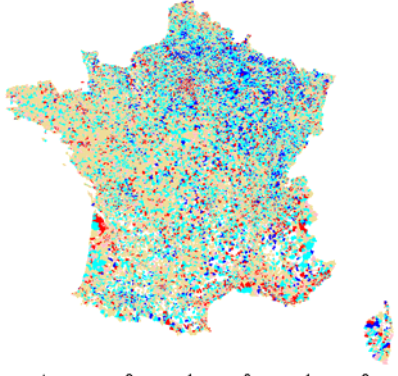
4.3 Comparaison des modèles

La déclinaison du modèle de base en sous-modèles dont les spécifications sont différentes a conduit à réaliser un ensemble de tests afin de retenir le plus efficient.

Modèle	Description	Résidus ¹
Global	<p>Une seule régression sur les 39 000 iris de Métropole ayant une surface urbaine dans CLCC.</p> <p>Problèmes d'auto-corrélations spatiales des résidus :</p> <p>Surestimation de la population dans le nord-est</p> <p>Sous-estimation dans le sud</p> <p>Les résidus ne sont pas répartis aléatoirement, donc en toute rigueur, le modèle n'est pas applicable.</p>	

¹ Résidus obtenus pour l'estimation de la population découpés en quatre classes, en bleu, population surestimée, en rouge population sous-estimée. Les cartes des résidus pour l'estimation du nombre de logements sont similaires

Détaillé	<p>Découpage par catégorie de communes (taille, département, littoral,...) 262 régressions à effectuer Encore des problèmes d'auto-corrélations spatiales.</p>	
Local	<p>Une régression par iris avec les voisins des trois couronnes, pas de pondération</p>  <p>Disparition presque complète des auto-corrélations. L'absence de pondération donne mécaniquement plus de poids aux iris éloignés.</p>	
Local pondéré 1	<p>Une régression par iris avec les voisins des trois couronnes, pondération égale pour chaque couronne (l'iris centre faisant partie de la première couronne)</p>	
Local pondéré 2	<p>Une régression par iris avec les voisins des trois couronnes, pondération 1,1/2 et 1/3 pour les couronnes</p>	

Local pondéré 3	Une régression par iris avec les voisins des trois couronnes, pondération 1,1/4 et 1/9 pour les couronnes	
-----------------	---	--

4.4 Tableau de synthèse, ensemble des Iris, tous modèles

Modèle	Erreur moyenne par Iris ² (%)	
	Population	Logement
Global	79%	87%
Détaillé	61%	70%
Local	57%	64%
Local pondéré 1	54%	59%
Local pondéré 2	51%	56%
Local pondéré 3	48%	53%

L'erreur moyenne par Iris, calculée avec les résidus, se réduit nettement entre le modèle simple au niveau France entière et le modèle local pondéré 3.

A noter : quel que soit le modèle, l'erreur moyenne est systématiquement plus faible pour l'estimation de la variable population.

4.5 Estimations en zones inondables, comparaison avec les données disponibles

Des estimations précises existent pour certaines zones inondables :

- Languedoc-Roussillon : avec bordereaux du recensement 1999 (source DIREN)
- Val d'Authion : avec bordereaux du recensement 1999 (source DIREN Pays de Loire)
- Loire moyenne, région Centre, communes de plus de 10000 habitants : avec un appariement RIL (Répertoire d'Immeubles Localisé), recensement 1999 (Test INSEE Centre / IFEN).

Ces estimations permettent de tester les différents modèles et de donner une indication sur le nombre d'iris à regrouper pour rendre le résultat acceptable, au vu de l'erreur commise.

4.5.1 Languedoc-Roussillon

4.5.1.1 Niveau régional, tous modèles

Modèle	Estimation population enZI ³	Erreur totale ⁴	Erreur moyenne par
--------	---	----------------------------	--------------------

$$^2 \text{ Erreur moyenne} = \frac{\sqrt{\sum_{i=1}^n (\hat{X}_i - X_i)^2}}{n}$$

, calculée à partir des estimations pour chaque iris i, le critère

du R² n'a pas été utilisé car il perd du sens dans un modèle linéaire sans constante.

³ Estimation calculée à partir du modèle de régression

			commune ⁵
Global	602 841	4%	180%
Détaillé	578 813	0%	61%
Local	582 277	1%	61%
Local pondéré 1	581 795	1%	61%
Local pondéré 2	580 820	0%	61%
Local pondéré 3	579 637	0%	61%

A noter : l'application du modèle a été effectuée sur les communes ayant fait l'objet d'une évaluation DIREN par les bordereaux M6 du recensement, sur un ensemble un peu réduit par rapport à celui de l'étude présentée dans ce dossier. Les valeurs estimées par le modèle apparaissent ainsi légèrement inférieures aux résultats précédents.

4.5.1.2 Niveau départemental, modèle local pondéré 3

Département	Population En ZI		Erreur globale	Erreur moyenne par commune
	DIREN	Estimation		
11	62 096	62 426	1%	46%
30	232 475	235 194	1%	42%
34	127 529	129 130	1%	115%
48	10 013	10 416	4%	88%
66	146 711	142 471	-3%	25%

Les écarts restent faibles mais l'écart moyen varie de 25% à 115%. Le mauvais chiffre pour le département de l'Hérault s'explique en partie par une mauvaise modélisation pour les communes touristiques. A l'intérieur de la région, les résultats varient beaucoup d'un département à l'autre, quel que soit le modèle.

4.5.1.3 Cas particulier des communes de plus de 10000 habitants, modèle local pondéré 3

L'erreur moyenne tombe à 21% et provient de quelques communes où l'estimation est mauvaise, en particulier pour Montpellier. Sans cette commune, l'erreur devient inférieure à 15%.

Commune	Population En ZI		Erreur par commune
	DIREN	Estimation	
11069	10 230	9 905	-3%
11076	250	330	32%
11262	16 700	16 874	1%
30007	18 070	17 247	-5%
30028	2 020	4 190	107%
30032	3 560	3 878	9%
30189	84 830	89 758	6%
30258	2 920	3 125	7%
30341	2 700	4 295	59%
30351	1 880	1 784	-5%
34003	6 200	4 244	-32%
34032	3 790	4 013	6%
34057	370	553	50%

⁴ Pour une zone donnée (région, département,...) :
$$\frac{(\text{estimation population} - \text{estimation DIREN})}{\text{estimation DIREN}}$$

⁵ Pour les estimations en zones inondables, l'erreur moyenne est calculée au niveau communal, les chiffrages des DIREN n'étant pas disponibles au niveau iris. Les estimations des communes avec plusieurs iris sont donc « moyennées » expliquant ainsi une erreur moyenne communale inférieure à l'erreur moyenne par iris du modèle.

34108	3 110	3 398	9%
34129	8 960	9 237	3%
34145	16 300	16 222	0%
34154	9 030	4 770	-47%
34172	2 940	10 640	262%
48095	2 300	2 823	23%
66037	2 000	1 462	-27%
66136	32 839	33 047	1%
Total	230 999	241 795	5%

4.5.2 Val d'Authion, tous modèles

Modèle	Estimation logements en ZI	Erreur globale	Erreur moyenne par commune
Global	11 543	9%	32%
Détaillé	11 885	12%	32%
Local	11 793	11%	31%
Local pondéré 1	11 824	11%	31%
Local pondéré 2	11 917	12%	32%
Local pondéré 3	12 004	13%	32%

Paradoxalement, le meilleur résultat global est obtenu avec le modèle le plus simple (France). Mais les écarts entre modèles sont minimes.

A noter : la comparaison porte sur les communes dont la zone inondable est définie par les 4 aléas, différente de l'enveloppe des zones inondables fournies dans l'outil cartorisque.

4.5.3 Centre (communes > 10000 habitants), tous modèles

Modèle	Estimation population en ZI	Erreur globale	Erreur moyenne par commune
Global	133 656	5%	11%
Détaillé	134 071	6%	11%
Local	133 958	5%	11%
Local pondéré 1	133 990	6%	11%
Local pondéré 2	134 005	6%	11%
Local pondéré 3	133 989	6%	11%

Les résultats sont très proches des données issues de l'appariement RIL/RP 1999 de l'Insee Centre, portant sur douze communes de plus de 10 000 habitants inondables par la Loire, quel que soit le modèle.

A noter : la comparaison porte sur les communes dont la zone inondable est définie par les 4 aléas, différente de l'enveloppe des zones inondables fournies dans cartorisque.

4.6 Bilan des modélisations

Les différentes variantes du modèle local semblent les plus fiables. Les tests statistiques et les comparaisons avec les résultats fournis dans les directions régionales ont conduit à retenir le modèle local pondéré (3), même si dans certains cas les écarts sont minimes.