

Aide à l'utilisation du code source Java de la modélisation ascendante d'un réseau de boucle locale optique mutualisée pour la tarification du dégroupage

Version 1.2 – avril 2020

1 Prise en main du projet Java

Les différents fichiers de classe composant le code Java brut sont publiés. Il est conseillé d'utiliser un environnement de développement (IDE) (tel que NetBeans sous Windows) pour pouvoir les ouvrir dans une interface appropriée, éventuellement les modifier, compiler et exécuter le code¹.

Les bibliothèques supplémentaires à télécharger sont les suivantes :

- GeoTools (<https://sourceforge.net/projects/geotools/files/GeoTools%2017%20Releases/17.1/>)²
- JGraphT (<http://jgrapht.org/>) ;
- KDTree (<http://home.wlu.edu/~levys/software/kd/>) ;
- Guava (<http://www.java2s.com/Code/Jar/g/Downloadguavajar.htm>).

2 Accès au modèle sur GitHub

Un dépôt Git permettant aux utilisateurs de récupérer le code et de repérer rapidement les modifications qui ont été apportées depuis les versions précédentes a été mis en place ici : <https://github.com/ARCEP-dev/BLOM.git>³.

3 Remarques préliminaires sur les fichiers manipulés par le modèle

Désormais, les emplacements de tous les **fichiers d'entrée** du modèle sont définis dans le fichier BLOM.ini, qui est lu à l'initialisation du modèle. Il est nécessaire de regrouper la plupart des fichiers d'entrée dans un répertoire commun nommé `inputFiles`, à définir également dans le fichier BLOM.ini. Les plus volumineux (tableau des lignes cuivre, fichiers de forme décrivant le génie civil et les routes) pourront quant à eux être disposés dans un emplacement différent (il est conseillé de les placer sur un disque local) ; l'utilisateur renseigne un chemin absolu vers leur emplacement dans le fichier BLOM.ini.

Les **fichiers intermédiaires** créés et réutilisés par le modèle sont stockés dans un répertoire commun nommé `cheminReseau`⁴, également à définir par l'utilisateur dans le fichier BLOM.ini.

¹ Sous NetBeans, créer un nouveau projet, choisir « Java » et « Java Application ». Ne pas créer de classe automatiquement mais insérer le dossier « arcep » publié en consultation publique dans le dossier « src » créé automatiquement par l'IDE. En cas d'utilisation avec Git, il est préférable d'utiliser l'option « Java Project with Existing Sources » lors de cette création.

² Il n'est pas nécessaire d'inclure dans le modèle le fichier « imageio-ext-imageio-readmt-1.1M012010.jar » qui figure par erreur dans le paquet à télécharger.

³ Ce dépôt distant est toutefois fourni sans garantie d'exactitude ni de disponibilité : le code publié sur le site web de l'Arcep en annexe de la consultation publique fait foi.

⁴ Il existe une exception néanmoins, les fichiers `LP/LP_[dpt].csv`, produits par le module de prétraitement, sont stockés avec les fichiers d'entrée du modèle.

Les **fichiers de résultats** sont regroupés pour chaque lancement du modèle dans un dossier commun identifié par la date et l'heure de sa création. Ces dossiers sont dans un répertoire commun nommé `racineResultats`, également à définir par l'utilisateur dans le fichier `BLOM.ini`.

Pour des raisons de lisibilité, il est fortement recommandé que les répertoires `inputFiles`, `cheminReseau` et `racineResultats` soient trois emplacements différents. Les fichiers `.csv` et `.txt` utilisent *systématiquement* des points-virgules pour délimiter les champs et la première ligne est une ligne d'en-tête dans tous les cas⁵.

Toutes les données géographiques (*shapefiles* et coordonnées des nœuds dans les fichiers CSV) doivent être fournies dans le même système de coordonnées de référence (pour une zone géographique⁶ donnée) ; il est préconisé d'utiliser **Lambert 93** (EPSG 2154) pour la France métropolitaine et les référentiels géodésiques habituels pour les départements et régions d'outre-mer.

Chemins à spécifier (BLOM.ini) :

- `cheminReseau`,
- `inputFiles`,
- `racineResultats`.

Remarque : les chemins indiqués pour les variables `inputFiles`, `cheminReseau` et `racineResultats` doivent tous se terminer par un « / ».

3.1 Fichiers texte issus d'Orange

Un fichier « `listeNRA.csv` » répertoriant l'identifiant et les coordonnées de tous les NRA : une ligne par NRA, 13 colonnes délimitées par des points-virgules dont seules la première (`CODE INSEE NRA`), la neuvième (`COX – coordonnée X`), la dixième (`COY – coordonnée Y`) et la onzième (`codage`, indiquant la référence EPSG du système de coordonnées) sont utilisés.

Un fichier « `listeSR.txt` » répertoriant l'identifiant propre, l'identifiant du NRA correspondant et les coordonnées de tous les sous-répartiteurs de 1^{er} niveau : une ligne par SR, 16 colonnes délimitées par des points-virgules dont seules la première (`CLE SR 1ER NIVEAU`), la deuxième (`CLE NRA`), la septième (`COORDONNEES X`) et la huitième (`COORDONNEES Y`) sont utilisées. Ce fichier n'est pas nécessaire pour faire fonctionner le modèle.

Un fichier « `LP_totalesT42019.csv` » obtenu à partir de la fusion des fichiers « `ND_banalises.csv` » et « `NB_suclocs_banalises.csv` » extraits de la base de données sous-jacente à l'application 42C, regroupant, pour toutes les lignes principales, identifiant du NRA, identifiant du SR, identifiant du PC, coordonnées du PC, code géographique (INSEE) de la commune du PC : une ligne par « ligne principale », 13 colonnes délimitées par des points-virgules dont seules la cinquième (`CLE_NRA`), la sixième (`CLE_SR`), la septième (`CLE_PC`), la huitième (`CODE_COMMUNE`), la onzième (`COORD_X`), la douzième (`COORD_Y`) et la treizième (`PCD_COORDINATE_SYSTEM`) sont utilisées.⁷ Ce fichier est à stocker en local.

⁵ Il existe deux exceptions néanmoins :

- les fichiers `NRO-XXkm/.../ListePC_[NRO].csv` présentent en 7^e (G1) et 8^e (H1) position de la première ligne les coordonnées X et Y du NRO ;
- Les fichiers `BLO[-GC][-routier]-[XX]km/.../Nœuds_[NRO].csv` présentent une première ligne donnant l'identifiant du nœud qui accueille le NRO et le type de zone de régulation ; l'en-tête des colonnes est à la deuxième ligne et les données tabulaires à proprement parler ne commencent donc qu'à la troisième ligne.

⁶ France métropolitaine, d'une part, et chacun des DROM, d'autre part.

⁷ Dans tout le modèle, l'identifiant d'un NRA est la clé à 8 caractères utilisée par Orange, composée de 5 chiffres du code INSEE de la commune d'implantation du NRA et de trois lettres inspirées du nom de cette même commune ; l'identifiant d'un SR est l'identifiant du NRA correspondant suffixé de quelques caractères identifiant le SR, et l'identifiant d'un PC est lui aussi

Un fichier « liens_collecte.csv » listant les liens de collecte⁸ entre les NRA d'Orange. Le fichier présente une ligne par lien avec 3 colonnes délimitées par des points-virgules qui sont toutes utilisées : la première donne la longueur du lien en mètres, la deuxième est le code du premier NRA du lien et la troisième le code du second NRA du lien.

Chemins à spécifier (BLOM.ini) :

- adresseLP,
- fichierSR,
- fichierNRA,
- fichierCollecte.

3.2 Fichiers de forme de génie civil et du réseau routier

Il existe un « fichier de forme » (*shapefile*) par département et par réseau (c'est-à-dire un groupe de fichiers d'extensions « .shp », « .shx », « .prj » etc.). Vu leur taille, il vaut mieux les stocker sur un disque local. Le fichier BLOM.ini permet de préciser l'emplacement et le nom de ces fichiers : le modèle lit des fichiers dont le chemin complet est défini comme suit :

adresseShapes[GC|Routes]/nameShape[GC|Routes].shp où :

- adresseShapes* désigne l'emplacement commun (termine obligatoirement par un « / ») ;
- nameShape* est un modèle de chemin relatif de fichier qui peut comporter la chaîne « [dpt] » à chaque fois qu'il sera nécessaire d'ajouter le numéro du département dans l'arborescence ou dans le nom du fichier.

À titre d'exemple, la spécification suivante :

```
adresseShapesRoutes=E:/donnees_2020/Routes 2020/  
nameShapeRoutes=[dpt]/TRONCON_DE_ROUTE_[dpt]
```

permettra de lire les fichiers suivants :

- E:/donnees_2020/Routes 2020/01/TRONCON_DE_ROUTE_01.shp
- E:/donnees_2020/Routes 2020/02/TRONCON_DE_ROUTE_02.shp
- etc.

L'emplacement et le radical commun doivent être spécifiés dans le fichier BLOM.ini pour les fichiers de forme décrivant le génie civil et pour ceux décrivant le réseau routier. Ces fichiers (lisibles en utilisant par exemple le logiciel libre QGIS) sont une collection d'objets (*features*) qui ont chacun une géométrie (*geometry*) et des champs (*attributes*) comparables à ceux d'une base de données non géographique.

Les fichiers de forme de génie civil doivent disposer *a minima* de l'attribut suivant : `MODE_POSE`. Cet attribut est supposé prendre des valeurs entières. Conformément à la structure des fichiers fournis par Orange, le code considère qu'une valeur entre 0 et 2 correspond à du génie civil aérien tandis que les valeurs 3 et 5 à 10 correspondent à du génie civil en conduite et la valeur 4 correspond à du génie civil de pleine terre (voir la section 3.4.5 de la consultation publique sur la version 1.2).

composé de l'identifiant du NRA correspondant suffixé de quelques caractères identifiant le PC. Le critère d'identification unique d'un PC utilisé par le code de la version 1.2 est le couple « `CLE_SR` / `CLE_PC` » conformément à la structure des fichiers fournis par Orange.

⁸ Sur support optique, cuivre ou par faisceaux hertziens.

Les fichiers de forme représentant les routes doivent disposer *a minima* de l'attribut suivant : `NATURE` (type : chaîne de caractères). Il est à noter que le modèle filtre les arêtes de ces fichiers de forme pour ne conserver que celle dont l'attribut `NATURE`⁹ :

- commence par « Route » ;
- ou vaut « Rond-point ».

Note : les fichiers fournis au téléchargement par l'IGN utilisent le système de coordonnées Lambert 93 pour les départements métropolitains.

Chemins à spécifier (BLOM.ini) :

- `adresseShapesGC,`
- `nameShapeGC,`
- `adresseShapesRoutes,`
- `nameShapeRoutes.`

3.3 Fichiers de géographie administrative

L'utilisation du modèle nécessite de fournir des fichiers de forme décrivant les formes des départements¹⁰ : il en existe un pour la métropole et un par département ultramarin. L'utilisateur est invité à entrer dans le fichier `BLOM.ini` les emplacements des fichiers et le radical commun des noms de fichier. Le modèle cherchera les codes et noms des départements dans les attributs 2 et 3 des fichiers. Pour les départements métropolitains, attention à bien convertir si nécessaire ces fichiers dans le SCR de travail auparavant (il est préconiser d'utiliser les SCR « Lambert 93 » pour la France métropolitaine).

Le modèle utilise également un fichier « `liste-de-departements-limitrophes-francais.csv` » indiquant pour chaque département la liste des codes des départements limitrophes : une ligne par départements et 9 colonnes séparées par des points-virgules : la première indique le numéro du département, les suivantes celles de ses voisins (tant qu'il en a – il existe au plus 8 voisins).

Chemins à spécifier (BLOM.ini) :

- `fichierDeptLimitrophes,`
- `adresseShapesDpts,`
- `nameShapeDpts.`

Remarques :

- le chemin `adresseShapesDpts` doit terminer par un « / » ;
- les fichiers de forme correspondant aux DROM seront dans les répertoires `adresseShapesDpts/971/`, `adresseShapesDpts/972/`, etc.

⁹ Concernant les fichiers publiés par l'IGN dans la Base « BD TOPO », cet attribut peut prendre les valeurs suivantes : Type autoroutier, Breteille, Route à 2 chaussées, Route à 1 chaussée, Route empierrée, Chemin, Escalier, Bac ou liaison maritime, Rond-point, Piste cyclable, Sentier. Se reporter à la documentation de la base pour davantage d'informations : https://geoservices.ign.fr/ressources_documentaires/Espace_documentaire/BASES_VECTORIELLES/BDTOPO/DC_BDTOPO_3-0.pdf

¹⁰ Pour l'exécution de la version 1.2, ce sont les fichiers « `DEPARTEMENT.shp` » publiés par l'IGN dans base de données « ADMIN-EXPRESS-COG » édition 2019 qui ont été utilisés. Ils sont accessibles publiquement à l'adresse suivante : <https://geoservices.ign.fr/documentation/diffusion/telechargement-donnees-libres.html#admin-express>

3.4 Fichiers en lien avec les zones de régulation

Le modèle utilise également des fichiers de forme décrivant les zones très denses et les zones moins denses d'initiative privée en métropole :

- le fichier choisi pour les zones très denses ne contient que deux objets géographiques (*features*) : l'union des poches de basse densité et l'union des zones très denses hors poches de basse densité. On utilise pour les distinguer leur attribut n°1, « Poche », qui peut prendre pour valeur uniquement « PHD » ou « PBD » ;
- le fichier utilisé pour les zones moins denses d'initiative privée contient un objet par commune de ces zones. Le premier champ doit être le code INSEE de la commune (débutant par le numéro du département).

Par ailleurs, il convient de fournir au modèle pour les départements auxquels on s'intéresse et leurs départements limitrophes un fichier par département (ex : « Communes/Communes_dpt.csv » où « dpt » est le numéro du département à deux ou trois caractères) contenant pour chaque code INSEE de commune du département la zone dont il fait partie : il y a une ligne par code INSEE et deux colonnes séparées par un point-virgule : la première contenant le « Code INSEE » (format : 5 chiffres), la seconde la « Zone » (parmi les trois zones retenues : « ZTD », « AMII » ou « RIP »).

Chemins à spécifier (BLOM.ini) :

- adresseShapeZTD,
- adresseShapeAMII,
- dossierCommunes.

4 Données d'entrée du module de déploiement (hors fichiers intermédiaires)

Le modèle peut utiliser un ou plusieurs fichier(s) au format texte délimité (fichiers « .csv ») de demande cible donnant pour chaque département français la demande cible retenue par zone (ZTD, ZMD privées, ZMD publiques). Ces fichiers comportent une ligne d'en-tête puis une ligne par département, et 4 colonnes séparées par des points-virgules : la première (« Département ») donne le code à 2 ou 3 chiffres du département, la deuxième (« ZTD ») la demande cible en ZTD, la troisième (« ZMD-AMII ») la demande cible en ZMD privées et la quatrième (« ZMD-RIP ») la demande cible en ZMD publiques. Il est possible d'indiquer deux fichiers dont les valeurs par zone et par département seront alors sommés par le modèle (deux fichiers ont été utilisés pour la version 1.2, le premier construit à partir du millésime 2016 de la base « Logements »¹¹ de l'INSEE et le second à partir du millésime 2015 de la base « Clap »¹² de l'INSEE)¹³.

Le modèle utilise un fichier nommé « immeubles.csv » donnant pour chaque NRA la distribution d'immeubles présents sur le périmètre de la zone arrière de NRA en fonction du nombre de logements de chacun d'entre eux. Ce fichier est constitué d'une ligne par NRA et de 99 colonnes séparées par des points-virgules : la première (« NRA ») donne l'identifiant à 8 caractères du NRA, les 98 autres (« NB IMM AVEC X » pour $1 \leq X \leq 98$) donnent le nombre d'immeubles avec X logements sur chaque ZANRA.

¹¹ Les données et la documentation de cette enquête sont disponibles publiquement à l'adresse suivante : <https://www.insee.fr/fr/statistiques/4171415?sommaire=4171436>

¹² Dispositif « Connaissance locale de l'appareil productif ». La documentation est accessible à l'adresse suivante : <https://www.insee.fr/fr/information/2044773>

¹³ À propos de la constitution de ces fichiers, le lecteur est invité à consulter la section 4.4 du document de consultation publique sur la version 1.2 du modèle.

Chemins à spécifier (BLOM.ini) :

- fichierDemandeCible1,
- fichierDemandeCible2,
- fichierImmeubles.

En outre, un fichier de coûts unitaires était initialement utilisé par le module de déploiement pour calculer les coûts totaux mais cette fonctionnalité a été séparée dans un module de coûts Excel pour faciliter la compréhension de son fonctionnement et permettre l'étude plus rapide des sensibilités aux paramètres de coûts. Ce fichier n'est donc plus utilisé (il est fourni par l'Arcep avec des valeurs nulles qui n'ont aucune signification particulière ; les valeurs de coûts unitaires utilisées sont celles renseignées dans le classeur Excel « BLOM - Module de coûts v1.2 (avril 2020).xlsx »).

Chemin à spécifier (BLOM.ini) :

- fichierCoutsUnitaires.

5 Autres paramètres

Le fichier BLOM.ini est également utilisé pour spécifier la valeur de paramètres numérique utilisés par le modèle.

- les poids relatifs des différents types d'arêtes d'infrastructure mobilisables pour construire le réseau (il faut noter que ces facteurs sont pris égaux à 1 s'ils ne sont pas spécifiés) :
 - o facteurConduite : poids linéique des arêtes de génie civil souterrain existant,
 - o facteurAerien : poids linéique des arêtes de génie civil aérien existant,
 - o facteurPleineTerre : poids linéique des arêtes à reconstruire (pleine terre et routes) ;
- le ratio entre la longueur « réelle » utilisée des arêtes construites par le modèle et celle de leur tracé à vol d'oiseau entre les deux nœuds qu'elles relient (coefVdO) ;
- la « graine » à partir de laquelle seront définis les nombres pseudo-aléatoires utilisés pour le tirage au sort des points de concentration (PC) sur lesquels la demande sera ajoutée (*seed*) (conserver le même *seed* permet de comparer des résultats du modèle, déploiement inclus, entre eux). Si elle n'est pas spécifiée, ou si elle est fixée à 0, le choix est effectué par le modèle mais n'est pas constant d'une exécution à l'autre ;
- le système de coordonnées de référence à utiliser pour les *shapefiles* écrits en sortie du modèle (CRS_Sortie). Par défaut, le SCR utilisé est Lambert 93.

6 Calcul parallèle avec le modèle de BLOM

L'exécution du module topologique du modèle de BLOM est assez gourmande en temps de calcul, en particulier le module C (module topologique, bouton « Module topo et sortie des fichiers intermédiaires « BLO » »). Il est néanmoins possible, et profitable sur une machine disposant de plusieurs cœurs, de lancer plusieurs exécutions parallèles de ce module car un nouveau *thread* est créé à chaque clic sur le bouton. On peut ainsi répartir les départements entre les différents *threads*, ce qui permet de réduire de façon substantielle le temps de calcul et de réaliser des calculs parallèles sans être contraint par la terminaison de l'exécution du premier *thread*. Par exemple :

- choisir les départements 1 à 10 (dans l'onglet « Choix des départements »), puis lancer le module topologique ;
- choisir les départements 11 à 2B, puis lancer le module topologique ;

- etc.

Il faut néanmoins noter qu'à ce stade du développement du modèle, il n'a pas été prévu une sortie de console pour chaque *thread*. Dans cette stratégie de calcul, la console est donc plus difficilement utilisable.