workflowDetail

ndiop

January 2021

1 Introduction

La demande dans MATSim renvoie aux activités journalières des agents de la population, caractérisées par un type (motif de déplacement), des coordonnées (position de l'activité) et une heure de fin (heure de départ) ainsi que leurs déplacements planifiés matérialisés par le mode choisi.

La génération de la demande au Nord-Pas-de-Calais repose sur les données de fichiers détails de l'INSEE portant sur le recensement de la population française en 2016. En effet, à partir de là sont obtenues les données sur la mobilité professionnelle et scolaire qui décrivent les déplacements quotidiens de la population française (communes de résidence et de travail ou d'étude, le mode principal de déplacement, l'âge, le sexe ...).

Ces données ont fait l'objet de traitement avec Python et Qgis pour extraire la partie Nord-Pasde-Calais pour générer les flux origine et destination en format xml compatible avec MATSim. Mais ces données bien qu'intéressantes pour notre cas d'étude qui se focalise sur un réseau à grande échelle, présentent quelques limites. Elles ne renseignent pas sur les différents types d'activités comme les loisirs et les achats.

Elles ont été complétées alors par les données de l'Enquête Régionale Mobilité et Déplacement 2016 (ERMD) fournies par le Conseil régional des Hauts-de-France.

Ces données représentent la concaténation des enquêtes effectuées par les collectivités de la région, à des dates différentes et qui ont été redressées par rapport à l'année 2016 et aux données de l'INSEE, pour pondérer les déplacements. Compte tenu du format initial des données brutes, les variables sont séparées en amont avec un éditeur et les données stockées dans le dossier datahdf. Les données ERMD sont composées de quatre fichiers textes : menage (Fig. 1) au nombre de 78251, person (Fig. 2) au nombre 76751, deplacement (Fig. 3) au nombre de 301760 et trajet (Fig. 4) au nombre de 246242 observés).

La méthodologie adoptée pour générer la population MATSim est schématisée dans DEMAND-GENERATION :

https://drive.google.com/file/d/1otVfADcOyOLeh8aHCrcnyqU3nhK3o93j/view.



Figure 1: données sur les ménages de l'ERMD 2016.

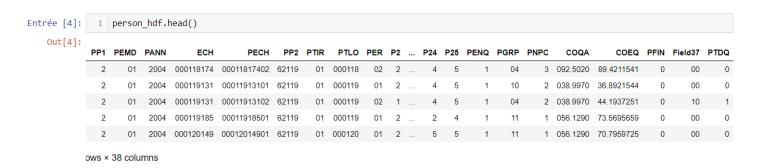


Figure 2: données sur les personnes de l'ERMD 2016.



Figure 3: données sur les déplacements de l'ERMD 2016.

trée [6]:	1	tra	ajet_h	df.hea	d()													
Out[6]:		TP1	TEMD	TANN	ECH	PECH	DECH	TP2	TTIR	TTLO	PER	NDEP	T1	T2	тз	Т6	T10	TFIN
	0	4	01	2004	000102147	00010214702	0001021470211	62119	01	000102	02	11	1	00	22	00	NaN	0
	1	4	01	2004	000102147	00010214702	0001021470212	62119	01	000102	02	12	1	00	22	00	NaN	0
	2	4	01	2004	000102165	00010216501	0001021650101	62119	01	000102	01	01	1	00	21	00	NaN	0
	3	4	01	2004	000102165	00010216501	0001021650102	62119	01	000102	01	02	1	00	21	00	2	0
	4	4	01	2004	000102165	00010216501	0001021650103	62119	01	000102	01	03	1	00	21	00	2	0

Figure 4: données sur les trajets de l'ERMD 2016.

2 Préparation des données

2.1 Concaténation des données d'enquête

Cette étape consiste à nettoyer la base de données ERMD (dataHdFpreparing.ipynb). Ce code fait la concaténation des clés avec le numéro et l'année d'enquête pour éviter les doublons. Il prend en entrée les données situées dans le dossier datahdf fourni par le conseil régional des Hauts-de-France et produit en sortie (exemple illustré à la Fig. 5) des données stockées dans le dossier inputhdf. Il est à noter qu'après la concaténation, une seule occurence est remarquée dans la base de déplacement. Elle correspond à une erreur de codage concernant la personne de clé 000101003. En effet, deux de ses déplacements on le même code 00010100301 alors que l'un est en réalité le dernier déplacement (04) si l'on regarde son trajet journalier. Du coup, ce lapsus a été corrigé avec les répercutions sur l'heure de départ à 16h30 pour être conforme à son heure d'arrivée de 17h00 à son domicile et d'une durée de trajet de 30mn.

[22]:		DP1	DEMD	DANN	ECH	PECH	DECH	DP2	DTIR	DTLO	PER	 D5AA	MODP	DGRP	DOIB	DIST
	0	3	01	2004	2004000101129	200400010112901	20040001011290101	62119	01	000101	01	 09	22	11	3174	4641
	1	3	01	2004	2004000101129	200400010112901	20040001011290102	62119	01	000101	01	 01	22	11	3174	4799
	2	3	01	2004	2004000101161	200400010116101	20040001011610101	62119	01	000101	01	 01	21	05	212507	248297
	3	3	01	2004	2004000101161	200400010116101	20040001011610102	62119	01	000101	01	 10	21	05	1964	4016
	4	3	01	2004	2004000101161	200400010116101	20040001011610103	62119	01	000101	01	 06	21	05	6822	10228
	301755	3	30	2014	20144560093087	2014456009308701	201445600930870103	80021	46	456009	01	 01	31	06	2703	4054
	301756	3	30	2014	20144560093629	2014456009362901	201445600936290101	80021	46	456009	01	 09	22	09	40899	45297
	301757	3	30	2014	20144560093629	2014456009362901	201445600936290102	80021	46	456009	01	 01	22	09	40899	45642
	301758	3	30	2014	20144560093672	2014456009367201	201445600936720101	80021	46	456009	01	 02	21	05	18713	22933
	301759	3	30	2014	20144560093672	2014456009367201	201445600936720102	80021	46	456009	01	 01	21	05	18713	23228

Figure 5: données nettoyées sur les deplacements de l'ERMD 2016.

Le code accessible ci-après, sort les données prêtes pour le traitement. DATAHDF__PREPARING : https://drive.google.com/file/d/1R9ggAUGAtCYYI4_rPDRFn1beNbVcV_gv/view?usp=sharing.

2.2 Récupération des coordonnées X,Y des communes

Cette étape, indépendante de la précédente, consiste à récupérer les coordonnées des centroides des communes et des codes commune et departement à partir des données cartographiques de la région fourni par l'IGN, à l'aide de Qgis. Cette étape est essentielle pour d'une part la conversion des codes INSEE en coordonnées X,Y et d'autre part de faire le filtrage des zones à étudier. La base de données est représentée à la Fig. 6

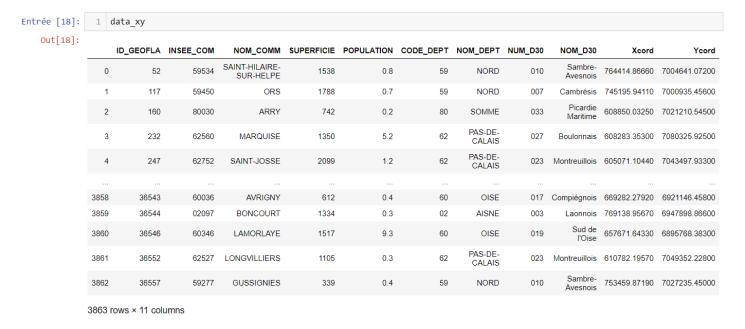


Figure 6: base de données sur les coordonnées des communes des HDF.

3 Déplacement des personnes éligibles en Haut-de-France

Cette étape permet de traiter les données de l'ERMD pour obtenir le fichier déplacement de la population enquêtée accessible sur : ERMD2016_TO_TRAVELDIARIES :

https://drive.google.com/file/d/1_-jjr_osuJDm20Yoku0Ff2iiL7D7RtYZ/view?usp=sharing. Ce code suit les sous-étapes suivantes :

3.1 Filtrage des grands mobiles

- 1. Il commence par étudier l'éligibilité des Grands Mobiles constitués de personnes âgées de plus de 11 ans qui ont réalisé au moins un déplacement de plus de 10km la veille de l'enquête. Cette définition des Grands Mobiles a été revue dans le cadre de cette étude pour l'harmonisation des données d'enquêtes (cf. rapport Les Grands Mobiles en Hauts-de-France (Hasiak et al., 2017)). Les variables MTDQ, PTDQ et DTDQ permettent de filtrer (=1 si éligibles) les données menage.txt, person.txt et deplacement.txt. Ce qui permet d'obtenir en sortie les obersvations éligibles (31083 ménages (40%), 36848 personnes (22%) et 116297 déplacements (39%) éligibles) stockées dans le dossier output_hdf/eligible.
- 2. Par la suite, le code fait la vérification d'éventuels doublons. Cette partie a été réalisée avec le code précédent (dataHdF_preparing) qui avait abouti à la concaténation des clés avec l'année d'enquête. Ceci a permis d'avoir les occurences des clés de chaque base toutes égales à 1.
- 3. Puis, il procède à la jointure des tables ménage, personne et déplacement. Ici, dans les tables menage et person, les observations éligibles sont pris en compte contrairement à la table deplacement que l'on prend intégralement pour tenir compte de tous les déplace-

ments d'une personne et ménage éligible.

Puisque ce compromis va prendre en compte des déplacement inéligibles, cela va engendrer des occurences égales à 2 sur les clés (concerne 11 déplacements). L'un des déplacements pour chaque doublon est extrait de la base en privilégiant les conditions d'éligibilité. Ces manipulations permettent d'obtenir 168623 déplacements comme illustré à la Fig. 7

40]:		PP1	PEMD	PANN	ECH_x	PECH	PP2	PTIR	PTLO	PER_x	P2	 D5AA	MODP	DGRP	DOIB	DIST	DISP	DFIN	Field
	0	2	01	2004	2004000119131	200400011913102	62119	01	000119	02	1	 02	21	04	615	1107	1107	0	
	1	2	01	2004	2004000119131	200400011913102	62119	01	000119	02	1	 01	21	04	615	1107	1107	0	
	2	2	01	2004	2004000119131	200400011913102	62119	01	000119	02	1	 02	21	04	615	1107	1107	0	
	3	2	01	2004	2004000119131	200400011913102	62119	01	000119	02	1	 01	21	04	615	1107	1107	0	
	4	2	01	2004	2004000119131	200400011913102	62119	01	000119	02	1	 10	21	04	3226	5114	5044	0	
	168630	2	30	2014	20144560093087	2014456009308701	80021	46	456009	01	2	 01	31	06	2703	4054	NaN	0	
	168631	2	30	2014	20144560093629	2014456009362901	80021	46	456009	01	2	 09	22	09	40899	45297	NaN	0	
	168632	2	30	2014	20144560093629	2014456009362901	80021	46	456009	01	2	 01	22	09	40899	45642	NaN	0	
	168633	2	30	2014	20144560093672	2014456009367201	80021	46	456009	01	1	 02	21	05	18713	22933	NaN	0	
	168634	2	30	2014	20144560093672	2014456009367201	80021	46	456009	01	1	01	21	05	18713	23228	NaN	0	

Figure 7: base de données des déplacements filtrés de GM en HdF.

3.2 Jointure des données avec les coordonnées X,Y

Cette fonction du code prend en entrée les données de déplacement filtrés et du traitement sur QGis qui a permis d'avoir pour chaque commune ses coorodnnées X,Y, les codes des départemets et de communes. Elle fait ensuite le mapping des X,Y; les départements avec les communes de la base de déplacements filtrés en fonction des codes INSEE comme clé de mapping. Ces mappings permettront par la suite de faire le filtre des périmètres d'étude (hauts-de-France et Nord-Pas-de-Calais). Avec une extraction des variables pertinentes pour l'étude, cette partie du code génère donc les données des déplacements des grands mobiles dans les HdF avec leurx coordonnées X,Y (Fig. 8). Il est à remarquer que la base générée et stockées dans le dossier output_hdf/travelDiaries, se présente sous deux formats : i) celle qui concerne tous les déplacements, qu'il proviennent des HdF ou pas (Fig. 8) ; ii) celle qui comprend les déplacements qui proviennent uniquement des HdF (obtenue grace à la limitation des données sur les communes de la France (Fig. 9)).

3.3 Déplacement des personnes éligibes au Nord-Pas-de-Calais

Après la génération de la base de déplacement des habitants des Hauts-de-France (Fig. 9), une méthode permet de filtrer les déplacements pour le Nord-Pas-de-Calais en tant qu'origine ou destination de flux de la région et de la générer dans le dossier output/travelDiaries (Fig. 10) nécessaires à la génération de la demande de transport.

t[53]:																
[33].		DECH	PECH	COEQ	DP2	D4A	D4B	D3	D3X	D3Y	D2AA	 P10	M5	DOIB	DIST	D8C
	0	20040001191310201	200400011913102	44.1937251	62119	07	45	62119	674606.16370	7048016.16600	01	 2	2	615	1107	0005
	1	20040001191310202	200400011913102	44.1937251	62119	12	00	62119	674606.16370	7048016.16600	02	 2	2	615	1107	0005
	2	20040001191310203	200400011913102	44.1937251	62119	13	20	62119	674606.16370	7048016.16600	01	 2	2	615	1107	0005
	3	20040001191310204	200400011913102	44.1937251	62119	16	30	62119	674606.16370	7048016.16600	02	 2	2	615	1107	0005
	4	20040001191310205	200400011913102	44.1937251	62119	17	30	62119	674606.16370	7048016.16600	01	 2	2	3226	5114	0010
	168630	201445600930870103	2014456009308701	8.93880200	80021	18	15	80021	648955.87620	6978257.48500	06	 NaN	0	2703	4054	0030
	168631	201445600936290101	2014456009362901	7.49811400	80021	18	00	80021	648955.87620	6978257.48500	01	 NaN	2	40899	45297	0040
	168632	201445600936290102	2014456009362901	7.49811400	80021	22	40	80685	684744.61030	6954979.24800	09	 NaN	2	40899	45642	0040
	168633	201445600936720101	2014456009367201	8.88031600	80021	80	30	80021	648955.87620	6978257.48500	01	 NaN	2	18713	22933	0030
	168634	201445600936720102	2014456009367201	8.88031600	80021	16	00	80570	662824.32860	6964266.62300	02	 NaN	2	18713	23228	0030

Figure 8: base de données des déplacements de GM en HdF avec leurx coordonnées X,Y.

e [54]:	1 tr	avelDiary_hdf														
ut[54]:		DECH	PECH	COEQ	DP2	D4A	D4B	D3	D3X	D3Y	D2AA	 P10	М5	DOIB	DIST	D8C
	0	20040001191310201	200400011913102	44.1937251	62119	07	45	62119	674606.16370	7048016.16600	01	 2	2	615	1107	0005
	1	20040001191310202	200400011913102	44.1937251	62119	12	00	62119	674606.16370	7048016.16600	02	 2	2	615	1107	0005
	2	20040001191310203	200400011913102	44.1937251	62119	13	20	62119	674606.16370	7048016.16600	01	 2	2	615	1107	0005
	3	20040001191310204	200400011913102	44.1937251	62119	16	30	62119	674606.16370	7048016.16600	02	 2	2	615	1107	0005
	4	20040001191310205	200400011913102	44.1937251	62119	17	30	62119	674606.16370	7048016.16600	01	 2	2	3226	5114	0010
	168630	201445600930870103	2014456009308701	8.93880200	80021	18	15	80021	648955.87620	6978257.48500	06	 NaN	0	2703	4054	0030
	168631	201445600936290101	2014456009362901	7.49811400	80021	18	00	80021	648955.87620	6978257.48500	01	 NaN	2	40899	45297	0040
	168632	201445600936290102	2014456009362901	7.49811400	80021	22	40	80685	684744.61030	6954979.24800	09	 NaN	2	40899	45642	0040
	168633	201445600936720101	2014456009367201	8.88031600	80021	80	30	80021	648955.87620	6978257.48500	01	 NaN	2	18713	22933	0030
	168634	201445600936720102	2014456009367201	8.88031600	80021	16	00	80570	662824.32860	6964266.62300	02	 NaN	2	18713	23228	003

Figure 9: base de données des déplacements de GM provenant des HdF.

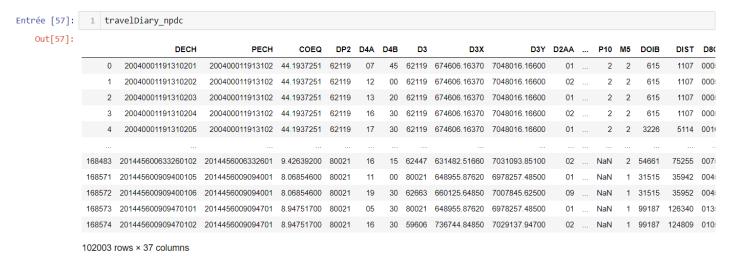


Figure 10: base de données des déplacements de GM provenant des HdF pour le NPdC.

4 Création de la population MATSim

4.1 Population synthétique pour le NPdC

Le code MATSIMDEMAND: https://drive.google.com/file/d/1nAGRgqdiIQrYobZ616cjxFNlimZIh_N9/view?usp=sharing prend en entrée, le fichier texte travelDiary_npdc (déplacement des agents) précédent et génère la population au format de MATSim (Fig. 11) en conformité avec les données de l'enquête et dont il conviendra de pondérer pour obtenir la population réelle.

Ce code commence par formater les variables aux codages utilisés dans MATSim sur les modes de transport (car, pt, walk, bike, ...) et les motifs de déplacement (home, work, leisure, education, ...).

Ensuite, le code fait les affectations des observations en regroupant les agents selon leur identifiant unique et suivant une organisation adaptée à MATSim.

La population illustrée à la Fig. 11 composée de 22479 agents, est une représentation simplifiée (1.86%) de la population réelle des Grands Mobiles en déplacement au NPdC. Chaque individu accompagné par son plan journalier est affecté, un poids qui signifie la probabilité que son plan se reproduise dans la base.

4.2 Population réelle en déplacement au NPDC

Cette population pondérée créée servira pour génèrer alors la population réelle du NPdC (1 209 213 agents) avec la création d'un échantillon souhaité selon la proportion à définir (fichier proportion.dat) tenant compte des poids de chaque individu. Une distribution sur 500 m des agents de même origine ou de destination est faite pour éviter qu'ils quittent le même point. Le code SAM-PLINGNPDC: https://drive.google.com/file/d/1U8BUYjZi1yzOPIvnrPO8K7vKjls7cfhu/view?usp=sharing se termine par une incrémentation des identifiants des agents pouré viter les doublons et fournit le fichier demande finale au format MATSim (samplePopNPdC.xml) illustré

```
?xml version="1.0" encoding="utf-8"?>
 <!DOCTYPE plans SYSTEM "http://www.matsim.org/files/dtd/plans_v4.dtd">
<person id="200400210303102">
   <!--; 35.399 ;-->
     <plan>
        <act type="home" x="667045.62600" y="7039803.07500" start time="00:00:00" end time="08:43:00"/>
        <leq mode="car"></leq>
        <act type="leisure" x="667045.62600" y="7039803.07500" start time="08:47:00" end time="08:55:00"/>
        <leg mode="car"></leg>
        <act type="work" x="668164.02800" y="7043751.36200" start time="09:12:00" end time="17:30:00"/>
       <leg mode="car"></leg>
        <act type="leisure" x="667045.62600" y="7039803.07500" start time="17:40:00" end time="17:43:00"/>
       <leg mode="car"></leg>
        <act type="home" x="667045.62600" y="7039803.07500" start time="17:46:00" end time="18:30:00"/>
        <leg mode="car"></leg>
        <act type="leisure" x="674606.16370" y="7048016.16600" start time="18:52:00" end time="20:30:00"/>
       <leg mode="car"></leg>
       <act type="home" x="667045.62600" y="7039803.07500" start time="20:55:00"/>
     </plan>
   </person>
  <person id="200400210304702">
   <!--; 31.154 ;-->
     <plan>
        <act type="home" x="667045.62600" y="7039803.07500" start time="00:00:00" end time="08:45:00"/>
        <leq mode="pt"></leq>
        <act type="work" x="674559.22410" y="7038731.15700" start time="09:00:00" end time="11:00:00"/>
        <leq mode="pt"></leg>
        <act type="leisure" x="674559.22410" y="7038731.15700" start time="11:10:00" end time="12:15:00"/>
        <leq mode="pt"></leg>
        <act type="work" x="674559.22410" y="7038731.15700" start time="12:25:00" end time="13:45:00"/>
       <leq mode="pt"></leq>
        <act type="home" x="667045.62600" y="7039803.07500" start_time="14:00:00" end_time="15:00:00"/>
        <leg mode="walk"></leg>
        <act type="leisure" x="667045.62600" y="7039803.07500" start time="15:02:00" end time="15:35:00"/>
       <leq mode="walk"></leq>
        <act type="leisure" x="667045.62600" y="7039803.07500" start time="15:40:00" end time="18:00:00"/>
       <leg mode="walk"></leg>
        <act type="home" x="667045.62600" y="7039803.07500" start time="18:05:00"/>
     </plan>
   </person>
```

Figure 11: population MATSim synhtétique du NPdC basée sur l'ERMD 2016.

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE plans SYSTEM "http://www.matsim.org/files/dtd/plans_v4.dtd">
<plans>
  <person id="200400010408203 0001">
     <plan>
        <act type="home" x="674514.93021" y="7047657.98572" start time="00:00:00" end time="06:45:00"/>
       <leq mode="car"></leq>
        <act type="work" x="712381.26609" y="7048154.48422" start_time="07:20:00" end_time="17:10:00"/>
        <leg mode="car"></leg>
        <act type="home" x="674514.93021" y="7047657.98572" start time="17:50:00" end time="18:30:00"/>
       <leg mode="car"></leg>
        <act type="work" x="712381.26609" y="7048154.48422" start time="19:00:00" end time="20:10:00"/>
       <leq mode="car"></leq>
       <act type="home" x="674514.93021" y="7047657.98572" start time="20:40:00"/>
     </plan>
  </person>
  <person id="200400010408203 0002">
     <plan>
        <act type="home" x="674594.67812" y="7047980.19650" start time="00:00:00" end time="06:45:00"/>
        <leg mode="car"></leg>
        <act type="work" x="712777.46269" y="7048028.38868" start time="07:20:00" end time="17:10:00"/>
       <leq mode="car"></leq>
       <act type="home" x="674594.67812" y="7047980.19650" start_time="17:50:00" end_time="18:30:00"/>
       <leg mode="car"></leg>
       <act type="work" x="712777.46269" y="7048028.38868" start time="19:00:00" end time="20:10:00"/>
       <leg mode="car"></leg>
        <act type="home" x="674594.67812" y="7047980.19650" start time="20:40:00"/>
    </plan>
  </person>
</plans>
```

Figure 12: Extrait de 1% de la Population MATSim du NPdC basée sur l'ERMD 2016.

La représentation des agents (10%) dans le Nord-Pas-de-Calais est illustrée à la Fig. 13.

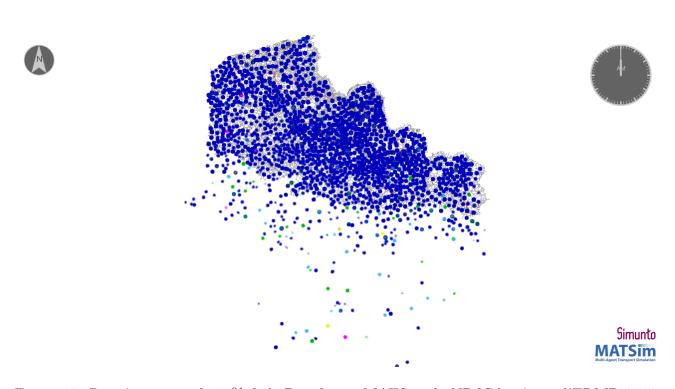


Figure 13: Représentation de 10% de la Population MATS im du NPdC basée sur l'ERMD 2016.