

École polytechnique de Louvain

Analyse du réseau belge de transports en commun

Étude de l'impact du réseau de transports sur les critères du décret Inscription

Auteurs : **Cédric DE BONTRIDDER, Céline EVERAERT**

Promoteurs : **Pierre SCHAU, Guillaume DERVAL**

Lecteur : **Alexandre DUBRAY**

Année académique 2021–2022

Master [120] : ingénieur civil en informatique

Master [120] : ingénieur civil en sciences des données

Remerciements

Nous tenons tout particulièrement à remercier nos promoteurs Pierre Schaus et Guillaume Derval, qui sont à l'origine de ce projet et qui nous ont guidés et encouragés durant toute la durée de ce mémoire. Ils nous ont également permis d'entrer en contact avec les instances politiques et ont communiqué notre rapport aux personnes concernées.

Merci également à nos relecteurs pour leurs avis constructifs sur cette rédaction. Grâce à eux, nous avons pu perfectionner notre travail et en améliorer la clarté ainsi que la pertinence.

Ensuite, nous remercions les personnes qui se sont intéressées à notre travail et qui ont pris le temps de nous rencontrer et d'échanger sur le sujet. Nos discussions nous ont permis d'envisager de nouvelles pistes de réflexion et ont contribué à rendre ce travail plus complet.

Enfin, nous remercions nos proches pour leur soutien durant toute cette année.

Résumé

En Fédération Wallonie-Bruxelles, les inscriptions des élèves entrant dans l'enseignement secondaire sont soumises au fameux décret Inscription. Celui-ci prévoit le calcul d'un indice composite pour chaque élève afin de départager les inscriptions. Cet indice est basé sur différents critères, dont certains ont fait couler beaucoup d'encre : des critères de distance à vol d'oiseau entre le domicile et l'établissement secondaire visé, mais aussi entre le domicile et l'école primaire fréquentée. Si de telles mesures sont effectivement faciles à mettre en œuvre, il n'est pas garanti qu'elles soient des plus pertinentes. En effet, une simple distance à vol d'oiseau entre deux points ne reflète pas nécessairement la durée du trajet entre ces deux lieux.

Par ailleurs, les quatre grandes compagnies de transports en commun belges (la SNCB, la STIB, les TEC et De Lijn) fournissent en libre accès des données concernant les horaires des trajets proposés ainsi que les informations géographiques qui y sont associées. Une telle source d'information pourrait donc tout à fait être mise à profit pour estimer l'accessibilité d'une école en termes de temps de trajet et pour améliorer la pertinence des critères actuellement utilisés qui s'expriment en termes de distance de trajet à vol d'oiseau.

L'objectif de ce mémoire est donc, premièrement, d'utiliser ces données afin d'analyser l'accessibilité des écoles en transports en commun.

De plus, nos analyses ont également pour but d'évaluer la valeur d'une mesure des distances à vol d'oiseau, en la comparant avec l'accessibilité réelle des écoles. Pour cela, nous avons évalué l'étendue de la discordance entre ces deux mesures en étudiant les différences de durée des trajets.

Enfin, ce mémoire a également pour vocation de proposer des améliorations relatives au calcul de l'indice composite, de façon à ce qu'il soit plus en accord avec les besoins des familles.

Ce mémoire a par ailleurs constitué l'opportunité de prendre contact avec les instances politiques afin de leur communiquer nos premiers résultats, dans un rapport qui a été rédigé à leur intention en février 2022. Cela nous a également permis de discuter des solutions qui pourraient être mises en place ainsi que des aspects légaux à prendre en considération.

Enfin, notre travail a mené à la création d'un outil de calcul permettant la comparaison entre l'indice composite actuellement utilisé et notre proposition d'amélioration.

Abstract

In the Wallonia-Brussels Federation, student's inscriptions to junior high-schools are ruled by the notorious Inscription decree. This decree defines a composite index that must be evaluated for each student in order to assess and sort out all inscriptions when the demand exceeds the number of available positions. The index is based on different criteria, some of which have caused distress within the political and academic communities and among students' parents : criteria such as distance ("as the crow flies") between residence and targeted school or even distance between residence and frequented primary schools. If such a metric is easy to establish, there is no guarantee that it is coherent and pertinent ; a simple Euclidean distance does not necessarily reflect the actual distance and commute time.

Moreover, the four Belgian public transportation authorities (SCNB, STIB, TEC and De Lijn) provide free access insights on their planned routes schedules along with associated geographic data. Such a source of information could easily be used to estimate a school's accessibility metric which would depend on the commute time using public transportation. This study aspires to use the open access data in order to estimate this metric.

Furthermore, our analysis aims to evaluate the quality of the "as the crow flies" metric by comparing it with each school's actual accessibility defined as a new metric. In order to proceed we've assessed the scope of the bias between the two metrics by studying difference in commute times between one another. This study's final goal will be to offer improvements on the manner of evaluating the composite index in such a way that it harmonises with a family's needs.

This thesis has also been an opportunity to contact political instances to share with them of our primary results detailed in a report dated February 2022. On this occasion, we have had the privilege of discussing potential solutions to implement along with legal aspects to be taken into account.

Finally, our study has brought to life a new computational tool allowing comparison between the composite index currently used and our improved solution.

Table des matières

Remerciements	i
Résumé	ii
Abstract	iii
Table des figures	viii
Liste des tableaux	ix
Abréviations	1
1 Introduction	2
1.1 Le décret Inscription	3
1.1.1 Description du décret	3
1.1.2 Réforme du décret en 2022	6
1.2 Motivations	6
1.3 Sources de données	7
1.3.1 Les données GTFS	7
1.3.2 Le fichier FASE	8
1.3.3 Les unités administratives	8
1.3.4 La grille de répartition de la population belge	8
1.4 Objectifs du mémoire	8
2 État de l’art et rappels théoriques	10
2.1 Le format GTFS	10
2.1.1 GTFS statique	10
2.1.2 GTFS en temps réel	10
2.2 Représentation du réseau de transport	11
2.2.1 Time-expanded model	11
2.2.2 Time-dependent model	11
2.3 Le problème des plus courts chemins	12
2.3.1 Single-source shortest paths	12
2.3.2 All-pair shortest paths	14
2.3.3 Comparaison des méthodes	16
2.4 Voronoï	16
2.4.1 Diagramme de Voronoï	16

2.4.2	Triangulation de Delaunay	17
2.4.3	Méthode de calcul	18
2.5	Autres considérations théoriques	19
2.5.1	Modélisation de la marche à pied	19
2.5.2	Calcul de distances à vol d'oiseau	19
2.5.3	Recherche du voisin le plus proche	20
3	Notre approche	22
3.1	Représentation du réseau	22
3.1.1	Modélisation du temps	23
3.1.2	Calcul du poids des arêtes	24
3.1.3	Déplacements autorisés	24
3.2	Calcul des plus courts chemins	25
3.3	Analyse des données	25
3.3.1	Première approximation : les zones de Voronoï	26
3.3.2	Prise en compte de plusieurs arrêts	27
3.4	Calculs pour la génération de nos figures	29
3.4.1	Calcul de population et temps relatifs	29
3.4.2	Calcul de l'indice composite	29
3.5	Le code	30
3.5.1	Organisation	30
3.5.2	Performances	33
4	Résultats	36
4.1	Résumé des hypothèses de travail	36
4.2	Analyse de l'accessibilité des établissements	38
4.2.1	Temps de trajet depuis l'école la plus proche	38
4.2.2	Discordance entre proximité et accessibilité	41
4.2.3	Conséquences pour la population	44
4.2.4	Évolution du phénomène avec le nombre d'écoles	48
4.3	Comparaison des résultats avec le rapport de février 2022	51
4.3.1	Relaxation de l'hypothèse de Voronoï	51
4.4	Étude de l'indice composite	53
4.4.1	Méthode de calcul du décret actuel : critère de distance	53
4.4.2	Impact d'un changement de critère : accessibilité	54
4.4.3	Discussion avec un expert politique	57
4.4.4	Proposition alternative : ajout d'un nouveau coefficient	58
4.5	Outil de visualisation	58
4.5.1	Application avec Streamlit	60
5	Discussion et Conclusion	63
5.1	Résumé de nos principaux résultats	63
5.2	Limites de notre travail et voies d'amélioration	64
5.3	Conclusion et perspectives	66
	Bibliographie	67

A	Lecture et interprétations de nos graphes	70
A.1	Boxplot	70
A.2	Histogramme	71
A.2.1	Variante : Histogramme cumulatif	71
A.3	Carte choroplèthes	73
A.3.1	Variante : Carte 1 km ²	73
A.4	Diagramme circulaire	74
B	Rapport remis aux instances politique (02/2022)	75
C	Guide d'utilisation de l'outil (Streamlit)	83
C.1	Lancement l'application	83
C.2	L'outil	83
C.3	Calcul de l'indice composite	85
C.4	Affichage de la carte	86
D	Description des scripts / organisation du code	88
D.1	params.py	88
D.2	utils.py	89
D.3	0_parse.py	89
D.4	1_precomputation.py	90
D.5	2_resolve.py	90
D.6	3a_schools_computations.py	90
D.7	3b_computation_grid.py	91
D.8	3c_computation_grid_bonus.py	92
D.9	4a_histograms.py, 4b_choropleth.py, 4c_pixel_map.py, 4d_ic.py .	92
D.10	6_tool.py	92
E	Détails sur les résultats	94
E.1	Comparaison des résultats par réseau	94
E.2	Distance moyenne avec l'école la plus proche	95
E.3	Temps de trajet moyen pour rentrer depuis l'école la plus proche	96
E.4	Cartes (au km ²) des temps de trajet pour revenir des écoles	97
E.5	Robustesse des résultats : listes de tailles deux à 10	98

Table des figures

1.1	Calcul du coefficient 7 (partenariat pédagogique)	5
2.1	Structure du format GTFS	11
2.2	Diagramme de Voronoï	16
2.3	Triangulation de Delaunay	17
2.4	Correspondance entre le diagramme de Voronoï et la triangulation de Delaunay	17
2.5	Ball tree	21
3.1	Création du graphe du réseau de transport à partir des fichiers GTFS . .	23
3.2	Exemple de calcul de la population par l'intersection avec les zones de Voronoï	26
3.3	Exemple de calcul : 3 arrêts autorisés	27
3.4	Pourcentage d'écoles les plus accessibles changeant avec l'ajout de 5 voisins.	28
3.5	Comparaison des temps absolu et relatif	29
3.6	Flowchart de notre code	31
3.7	Analyse de performances : Grille 1 km ²	34
3.8	Analyse de performances : Plus courts chemins	35
a	Évolution du temps d'exécution moyen en fonction du nombre de nœuds dans le graphe	35
b	Boxplot	35
4.1	Temps de trajet pour rentrer au centre de sa cellule depuis l'école la plus proche	39
a	Distribution des temps de trajet par tranche de 5 minutes	39
b	Répartition des temps de trajet sur le territoire belge	39
4.2	Temps de transport pour rentrer depuis l'école la plus proche	40
4.3	L'école la plus proche est-elle également la plus accessible ?	42
4.4	Population affectée par commune	43
4.5	Différence de temps de trajet : école proche vs école accessible	46
a	Durée moyenne du trajet supplémentaire par commune	46
b	Population impactée par minute de trajet supplémentaire	46
4.6	Différence de temps de trajet : école proche vs école accessible	47
a	Différence de la durée du trajet (école la plus proche VS école la plus accessible)	47
b	Population impactée par pourcentage de temps de trajet supplémentaire (école proche vs école accessible)	47
4.7	Exemple : calcul de la similitude entre écoles proches et accessibles . . .	48
4.8	Comparaison des 5 écoles les plus proches vs les 5 plus accessibles	49

a	Degré de similitude	49
b	Population concernée par degré de similitude	49
4.9	Comparaison de résultats : hypothèse de Voronoï vs plusieurs voisins autorisés	52
a	Temps de trajet supplémentaire	52
b	Temps de trajet supplémentaire	52
4.10	Pourcentage de population par valeur d'indice composite (ancienne méthode)	55
a	Valeurs pour l'école la plus accessible mais pas la plus proche . . .	55
b	Valeurs pour les écoles les plus accessibles qui sont aussi les plus proches	55
4.11	Indice composite calculé avec la nouvelle méthode : ratio entre le nouvel indice et l'ancien	56
a	Écoles accessibles	56
b	Écoles proches	56
4.12	Outil de visualisation : carte interactive (Folium)	59
a	Organisation de la carte	59
b	Informations pour chaque station	59
4.13	Outil de visualisation : section résultats	61
4.14	Outil de visualisation : carte et données de l'élève	62
A.1	Exemple de boxplot	70
A.2	Exemple d'histogramme	71
A.3	Exemple d'histogramme cumulatif inversé	72
A.4	Exemple de carte choroplèthe	73
A.5	Exemple de diagramme circulaire	74
C.1	Affichage par défaut	84
C.2	Section résultats de l'outil	85
C.3	La carte affichée	87
D.1	Exemple d'application de la fonction clip	92
E.1	Population impactée par minute de trajet supplémentaire : réseau confes- sionnel (orange) VS réseau non confessionnel (bleu)	94
E.2	Distance moyenne par commune [km]	95
E.3	Population concernée par valeur de distance (réparties par tranche de 500m)	95
E.4	Temps de trajet [min]	96
E.5	Temps de trajet depuis l'école la plus proche	97
E.6	Temps de trajet depuis l'école la plus accessible	97
E.7	Liste de taille 2	98
E.8	Liste de taille 3	98
E.9	Liste de taille 4	99
E.10	Liste de taille 5	99
E.11	Liste de taille 6	100
E.12	Liste de taille 7	100
E.13	Liste de taille 8	101
E.14	Liste de taille 9	101
E.15	Liste de taille 10	102

Liste des tableaux

1.1	Coefficient 1 : Préférence	3
1.2	Coefficient 2 : Distance avec l'école primaire	3
1.3	Coefficient 3 : Distance avec l'établissement secondaire	4
1.4	Coefficient 4 : Distance entre les deux établissements	4
1.5	Mise à jour du Coefficient 2 : Distance avec l'école primaire	6
2.1	Tableau récapitulatif des algorithmes	16
3.1	Graphes utilisés pour l'analyse de performance	33
4.1	Répartition des écoles par province	38
4.2	Vue d'ensemble par province : distance à vol d'oiseau de l'école la plus proche et temps de trajet minimum pour en revenir	41
4.3	Résumé par province : fraction de la population pour laquelle l'école la plus proche n'est pas la plus accessible + impact sur le temps de trajet absolu et relatif	48

Abréviations

AGPD Administration générale de la Documentation patrimoniale

APSP All-Pairs Shortest Paths

CoGI Commission de Gouvernance des inscriptions

DAG Directed Acyclic Graph

FASE Fichier signalétique des établissements d'enseignement de la Fédération Wallonie-Bruxelles

FWB Fédération Wallonie-Bruxelles

GTFS General Transit Feed Specification

ILI Instance locale des inscriptions

JSON JavaScript Object Notation

SNCB Société nationale des chemins de fer belges

SSSP Single-Source Shortest Paths

STIB Société des transports intercommunaux belges

TEC Transports en communs

Chapitre 1

Introduction

Ce premier chapitre a pour but de présenter le contexte dans lequel s'inscrit notre étude, ainsi que de la motiver et d'en présenter les objectifs. Pour commencer, nous expliquerons ce qu'est le décret Inscription ainsi que les changements apportés par la réforme approuvée en ce début d'année 2022. Des détails sur le calcul de l'indice composite¹ seront également fournis puisqu'ils seront utilisés pour une partie de nos résultats. Nous présenterons ensuite les raisons qui motivent notre travail ainsi que les objectifs poursuivis. Enfin, ce chapitre sera également l'occasion de décrire les données que nous avons utilisées.

Nous présenterons ensuite au chapitre 2 les aspects théoriques importants qui ont été utilisés pour ce mémoire, tels que la façon de représenter un réseau à partir des informations sur les horaires des trajets, ou encore les algorithmes existants pour calculer les chemins les plus courts entre un point A et un point B. Nous présenterons également un rapide état de l'art concernant l'utilisation des données GTFS².

Après ces considérations théoriques, nous détaillerons nos méthodes de calcul dans le chapitre 3. Une vue d'ensemble sur notre code y sera également présentée (l'annexe D en contient la description complète).

Ensuite, nous présenterons les résultats de nos analyses au chapitre 4. Notez qu'un rapport avait déjà été rédigé en février 2022 et communiqué aux cabinets de la Ministre de l'Éducation et du Ministre-Président de la Fédération Wallonie-Bruxelles. Celui-ci est disponible dans l'annexe B. Dans ce chapitre dédié à nos résultats, nous parlerons également de l'outil que nous avons créé. Celui-ci a pour but de fournir une estimation de l'indice composite pour l'école choisie, ainsi qu'une comparaison avec une autre version de cet indice, prenant en compte le réseau de transport.

Enfin, ce mémoire se conclura par une discussion de nos résultats et des perspectives pour ce travail (chapitre 5). Nous y aborderons également les limites de notre travail ainsi que les pistes d'amélioration.

1. L'indice composite est une valeur attribuée à chaque élève pour un établissement donné, utilisée pour réguler les inscriptions.

2. General Transit Feed Specification, un format standard utilisé par les compagnies de transports en commun (plus de détails sont fournis au chapitre 2).

1.1 Le décret Inscription

1.1.1 Description du décret

En Belgique, lors du passage de l'enseignement primaire à l'enseignement secondaire, les inscriptions des élèves sont soumises au décret Inscription, présent au cœur de nombreux débats. Celui-ci est entré en vigueur en 2010 et concerne l'enseignement fondamental ainsi que l'enseignement secondaire ordinaire et spécialisé organisés ou subventionnés par la Communauté française [1] (le décret est donc d'application dans les régions wallonne et bruxelloise). L'un de ses trois principaux objectifs est la régulation équitable des inscriptions en première année secondaire, puisque la demande de places excède régulièrement l'offre de certains établissements. Le second objectif est d'établir une procédure d'inscription plus transparente et centralisée. Ceci permet d'assurer une égalité de traitement d'un établissement à l'autre. Enfin, le dernier objectif annoncé est de « *promouvoir la lutte contre l'échec scolaire, améliorer les performances de chaque enfant, lutter contre les mécanismes de relégation en soutenant la mixité sociale, culturelle et académique* » [2].

Afin de remplir au mieux ces objectifs, la régulation des inscriptions est basée sur le calcul d'un indice composite à partir de 7 critères décrits ci-dessous. Cet indice a pour but de départager les candidatures dans un établissement donné lorsque le nombre de places disponibles est inférieur à la demande [3]. Dans sa version actuelle, le décret calcule l'indice composite via la multiplication de 7 coefficients (un par critère) [4] :

Coefficient 1 : La préférence Ce premier coefficient reflète l'ordre de préférence des parents. Le coefficient est d'autant plus élevé que l'école considérée est haute dans leur classement, les valeurs possibles sont résumées dans le tableau 1.1.

TABLE 1.1 – Coefficient 1 : Préférence

Préférence	1ère	2e	3e	4e	5e	6e à 10e
Coefficient	1.5	1.4	1.3	1.2	1.1	1

Ainsi, le critère de préférence a pour but de respecter autant que possible le choix des parents pour l'inscription dans un établissement de l'enseignement secondaire.

Coefficient 2 : La distance entre le domicile et l'école primaire fréquentée Ce coefficient dépend de la distance à vol d'oiseau entre le domicile de l'enfant et l'école primaire qu'il a fréquentée. Le coefficient est d'autant plus grand que l'école est proche, comme indiqué dans le tableau 1.2. Notons que ce coefficient prendra moins d'importance dans les futures versions du décret, afin de donner plus de poids au choix des parents et à la proximité de l'établissement secondaire (voir sous-section 1.1.2).

TABLE 1.2 – Coefficient 2 : Distance avec l'école primaire

Proximité	1ère	2e	3e	4e	5e	6e et +
Coefficient	2	1.81	1.61	1.41	1.21	1

Coefficient 3 : La distance entre le domicile et l'établissement secondaire Ce facteur reflète la distance à vol d'oiseau entre le domicile de l'enfant et l'établissement secondaire souhaité. Le coefficient est d'autant plus grand que l'établissement est proche, comme indiqué dans le tableau 1.3.

TABLE 1.3 – Coefficient 3 : Distance avec l'établissement secondaire

Proximité	1ère	2e	3e	4e	5e	6e et +
Coefficient	1.98	1.79	1.59	1.39	1.19	1

Ces deux critères basés sur la distance ont pour but d'encourager l'inscription des enfants dans une école proche de leur quartier [3]. Notons également que l'ordre de proximité est déterminé en comparant les écoles appartenant à un même réseau (confessionnel ou non confessionnel).

Coefficient 4 : La distance entre l'école primaire et l'établissement secondaire Ce coefficient est calculé sur base des deux précédents selon le tableau 1.4. Ce critère a pour effet de rehausser la valeur totale de l'indice composite dans le cas où l'établissement secondaire sélectionné se trouve loin du domicile, mais proche de l'école primaire qui était fréquentée. Le but poursuivi est de prendre en compte un éventuel choix des parents comme, par exemple, la sélection d'un établissement proche d'un lieu de travail [3].

TABLE 1.4 – Coefficient 4 : Distance entre les deux établissements

		Coefficient 3					
		1.98	1.79	1.59	1.39	1.19	1
Coefficient 2	2	1	1.054	1.108	1.162	1.216	1.27
	1.81	1.054	1.108	1.162	1.216	1.27	1.324
	1.61	1.108	1.162	1.216	1.27	1.324	1.378
	1.41	1.162	1.216	1.27	1.324	1.378	1.432
	1.21	1.216	1.27	1.324	1.378	1.432	1.486
	1	1.27	1.324	1.378	1.432	1.486	1.54

Coefficient 5 : L'immersion Un bonus de 1.18 est accordé si l'élève souhaite s'inscrire en immersion. Il faut cependant remplir la condition d'être déjà inscrit en immersion dans la même langue depuis la troisième année primaire au moins afin de bénéficier de ce coefficient bonus. Les élèves ne remplissant pas cette condition ou ne souhaitant pas s'inscrire en immersion reçoivent un coefficient neutre de 1.

Coefficient 6 : L'offre scolaire dans la commune de l'école primaire fréquentée Un coefficient de 1.51 est appliqué si des établissements secondaires confessionnels et non confessionnels sont présents dans la commune de l'école primaire fréquentée. Si aucun établissement secondaire n'est présent dans la commune de l'école primaire fréquentée, ou seulement des établissements confessionnels, ou seulement des établissements non confessionnels, alors un coefficient neutre de 1 est appliqué.

Coefficient 7 : Le partenariat pédagogique Ce dernier critère n'est d'application que dans le cas où le coefficient concernant l'offre scolaire vaut 1 (il s'agit donc d'un cas où les établissements secondaires présents dans la commune appartiennent au même réseau). Dans ce cas, un coefficient de 1.51 est accordé si deux conditions sont réunies :

- L'établissement secondaire concerné a conclu un partenariat pédagogique avec au moins 3 écoles primaires.
- L'élève vient d'une école primaire remplissant l'une des trois conditions suivantes :
 - L'école primaire est l'une de celles qui ont conclu le partenariat pédagogique avec l'établissement secondaire souhaité.
 - L'école primaire n'a conclu aucun partenariat pédagogique.
 - L'école primaire a conclu une convention avec un autre établissement secondaire mais l'école primaire d'origine de l'élève est la plus proche du domicile ou la date de conclusion de la convention de partenariat est postérieure à l'inscription de l'élève en primaire.

Si ce n'est pas le cas, le coefficient du partenariat pédagogique sera de 1. La procédure de calcul est résumée à la figure 1.1.

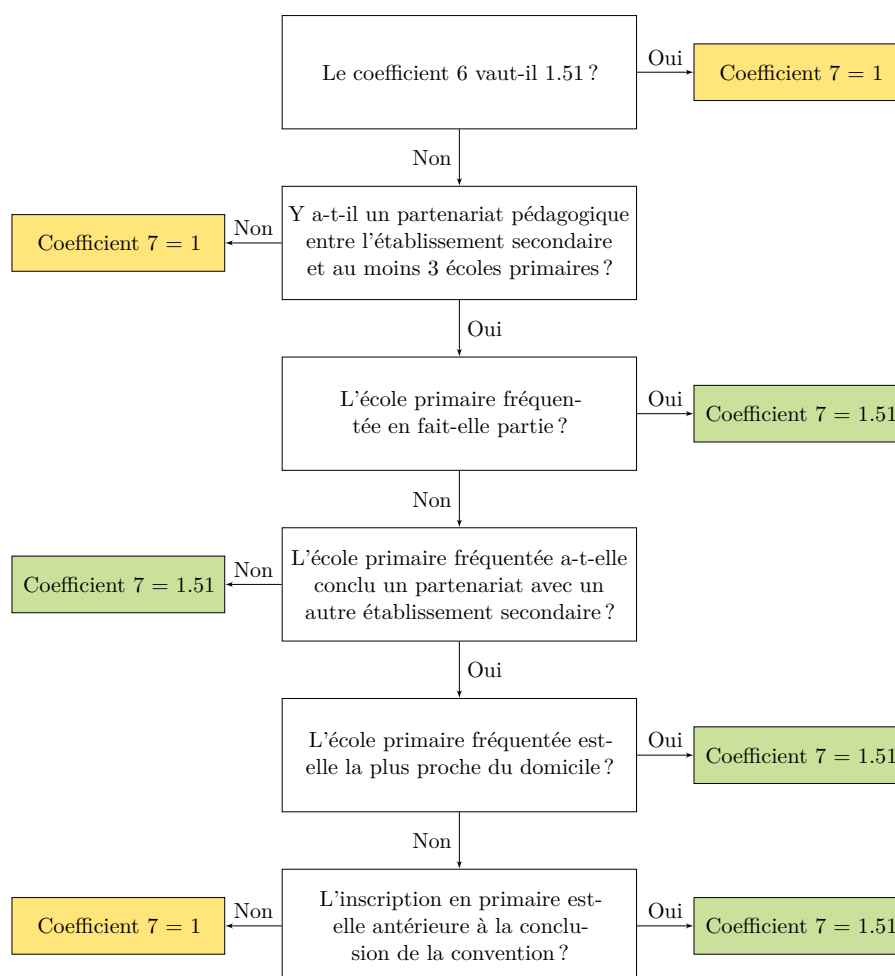


FIGURE 1.1 – Calcul du coefficient 7 (partenariat pédagogique)

1.1.2 Réforme du décret en 2022

Une nouvelle version du décret Inscription a été approuvée par le Parlement de la Fédération Wallonie-Bruxelles le 12 janvier 2022 [5]. Certaines nouveautés de la réforme entreront en vigueur dès la rentrée 2022. Cela concerne l’inscription dans les établissements présumés incomplets³, pour lesquels l’inscription sera validée plus rapidement. Par ailleurs, les élèves ayant suivi une première année différenciée et souhaitant s’inscrire en première année commune pourront également participer à la procédure d’inscription avec les mêmes priorités que les élèves de sixième primaire.

Les autres nouveautés de la réforme entreront en vigueur lors de la rentrée 2023. Cela concerne la création de la CoGI (Commission de Gouvernance des inscriptions), la création des ILI (Instance Locale des inscriptions), la modification du calcul de l’indice composite, la modification des conditions pour l’octroi d’une priorité, et la modification des classements des écoles et de la CoGI. [6]

La modification du calcul de l’indice composite comprend elle-même une modification de la pondération du coefficient 2, la création d’un coefficient 8 lié à l’encadrement différencié, et une modification du coefficient 6 dès 2025. La modification qui nous intéresse ici est celle de la pondération du critère 2 concernant la distance entre le domicile de l’enfant et l’école primaire qu’il fréquente. En effet, l’importance de ce critère devrait être sensiblement diminuée, donnant ainsi plus d’importance à la distance avec l’établissement secondaire [7]. Les nouvelles valeurs de ce coefficient sont présentées dans le tableau 1.5.

TABLE 1.5 – Mise à jour du Coefficient 2 : Distance avec l’école primaire

Proximité	1ère	2e	3e	4e	5e	6e et +
Coefficient	1.3	1.24	1.18	1.12	1.06	1

1.2 Motivations

Des coefficients souvent contestés Les coefficients utilisés par ce décret Inscription ont fait l’objet de nombreuses critiques de la part des parents d’élèves, et ce depuis sa mise en place [8, 9]. Les coefficients concernant les distances géographiques sont les plus concernés. En effet, il est reproché à la distance à vol d’oiseau de ne pas refléter le trajet effectivement parcouru par les élèves, que celui-ci soit effectué à pied ou en transports [3].

De plus, la trop grande prépondérance accordée à la distance domicile-école primaire pose également problème. Si ce dernier point a été amélioré par la réforme 2022, les distances à vol d’oiseau sont toujours utilisées pour trois des coefficients. Ceci est d’autant plus regrettable que les données concernant les transports en commun disponibles sont accessibles et exploitables.

La mobilité des jeunes Il est également important de souligner que la plupart des parents sont encore sur leur lieu de travail à l’heure où les cours se terminent (typiquement

3. Un établissement est « présumé incomplet » s’il n’a pas été complet pendant les 3 années précédentes, et que le nombre de places disponibles n’a pas diminué dans cet intervalle de temps. De plus, le nombre de demandes reçues ne peut pas non plus avoir excédé ce nombre de places.

aux alentours de 16h, même si cela varie d’une école à l’autre), et sont donc dans l’incapacité d’aller chercher leur enfant à ce moment-là. Les élèves entrant en enseignement secondaire sont cependant en âge de se déplacer seuls dans les transports en commun sur un trajet qu’ils connaissent (ce qui est le cas du trajet entre leur école et leur domicile).

Les résultats de l’enquête MONITOR [10] (une enquête menée par le Service public fédéral Mobilité et Transports en collaboration avec l’Institut Belge pour la Sécurité Routière et portant sur la mobilité des Belges menée en 2017) nous indiquent d’ailleurs que plus de la moitié des déplacements chez les moins de 18 ans concernent un trajet domicile-école. 21 % de ces déplacements sont effectués en transports en commun et durent en moyenne 20 minutes, pour une distance de 8 km. La mise en valeur de l’accessibilité d’une école plutôt que sa proximité est donc essentielle pour les familles.

La question écologique Une autre source de motivation pour cette étude est bien sûr l’argument écologique. Depuis un certain nombre d’années déjà, l’on s’inquiète de l’impact que peuvent avoir les trajets effectués en voiture sur les changements climatiques et l’utilisation des transports en commun est de plus en plus encouragée. La part de trajets domicile-école effectuée en voiture est de 35 %, ce qui reste supérieur à la proportion de trajets effectuée en transports en commun [10]. Ainsi, une modification du décret afin de mettre en valeur l’accessibilité des écoles pourrait permettre de promouvoir une plus grande utilisation des transports publics pour se rendre à l’école.

Pour toutes ces raisons, il nous semble important d’étudier et de quantifier l’impact pour la population des critères actuels du décret Inscription ainsi que de proposer des alternatives permettant d’améliorer le décret. Prendre en compte la durée des trajets en transports en commun dans le calcul de l’indice composite, afin de favoriser les écoles les plus accessibles nous paraît tout à fait pertinent et permettrait à la fois de faciliter la vie des familles et d’encourager l’utilisation des transports publics.

1.3 Sources de données

Les sources de données que nous avons utilisées dans ce travail sont toutes disponibles en *open source*, certaines nécessitant toutefois d’introduire une demande auprès de la société concernée. Cette section présente une vue d’ensemble de ces données ainsi que leur description.

1.3.1 Les données GTFS

Le format GTFS (General Transit Feed Specification) sera décrit plus en détail dans le chapitre 2 dans une section dédiée. Il s’agit d’informations fournies par tous les opérateurs belges de transports en commun et décrivant les informations géographiques de leurs services ainsi que les horaires associés. Ainsi, les données sont composées d’une collection de fichiers textes contenant tous les arrêts des opérateurs belges avec leur nom et emplacement, les lignes desservies ainsi que les horaires de passage des véhicules [11, 12, 13, 14].

Nous avons utilisé ces données afin de calculer tous les temps de trajet entre les écoles de la Fédération Wallonie-Bruxelles et toutes les stations de Belgique (voir la description de notre méthode au chapitre 3).

1.3.2 Le fichier FASE

Le « Fichier signalétique des établissements d'enseignement de la Fédération Wallonie-Bruxelles » (FASE) est publié par la Fédération Wallonie-Bruxelles. Il est disponible librement au format JSON (JavaScript Object Notation) et contient diverses informations à propos des établissements scolaires recensés, telles que le numéro FASE (utilisé par l'administration) de l'établissement, son nom, son adresse et ses coordonnées géographiques, son genre, son réseau, son niveau, son pouvoir organisateur, etc [15].

Ces données nous ont donc permis de sélectionner les écoles que nous voulions prendre en compte (enseignement ordinaire) ainsi que de récupérer leur nom, leur position et leurs caractéristiques (nom, réseau, type d'enseignement).

1.3.3 Les unités administratives

L'Administration générale de la Documentation patrimoniale (AGDP) publie un ensemble de cartes des différentes limites administratives belges. Elles incluent les communes, les arrondissements, les provinces, les régions et le pays. Chaque zone contient son nom dans toutes les langues nationales, ainsi que la forme de la zone [16].

Nous avons utilisé ces données pour identifier la région de chaque école et de chaque arrêt de transports en commun, ainsi que pour dessiner les limites des régions et provinces sur les cartes que nous avons générées.

1.3.4 La grille de répartition de la population belge

La « Répartition de la population sur une grille de cellules de 1 km² » est une carte divisant la Belgique en cellules carrées de 1 km de côté. Chaque cellule est accompagnée de sa population totale, des populations masculine et féminine, ainsi que des groupes de population de 0 à 14 ans, de 15 à 64 ans et de plus de 65 ans. Le recensement date de 2018 [17].

Nous avons utilisé ces données pour calculer la population impactée par des temps de trajet supplémentaires pour revenir de l'établissement scolaire (voir la description de l'hypothèse de Voronoï).

1.4 Objectifs du mémoire

Ce travail de fin d'études a pour objectif de mettre à profit les données décrites précédemment afin d'étudier différentes questions :

- L'accessibilité des écoles en transports en commun.
- La pertinence de l'indice composite utilisé par le décret Inscription. Plus précisément, nous souhaitons explorer la discordance entre distance à vol d'oiseau et temps de parcours en transports en commun.
- La fraction de population concernée par cette discordance entre les critères ainsi que les conséquences qui en découlent (trajet supplémentaire, impact sur l'indice composite, etc.).

Au-delà de cette analyse, l'un des objectifs de ce travail est de proposer des voies d'amélioration pour le décret Inscription (modification de coefficients existants, ajout d'un nouveau coefficient, etc.).

Enfin, l'un des principaux objectifs du mémoire est, bien sûr, la communication de nos résultats aux instances politiques concernées, afin de leur faire prendre conscience des problèmes observés et de discuter de possibles solutions. Ceci inclut la rédaction d'un premier rapport qui avait été envoyé aux cabinets du Ministre-Président de la Fédération Wallonie-Bruxelles et de la Ministre de l'Éducation en février 2022. Celui-ci est disponible dans l'annexe B. Nous avons également eu l'occasion de discuter de ces résultats avec un conseiller du ministère (voir la discussion au chapitre 5). Par ailleurs, nous avons aussi créé un outil permettant de visualiser l'effet d'un changement de coefficients sur la valeur de l'indice composite.

Chapitre 2

État de l’art et rappels théoriques

Dans ce chapitre, nous présenterons les notions théoriques importantes utilisées dans le cadre de notre travail. Par ailleurs, ce chapitre sera également l’occasion de présenter un résumé de l’état de l’art sur le sujet. Cependant, la littérature autour des données GTFS reste à ce jour plutôt limitée et cette partie sera par conséquent assez générale. Par ailleurs, il n’y a, à notre connaissance, pas de publication sur une application de l’utilisation des données GTFS concernant le décret Inscription, et l’étude de l’existant ne servira donc qu’aux parties « représentation du réseau » et « calcul des plus courts chemins » de notre démarche (voir descriptions au chapitre 3).

2.1 Le format GTFS

Comme brièvement discuté dans le chapitre précédent, le format GTFS (*General Transit Feed Specification*) est le format standard utilisé par les sociétés de transports en commun pour publier, entre autres, les horaires des voyages proposés [18, 19]. En Belgique, tous les opérateurs de transports en commun utilisent ce format GTFS. Il existe deux types de formats GTFS : le format GTFS statique et le format GTFS en temps réel.

2.1.1 GTFS statique

Le format GTFS *statique* contient les horaires théoriques des transports en commun ainsi que les informations géographiques qui y sont associées (emplacement des arrêts et des lignes, horaires de départ et d’arrivée aux arrêts, etc.). Ces données statiques sont fournies sous la forme de plusieurs fichiers textes rassemblés dans un fichier ZIP. Le format GTFS requiert au moins 6 fichiers différents dont le lien est illustré à la figure 2.1. Par ailleurs, l’ajout de certains fichiers facultatifs peut permettre d’apporter plus de précisions sur les services proposés [20] (voir les fichiers bleus sur la figure 2.1).

Ces fichiers statiques sont ceux que nous avons utilisés dans le cadre de cette étude.

2.1.2 GTFS en temps réel

Le format GTFS-realtime est une extension permettant de fournir des mises à jour en temps réel de l’état du réseau. Ceci inclut les mises à jour des trajets (retards, annulations et itinéraires modifiés), les alertes de service (arrêts déplacés et événements imprévus

affectant une station, un itinéraire ou l'ensemble du réseau), et la position du véhicule (informations sur les véhicules, y compris leur localisation et la densité du trafic).

Ces données n'ont toutefois pas été exploitées dans ce travail.

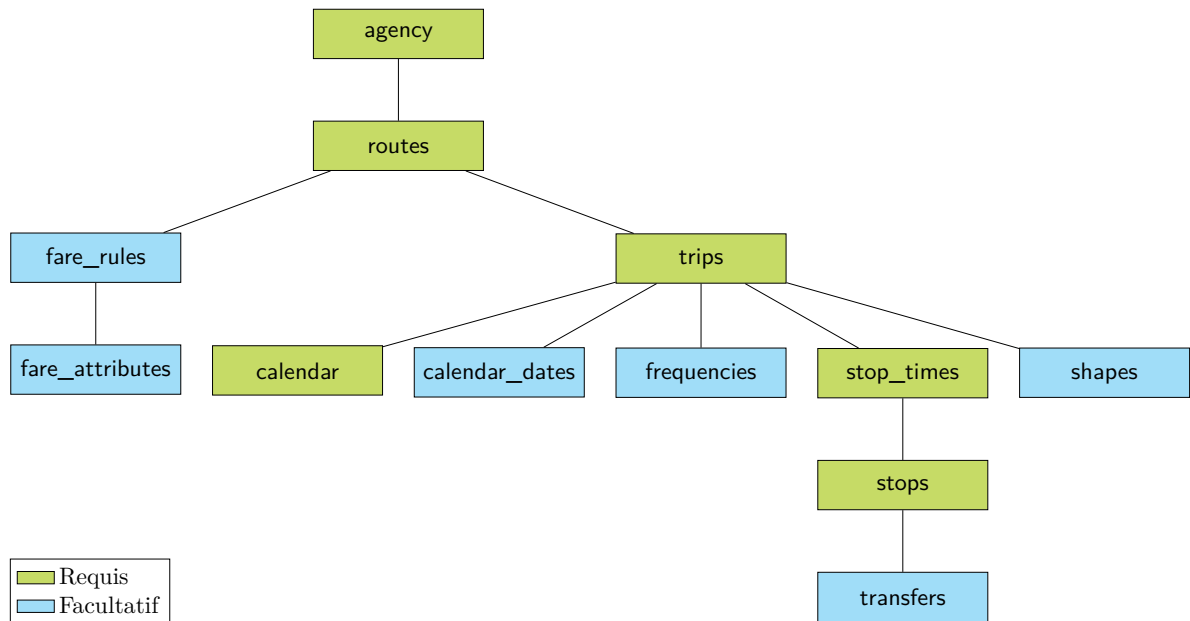


FIGURE 2.1 – Structure du format GTFS (tiré de [20])

2.2 Représentation du réseau de transport

Représenter le réseau de transports en commun sous forme d'un graphe permet de traduire la question de savoir « à quelle heure un utilisateur peut-il arriver au point *A* en partant du point *B* à l'heure *H* ? » en un problème des plus courts chemins. Il existe plusieurs alternatives pour une telle représentation. Dans la littérature, deux grandes approches sont typiquement utilisées : le *time-expanded model* et le *time-dependant model* [21, 22].

2.2.1 Time-expanded model

Le principe de cette approche est de représenter le réseau par un graphe statique, dans lequel chaque nœud correspond à un événement précis (une station à un horaire).

Ainsi, une même station est représentée par plusieurs nœuds dans le graphe. Les arêtes peuvent représenter une ligne de transport reliant deux arrêts, mais aussi un temps d'attente à un même arrêt. Le graphique ainsi construit est donc considérablement plus grand que pour une approche utilisant un graphe dynamique. D'un autre côté, les nœuds ont un degré bien inférieur.

2.2.2 Time-dependent model

Cette seconde approche modélise le réseau de transport comme un graphe dynamique, c'est-à-dire que les poids des arêtes de ce graphe dépendent du temps. Ainsi, contrairement

au cas du premier modèle, chaque station du réseau n'est représentée qu'une seule fois. Il est possible d'attendre à un nœud du graphe. Cela entraîne un certain coût qui est également dépendant du temps (nous pouvons par exemple imaginer qu'attendre à un arrêt pour lequel le prochain train passe 3 h plus tard correspond à un poids élevé, contrairement au cas où le temps d'attente serait de quelques minutes).

Une fois un tel graphe obtenu, le problème du calcul des temps de trajet revient à un problème des plus courts chemins, que nous explorons dans la section suivante.

2.3 Le problème des plus courts chemins

Il existe principalement deux variantes du problème : le *single-source shortest paths*, correspondant donc à une recherche de trajet optimal depuis une seule source vers tous les autres nœuds du graphe, et le *all-pair shortest paths*, où l'on s'intéresse au plus court chemin entre chaque paire de nœuds possible. Les algorithmes utilisés pour chacun des cas sont différents.

Pour la suite de cette section, nous considérons un graphe G , avec

- V , l'ensemble de ses nœuds
- E , l'ensemble de ses arêtes

2.3.1 Single-source shortest paths

Dijkstra

L'algorithme de Dijkstra (voir algorithme 1) est typiquement utilisé en théorie des graphes pour résoudre le problème des plus courts chemins entre un nœud-source et tout autre nœud du graphe. Il s'applique à des graphes orientés ou non, et pondérés par des poids positifs [23, 24].

Dans cet algorithme, on part d'un nœud de départ, dont la distance par rapport au point de départ (lui-même) est nulle. Tous les autres nœuds ont une distance infinie, exprimant qu'ils sont inaccessibles depuis le point source. Les nœuds qui sont accessibles et qui n'ont pas encore été traités par l'algorithme sont stockés dans une *priority queue* en fonction de leur distance depuis le nœud-source. À chaque itération, la distance depuis le nœud-source jusque chaque voisin du nœud considéré est calculée en additionnant le poids de l'arête à la distance déjà calculée vers le nœud considéré. Chaque voisin est donc ajouté à la *priority queue*. Ensuite, le nœud considéré est mis à jour sur le nœud le plus proche du nœud-source, c'est-à-dire le premier nœud de la file de priorité.

Cette itération est répétée jusqu'à l'arrivée au nœud-destination, ou jusqu'à épuisement de tous les nœuds accessibles si le nœud-destination n'en fait pas partie.

L'une des limitations de cet algorithme est qu'il ne prend pas en charge les graphes contenant des poids négatifs. En effet, la présence d'une arête négative peut empêcher l'algorithme de Dijkstra de trouver le chemin le plus court à cause de son approche gloutonne du problème. Une fois qu'un nœud a été atteint, il n'est plus possible de modifier sa valeur, même si un autre chemin permet ensuite d'y arriver avec un coût moindre. Ce cas ne peut se produire qu'avec des arêtes à poids négatif.

Algorithme 1 : Dijkstra (problème des plus courts chemins depuis une source)

Data : G a weighted graph
 s the source node

Result : $dist$ the vector containing the distances between s and all nodes of G

$dist \leftarrow |V| \times 1$ array initialised to $+\infty$

$dist[s] \leftarrow 0$

$Q \leftarrow$ priority queue containing all vertices of G sorted according to their $dist$

while Q not empty **do**

$u \leftarrow$ vertex from Q with min $dist[u]$

$Q \leftarrow Q \setminus u$

foreach neighbour v of u **do**

if $v \in Q$ **then**

if $dist[v] > dist[u] + weight(u, v)$ **then**

$dist[v] \leftarrow dist[u] + weight(u, v)$

Complexité L'algorithme de Dijkstra possède une complexité temporelle en $O((|E| + |V|) \log |V|)$, et une complexité spatiale en $O(V)$.

Variation Dans le cas particulier où le graphe est acyclique (il s'agit alors d'un DAG, *directed acyclic graph*), on peut utiliser une variation simplifiée de l'algorithme de Dijkstra. En effet, il n'y a plus de risque de créer un circuit en parcourant le graphe, et la *priority queue* n'est alors plus nécessaire. Il suffit de parcourir le graphe dans l'ordre chronologique depuis le nœud-source pour trouver la distance minimale vers tous les autres nœuds du graphe en un seul passage.

Complexité La complexité temporelle devient alors du $O(|E| + |V|)$.

Bien sûr, Dijkstra n'est pas le seul algorithme existant pour calculer le chemin le plus court entre une source et un point d'arrivée.

Bellman-Ford

Pour un problème de type *single-source*, outre l'algorithme de Dijkstra, valable dans les graphes dont tous les poids sont positifs, on peut citer l'algorithme de Bellman-Ford (algorithme 2), qui fonctionne également dans les graphes dont les poids sont quelconques, à condition qu'il n'y ait pas de circuit absorbant (c'est-à-dire de circuit de poids négatif) [25, 26, 27]. Plutôt que d'utiliser une *priority queue* comme l'algorithme de Dijkstra, celui de Bellman-Ford essaye de placer chaque arête du graphe si elle permet une diminution de la distance. Cela est répété $|V| - 1$ fois, les distances ainsi calculées diminuent à chaque itération.

Complexité L'algorithme de Bellman-Ford a une complexité temporelle en $O(|E| |V|)$, et une complexité spatiale en $O(|V|)$. La complexité plus élevée est due au fait que

l'algorithme de Bellman-Ford effectue une vérification de l'existence de circuits absorbants.

Algorithme 2 : Bellman-Ford (*single-source*)

Data : G a weighted graph
 s the source node

Result : $dist$ the vector containing the distances between s and all nodes of G

$dist \leftarrow |V| \times 1$ array initialised to $+\infty$
 $dist[s] \leftarrow 0$

repeat $|V| - 1$ **times**

foreach $edge(u, v, w)$ **in** G **do**

if $dist[v] > dist[u] + w$ **then**

$dist[v] \leftarrow dist[u] + w$

foreach $edge(u, v, w)$ **in** G **do**

if $dist[v] > dist[u] + w$ **then**

$negativeloop \leftarrow [v, u]$

repeat $|V| - 1$ **times**

$u \leftarrow negativeloop[0]$

foreach $edge(u, v, w)$ **in** G **do**

if $dist[v] > dist[u] + w$ **then**

$negativeloop \leftarrow [v] + negativeloop$

if *cycle in negativeloop* **then**

error "Graph contains a negative-weight cycle"

2.3.2 All-pair shortest paths

Pour un problème de type *all-pairs*, on peut citer les algorithmes de Floyd-Warshall et de Johnson.

Floyd-Warshall

L'algorithme de Floyd-Warshall (algorithme 3) est utilisable sur des graphes orientés avec des poids quelconques, mais il ne peut y avoir de circuit absorbant. Cet algorithme tente d'insérer chaque nœud entre deux autres pour voir si un chemin passant par ce nœud serait plus court que celui n'y passant pas [28, 29, 30].

Complexité Il a une complexité temporelle de $O(|V|^3)$, et une complexité spatiale de $O(|V|^2)$.

Algorithme 3 : Floyd-Warshall (*all-pairs*)

Data : G a weighted graph
Result : $dist$ the array containing the distances between all pairs of nodes
 $dist \leftarrow |V| \times |V|$ array initialised to $+\infty$
foreach $edge (u, v, w)$ **in** G **do**
 $dist[u][v] \leftarrow w$
foreach $vertex v$ **in** G **do**
 $dist[v][v] \leftarrow 0$
for $k \in 1..|V|$ **do**
 for $i \in 1..|V|$ **do**
 for $j \in 1..|V|$ **do**
 if $dist[i][j] > dist[i][k] + dist[k][j]$ **then**
 $dist[i][j] \leftarrow dist[i][k] + dist[k][j]$

Johnson

L'algorithme de Johnson (algorithme 4) est également utilisable sur des graphes orientés avec des poids quelconques sans circuit absorbant. Cet algorithme utilise l'algorithme de Bellman-Ford pour transformer le graphe de départ en lui retirant ses poids négatifs. L'algorithme de Dijkstra peut alors être appliqué sur le graphe ainsi repondéré [31].

Algorithme 4 : Johnson (*all-pairs*)

Data : G a weighted graph with w the weights
Result : $dist$ the array containing the distances between all pairs of nodes in the graph
1. add a vertex q in the graph and connect it to all other vertices with 0-weighted edges
2. apply Bellman-Ford algorithm starting from vertex q to find for each vertex v the minimum weight $h(v)$ from q to v
3. recompute weights : $w(u, v) \leftarrow w(u, v) + h(u) - h(v)$
4. remove q and apply Dijkstra algorithm from each vertex s to every other vertex
5. add $h(v) - h(u)$ to each distance $dist(u, v)$ found with Dijkstra

Complexité Il a une complexité de $O(|V|^2 \log |V| + |V||E|)$, ce qui fait qu'il est plus rapide que celui de Floyd-Warshall sur les graphes creux.

2.3.3 Comparaison des méthodes

Le tableau 2.1 résume les caractéristiques des algorithmes présentés précédemment. L'approche qui sera utilisée pour le calcul des temps de trajet est indiquée en couleur.

TABLE 2.1 – Tableau récapitulatif des algorithmes

Algorithme	Type	Poids négatifs	Complexité temporelle	Complexité spatiale
Dijkstra	<i>single-source</i>	Non	$O((E + V) \log V)$	$O(V)$
Shortest path in DAG	<i>single-source</i>	Oui	$O(E + V)$	
Bellman-Ford	<i>single-source</i>	Oui	$O(V E)$	$O(V)$
Floyd-Warshall	<i>all-pairs</i>	Oui	$O(V ^3)$	$O(V ^2)$
Johnson	<i>all-pairs</i>	Oui	$O(V ^2 \log V + V E)$	

2.4 Voronoï

2.4.1 Diagramme de Voronoï

« En mathématiques, un diagramme de Voronoï est un pavage (découpage) du plan en cellules (régions adjacentes) à partir d'un ensemble discret de points appelés "germes". Chaque cellule enferme un seul germe, et forme l'ensemble des points du plan plus proches de ce germe que d'aucun autre. » [32]

Définition Soit S un ensemble de n sites de l'espace euclidien en dimension d . Pour chaque site p de S , la cellule de Voronoï $V(p)$ de p est l'ensemble des points de l'espace qui sont plus proches de p que de tous les autres sites de S . Le diagramme de Voronoï de $V(S)$ est la décomposition de l'espace formé par les cellules de Voronoï des sites.

Ainsi, sur la figure 2.2, chaque région de Voronoï, délimitée par les lignes bleues, correspond à l'ensemble des points du plan qui sont les plus proches des centres rouges.

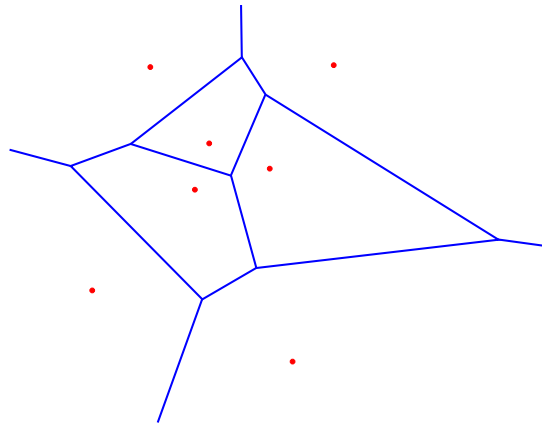


FIGURE 2.2 – Diagramme de Voronoï [33]

2.4.2 Triangulation de Delaunay

« En mathématiques et plus particulièrement en géométrie algorithmique, la triangulation de Delaunay d'un ensemble P de points du plan est une triangulation $DT(P)$ telle qu'aucun point de P n'est à l'intérieur du cercle circonscrit d'un des triangles de $DT(P)$. » [34]

Définition La triangulation de Delaunay de S est le dual géométrique du diagramme de Voronoï de S : deux points de S sont reliés par une arête dans la triangulation de Delaunay si et seulement si leurs cellules sont adjacentes dans le diagramme de Voronoï de S .

Ainsi, sur la figure 2.3, chaque triangle de Delaunay, délimité par les lignes noires, est inscrit dans un cercle gris, dont le centre est marqué par un point rouge. Aucun point noir ne se trouve dans d'autres cercles que ceux qui circonscrivent les triangles dont il fait partie. La figure 2.4 illustre le lien avec les diagrammes de Voronoï.

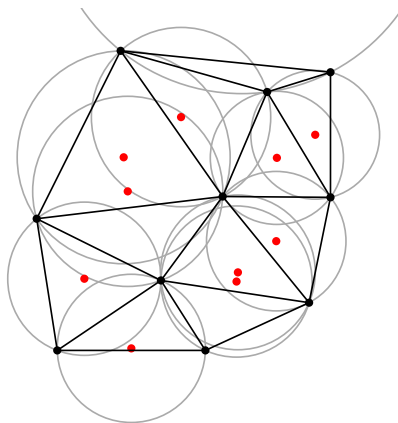


FIGURE 2.3 – Triangulation de Delaunay

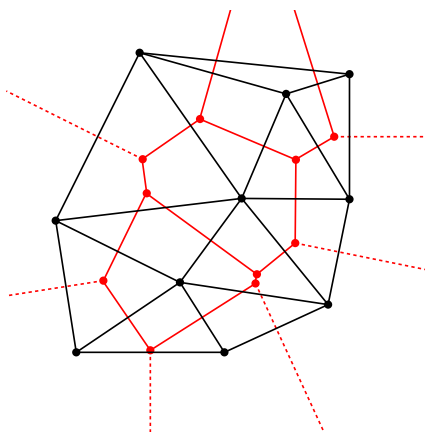


FIGURE 2.4 – Correspondance entre le diagramme de Voronoï (en rouge) et la triangulation de Delaunay (en noir)

2.4.3 Méthode de calcul

Il existe plusieurs manières de calculer les zones de Voronoï.

Algorithme de Green et Sibson

Il s'agit d'un algorithme incrémental. Il se base sur une propriété intéressante des diagrammes de Voronoï : l'ajout ou la suppression d'un point ne nécessite pas la modification complète de la triangulation. L'insertion d'un nouveau point ne modifie donc le diagramme que localement [35].

Avec cette méthode, tous les points sont ajoutés un à un à un diagramme initialement vide, pour obtenir un diagramme de Voronoï complet.

Cet algorithme possède une complexité temporelle de $O(n^2)$.

Algorithme de Shamos et Hoey

Il s'agit d'un algorithme de type « diviser pour régner ». Il se base sur une méthode récursive, consistant à séparer l'ensemble des points en deux, calculer les deux diagrammes, puis les réassembler [36].

La difficulté de cet algorithme réside dans la fusion entre les deux diagrammes ainsi calculés. Pour établir la nouvelle séparation des zones précédemment obtenues, il faut ainsi suivre les médiatrices des arêtes reliant les deux sous-graphes. Cette opération est assez difficile à implémenter, et la méthode est donc très peu utilisée en pratique.

Cet algorithme possède une complexité temporelle de $O(n \log n)$.

Algorithme de Fortune

Il s'agit d'un algorithme de type *sweep line*. Il balaie l'ensemble des points dans une dimension en mettant à jour le diagramme au fur et à mesure qu'il le parcourt [37].

Lorsque la droite D parcourant le diagramme rencontre un point p , la zone plus proche du point que de la droite forme une parabole dont p est le foyer et D la directrice. Lorsque d'autres points sont rencontrés, d'autres paraboles sont créées, et les intersections entre ces paraboles déterminent les frontières entre les zones de Voronoï.

Ces paraboles correspondent en fait à l'intersection du plan avec des cônes dans une dimension supplémentaire. Ces cônes ont pour effet d'anticiper les modifications du diagramme qui n'ont pas encore été rencontrées par l'algorithme.

Cet algorithme possède une complexité temporelle de $O(n \log n)$ et une complexité spatiale de $O(n)$.

Construction duale : triangulation de Delaunay

Une autre manière de calculer les zones de Voronoï consiste à calculer la solution du problème dual, puis transformer la solution obtenue pour revenir au problème initial.

Problème dual oblige, des méthodes similaires existent pour le résoudre (méthodes incrémentale, diviser pour régner, *sweep line*).

Il est également possible de créer une triangulation quelconque, et de la transformer en triangulation de Delaunay en changeant la direction d'une arête séparant deux triangles, jusqu'à l'obtention d'une triangulation valide.

Il est encore possible de transformer le problème qui consiste à trouver la triangulation de Delaunay en un autre problème, à savoir le calcul de l'enveloppe convexe. Pour cela, il faut ajouter une dimension valant la somme des carrés des coordonnées pour chaque point. Le fond de l'enveloppe convexe, une fois la coordonnée supplémentaire retirée, donne la triangulation de Delaunay du problème. Il existe de nombreux algorithmes permettant de trouver l'enveloppe convexe, que nous ne décrivons pas ici.

2.5 Autres considérations théoriques

2.5.1 Modélisation de la marche à pied

Le temps de marche à pied lorsque celle-ci intervient dans le trajet constitue également un point important à modéliser. Pour calculer ce temps, la méthode la plus correcte serait de le calculer en tenant compte du réseau routier pour trouver l'itinéraire qui permet de relier le point de départ au point d'arrivée. Connaissant la distance réelle de marche, et avec une estimation de la vitesse de marche à pied, une valeur assez précise du temps de marche à pied peut être établie. Mais cela n'est pas toujours réaliste d'effectuer ces calculs sur le réseau routier complet. Pour simplifier cela, il est possible de trouver une estimation du temps de marche à pied qui est uniquement fonction de la distance à vol d'oiseau entre le point de départ et celui d'arrivée.

Une telle analyse du détour moyen a été effectuée par Héran : le détour moyen effectué par les piétons par rapport au trajet à vol d'oiseau est d'environ 20 % en centre-ville, et d'environ 30 % en périphérie [38].

2.5.2 Calcul de distances à vol d'oiseau

La formule de haversine permet de déterminer la distance orthodromique entre deux points, c'est-à-dire la distance la plus courte entre ces points sur une sphère, à partir de leurs longitude et latitude. Cette formule permet donc de calculer les distances à vol d'oiseau entre différents points en prenant en compte la courbure de la Terre. La distance est calculée selon la formule suivante [39] :

$$d = 2r \arcsin \left(\sqrt{\sin^2 \left(\frac{\phi_2 - \phi_1}{2} \right) + \cos(\phi_1) \cos(\phi_2) \sin^2 \left(\frac{\lambda_2 - \lambda_1}{2} \right)} \right)$$

où d est la distance du grand cercle entre les deux points (en km), r est le rayon de la sphère (donc ici, de la Terre : 6367 km), ϕ_1 et ϕ_2 sont les latitudes des points 1 et 2 (en radians), λ_1 et λ_2 sont les longitudes des points 1 et 2 (en radians).

2.5.3 Recherche du voisin le plus proche

La recherche des voisins les plus proches nécessite d'avoir des algorithmes performants pour éviter de devoir tester tous les nœuds un à un pour trouver les plus proches. Une manière efficace pour ce genre de recherche est d'utiliser un *ball tree*. Il s'agit d'une structure de données dans laquelle les points sont placés dans un arbre binaire en fonction de la distance qui les sépare. Les points sont rassemblés dans des hypersphères (appelées boules) de plus en plus petites imbriquées les unes dans les autres, avec une boule à chaque nœud de l'arbre. La précision des boules augmente donc au fur et à mesure que l'on descend dans l'arbre. Cela permet de rechercher efficacement les voisins, sans avoir à effectuer une recherche par force brute qui est particulièrement inefficace lorsque le nombre de voisins à tester est important.

Un exemple de *ball tree* est visible à la figure 2.5. Sur la figure 2.5a se trouvent 9 points, numérotés de 1 à 9, et 7 boules, numérotées de A à G. Une première grande boule (A) contient tous les points. L'espace est ensuite subdivisé en boules de plus en plus petites. Chaque boule (sauf la première) ne contient des nœuds que d'une seule boule de taille supérieure. La figure 2.5b représente l'arbre qui organise les boules. Chaque nœud correspond à une boule. À la racine de l'arbre se trouve la plus grande boule, celle qui englobe tous les points. Au fur et à mesure que l'on descend dans l'arbre, les boules sont de plus en plus petites.

Lors de la recherche des voisins les plus proches, les points contenus dans la feuille de l'arbre du point-source sont testés pour trouver le plus proche, puis l'algorithme remonte l'arbre jusqu'à la racine. À chaque nœud en remontant l'arbre, l'algorithme évalue s'il est possible de trouver un point plus proche que le point déjà trouvé. Ceci est fait en calculant la distance entre le point-source et la boule au niveau du nœud considéré, et en comparant celle-ci avec la distance entre le point-source et le point le plus proche déjà trouvé. Si c'est le cas, une recherche similaire à celle qui a été faite initialement est exécutée en descendant sur les enfants du nœud concerné, jusqu'à trouver un point plus proche, ou jusqu'à ce que la boule d'un nœud ne permette plus de trouver un point plus proche que celui déjà trouvé. Sinon, la recherche continue en remontant l'arbre. Une fois que la racine du *ball tree* est atteinte, on a la certitude que le nœud trouvé est bien le nœud le plus proche du nœud-source.

Une généralisation de cette recherche est possible pour obtenir les k voisins les plus proches : il suffit, au lieu de ne garder que le nœud le plus proche, d'utiliser une *priority queue* limitée à k éléments dont la clé serait la distance du nœud-source.

Il est aussi possible de récupérer tous les nœuds dans un certain rayon : dans ce cas, tous les nœuds qui satisfont à la condition de proximité sont gardés. De nouveau, l'algorithme ne redescendra dans l'arbre que si la boule montre qu'il pourrait y avoir des nœuds satisfaisants. Le nombre de nœuds explorés peut alors être très inférieur au nombre total de points, augmentant les performances par rapport à une recherche par force brute. L'algorithme de cette recherche est visible dans la fonction `knnSearch`.

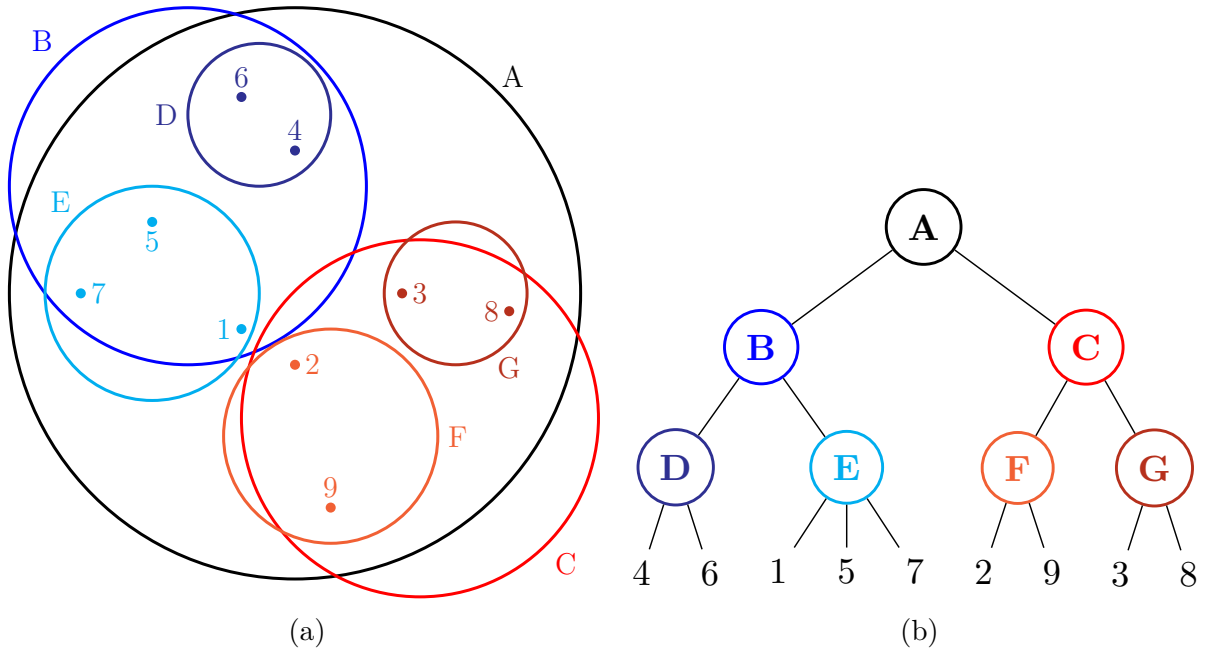


FIGURE 2.5 – Ball tree

Function knnSearch(t, k, Q, child)

Data : t the target point for the query
 k the number of nearest neighbours of t to search for
 Q max-first priority queue containing at most k points
 B a node, or ball, in the tree

Result : Q containing the k nearest neighbours from within B

if $\text{distance}(t, B.\text{pivot}) - B.\text{radius} \geq \text{distance}(t, Q.\text{first})$ **then**

return Q unchanged

else if B is a leaf node **then**

foreach point p in B **do**

if $\text{distance}(t, p) < \text{distance}(t, Q.\text{first})$ **then**

$Q \leftarrow Q \cup p$

if $\text{size}(Q) > k$ **then**

$Q \leftarrow Q \setminus Q.\text{first}$

else

$\text{child}_1 \leftarrow$ child node closest to t

$\text{child}_2 \leftarrow$ child node furthest from t

$Q \leftarrow \text{knnSearch}(t, k, Q, \text{child}_1)$

$Q \leftarrow \text{knnSearch}(t, k, Q, \text{child}_2)$

return Q

Chapitre 3

Notre approche

Dans ce chapitre, nous allons présenter les principales phases de calcul par lesquelles nous sommes passés pour générer nos résultats (ces derniers seront présentés dans le chapitre 4). Ceci inclut donc la construction du graphe représentant le réseau de transport, le calcul des temps de trajet entre chaque école et le domicile des élèves, ainsi que l'estimation de la population par arrêt ou encore les calculs effectués pour obtenir des mesures par km^2 sur la carte de la Belgique. Par ailleurs, nous présenterons également une vue d'ensemble sur l'organisation de notre code et sur son fonctionnement (des explications plus détaillées à ce sujet sont disponibles en annexe D). Enfin, ce chapitre contient une analyse des performances de notre code.

3.1 Représentation du réseau

Calculer les temps de trajet en transports en commun entre une école donnée et un point d'arrivée constitue une application du problème des plus courts chemins, décrit dans le chapitre 2. La première étape de ce travail consiste donc à construire un graphe représentant le réseau de transports en commun à partir des données GTFS fournies par les sociétés de transports en commun. La figure 3.1 schématise notre procédure. Nous avons utilisé un modèle de type *time-expanded* (défini dans la section 2.2). Ainsi, notre graphe est composé de nœuds représentant chacun une station à un horaire donné. Une même station peut donc être représentée plusieurs fois : le graphe contient autant de nœuds la concernant que d'heures d'arrivée ou de départ à cette station. Les arêtes du graphe correspondent à un trajet (à pied ou en transports en commun) entre deux stations et le poids de chacune d'entre elles est donc égal à la durée du trajet. Ces calculs sont détaillés dans les sections qui suivent. De plus, une arête peut également relier la même station à des temps différents, son poids correspond dans ce cas au temps d'attente à l'arrêt en question.

Le graphe ainsi construit est un graphe orienté et acyclique puisqu'il est évidemment impossible de remonter le temps. Notre graphe a été construit sur base des données GTFS concernant les trajets prévus le mardi 26 octobre 2021 pour la Belgique entière. Il contient au total 2 773 888 nœuds et 5 791 196 arêtes.

Données GTFS

stop_times.txt			
trip_id	arrival_time	departure_time	stop_id
trip1	10:15:00	10:15:00	id1
trip1	10:25:00	10:25:00	id2
trip1	10:27:00	10:27:00	id3

stops.txt				
stop_id	stop_code	stop_name	stop_lon	stop_lat
id1	Arr1	Arrêt 1	51.17	6.10
id2	Arr2	Arrêt 2	50.77	6.13



Graphe du réseau

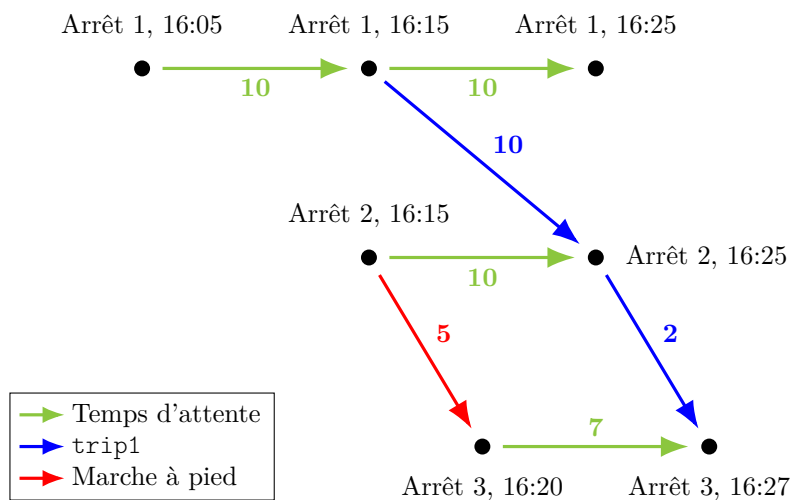


FIGURE 3.1 – Création du graphe du réseau de transport à partir des fichiers GTFS

3.1.1 Modélisation du temps

Comme expliqué dans la section 2.2, une représentation de type *time-expanded* utilise un graphe statique nécessitant une représentation de chaque lieu à différents intervalles de temps.

Ainsi, le temps discret dans notre réseau est représenté en dizaines de secondes. Les conversions en minutes et en heures seront effectuées plus tard lors de l'analyse de nos données. Le temps zéro correspond donc à minuit le 26 octobre 2021. De plus, puisque nos calculs ne s'étendent que sur une journée, la valeur de temps maximale est donc de 8640, ce qui correspond à minuit le jour suivant.

3.1.2 Calcul du poids des arêtes

Les temps de marche à pied Pour la modélisation du temps de marche à pied, il nous a semblé non réaliste de calculer celui-ci sur base du réseau routier, données que nous n'utilisons pas ailleurs dans ce travail. Nous avons donc suivi une estimation similaire à celle de Schorochoff [40], elle-même basée sur l'analyse effectuée par Héran [38], et détaillée à la sous-section 2.5.1. Ainsi, le temps de trajet à pied entre deux stations est estimé par le temps de parcours à vitesse réduite de la distance à vol d'oiseau qui les sépare.

La distance à vol d'oiseau entre deux points a été calculée à partir de leurs coordonnées en utilisant la formule haversine (voir sous-section 2.5.2).

Pour déterminer la vitesse réduite à utiliser, nous avons formulé deux hypothèses :

1. **La vitesse de marche des élèves** : Les piétons marchent à une vitesse d'environ 4 à 5 km/h. Sachant que nous nous intéressons ici à des trajets parcourus par des enfants, nous avons considéré la valeur la plus basse de cet intervalle.
2. **Le détour moyen** : Notre analyse concerne la totalité de la Fédération Wallonie-Bruxelles et englobe de ce fait à la fois des zones citadines (détour moyen : 30 %) et rurales (détour moyen : 20 %). Pour cette raison, nous avons choisi d'approximer le détour moyen effectué par les piétons comme étant de 25 % (la médiane des valeurs) par rapport au trajet à vol d'oiseau.

Ainsi, nous obtenons la vitesse de marche que nous avons finalement utilisée, qui vaut :

$$\frac{4 \text{ km/h}}{1.25} = 3 \text{ km/h}$$

Les trajets en transports en commun Le poids des arrêtes représentant un trajet en transports est directement calculé sur base des heures de départ et d'arrivée contenues dans les fichiers GTFS. Par exemple, à la figure 3.1, le temps de trajet entre l'arrêt 1 et l'arrêt 2 est de 10 minutes si l'on prend la ligne *trip1* (bien sûr, cette valeur est exprimée en dizaines de secondes dans notre graphe).

3.1.3 Déplacements autorisés

Afin de limiter la taille finale de notre graphe et d'éviter les trajets irréalistes, nous avons également imposé certaines limites sur les trajets autorisés. Ainsi, prendre une correspondance est autorisé si le trajet à pied entre les deux stations respecte certaines conditions :

1. Le trajet doit durer au maximum 10 minutes (l'arrêt à rejoindre se trouve donc dans un rayon de 500 m du point de départ). En effet, la distance parcourue à pied ne peut pas être trop importante, puisque les enfants n'ont pas la même endurance que les adultes.
2. Pour la même raison, les élèves ne peuvent partir que des arrêts se situant dans un rayon de 1.5 km à vol d'oiseau (30 min à pied) de leur école.
3. Afin d'éviter de se retrouver avec un trop grand nombre d'arrêts considérés dans les zones densément desservies, nous avons également limité le nombre de stations pouvant être atteintes à pied à 60 (voir à ce sujet la discussion à la section 3.3.2).

3.2 Calcul des plus courts chemins

Une fois le graphe construit, le calcul du temps des trajets peut être effectué. Étant donné les nombreuses possibilités de trajets, nous avons considéré que chaque élève cherche à minimiser son temps de trajet (seuls les temps de trajet minimaux entre deux points seront donc considérés). Il s’agit donc d’un problème des plus courts chemins et l’algorithme de Dijkstra, présenté dans la sous-section 2.3.1, est typiquement d’application. Plus précisément, il s’agit d’un problème de type *single-source shortest paths*, tel que décrit à la section 2.3. En effet, nous cherchons à évaluer, pour chaque école (la source), les temps de trajet vers toutes les stations de Belgique (les destinations). Cependant, les temps de trajet d’une station à l’autre ne nous intéressent pas. Il est donc plus intéressant de résoudre le problème *single-source shortest paths* pour chacune des écoles plutôt que de résoudre une fois le problème *all-pair shortest paths* et d’en filtrer les résultats. Puisque le graphe contient de nombreux nœuds, il est préférable d’avoir une complexité minimale par rapport à $|V|$ (les détails des complexités sont disponibles en section 2.3).

Ainsi, nous avons utilisé la version simplifiée de l’algorithme de Dijkstra que nous avons présentée dans la même sous-section, puisque nous travaillons dans un graphe orienté acyclique. En effet, nous savons par construction qu’aucun circuit n’est possible puisque chaque nœud n’est relié qu’à des nœuds correspondant à des événements ultérieurs. Ainsi, il n’est plus nécessaire de maintenir une *priority queue* contenant les nœuds : il suffit de simplement parcourir l’ensemble des nœuds par ordre chronologique.

Notons également que nous souhaitons obtenir la durée des trajets pour toutes les destinations possibles depuis un nœud-source. Par conséquent, les trajets sont tous calculés en une fois et l’algorithme n’est utilisé qu’une seule fois par source (plutôt que de l’utiliser pour trouver le chemin optimal pour chacune des paires source-destination possibles). Les calculs sont effectués en utilisant les écoles secondaires de la Fédération Wallonie-Bruxelles comme points de départ. L’algorithme retourne comme résultat une grande matrice contenant les temps de trajet vers tous les arrêts de la Fédération Wallonie-Bruxelles, en partant de chacune des écoles à 16 h.

Cette version simplifiée de l’algorithme de Dijkstra permet d’en réduire la complexité. Le problème des plus courts chemins peut donc être résolu en temps linéaire par rapport au nombre de nœuds et d’arêtes dans le graphe. Nous passons donc d’une complexité en $O((|E| + |V|) \log |V|)$ à du $O(|E| + |V|)$, avec V le nombre de nœuds (chaque arrêt de transports en commun à chaque horaire possible) et E , le nombre d’arêtes (parties de trajets, à pied ou en transports, reliant deux arrêts distincts, ou attente à un même arrêt).

3.3 Analyse des données

Les précédentes sections concernent le passage des données GTFS à un temps de trajet école-arrêt. Dans cette section, nous détaillons l’utilisation de ces premiers résultats pour générer les mesures qui nous intéressent : évaluation de la population concernée, identification des écoles les plus proches et accessibles, calcul des mesures pour une zone entière, etc.

3.3.1 Première approximation : les zones de Voronoï

Afin d'estimer la population concernée pour les différents temps de trajet calculés, nous avons posé comme première hypothèse que chaque personne se rend à l'arrêt le plus proche de son domicile. Nous avons donc calculé les zones de Voronoï (définies à la section 2.4) autour des arrêts de transports en commun, chaque zone correspondant donc à la partie du territoire belge qui est la plus proche de l'arrêt en question.

Afin de calculer la population correspondante, nous avons ensuite utilisé la grille de répartition de la population belge par cellule d'1 km² décrite à la sous-section 1.3.4. Une seconde hypothèse que nous avons formulée est la répartition uniforme de la population par cellule. Ainsi, nous pouvons estimer la population présente dans une zone de Voronoï en calculant son intersection avec les cellules. La population totale est estimée selon la formule suivante :

$$\text{PopAffectée}_{\text{stop}} = \sum_{\forall \text{ cellule}} \frac{\text{aire}_{\text{cellule} \cap \text{stop}}}{\text{aire}_{\text{cellule}}} \cdot \text{pop}_{\text{cellule}}$$

où $\text{cellule} \cap \text{stop}$ dénote la forme résultant de l'intersection de la zone de Voronoï d'influence de l'arrêt avec la cellule en question.

Concrètement, nous estimons donc que la fraction de population concernée par un arrêt est proportionnelle à la portion de cellule comprise dans la zone de Voronoï autour de cet arrêt. Si nous prenons l'exemple représenté à la figure 3.2, où une cellule comptant 1000 habitants est présente sur deux zones distinctes, nous estimerons que 75 % de la population de la cellule se rend à l'arrêt 1 et est donc concernée par les temps de trajet correspondants. Le reste de la population est assignée à l'arrêt 2.

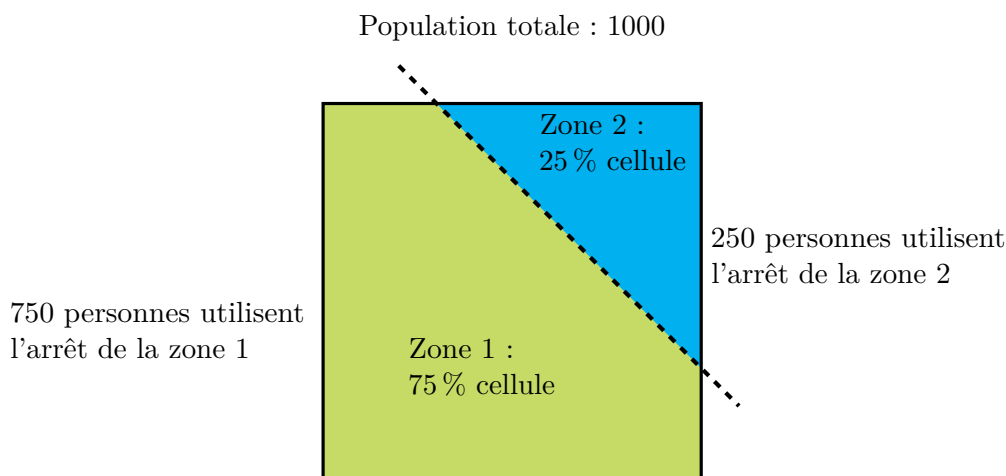


FIGURE 3.2 – Exemple de calcul de la population par l'intersection avec les zones de Voronoï

Bien que cette première hypothèse soit assez forte, elle nous a permis de générer une première estimation de la population impactée, rendant nos résultats plus pertinents que lorsque nous considérons uniquement le nombre d'arrêts concernés. C'est cette hypothèse qui a été utilisée dans le cadre de notre rapport de février 2022 (annexe B).

3.3.2 Prise en compte de plusieurs arrêts

Comme expliqué précédemment, notre première hypothèse impliquait que chaque personne cherche à se rendre à l'arrêt le plus proche de chez elle en revenant de son école, et le temps de trajet à pied pour aller jusqu'au domicile n'était pas pris en compte.

L'étape suivante pour affiner nos résultats est donc d'autoriser les élèves à choisir la station qui leur convient le mieux parmi celles qui sont proches du domicile, afin d'obtenir des résultats reflétant mieux la réalité. En effet, si un arrêt n'est pas le plus proche du domicile mais permet un temps de trajet total significativement plus court, il serait absurde de ne pas en tenir compte. Ceci est d'autant plus vrai à Bruxelles par exemple, où les arrêts sont nombreux et proches les uns des autres.

De plus, cela nous permet également d'approximer le temps de trajet à pied entre le domicile et l'arrêt, ce qui avait jusqu'ici été négligé (puisque seul le plus proche des arrêts était considéré). Notons tout de même que le domicile exact des habitants n'est pas pris en compte (et n'est d'ailleurs pas connu). Puisque les calculs effectués concernent l'ensemble de la population de la cellule, nous avons dû faire une hypothèse sur le domicile considéré. Ainsi, nous supposons que la population habite au centre de la cellule.

Cette approximation ne semble toutefois pas trop excessive étant donné la taille des cellules. En effet, l'erreur sur le domicile est, dans le pire des cas, d'environ 700 mètres (c'est-à-dire dans le cas où le domicile se trouve en fait dans l'un des coins de la cellule). En termes de temps de trajet, l'erreur est donc d'au maximum 14 minutes (temps de trajet à pied). De plus, il s'agit d'une borne supérieure de l'erreur qui représente un trajet passant par le centre de la cellule avant de rejoindre l'arrêt considéré. En pratique, il est peu probable qu'un tel détour soit nécessaire, et l'erreur devrait donc être bien inférieure.

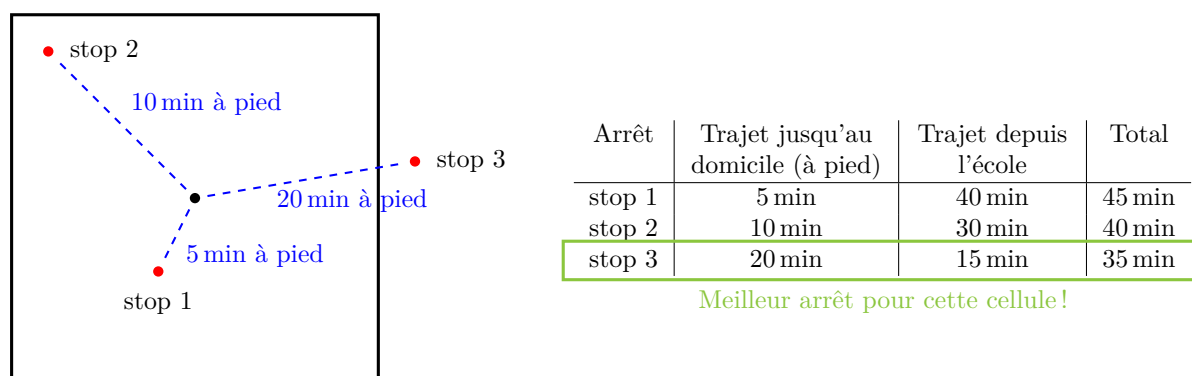


FIGURE 3.3 – Exemple de calcul : 3 arrêts autorisés

La figure 3.3 représente la démarche que nous avons suivie. Dans un premier temps, les temps de trajet à pied depuis les n arrêts les plus proches sont calculés. Ceci est effectué, comme précédemment, en estimant que la distance à vol d'oiseau entre l'arrêt et le centre de la cellule est parcourue à une vitesse réduite de 3 km/h. De plus, comme illustré à la figure 3.3, nous ne limitons pas les arrêts considérés à ceux qui sont présents sur la cellule. Le temps de trajet à pied ainsi obtenu est additionné au temps de trajet nécessaire pour revenir à cet arrêt depuis l'école considérée, calculé à la section 3.2. Enfin, nous considérons que l'entièreté de la population de la cellule utilise l'arrêt correspondant à un temps de trajet total minimum. Par exemple, pour la cellule de la figure 3.3, l'ensemble de la cellule utilisera l'arrêt stop 3, puisqu'il correspond au trajet total le plus court. Pour

cette cellule, nous estimons donc que le domicile des personnes se trouve à 35 minutes de trajet depuis l'école considérée.

L'effet du nombre de voisins Un paramètre important à choisir pour cette partie était le nombre d'arrêts voisins à considérer. Pour déterminer cela, nous avons comparé nos résultats pour différents nombres d'arrêts considérés. Plus précisément, nous avons analysé la différence en termes d'école la plus accessible lorsque plus de voisins étaient ajoutés. La figure 3.4 représente la proportion de résultats ayant changé. Notons déjà que cette proportion est relativement faible puisqu'elle est de maximum 8% selon nos observations. Les résultats semblent se stabiliser au-delà de 40 voisins et les différences de résultats atteignent véritablement 0 à partir de 60 voisins.

Ainsi, pour la suite de nos calculs, nous avons choisi d'autoriser jusqu'à 60 des plus proches voisins pour chaque cellule. Ceci reste une borne supérieure. En effet, nous avons également imposé une distance maximale de 1.5 km des stations voisines par rapport au centre de la cellule, ce qui correspond à 30 min de marche à pied à vitesse réduite.

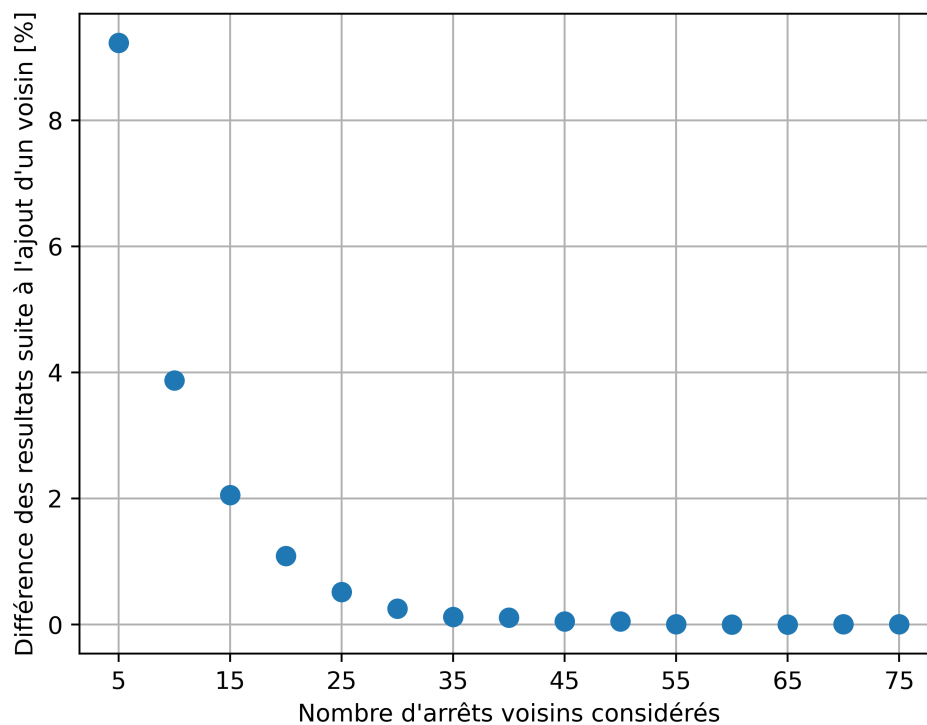


FIGURE 3.4 – Pourcentage d'écoles les plus accessibles changeant avec l'ajout de 5 voisins.

3.4 Calculs pour la génération de nos figures

3.4.1 Calcul de population et temps relatifs

À partir des calculs décrits précédemment, nous avons pu générer des graphes et histogrammes illustrant la fraction de population affectée par temps de trajet supplémentaire. Pour calculer le pourcentage de population impacté par arrêt, nous n'avons considéré que la population présente en Wallonie et à Bruxelles.

Afin de mieux évaluer l'importance d'un temps de trajet supplémentaire, nous avons également travaillé en termes de temps de trajet « relatif ». Ceci correspond à un pourcentage par rapport au temps de trajet optimal depuis une école. Prenons l'exemple d'un arrêt situé à respectivement 35 et 20 minutes des écoles A et B. Le trajet optimal depuis une école est de 20 minutes, c'est ce trajet qui sera utilisé comme référence pour calculer les pourcentages. La différence de temps de trajet entre les deux écoles est de 15 minutes : en termes de temps de trajet relatif, le chemin depuis l'école A représente une augmentation de 75 % par rapport au trajet optimal. Cette procédure est schématisée à la figure 3.5.

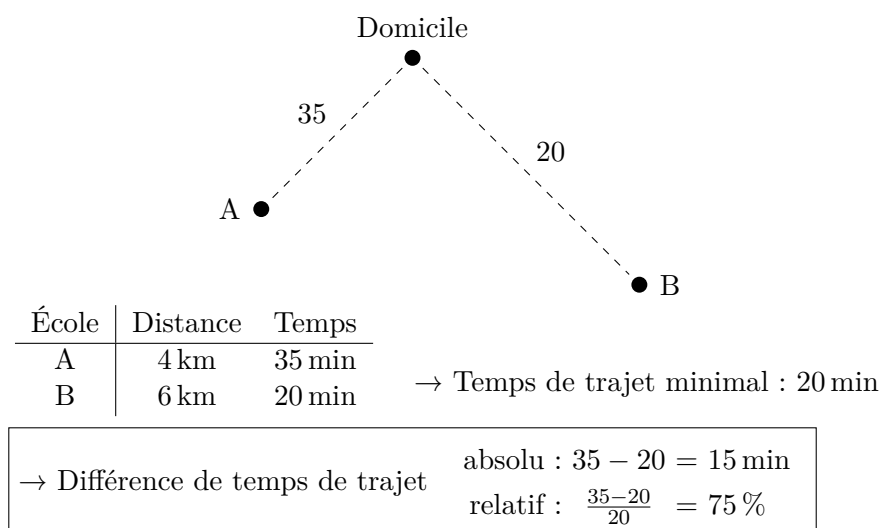


FIGURE 3.5 – Comparaison des temps absolu et relatif

3.4.2 Calcul de l'indice composite

En ce qui concerne l'évaluation de l'indice composite d'une école donnée, la majorité des coefficients a été prise en compte. Seuls le partenariat pédagogique et l'immersion ont été négligés au vu du manque de données utilisables. Cependant, comme vous le verrez à la sous-section 4.5.1, l'immersion peut être prise en compte dans le cadre de l'outil de visualisation, selon les préférences de l'utilisateur. Nos calculs utilisent les valeurs des coefficients présentés au chapitre 1. Nous avons utilisé la dernière version du décret, le coefficient 2 suit donc la pondération du tableau 1.5. Par ailleurs, nous avons dû poser plusieurs hypothèses plus ou moins fortes pour cette partie. Le but n'est bien sûr pas d'avoir une évaluation exacte des indices composites attribués à chacun, mais plutôt d'avoir une première estimation des divergences possibles entre les valeurs d'indice composite en fonction du type de distance considérée (temps de transport ou à vol d'oiseau).

Coefficient 1 : L'ordre de préférence Pour calculer le premier coefficient, nous avons posé comme hypothèse que la préférence des parents pour une école reflète son accessibilité. Ainsi, pour chaque cellule de 1 km^2 , la liste des établissements secondaires préférés correspond à la liste des 5 écoles les plus accessibles. Dans le cas d'écoles ayant une même accessibilité, leur ordre est déterminé par leur ordre d'apparition dans les voisins.

Coefficient 2 : L'école primaire considérée De façon similaire, nous avons estimé que l'école primaire fréquentée est la plus accessible depuis le centre de la cellule pour laquelle on effectue les calculs.

Comparaison d'indices Enfin, en plus de calculer l'indice composite prévu par le décret, nous en avons également calculé une version modifiée dont les coefficients 2, 3 et 4 ont été remplacés par des coefficients représentant l'accessibilité de l'établissement plutôt que sa proximité. L'objectif est de comparer les deux versions de l'indice afin de quantifier leur différence.

3.5 Le code

Cette section concerne l'implémentation de notre approche. Nous avons décidé d'utiliser le langage Python pour ce travail. Pour plus de lisibilité, notre code est divisé en plusieurs scripts, chacun chargé d'une des grandes phases de calcul décrites plus haut.

Dans la première partie, nous allons donner une vue d'ensemble sur l'organisation de nos scripts et leur fonctionnement. Une description plus complète de l'ensemble du code est fournie en annexe D. La deuxième partie est dédiée à une étude des performances de notre code.

3.5.1 Organisation

La structure globale de notre programme est schématisée dans la figure 3.6. Nous y avons également indiqué, pour chaque script, les données entrantes ainsi que les principaux fichiers générés. Comme représenté sur ce flowchart, nous avons rassemblé nos fichiers en quatre grandes phases :

Phase I – Représentation du réseau

Cette première phase consiste en la création du graphe représentant le réseau de transport à partir des données publiques dont nous disposons. Elle est composée de deux scripts : le premier (`0_parse.py`) effectue le *parsing* des données proprement dit et extrait les données d'intérêt pour chaque nœud (position géographique, horaire, code, etc.), tandis que le second (`1_precomputation.py`) réalise des calculs tels que l'évaluation de la population affectée à chaque arrêt ou encore le calcul du temps de parcours à pied entre deux stations. Enfin, le graphe du réseau utilisé en Phase II est généré.

Ces deux scripts sont majoritairement repris du travail déjà réalisé par Guillaume Derval [41]. Nous y avons ajouté la partie chargée du *parsing* des écoles ainsi que le calcul de la population affectée par arrêt (Voronoi).

Données GTFS Fichier FASE

PHASE I – REPRÉSENTATION DU RÉSEAU

0_parse.py

Création des nœuds du graphe :

- Un par école à l'heure de départ
- Un par arrêt à une heure de départ/arrivée d'une ligne

Output (format GeoJSON)

Données sur chaque nœud (voisins, latitude, longitude, nom des lignes, réseau des écoles, etc.)

1_precomputation.py

- Calcul de la population par arrêt
- Calcul des temps de trajet à pied vers les voisins

Output (format JSON)

Distances à pied entre voisins, données sur chaque nœud (population par arrêt, temps de marche entre voisins, etc.)

PHASE II – CALCUL DES PLUS COURTS CHEMINS

2_resolve.py

- Calcul des temps de trajet vers tous les arrêts depuis chaque école
- Calcul des distances à vol d'oiseau pour chaque paire école-station

Output (format NumPy)

Temps de trajet depuis les écoles vers tous les arrêts, distances à vol d'oiseau (données calculées pour les écoles secondaires et primaires)

PHASE III – ANALYSE DES DONNÉES (COMPARAISON DISTANCE VS TRANSPORT)

3a_schools_computation.py

Calcul par arrêt → Voronoï

- Calcul des n écoles les plus accessibles
- Comparaison avec les n les plus proches
- Calcul des indices composites

3b_computation_grid.py

Calcul par cellule → plusieurs voisins

- Calcul des n écoles les plus proches
- Calcul des écoles les plus accessibles (*prise en compte de la marche jusqu'aux arrêts*)
- Comparaison avec les n les plus proches

Output (format GeoJSON)

Données sur les écoles les plus proches/accessibles (temps de trajet, distance, nom des écoles, différence de temps de trajet pour le top 1, mesure de la similarité des listes, etc.)

PHASE IV – PRÉSENTATION DES RÉSULTATS : FIGURES + OUTIL

4a_histograms.py

4b_choropleths.py

4c_pixel_maps.py

6_tool.py

FIGURE 3.6 – Flowchart de notre code

Phase II – Calcul des plus courts chemins

Cette phase correspond au calcul des temps de trajet effectués par les élèves, sur base des données analysées dans la phase précédente. Par ailleurs, ce script calcule également les distances à vol d’oiseau entre chaque école et chaque arrêt. Les résultats sont sauvegardés sous forme de deux matrices (une pour les distances et une pour les temps de trajet) qui seront exploitées dans la suite des calculs. Ces matrices sont de taille (nombre d’arrêts $\times$ nombre d’écoles). Ainsi, l’entrée (i, j) de la matrice des temps de trajet correspond au temps de transport nécessaire pour revenir à l’arrêt i en partant de l’école j à 16 h (ce temps de trajet tient compte à la fois des éventuels temps d’attente, du temps de trajet en transports et des déplacements à pied).

Ce script est également repris du travail de G. Derval [41]. Nous y avons ajouté le calcul des distances à vol d’oiseau, ainsi que la récupération des temps de trajet sous forme d’une matrice (utilisée pour les scripts suivants).

Phase III – Analyse des données (comparaison distance vs transport)

Nous avons implémenté la totalité de cette phase de calcul, qui correspond donc à la partie d’analyse de nos résultats à proprement parler. Elle compte deux scripts, correspondant aux deux hypothèses explorées dans ce travail :

- `3a_schools_computations.py` correspond à notre première approximation, selon laquelle chaque personne se rend à l’arrêt le plus proche de son domicile. Les calculs sont donc effectués pour chaque arrêt des régions wallonne et bruxelloise.
- `3b_computation_grid.py` correspond à un modèle plus raffiné, effectuant les calculs pour chaque cellule d’1 km². Dans ce cas, les élèves quittant l’école à 16 h ont la possibilité de se rendre à un des 60 arrêts les plus proches du centre de la cellule (et situés dans un rayon de 1.5 km).

Dans les deux cas, les valeurs de temps et de distance (calculées lors de la phase de calcul précédente) sont utilisées pour identifier les n écoles les plus proches et les n plus accessibles de la zone considérée (dans notre cas, nous avons calculé des top 5 d’écoles, mais les fonctions des scripts sont génériques). Les indices composites sont également calculés et le tout est sauvegardé pour la dernière phase. Par ailleurs, la différence en termes de temps de trajet entre l’école la plus proche et l’école la plus accessible est également estimée.

Phase IV – Présentation des résultats : figures et outil

Cette dernière phase consiste en la visualisation et présentation de nos résultats et a également été entièrement écrite par nos soins.

Différents scripts ont été créés afin de gérer les différents visuels souhaités (graphiques, histogrammes, cartes choroplèthes, cartes sur base des cellules au 1 km²). Ces scripts génèrent donc l’intégralité des figures présentées dans le chapitre 4 de ce mémoire, ainsi que les figures de notre rapport de février 2022 (annexe B). Enfin, cette phase englobe également le script dédié à la génération de notre outil de visualisation, qui présente pour des coordonnées données (le domicile de l’utilisateur), les écoles les plus proches et les plus accessibles, les temps de trajet ainsi qu’une comparaison de l’indice composite prévu par le décret Inscription avec notre version, pour l’établissement secondaire souhaité.

3.5.2 Performances

Représentation du réseau Pour commencer, notre premier script demande 15 s pour effectuer le *parsing* des écoles. L'utilisation d'un *ball tree* (voir fonction `knnSearch`) pour cette étape a permis de significativement améliorer les performances du script, étant donné que l'usage d'un algorithme de type *brute force* nécessitait 380 s, soit 25 fois plus de temps.

Le calcul des zones de Voronoï (et de la population attribuée à chacune d'entre elles) dure quant à lui une vingtaine de minutes pour l'entièreté de la Belgique. La documentation de la librairie que nous avons utilisée nous apprend que la méthode de calcul de ces zones est une résolution du problème dual : la triangulation de Delaunay. Pour cela, la librairie calcule l'enveloppe convexe (voir section 2.4.3).

Calcul des plus courts chemins Les performances les plus intéressantes à analyser concernent évidemment la partie de notre programme calculant les plus courts chemins entre les écoles et tous les arrêts de transports en commun en Belgique. Nous avons donc mesuré le temps d'exécution du code pour des graphes (zones de Belgique considérées) de différentes tailles. Le tableau 3.1 résume les données relatives à ces différents graphiques. Nous avons effectué des tests sur la Belgique entière, uniquement la Fédération Wallonie-Bruxelles ainsi que sur plusieurs provinces séparément (sélectionnées arbitrairement afin de couvrir plusieurs tailles de graphe).

TABLE 3.1 – Graphes utilisés pour l'analyse de performance

Zone concernée	Code	Nombres de nœuds	Nombres d'arêtes	Nombres d'écoles
Belgique entière	BEL	2 773 888	5 791 196	709
Fédération Wallonie-Bruxelles	FWB	1 175 732	2 308 967	708
Wallonie	WAL	717 188	1 398 835	554
Bruxelles	BRU	458 497	909 990	154
Liège	WLG	291 992	576 777	159
Hainaut	WHT	234 072	455 146	208
Namur	WNA	83 084	160 018	87
Brabant wallon	WBR	55 293	105 574	45
Luxembourg	WLX	51 606	97 073	55

Les résultats sont illustrés à la figure 3.8. La figure 3.8a illustre le temps total nécessaire à l'exécution du script `2_resolve` en fonction de la taille du graphe considéré. Au vu des différences de taille entre le plus grand et le plus petit graphe, nous avons utilisé une échelle algorithmique pour les deux axes. Nous pouvons observer que le temps d'exécution semble être linéaire par rapport à la taille du graphe. Cela correspond aux complexités évoquées au chapitre 2. Notons tout de même que cette évolution n'est pas exactement linéaire puisqu'elle dépend également du nombre d'écoles. Ceci reste toutefois négligeable au vu du nombre important de nœuds. Par ailleurs, la figure 3.8b nous montre que, comme attendu, la moyenne du temps d'exécution par station tend à augmenter avec la taille du graphe.

Calcul des écoles proches et accessibles Les performances du calcul des écoles proches et accessibles, effectuées par le script `3b_grid_computation` sont représentées sur la figure 3.7. Nous avons mesuré séparément le temps d'exécution pour le calcul des stations les plus proches du domicile (approximé comme étant le centre de la cellule) et pour le calcul des temps de trajet. On observe une certaine linéarité dans le premier cas. En ce qui concerne le calcul des temps de trajet, le temps d'exécution semble plafonner aux alentours de 40 à 50 voisins. Ceci correspond au phénomène que nous avons observé à la figure 3.4. En effet, c'était aux alentours de 40 voisins autorisés par cellule que les résultats (écoles les plus accessibles) ne variaient plus.

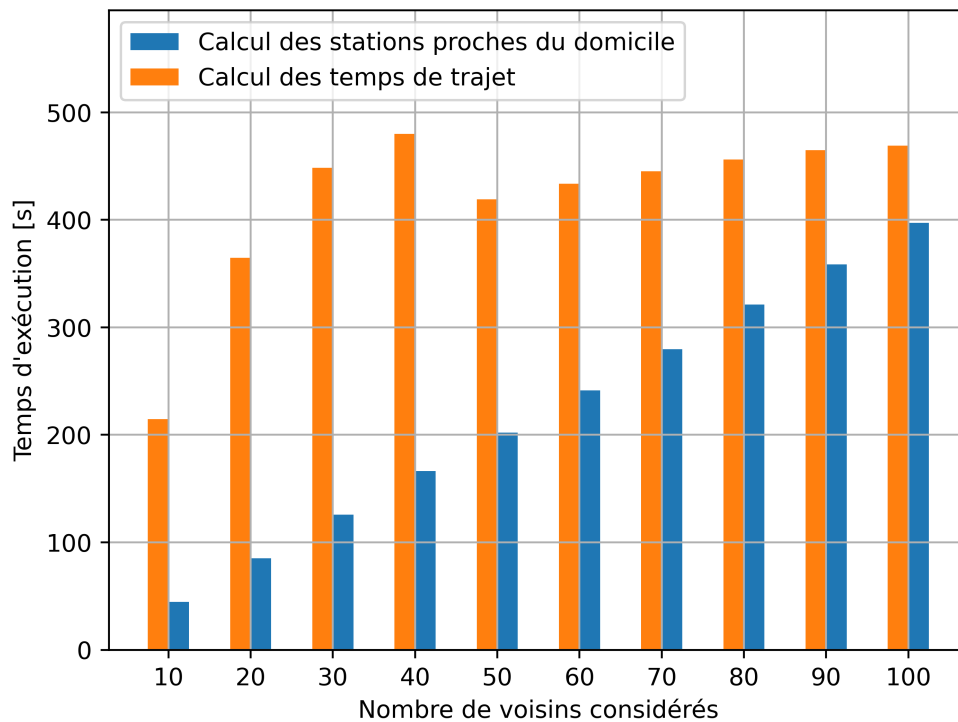
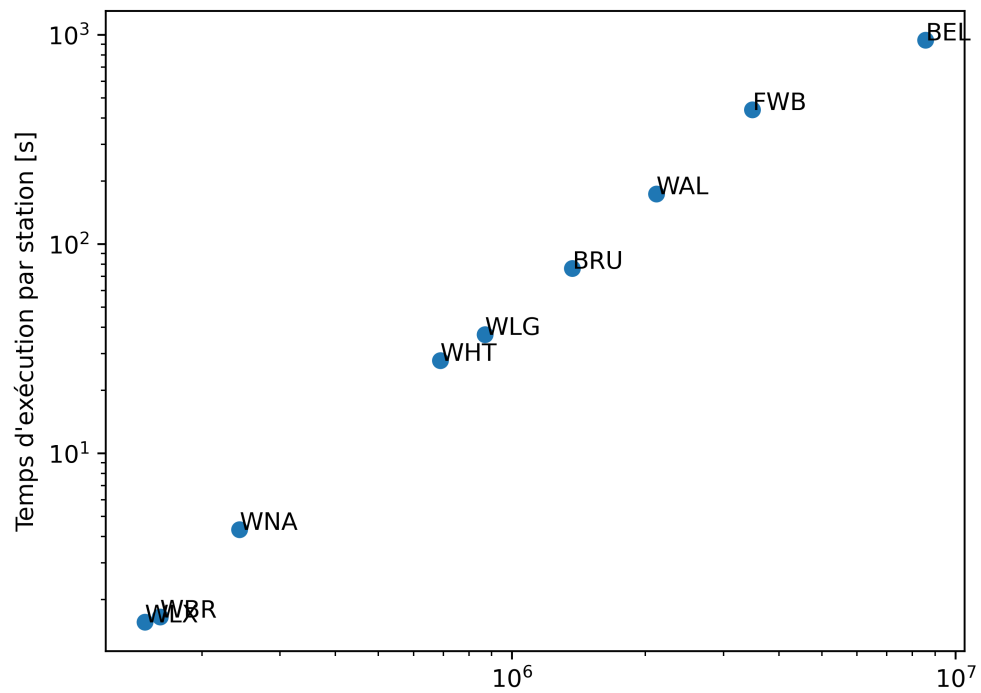
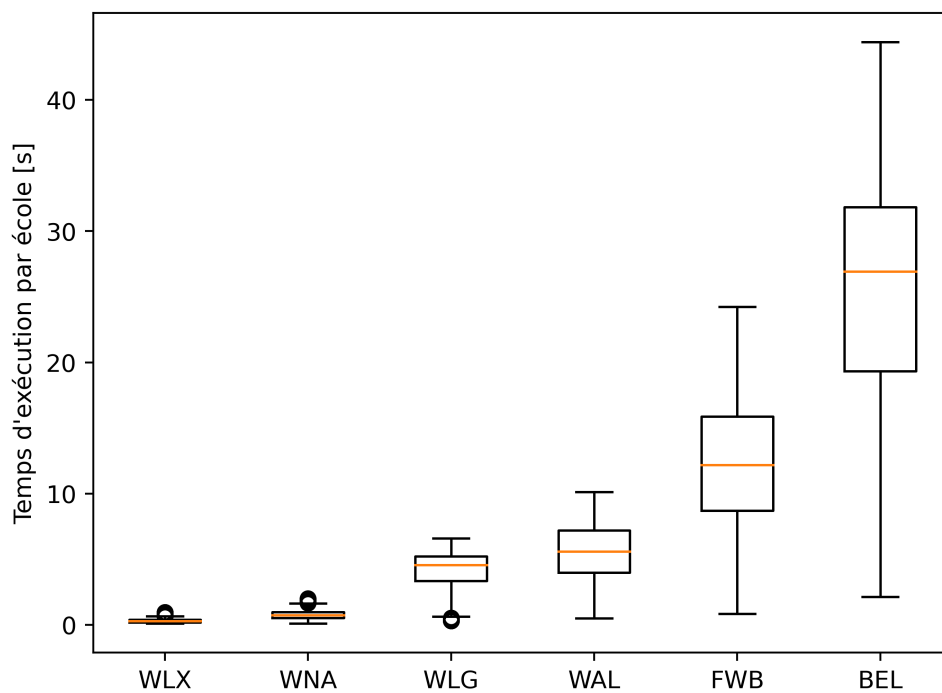


FIGURE 3.7 – Analyse de performances : Temps d'exécution sur la grille de 1 km^2



(a) Évolution du temps d'exécution moyen en fonction du nombre de nœuds dans le graphe



(b) Boxplot : temps d'exécution pour différentes tailles de graphe

FIGURE 3.8 – Analyse de performances : Temps d'exécution du calcul des plus courts chemins

Chapitre 4

Résultats

Ce chapitre est dédié à la présentation de nos résultats, basés sur les horaires de transports organisés le 26 octobre 2021. Dans un premier temps, nous résumerons les hypothèses de travail que nous avons dû imposer pour nos calculs. Pour plus de clarté, nous ferons désormais la distinction entre les écoles « accessibles », c'est-à-dire les écoles correspondant à un temps de trajet minimal vers le domicile, et les écoles « proches », correspondant aux écoles les plus proches du domicile en termes de distance à vol d'oiseau.

Nous présenterons ensuite nos investigations concernant le temps de trajet nécessaire pour se rendre de l'école la plus proche jusqu'au domicile et nos conclusions sur l'existence ou non d'une discordance entre les critères de proximité (à vol d'oiseau) et d'accessibilité (en transports en commun). Nous tenterons également d'évaluer l'ampleur de cette discordance, en analysant la population impactée et les différences de durée des trajets parcourus. Nous étudierons également les répercussions sur la valeur de l'indice composite attribué aux écoles. Nous discuterons enfin de la robustesse de nos observations (sont-elles toujours valables lorsque plus d'écoles sont prises en compte pour comparaison ?) et des différences de résultats obtenus par rapport à notre rapport de février 2022 (annexe B), depuis notre amélioration de l'hypothèse sur le point d'arrivée.

La dernière partie de ce chapitre sera consacrée à la présentation de notre outil de visualisation. Celui-ci est un tableau de bord développé à destination des parents d'élèves et a pour but de fournir des informations sur l'établissement secondaire visé (temps de trajet, distance à vol d'oiseau, etc.). Il permet également de comparer la valeur de l'indice composite calculé selon la méthode prévue par le décret Inscription avec un indice composite alternatif, basé sur l'accessibilité des écoles plutôt que sur leur proximité.

4.1 Résumé des hypothèses de travail

Les écoles Les résultats présentés dans ce travail ne prennent en compte que les écoles de l'enseignement secondaire ordinaire de la Fédération Wallonie-Bruxelles documentées dans le fichier FASE décrit dans la sous-section 1.3.2. Nous avons à la fois considéré les réseaux confessionnel et non confessionnel. Pour les coefficients nécessitant une école primaire, nous avons uniquement pris en compte les écoles primaires ordinaires présentées dans ce même fichier FASE. De plus, ce fichier ne permettant pas d'identifier les implantations appartenant à un même établissement, nous avons considéré chaque implantation comme un établissement à part entière.

Départ de l'école et arrêts considérés Nous avons considéré que les élèves quittent leur école à 16 h. En effet, il s'agit d'une heure typique pour la fin des cours et c'est à ce moment que l'on observe un pic dans le nombre de trajets domicile-école effectués [10]. Pour quitter leur école, nous avons estimé que les élèves choisiraient de partir depuis l'arrêt le plus avantageux pour eux (c'est-à-dire, correspondant à un temps de trajet de retour minimal), dans un rayon de 1.5 km autour de l'école et avec une limite de 60 arrêts.

Le domicile Comme expliqué au chapitre 3, deux hypothèses sur le domicile des enfants ont été envisagées. Notre première approximation imposait que chaque élève cherche à se rendre à l'arrêt le plus proche de chez lui. Cela revenait donc à supposer que chaque personne habite « sur » cet arrêt le plus proche. La seconde hypothèse suppose que la population de chaque cellule de 1 km² habite en son centre. Nous considérons donc que les élèves cherchent à se rendre à l'un des arrêts les plus proches du centre de leur cellule, en choisissant la possibilité qui minimise la durée de leur trajet. Jusqu'à 60 arrêts (voir discussion à la section 3.3.2) proches du centre de la cellule ont été autorisés, seuls ceux qui se trouvent dans un rayon de 1.5 km sont pris en compte. Afin d'éviter une certaine répétition, nous n'allons, dans ce chapitre, nous intéresser qu'aux résultats obtenus sous cette seconde hypothèse, plus générale que la première. Nous allons toutefois discuter à la section 4.3 des différences entre les résultats de ce chapitre et ceux du rapport de février.

Les déplacements autorisés Seuls les déplacements à pied et en transports en commun ont été étudiés (le temps de trajet total inclut donc le temps total de marche, les éventuels temps d'attente aux arrêts, ainsi que le temps effectivement passé dans les transports). La vitesse de marche à pied a été imposée à 3 km/h afin de prendre en compte les détours par rapport à la distance à vol d'oiseau, comme expliqué à la section 3.1.2. De plus, la marche vers un arrêt voisin est autorisée si celui-ci se trouve à moins de 10 minutes de marche (500 m) du point de départ. Ceci permet d'autoriser une correspondance si nécessaire. Une limite de 60 arrêts voisins a également été imposée.

Répartition de la population Nous avons considéré que la population est répartie uniformément au sein de chaque cellule d'1 km². N'ayant pas d'informations précises sur la proportion d'élèves du secondaire par cellule, nous avons supposé que le nombre d'enfants est proportionnel à la densité de population de la cellule. Ainsi, lorsque nous étudierons des pourcentages de population concernée par certaines mesures, nous supposerons que les observations reflètent également la part d'élèves concernés.

Calcul de l'indice composite Pour commencer, nous n'avons pas tenu compte du nombre limité de places disponibles par établissement. De plus, la priorité accordée aux élèves dont un frère ou une sœur fréquente déjà l'école secondaire a été négligée. Ensuite, en ce qui concerne le premier coefficient de l'indice composite, c'est-à-dire celui qui reflète la préférence des parents, nous posons l'hypothèse que l'ordre de préférence est identique à l'ordre d'accessibilité des écoles. De même, le troisième coefficient concerne la distance par rapport à l'école primaire fréquentée. N'ayant pas de données sur le sujet, nous posons l'hypothèse que chaque élève a fréquenté l'école primaire la plus accessible pour son domicile. Enfin, le coefficient 7 (partenariat pédagogique) n'a pas été calculé au vu du manque de données utilisables.

4.2 Analyse de l’accessibilité des établissements

Les résultats présentés dans cette section prennent en compte l’ensemble des établissements secondaires de la Fédération Wallonie-Bruxelles, sans distinction de réseau. Toutefois, le décret Inscription est appliqué de façon à traiter indépendamment les réseaux confessionnel et non confessionnel. Afin de confirmer la validité de nos résultats, nous avons donc également effectué séparément les calculs pour chacun des réseaux. Nous n’avons pas observé de changement significatif par rapport aux résultats obtenus pour l’ensemble des écoles. Pour plus de facilité, cette section se concentrera donc sur l’analyse de l’ensemble des écoles, tous réseaux confondus. Une présentation des résultats par réseau est fournie en annexe E.

Avant de détailler nos résultats, notons également qu’un total de 709 établissements secondaires ont été pris en compte dans nos calculs, dont 154 situés à Bruxelles. La densité d’écoles y est radicalement plus élevée qu’ailleurs en Belgique (les valeurs exactes par province sont résumées au tableau 4.1).

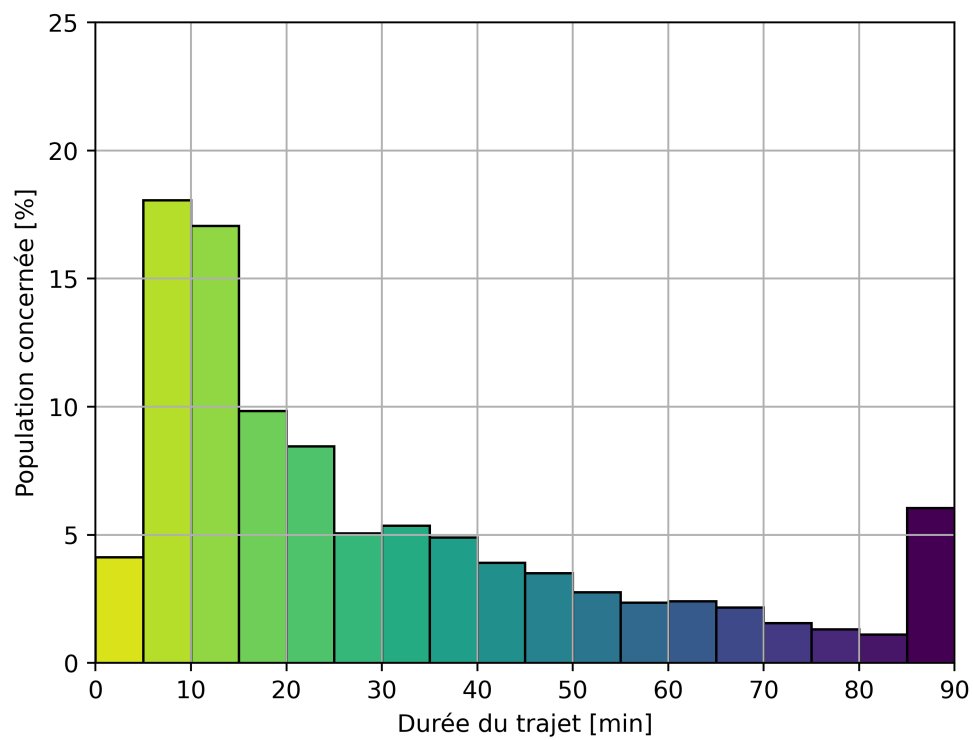
TABLE 4.1 – Répartition des écoles par province

Province	Nombre d’écoles par km ²	Nombre total d’écoles
Bruxelles	0.948	154
Hainaut	0.055	208
Liège	0.041	159
Brabant wallon	0.041	45
Namur	0.024	87
Luxembourg	0.012	55

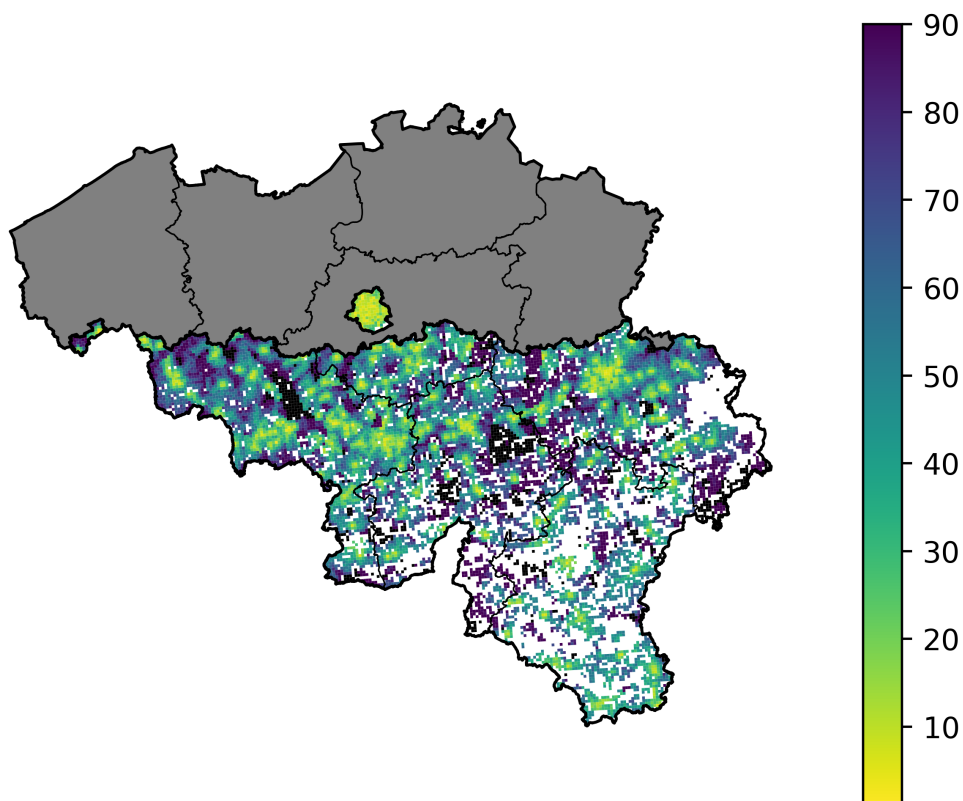
4.2.1 Temps de trajet depuis l’école la plus proche

Étant donné que les coefficients de l’indice composite valorisent la proximité à vol d’oiseau des écoles, nous allons commencer par nous intéresser à l’accessibilité de ces écoles les plus proches. Pour chaque cellule de 1 km² des régions wallonne et bruxelloise, nous avons calculé le temps de trajet minimal pour un retour au centre de la cellule depuis l’école la plus proche (voir détails de la méthode au chapitre 3). Les résultats sont visibles à la figure 4.1.

La figure 4.1a est un histogramme indiquant la proportion d’élèves concernés par durée de trajet. Pour plus de lisibilité, les temps de trajet sont groupés par tranches de 5 minutes. Ainsi, un peu moins de 10 % des élèves mettraient entre 15 et 20 minutes pour revenir de l’école la plus proche de chez eux et la majorité semble mettre moins de 30 minutes au total. Le pic observé pour une durée de 90 minutes est dû au fait que nous avons imposé le temps de trajet maximal à cette valeur. Ainsi, ce pic représente tous les trajets d’1 h 30 ou plus. Ces observations peuvent être complétées par l’analyse de la version cumulative de cet histogramme, représenté à la figure 4.2. Cette fois, les trajets sont groupés par minute. L’on peut donc y observer qu’environ 50 % des élèves ont un temps de trajet d’au moins 20 minutes pour revenir de l’école la plus proche de leur domicile.



(a) Distribution des temps de trajet par tranche de 5 minutes



(b) Répartition des temps de trajet sur le territoire belge (légende : durée du trajet [min])

FIGURE 4.1 – Temps de trajet pour rentrer au centre de sa cellule depuis l'école la plus proche (départ à 16 h)

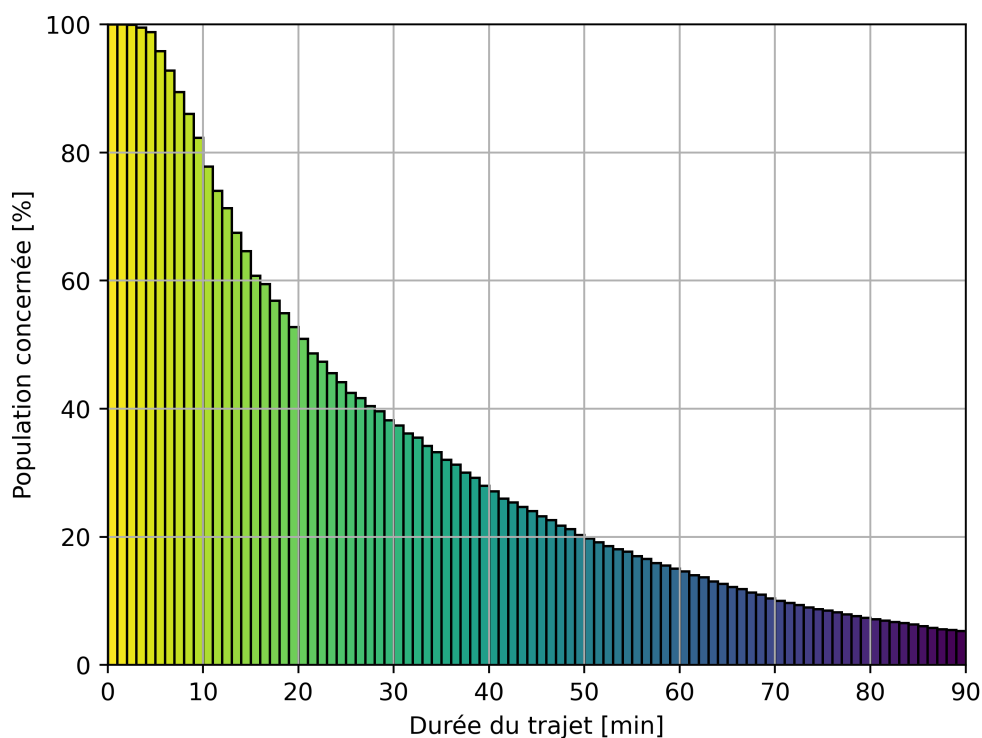


FIGURE 4.2 – Temps de transport pour rentrer depuis l'école la plus proche

Ceci est d'ailleurs cohérent avec les résultats de l'enquête MONITOR dont nous avons brièvement parlé dans notre introduction. De même, il apparaît que plus de 15 % des personnes, soit environ une personne sur six, ont un long trajet de plus d'une heure à parcourir. Si l'on fait abstraction des temps de trajet supérieurs à 1 h 30, les élèves mettent en moyenne une quarantaine de minutes pour rentrer chez eux. Notons également qu'il ne s'agit ici que du chemin pour revenir de l'école et que, par conséquent, le temps de trajet total sur la journée sera donc plus élevé que ces valeurs. Cette moyenne concerne toutefois l'ensemble de la Fédération Wallonie-Bruxelles, il serait donc intéressant d'analyser les différences qui peuvent apparaître selon la zone considérée.

La figure 4.1b illustre la répartition de la durée des trajets depuis l'école la plus proche sur le territoire belge (divisé ici en cellules de 1 km^2). Les zones blanches de la carte correspondent à des cellules pour lesquelles il n'y a pas de données sur la population (soit parce qu'elles sont très peu ou pas du tout habitées, soit parce que les données sont manquantes).

En ce qui concerne les autres cellules, leur couleur est attribuée selon la durée du trajet optimal pour un retour de l'école à 16 heures. Notons également qu'une série de cellules situées en région wallonne sont colorées en noir. Ceci indique qu'il est impossible d'y revenir depuis l'école la plus proche en utilisant les transports en commun (pour l'heure de départ considérée). L'inaccessibilité d'une cellule peut être due à l'absence de ligne disponible la reliant à son école la plus proche, ou encore à des trajets à pied trop longs pour rejoindre une station. Ces zones inaccessibles concernent environ 2 % de la population de la Fédération Wallonie-Bruxelles, soit 104 811 personnes.

La première chose à observer sur cette figure 4.1b est la différence de durée des trajets

selon la région considérée. En effet, la région de Bruxelles-Capitale semble majoritairement compter des trajets inférieurs à 30 minutes, là où la Wallonie est concernée par des trajets plus longs, souvent de l'ordre de l'heure. En moyenne, le temps de trajet pour rentrer de l'école la plus proche est d'environ 14 minutes à Bruxelles. Ceci est bien inférieur à la moyenne en région wallonne qui est de l'ordre de 50 minutes. La durée des trajets peut d'ailleurs s'y étendre à plus d'1 h 30 alors que la valeur maximale observée à Bruxelles est de 58 minutes.

Ces différences sont assez logiques puisque la densité d'arrêts par cellule est beaucoup plus élevée en région bruxelloise (en moyenne, 26 arrêts par km² habitée contre seulement 3 arrêts en région wallonne). Comme précédemment expliqué, c'est également à Bruxelles que la densité d'écoles est la plus élevée, la distance moyenne avec l'école la plus proche est par conséquent bien moins importante qu'ailleurs en Belgique : 670 m en moyenne alors que les autres régions ont toutes une distance moyenne de plus de 3 km.

Les chiffres détaillés par province de la distance à vol d'oiseau avec l'école la plus proche et de la durée moyenne du trajet de retour sont résumés au tableau 4.2. Notons que nous avons imposé la durée des trajets à un maximum de 90 minutes, il est donc tout à fait normal que le même maximum soit observé pour la majorité des provinces.

Des cartes choroplèthes¹ détaillant la distance moyenne par commune entre le domicile et l'école la plus proche, ainsi que le temps de trajet pour rentrer à 16 h de l'école la plus proche sont disponibles en annexe (section E.2).

TABLE 4.2 – Vue d'ensemble par province : distance à vol d'oiseau de l'école la plus proche et temps de trajet minimum pour en revenir

Province	Distance de l'école [km]			Temps de trajet [min]		
	Moyenne	Min	Max	Moyenne	Min	Max
Luxembourg	6.43	0.33	17.06	55	11	90
Namur	4.01	0.16	18.14	59	9	90
Brabant wallon	3.81	0.26	12.20	57	9	90
Liège	3.04	0.20	17.56	46	7	90
Hainaut	3.01	0.17	14.97	48	7	90
Bruxelles	0.67	0.17	4.16	14	7	58

4.2.2 Discordance entre proximité et accessibilité

Maintenant que nous avons étudié les temps de trajet typiques pour revenir de l'école la plus proche, nous allons nous intéresser à la proportion de ces trajets qui sont effectivement des trajets de retour optimaux.

Pour chaque cellule située dans les régions wallonne et bruxelloise, nous avons calculé l'école la plus accessible et comparé celle-ci à l'école la plus proche. Les cellules pour lesquelles les deux écoles étaient différentes ont été colorées en rouge, tandis que les autres

1. « Il s'agit d'une carte thématique où les régions sont colorées selon une mesure statistique, par exemple la densité de population. Ce type de carte facilite la comparaison de la mesure d'une région à l'autre ou montre sa variabilité pour une région donnée. » [42]

sont colorées en vert. Ceci a donné la figure 4.3 (les cellules noires sont inaccessibles en transports en commun). Ces cellules rouges représentent environ 30 % de la population de la Fédération Wallonie-Bruxelles. Pour ces personnes, la version actuelle du calcul de l'indice composite favorise donc des trajets plus longs.

Contrairement à la répartition des temps de trajet étudiée précédemment, nous pouvons observer que la tendance reste globalement la même à Bruxelles et en Wallonie. Plus précisément, 26.7 % de la population en région bruxelloise est concernée, alors que le pourcentage en région wallonne monte à 34.5 %.

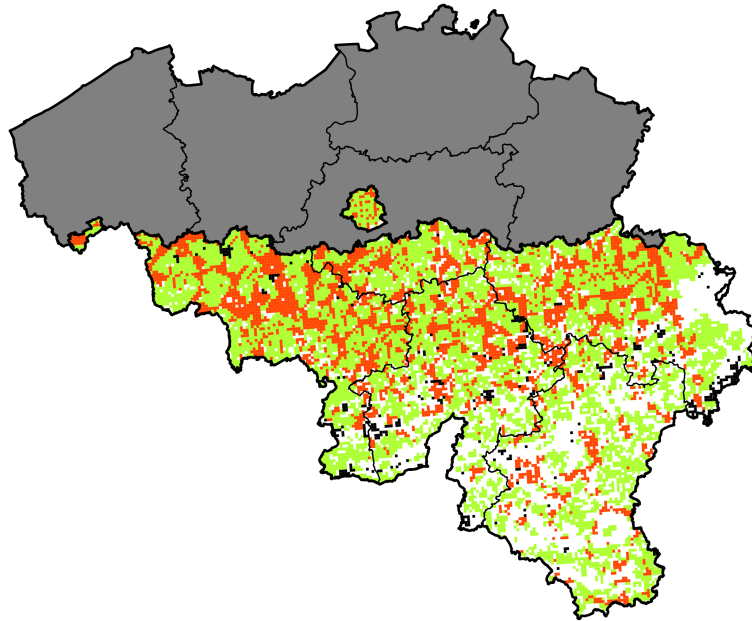


FIGURE 4.3 – L'école la plus proche est-elle également la plus accessible ? (cellules vertes : oui, cellules rouges : non, cellules noires : inaccessibles en transports en commun)

Afin d'identifier les zones les plus problématiques, nous avons également calculé l'impact par commune et par province. Les valeurs par province sont renseignées au tableau 4.3. La figure 4.4 est une carte choroplèthe illustrant la part de population concernée par commune. Les communes sont d'autant plus foncées que la proportion de personnes habitant cette commune et se situant dans une « cellule rouge » de la figure 4.3 est grande. Les communes blanches correspondent à des zones pour lesquelles nous n'avons pas d'information sur la population.

Pires communes à Bruxelles Les communes les plus concernées en région bruxelloise sont, par ordre décroissant :

1. Watermael-Boitsfort, avec 44.36 % de sa population concernée
2. Berchem-Sainte-Agathe (42.46 %)
3. Woluwe-Saint-Lambert (41.95 %)
4. Woluwe-Saint-Pierre (41.51 %)
5. Forest (39.87 %)

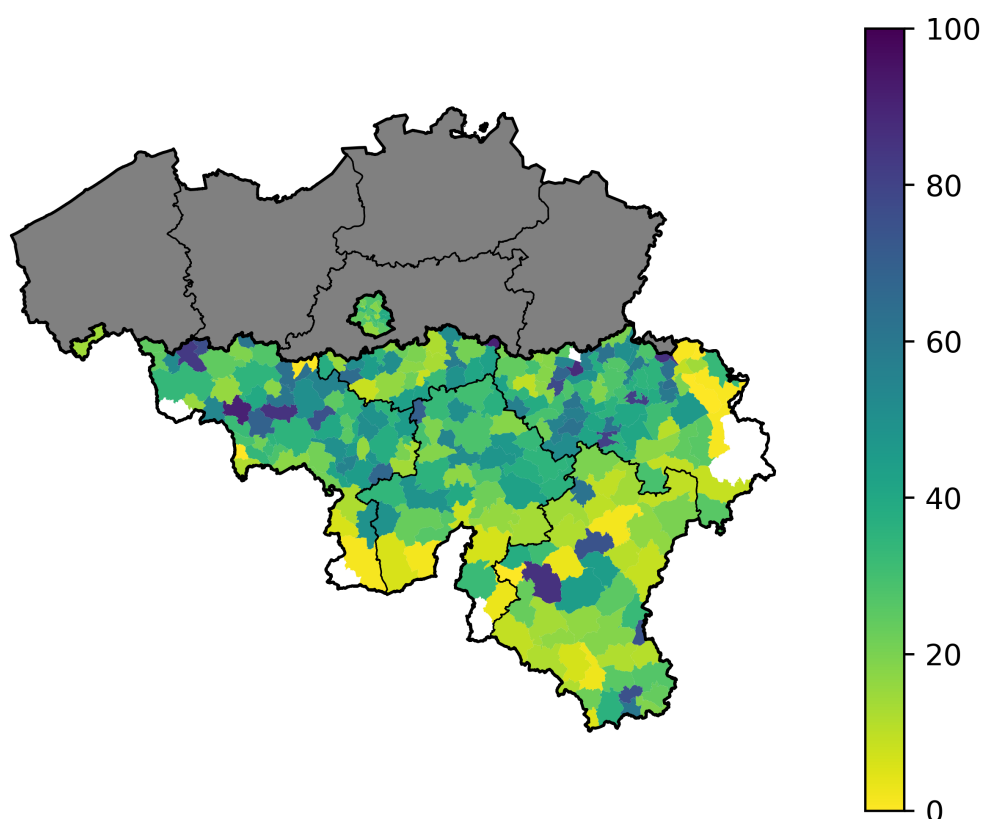


FIGURE 4.4 – Population affectée par commune (légende : population impactée [%])

Pires communes en Wallonie En région wallonne, les communes les plus pénalisées par les critères de distance sont :

- | | |
|------------------------------------|----------------------------------|
| 1. Berloz (100 %) | 6. Gesves (96.23 %) |
| 2. Geer (99.97 %) | 7. Orp-Jauche (95.16 %) |
| 3. Fexhe-le-Haut-Clocher (97.56 %) | 8. Ittre (93.92 %) |
| 4. Belœil (97.05 %) | 9. Aubel (93.68 %) |
| 5. Hélécine (97.01 %) | 10. Thimister-Clermont (93.29 %) |

Ces cartes mettent en évidence qu'il y a bien un décalage entre la proximité d'une école en termes de distance à vol d'oiseau et son accessibilité en transports en commun, et qu'une part non négligeable de la population est concernée par cette différence. Ce problème est visiblement présent sur l'entièreté du territoire belge, et non spécifique à certaines zones (toutefois, nous pouvons identifier des communes plus touchées que d'autres). Notons également qu'une seule école proche a été considérée pour cette analyse. L'évolution de ces résultats lorsque plusieurs écoles sont prises en compte sera explorée en sous-section 4.2.4. Par ailleurs, nous avons pour le moment souligné le manque de consistance entre une distance à vol d'oiseau et la durée des trajets, mais nous n'avons encore aucune information sur le réel impact pour les élèves. C'est ce que nous allons explorer dans la section suivante.

4.2.3 Conséquences pour la population

Dans cette section, nous allons essayer de quantifier les différences entre ces deux critères, en nous intéressant à la durée des trajets parcourus par les élèves.

Les cartes illustrant la répartition précise des temps de trajet par cellule de 1 km² sont comparées en annexe (section E.4). Nous pouvons y observer une diminution générale de la durée des trajets lors d'un départ de l'école la plus accessible. Par ailleurs, la part de population se trouvant dans une zone inaccessible a été réduite à environ 0.4 % (19 846 personnes). Ceci est bien inférieur aux 104 811 personnes pour lesquelles l'école la plus proche est inaccessible. Ainsi donc, pour environ 85 000 personnes (un peu plus d'1 personne sur 100), le décret Inscription favorise une école proche mais inaccessible au détriment d'écoles plus éloignées dont il est possible de revenir en transports.

Conséquence sur la durée absolue des trajets Pour plus de lisibilité, nous allons, dans cette section, uniquement nous intéresser à la carte de la figure 4.5a. Celle-ci représente la différence moyenne (en minutes) par commune entre un retour de l'école la plus proche, et un retour depuis l'école la plus accessible. Cette carte est par ailleurs complétée par la figure 4.5b, un histogramme qui indique le pourcentage de population affectée par temps de trajet supplémentaire (ces durées sont regroupées par tranche d'1 min). L'histogramme est cumulatif, ce qui veut dire que chaque barre indique la part de population concernée par un trajet d'au moins la valeur concernée. Ainsi, presque 10 % des personnes voient leur temps de trajet augmenté d'au moins 20 minutes.

À nouveau, il est clair au vu de la figure 4.5a que la tendance à Bruxelles est moins marquée que pour le reste de la Wallonie. Si l'on analyse les chiffres précis par région, nous apprenons que la différence moyenne de temps de trajet à Bruxelles est d'un peu plus de 3 minutes et que la plus grande différence observée est de 57 minutes. En région wallonne, par contre, la moyenne est un peu plus élevée : environ 22 minutes de différence. Notons tout de même qu'il existe certains cas dépassant 1 h 30 de différence. Les valeurs de différence de temps de trajet détaillées par région sont résumées dans le tableau 4.3. Les communes les plus touchées sont listées ci-dessous (valeurs arrondies à l'excès).

Pires communes à Bruxelles

1. Evere, avec une moyenne de 10 minutes de différence
2. Watermael-Boitsfort (7 min)
3. Uccle (7 min)
4. Woluwe-Saint-Lambert (6 min)
5. Auderghem (5 min)

Pires communes en Wallonie

- | | |
|---------------------------------|-------------------------|
| 1. La Roche-en-Ardenne (90 min) | 6. Donceel (82 min) |
| 2. Onhayé (89 min) | 7. Assesse (81 min) |
| 3. Gesves (88 min) | 8. Houffalize (78 min) |
| 4. Lens (87 min) | 9. Bertogne (75 min) |
| 5. Braives (85 min) | 10. Saint-Vith (71 min) |

Conséquence sur la durée relative des trajets L'analyse précédente a ses limites en termes d'interprétation. En effet, une différence de 20 minutes n'a pas le même impact pour quelqu'un ayant un temps de trajet d'une dizaine de minutes que d'une cinquantaine de minutes. Nous avons donc effectué une analyse similaire, mais prenant en compte les temps de trajet relatifs cette fois (figure 4.6b). Pour rappel, le pourcentage est exprimé en fonction du temps de trajet depuis l'école la plus accessible. À nouveau, les valeurs par province sont résumées au tableau 4.3.

Sur cette figure 4.6b, nous pouvons donc observer que pour environ 12 % des personnes (soit une personne sur huit), revenir de l'école la plus proche augmente au moins de moitié la durée de trajet, par rapport au trajet le plus rapide. Il s'avère aussi que, pour un peu plus d'une personne sur vingt, le trajet de retour depuis l'école la plus proche est deux fois plus long que le trajet depuis l'école la plus accessible.

À nouveau, nous pouvons étudier séparément les chiffres en fonction des régions. Cette fois encore, les chiffres pour la région de Bruxelles-Capitale qui sont les moins élevés avec une moyenne de 34 % de temps de trajet en plus, contre plus de 50 % de trajet supplémentaire en Wallonie. Notons également que pour les deux régions, une fraction de la population est concernée par une augmentation du temps de trajet de plus de 100 %. Les pires communes pour les régions bruxelloise et wallonne sont reprises dans la liste ci-dessous.

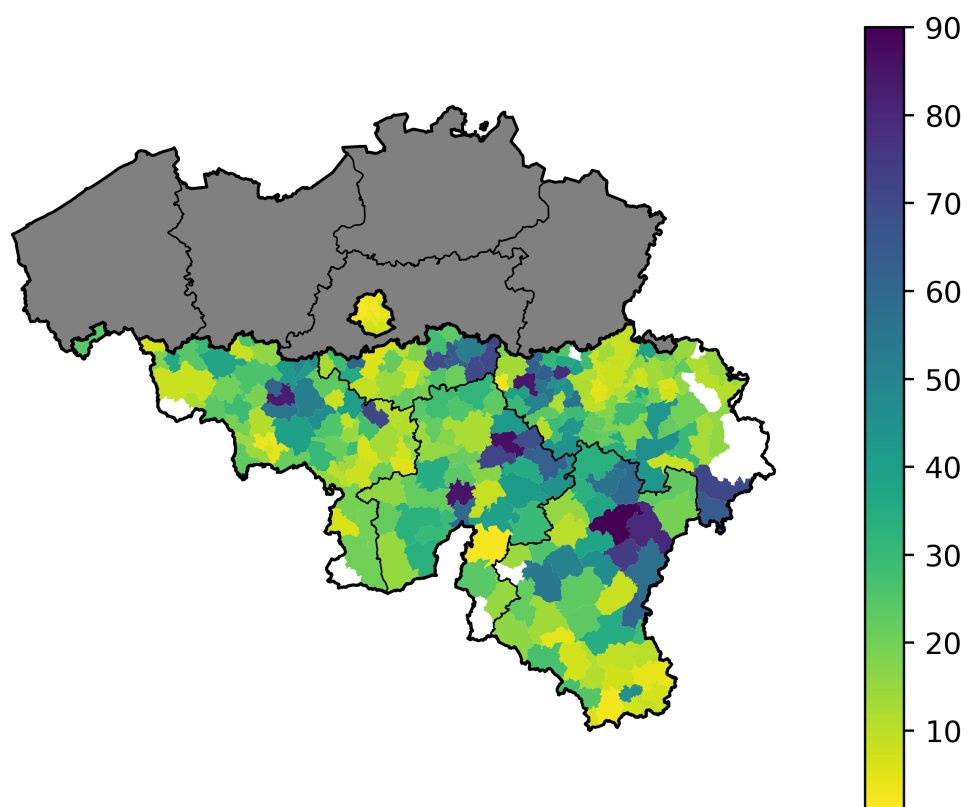
Pires communes à Bruxelles

1. Evere, avec en moyenne un trajet depuis l'école la plus proche 123.17 % plus long qu'un retour de l'école la plus accessible
2. Auderghem (70.39 %)
3. Uccle (56.85 %)
4. Woluwe-Saint-Lambert (52.30 %)
5. Ixelles (43.07 %)

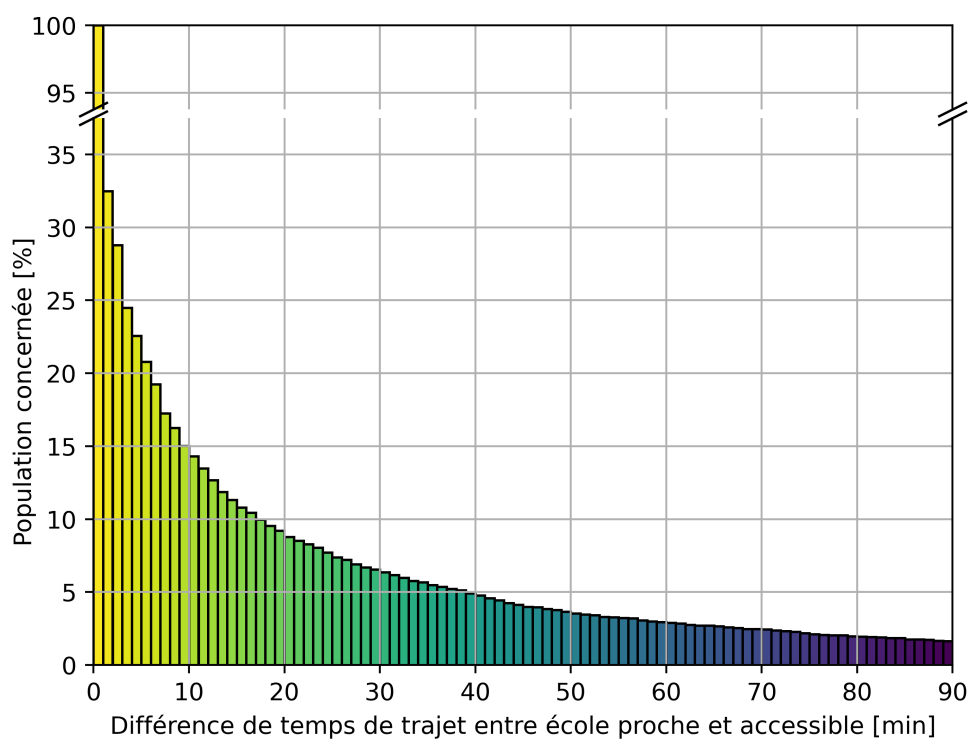
Pires communes en Wallonie

- | | |
|-----------------------------------|------------------------|
| 1. La Roche-en-Ardenne (150.00 %) | 6. Geer (142.45 %) |
| 2. Onhaye (147.77 %) | 7. Donceel (140.96 %) |
| 3. Saint-Hubert (147.14 %) | 8. Gesves (140.17 %) |
| 4. Braives (147.10 %) | 9. Assesse (133.93 %) |
| 5. Lens (144.19 %) | 10. Incourt (133.38 %) |

Complémentarité des temps absolus et relatifs Le problème de cette analyse en temps relatif est que nous perdons la valeur réelle de temps ajouté au trajet. En effet, une durée de trajet doublée peut vouloir dire qu'une personne doit voyager 6 minutes pour revenir de l'école la plus proche, alors que la plus accessible se trouve à 3 minutes de chez elle. Toutefois, une différence de 3 minutes de trajet n'est pas réellement problématique pour les enfants. Cet exemple montre bien l'importance d'étudier ces deux approches qui offrent des résultats complémentaires.

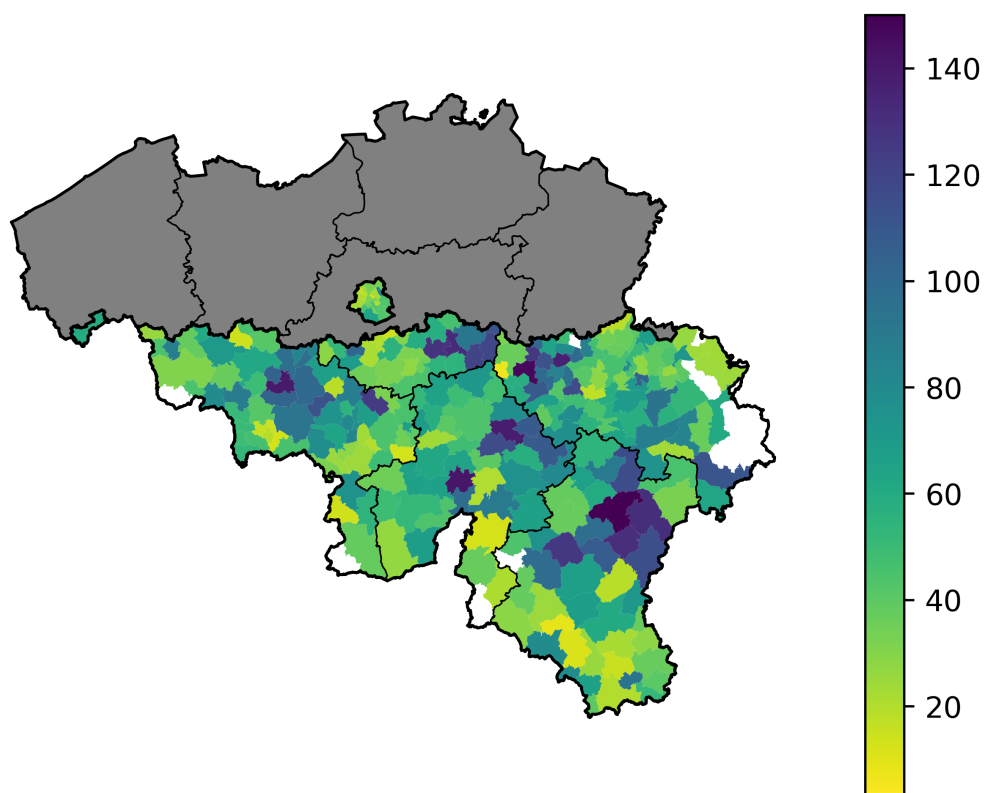


(a) Durée moyenne du trajet supplémentaire par commune

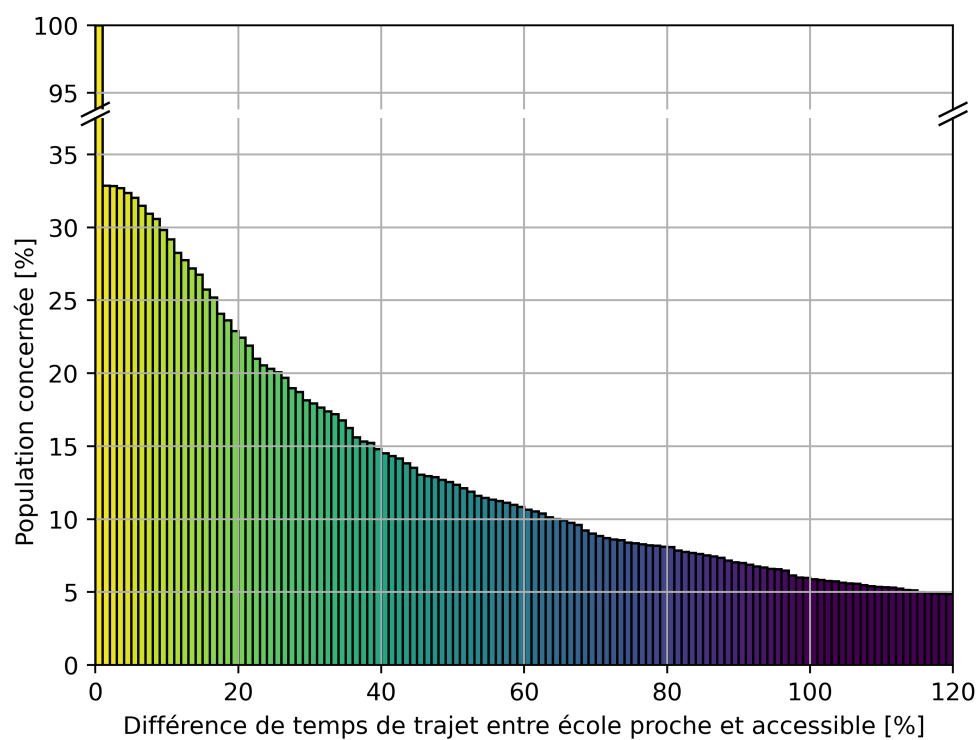


(b) Population impactée par minute de trajet supplémentaire

FIGURE 4.5 – Différence de temps de trajet : école proche vs école accessible [min]



(a) Différence de la durée du trajet (école la plus proche vs école la plus accessible) [%]



(b) Population impactée par pourcentage de temps de trajet supplémentaire (école proche vs école accessible)

FIGURE 4.6 – Différence de temps de trajet : école proche vs école accessible [%]

Tableau récapitulatif par province Le tableau 4.3 résume, pour chaque province, la part de population pour laquelle l'école la plus proche n'est pas la plus accessible ainsi que les différences absolues et relatives de la durée des trajets. Les moyennes sont bien sûr calculées uniquement pour cette part de population concernée (les différences de temps de 0 (minutes ou 0 %) ne sont donc pas prises en compte). Nous avons également imposé un maximum pour le trajet supplémentaire, afin d'éviter que les trop grandes valeurs ne faussent les moyennes. Ainsi, les différences de durée des trajets sont de maximum 90 minutes ou 150 % (valeurs arbitrairement choisies). Il est donc tout à fait normal que les maximums soient les mêmes pour la majorité des provinces.

TABLE 4.3 – Résumé par province : fraction de la population pour laquelle l'école la plus proche n'est pas la plus accessible + impact sur le temps de trajet absolu et relatif

Province	Population impactée [%]	Trajet supplémentaire [min]			Trajet supplémentaire [%]		
		Moyenne	Min	Max	Moyenne	Min	Max
Hainaut	36.43	20.71	1	90	57.68	1	150
Brabant wallon	36.34	26.46	1	90	55.56	1	150
Liège	35.28	18.15	1	90	51.55	1	150
Namur	32.12	28.21	1	90	62.10	1	150
Bruxelles	26.67	3.52	1	57	33.80	3	150
Luxembourg	22.81	23.91	1	90	55.87	1	150

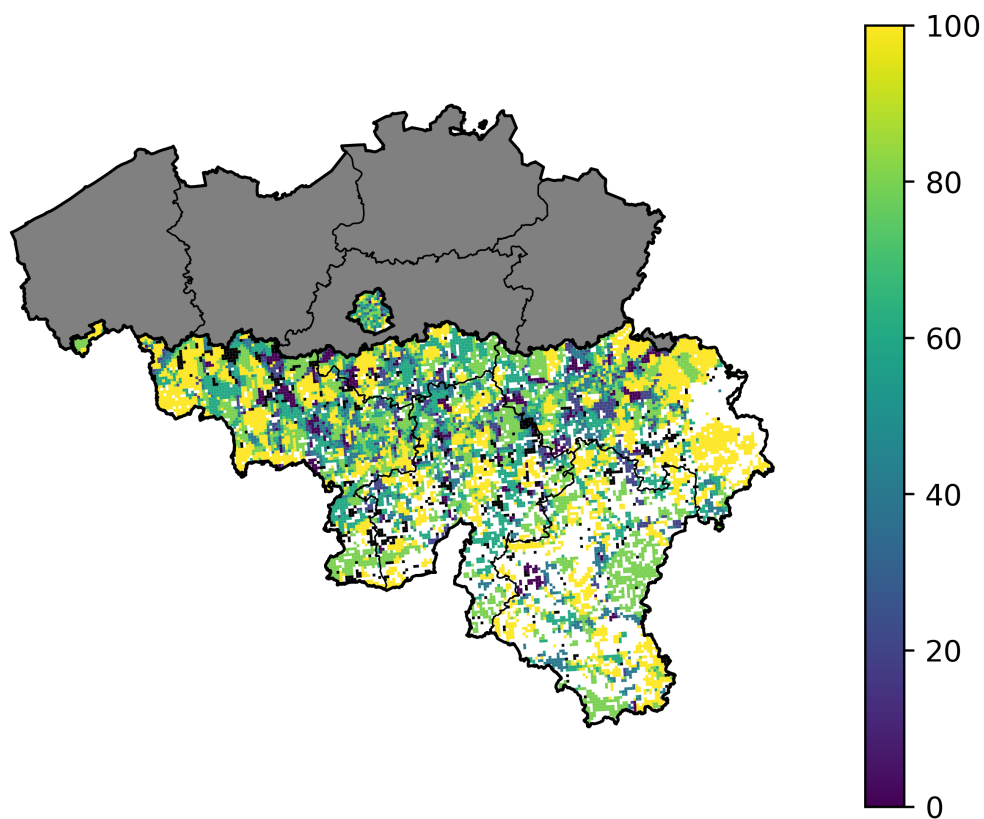
4.2.4 Évolution du phénomène avec le nombre d'écoles

Les analyses précédentes n'ont évalué les différences entre proximité et accessibilité qu'en comparant une école (la meilleure) pour chaque critère : l'école la plus accessible est-elle la même que l'école la plus proche. Nous pouvons donc nous demander comment ces différences évoluent lorsque plus d'écoles sont prises en compte pour la comparaison : la discordance entre les deux critères sera-t-elle atténuée ou au contraire augmentée ? C'est la question qui sera adressée dans cette section. Pour y répondre, nous nous sommes intéressés à la comparaison entre une liste des n écoles les plus accessibles avec les n plus proches. Afin de comparer ces listes, nous avons voulu exprimer leur degré de similitude selon la méthode schématisée à la figure 4.7.

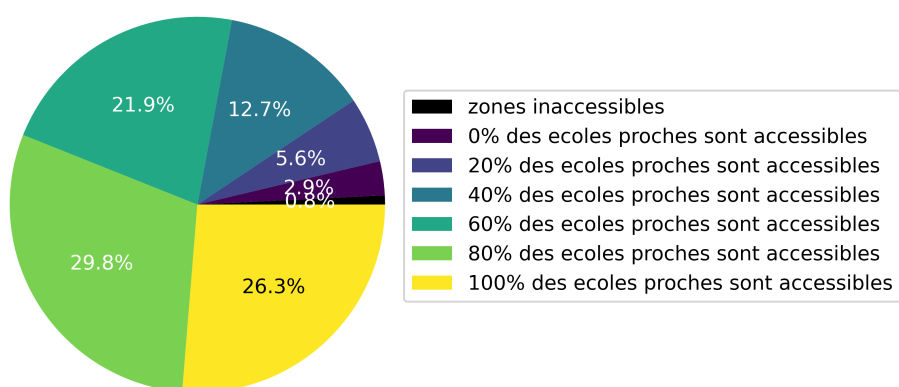
Écoles proches	A	B	C	D	E	
distance [km]	1	1.3	1.5	1.9	2.3	
Écoles accessibles	C	F	A	D	G	H
temps de trajet [min]	6	6	8	10	10	10

$$\rightarrow \text{similitude} = \frac{3}{5} = 66.7\%$$

FIGURE 4.7 – Exemple : calcul de la similitude entre écoles proches et accessibles



(a) Les 5 écoles les plus proches (échelle : similitude des critères [%])



(b) Population concernée par degré de similitude

FIGURE 4.8 – Comparaison des 5 écoles les plus proches vs les 5 plus accessibles

Nous obtenons cette similitude en calculant la proportion d'écoles proches qui sont également présentes dans la liste des écoles accessibles. Dans ce cas-ci, 3 écoles proches (A, C et D) sur 5 sont accessibles ; nous estimons donc la similitude des listes à 66.7 %. Ces calculs ont été effectués pour n allant de 1 à 5, et $n = 10$.

Nous ne présentons ici que le cas où $n = 5$, les résultats pour les autres valeurs sont fournis en section E.5. Notons également que, comme nous pouvons le voir dans la figure 4.7, la liste des écoles les plus accessibles peut être légèrement plus longue que la liste des écoles les plus proches. Ceci est dû au fait que plusieurs établissements peuvent correspondre à des temps de trajet équivalents, contrairement à leur distance à vol d'oiseau qui a peu de chance d'être la même. Nous avons donc fait le choix de conserver dans cette liste toutes les écoles ayant des temps de trajet égaux. Dans cet exemple, nous avons généré une liste de 5 écoles les plus accessibles. Toutefois, la sixième école, l'école H, a un temps de trajet égal à celui d'écoles comprises dans les 5 écoles les plus accessibles. Ainsi, H sera également considéré comme l'une des 5 écoles les plus accessibles. Cela n'affectera pas notre évaluation de la similitude des listes puisque le pourcentage est exprimé par rapport à la liste des écoles proches. Ainsi, si les 5 écoles les plus proches sont également reprises dans la liste des plus accessibles, le pourcentage calculé sera de 100 %, même si la liste des plus accessibles est beaucoup plus longue.

La figure 4.8 présente les résultats obtenus pour le cas $n = 5$ écoles proches. La répartition de ces valeurs sur le territoire belge est représentée à la figure 4.8a, tandis que la population concernée par degré de similitude est indiquée à la figure E.10. La carte peut être interprétée comme une version étendue de la figure 4.3, qui affichait des valeurs binaires étant donné que la comparaison concernait des listes d'un seul établissement. Ainsi, nous pouvons observer que même en étendant la comparaison à une liste de 5 écoles, il reste des zones pour lesquelles le critère de distance n'est absolument pas le reflet de l'accessibilité des écoles (0 % d'écoles communes). Par ailleurs, nous pouvons voir sur la figure E.10 que cela concerne presque 3 % des habitants de la Fédération Wallonie-Bruxelles. De plus, pour ce « top 5 », à peine plus d'une personne sur quatre se situe dans une zone où les critères de proximité et d'accessibilité sont équivalents (c'est-à-dire que les 5 écoles les plus proches font également partie des plus accessibles). Notons enfin que 22 % de la population a moins de 50 % de similitude entre sa liste d'écoles proches et sa liste d'écoles accessibles. Pour cette part de la population, le critère de distance ne reflète donc pas bien l'accessibilité des écoles malgré le fait que le nombre d'écoles utilisé dans la comparaison ait augmenté. Un rapide coup d'œil à la section E.5 de l'annexe E nous apprend que dans le cas d'une comparaison de listes de 10 écoles (proches et accessibles), les chiffres restent comparables puisque 18.1 % des personnes ont une similitude entre les coefficients inférieure ou égale à 50 %.

Ces résultats nous montrent donc que la divergence entre proximité et accessibilité est observée même lorsque la liste d'écoles à comparer grandit. Les critères du décret Inscription accordent des coefficients selon la place d'une école dans la liste des cinq plus proches. Au vu des différences qui subsistent entre proximité et accessibilité même pour une liste de 5 écoles, il semble opportun de proposer une modification du décret visant à valoriser l'accessibilité des écoles.

4.3 Comparaison des résultats avec le rapport de février 2022

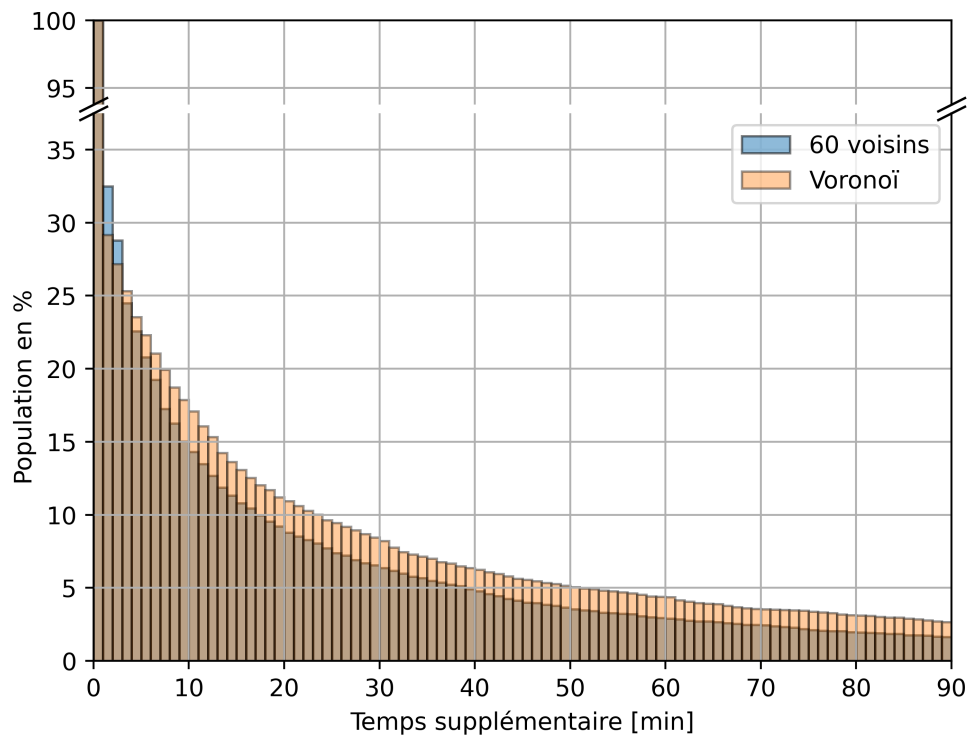
Les résultats présentés dans ce chapitre diffèrent légèrement de ceux présentés dans notre rapport de février 2022 (annexe B). Plusieurs raisons sont en cause.

La première est le changement d'hypothèse en ce qui concerne le domicile des élèves. D'autre part, nous avons également modifié le temps de marche autorisé entre des arrêts. En effet, nous l'avions initialement fixé à 30 minutes. Si l'autorisation d'un tel temps de marche est nécessaire pour prendre en compte les correspondances possibles sur un trajet, nous nous sommes rendu compte qu'autoriser une marche de 30 minutes semblait long pour des enfants. En effet, il ne s'agit que d'une partie de leur trajet, et un temps de marche entre l'école et l'arrêt de départ ainsi qu'entre le domicile et l'arrêt d'arrivée est déjà à prendre en compte. Nous étudions le trajet d'enfants qui peuvent se fatiguer plus facilement et sont moins susceptibles de couvrir de grandes distances à pied. Nous avons donc réduit ce maximum à 10 minutes de marche.

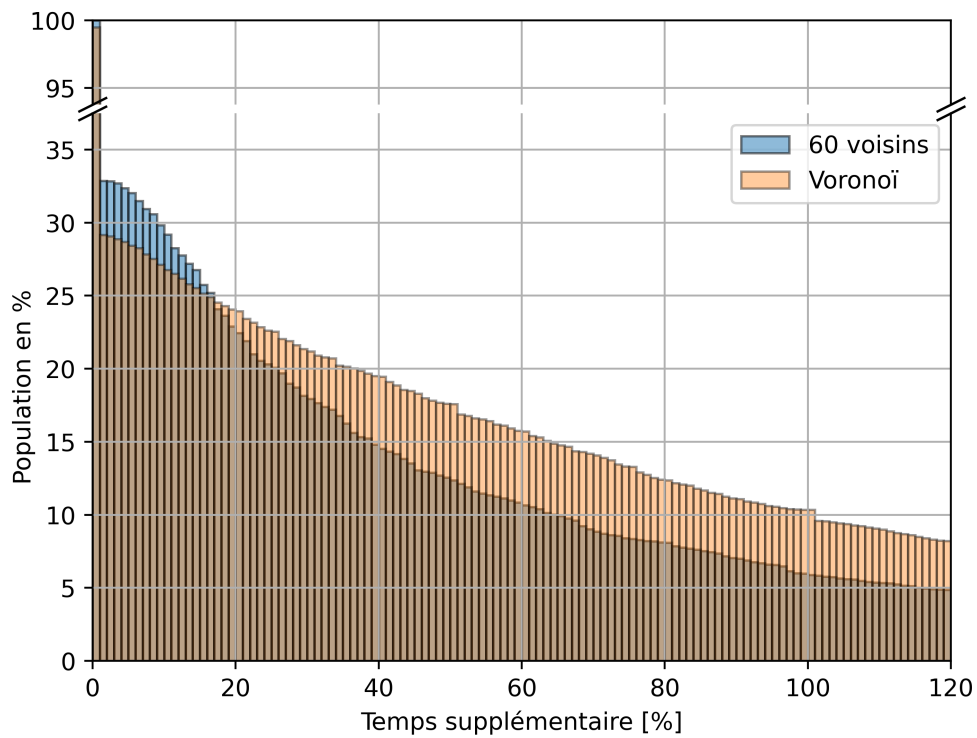
4.3.1 Relaxation de l'hypothèse de Voronoï

Comme expliqué au chapitre 3, nous avons exploré deux grandes hypothèses pour estimer la population concernée. Les premiers résultats ont été générés sous l'hypothèse que chaque personne cherche à rejoindre l'arrêt le plus proche de chez elle (voir hypothèse de Voronoï à la sous-section 3.3.1). Cela revenait à supposer que chaque élève habitait « sur son arrêt le plus proche ». Par la suite, l'hypothèse a été assouplie afin d'autoriser plusieurs stations proches du domicile. Les résultats obtenus par ces deux approches sont comparés à la figure 4.9. Les graphiques ont été générés avec la modification du temps de trajet maximum à pied présentée plus haut, celui-ci étant à présent fixé à 10 minutes. Les deux graphiques présentés dans notre rapport de février 2022 ont été comparés à leur nouvelle version. Il s'agit des histogrammes cumulatifs représentant les différences de temps de trajets entre un retour depuis l'école la plus accessible et l'école la plus proche. Ces différences sont exprimées respectivement en minutes (figure 4.9a) et en pourcentage de temps de trajet (figure 4.9b). Plusieurs observations peuvent être faites :

1. Dans les deux cas, la distribution obtenue est similaire. Ceci est particulièrement visible à la figure 4.9a et semble moins exact dans le second cas.
2. L'hypothèse de Voronoï mène à une sous-estimation de la population impactée. En effet, une différence de l'ordre de 4 % (de personnes concernées) est observée. Cette différence en termes de population concernée peut tout d'abord s'expliquer par le fait que, dans le cas de Voronoï, la population des cellules de 1 km^2 se trouvant sur deux zones différentes a été répartie uniformément entre ces zones (voir détails sous-section 3.3.1). Par contre, dans le cas de notre seconde hypothèse, nous avons supposé que l'entièreté de la population de la cellule utilisera le meilleur arrêt parmi les voisins du centre de la cellule.
3. L'hypothèse de Voronoï surestime la population impactée par des longs temps de trajet. Ceci est assez logique puisque les élèves sont obligés d'utiliser l'arrêt le plus proche de chez eux dans ce cas. Le passage à notre seconde hypothèse a pour conséquence de laisser aux élèves le choix de plusieurs trajets. Il est donc



(a) Résultat 1 : Temps de trajet supplémentaire [min]. Voronoï surestime la population impactée par les différences importantes et sous-estime la population impactée totale.



(b) Résultat 2 : Temps de trajet supplémentaire [%]. Voronoï surestime la population impactée par les différences importantes et sous-estime la population impactée totale.

FIGURE 4.9 – Comparaison de résultats : hypothèse de Voronoï vs plusieurs voisins autorisés

plus probable qu'ils trouvent un trajet plus court pour rentrer chez eux, et ainsi la proportion de personnes impactées par de « longs » trajets diminue.

Notons également que l'estimation des temps de trajet était moins précise sous l'hypothèse de Voronoï, étant donné que le temps de marche arrêt-domicile n'était pas du tout estimé, alors que dans sous la nouvelle hypothèse, une distance arrêt-domicile est approximée par la distance entre le centre de la cellule et l'arrêt.

4.4 Étude de l'indice composite

Jusqu'ici, nous avons montré que la proximité d'un établissement ne reflétait pas toujours son accessibilité. Nous avons également quantifié les conséquences d'une valorisation de la proximité, en évaluant le temps supplémentaire nécessaire pour se rendre à une école proche plutôt qu'une école accessible.

Dans cette section, nous allons explorer à quel point les critères de distance peuvent pénaliser des personnes souhaitant s'inscrire dans un établissement accessible mais pas proche du domicile.

4.4.1 Méthode de calcul du décret actuel : critère de distance

L'indice composite actuel pénalise une partie des écoles accessibles Afin de déterminer si la différence entre accessibilité et proximité avait entraîné des répercussions sur le calcul des indices composites, nous avons calculé pour chaque zone la valeur d'indice composite que le décret Inscription lui attribuerait. Nous avons ensuite exprimé cette valeur comme un pourcentage par rapport au plus grand indice composite rencontré. Le résultat est présenté sur les diagrammes circulaires de la figure 4.11, où la part du diagramme occupée par une valeur d'indice correspond au pourcentage de la population se voyant attribuer cet indice. Nous avons séparé les résultats en deux groupes :

- La partie de la population pour laquelle l'école la plus accessible n'est pas la plus proche (figure 4.10a). Cela correspond aux cellules rouges de la figure 4.3 et concerne un peu plus de 30 % de la population de la Fédération Wallonie-Bruxelles.
- La partie de la population pour laquelle l'école la plus accessible est également la plus proche (figure 4.10b). Il s'agit des cellules vertes de la figure 4.3 (un peu plus de 65 % de la population de la Fédération Wallonie-Bruxelles).

En comparant ces deux graphiques, il est clair que les valeurs d'indices sont généralement plus basses lorsque l'on s'intéresse aux écoles accessibles mais moins proches. En effet, comme l'on peut le voir à la figure 4.10b, il s'avère que les écoles accessibles et proches ont un indice composite représentant au moins 60 % de la valeur maximale. À la figure 4.10a, l'on peut pourtant observer qu'à peu près la moitié des personnes se trouvant « en zone rouge » se voit attribuer un indice plus faible, et a donc moins de chance d'obtenir une place dans cet établissement. Nous pouvons également observer que la proportion d'indices ayant une valeur maximale est 3 fois plus élevée dans le second groupe.

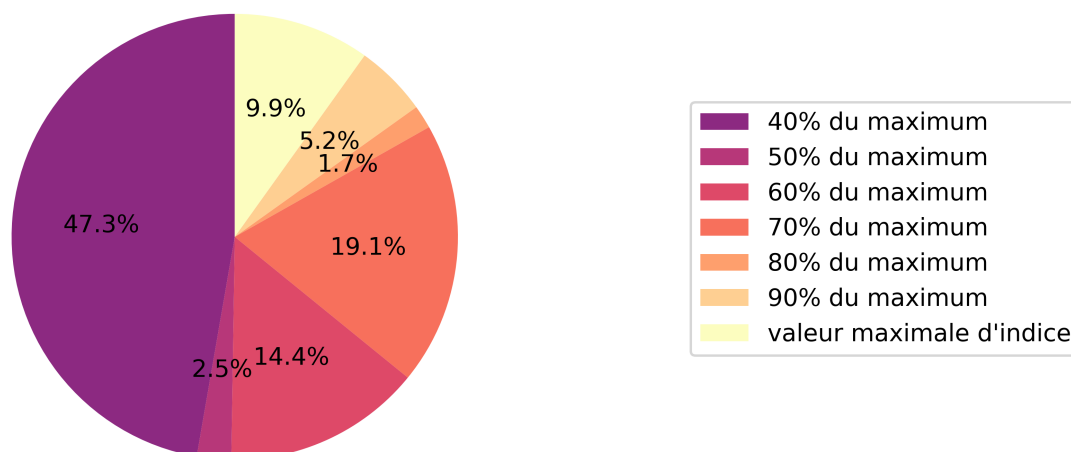
4.4.2 Impact d'un changement de critère : accessibilité

Un calcul basé sur des critères d'accessibilité augmente significativement la valeur de l'indice composite pour les écoles accessibles Nous allons maintenant analyser les différences entre l'indice composite actuellement utilisé par le décret Inscription (vol d'oiseau) avec les valeurs de cet indice lorsque les coefficients 2, 3 et 4 reflètent l'accessibilité de l'établissement. Afin de comparer les deux valeurs, nous avons choisi d'en calculer le ratio. Par exemple, pour une personne ayant un indice composite de 3.41 avec le critère de proximité et de 5.45 avec le critère d'accessibilité, nous dirons que la nouvelle valeur de l'indice est égale à $\frac{5.45}{3.41} = 160\%$ de l'ancien indice composite. Nous avons calculé le pourcentage de la population concernée pour chacun de ces ratios. Le résultat est illustré sous forme de diagramme circulaire à la figure 4.11a.

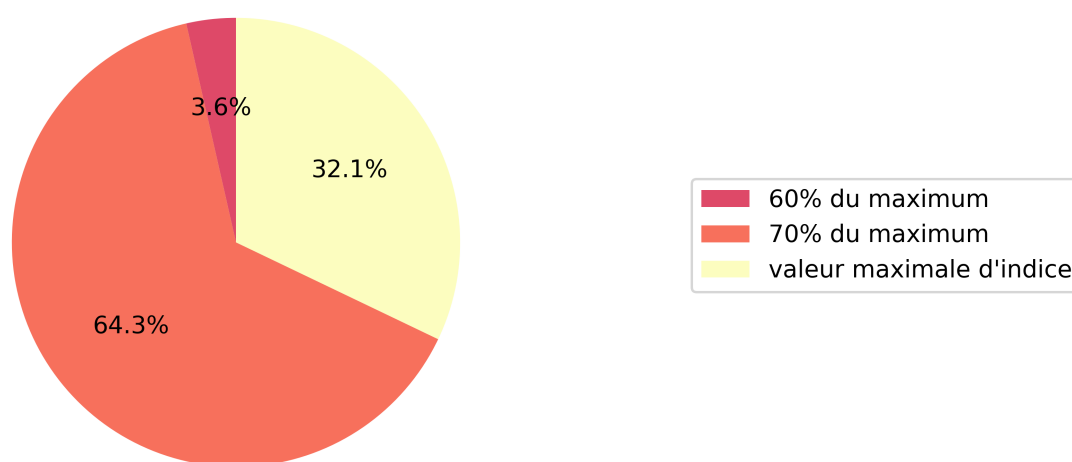
La majorité de la population située en zone rouge verrait une augmentation de son indice composite avec ce changement de méthode. Par ailleurs, pour plus de la moitié de ces personnes (57.3 % plus exactement), ce changement représenterait une augmentation de plus de 50 %. Des différences aussi importantes prouvent que le calcul actuel de l'indice composite est effectivement préjudiciable pour les personnes qui souhaitent obtenir une place dans un établissement facile d'accès même s'il n'est pas le plus proche de son domicile et les critères de distance diminuent leur chance d'y obtenir une place.

Valeur de l'indice composite des écoles proches mais non accessibles avec ces nouveaux critères Après avoir étudié à quel point les indices composites des établissements accessibles mais distants du domicile sont réduits en raison de ces critères de distance, nous avons calculé un indice composite basé uniquement sur des critères de proximité. Afin d'évaluer l'impact de ce changement de critère, nous avons à nouveau calculé un ratio. Ainsi, pour chaque cellule, le rapport entre ce nouvel indice composite basé sur l'accessibilité, et l'indice composite calculé sur base de distances a été calculé. La figure 4.11a représente la part de population concernée par valeur de ratio pour les écoles accessibles. L'augmentation est impressionnante, plus de la moitié des personnes ont un indice composite valant plus de 150 % de la valeur attribuée par le décret Inscription.

Pour ces personnes, les critères actuels sont donc particulièrement injustes. La figure 4.11b représente le même graphe, mais l'indice composite est cette fois calculé pour l'école la plus proche du domicile. Nous voyons donc que la plupart des indices composites sont revus à la baisse avec cette nouvelle méthode de calcul. Cependant, le changement est moins marqué que dans la figure 4.11a. En effet, les différences les plus marquées restent tout de même égales à 60 % de la valeur attribuée par le décret Inscription (donc une différence de 40 %), alors qu'en figure 4.11a, le changement maximal représentait une augmentation de 70 % de la valeur attribuée par le décret.

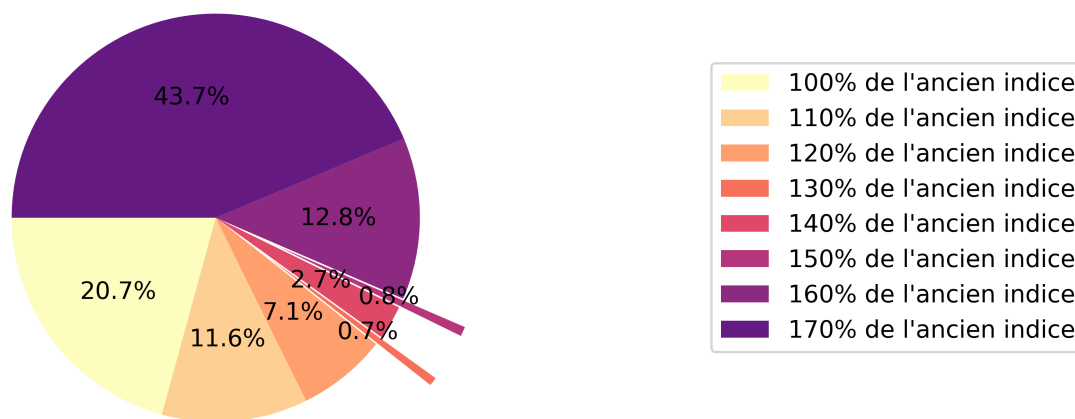


(a) Valeurs pour l'école la plus accessible mais pas la plus proche. Ainsi, pour 47.3% de la population en zone rouge, l'indice composite attribué vaut 40% de la meilleure valeur d'indice observée.

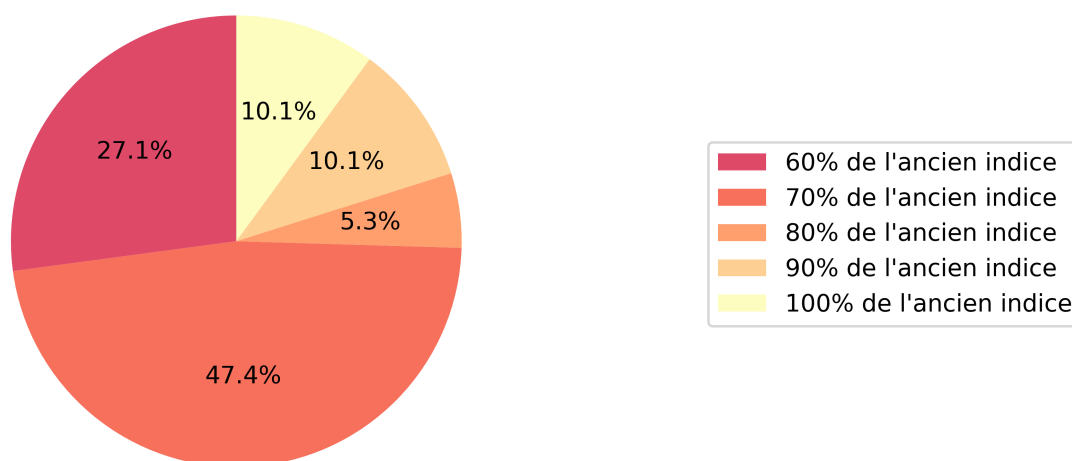


(b) Valeurs pour les écoles les plus accessibles qui sont aussi les plus proches. Ainsi, pour 64.3% de la population en zone verte, l'indice composite attribué vaut 70% de la meilleure valeur d'indice observée.

FIGURE 4.10 – Pourcentage de population par valeur d'indice composite (ancienne méthode, l'indice composite est exprimé en pourcentage par rapport à la valeur maximale observée)



(a) Pourcentage de population par valeur d'indice composite (nouvelle méthode de calcul) – écoles accessibles. Ainsi, pour 47.3 % de la population, un indice composite basé sur l'accessibilité vaut 70 % de plus que l'indice attribué sur base de la proximité.



(b) Pourcentage de population par valeur d'indice composite (nouvelle méthode de calcul) – écoles proches. Ainsi, pour 47.4 % de la population, un indice composite basé sur l'accessibilité vaut 30 % de moins que l'indice attribué sur base de la proximité.

FIGURE 4.11 – Indice composite calculé avec la nouvelle méthode : ratio entre le nouvel indice et l'ancien

4.4.3 Discussion avec un expert politique

Comme expliqué à plusieurs reprises dans ce travail, nous avons rédigé en février 2022 un rapport contenant nos premiers résultats. Celui-ci a été communiqué aux cabinets de la Ministre de l'Éducation et du Ministre-Président de la Fédération Wallonie-Bruxelles afin de prendre contact avec eux et éventuellement d'initier une collaboration afin d'approfondir au maximum nos recherches. Par la suite, nous avons eu l'occasion d'échanger avec un membre du cabinet du Ministre-Président lors d'une réunion le mardi 15 mars 2022. Au cours de cette discussion, nous avons pu lui présenter nos conclusions sur le manque de lien entre proximité et accessibilité et avons évoqué l'idée de remplacer la méthode de calcul des coefficients 2,3 et 4 par une approche basée sur l'accessibilité plutôt que sur les distances à vol d'oiseau. Notre interlocuteur a souligné que l'une des principales oppositions à un changement de calcul de l'indice composite visant à valoriser l'accessibilité est d'ordre légal. Les explications ci-dessous rapportent ce que nous avons personnellement compris et retenu de cette réunion, elles peuvent donc contenir des erreurs et/ou des imprécisions.

Les problèmes légaux d'un critère basé sur l'accessibilité Même si la mise en valeur de l'accessibilité d'une école semble être nécessaire pour une part non négligeable de la population, plusieurs points posent problème et empêchent donc son utilisation en pratique :

1. **Le critère d'accessibilité ne place pas tous les établissements secondaires sur un même pied d'égalité.** En effet, mettre en valeur l'accessibilité en transports d'une école peut pénaliser celles qui sont mal desservies. De même, des élèves résidant dans une zone comptant peu ou pas de transports en commun se verraient donc désavantagés par rapport aux autres. Ce point est particulièrement problématique puisque les coefficients utilisés doivent être égaux pour tous. La distance à vol d'oiseau a le mérite d'être une mesure totalement objective et de placer les élèves sur un pied d'égalité, même si ce critère est visiblement limité pour refléter la durée réelle des trajets.
2. **Le caractère subjectif des paramètres du modèle.** Comme détaillé au chapitre 3 et résumé en section 4.1, nous avons dû, pour obtenir nos résultats, imposer des hypothèses de travail dont une partie résulte de choix arbitraires. Par exemple, le choix de la vitesse de marche à considérer, la limite fixée pour la durée de marche vers un arrêt, le fait de ne considérer que les déplacements à pied et en transports sont toutes des hypothèses qui peuvent être contestées et mener à une requête en annulation auprès du Conseil d'État (pourquoi ne pas prendre en compte la possibilité d'utilisation d'un vélo ? pourquoi ne considérer qu'un départ à 16 h ? etc.)
3. **Le besoin régulier de mises à jour.** Ce dernier point est dû au fait que les horaires de transports en commun sont variables, contrairement à une distance à vol d'oiseau. Les trajets sont généralement mis à jour à raison d'une fois par an. Ainsi, la méthode de calcul d'accessibilité devrait être continuellement mise à jour. De plus, le choix de la date à partir de laquelle les nouveaux trajets sont pris en compte pourrait également être l'objet de débats et la source d'injustices (cela rejoint donc la discussion du point précédent). En effet, il est tout à fait possible qu'une école accessible au moment de l'inscription devienne inaccessible ultérieurement, ou inversement.

Conclusion de l'échange et possibilité de réévaluation de la réforme Il ressort de cette discussion que l'idée de remplacer les critères basés sur des distances à vol d'oiseau par des critères prenant en compte la durée des trajets a déjà été envisagée et étudiée par le cabinet. Cependant, elle a été rejetée au vu de la difficulté de mettre en place un coefficient objectif et inattaquable d'un point de vue juridique. En l'état actuel, il n'est donc pas question de modifier le décret Inscription dans ce sens. Toutefois, l'idée n'est pas écartée de manière définitive puisqu'une révision du décret est prévue pour 2026. La proposition pourrait donc être réévaluée à cette occasion.

4.4.4 Proposition alternative : ajout d'un nouveau coefficient

Au vu de la discussion du point précédent, il semble difficile d'envisager un remplacement total des coefficients de l'indice composite mesurant la distance avec les écoles.

Cependant, nos résultats ont également montré que les chances d'obtenir une place dans un établissement secondaire accessible sont fortement diminuées s'il n'est pas l'un des plus proches, car l'indice composite attribué est dans ce cas beaucoup plus bas. Plus de 30 % de la population de la Fédération Wallonie-Bruxelles est concernée. Il semble donc essentiel de corriger le décret Inscription afin que l'accessibilité soit prise en compte.

Ainsi, plutôt qu'un remplacement des coefficients existants, nous pourrions proposer l'ajout d'un nouveau coefficient, récompensant l'accessibilité. Ce coefficient ne serait d'application que si la valeur du coefficient 2 (distance avec l'école primaire) n'est pas maximale. Dans ce cas, un bonus de 1.7 serait attribué si l'école sélectionnée est la plus accessible (cette valeur a été choisie au vu des résultats de la figure 4.11a). Un tel critère ne semble pas être discriminant envers certains établissements puisqu'il n'est d'application que si le critère de distance n'est pas élevé. La problématique de la subjectivité des paramètres demeure, mais cette approche nous semble prometteuse et la plus susceptible d'aboutir à une amélioration satisfaisante du décret.

4.5 Outil de visualisation

La dernière partie de notre travail a pour but de créer un outil à destination des parents d'élèves et permettant de visualiser le résultat de nos analyses pour leur domicile.

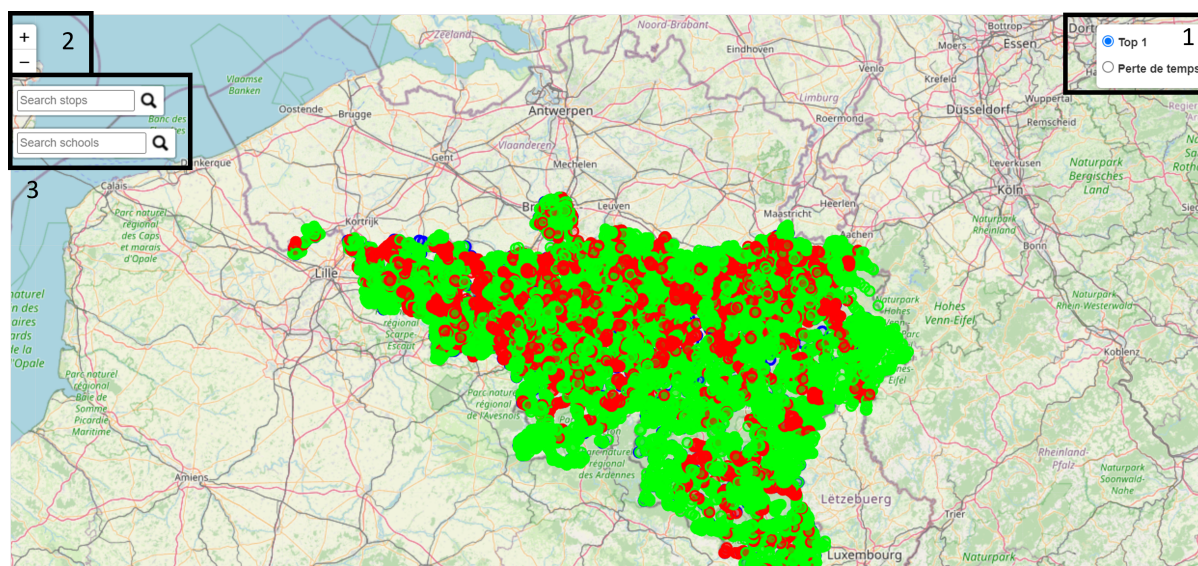
Notre première (infructueuse) approche a été de présenter nos résultats sous la forme d'une carte interactive au format HTML. Cela n'a malheureusement pas donné de résultat concluant. Pour cela, nous avons utilisé la librairie Folium² de Python. Le résultat est visible à la figure 4.12.

Le zoom initial de la carte est choisi initialement pour que l'entière de la Belgique soit visible (figure 4.12a). Les écoles secondaires ordinaires de la Fédération Wallonie-Bruxelles y sont indiquées en bleu. De plus, tous les arrêts y sont affichés. L'utilisateur peut choisir la métrique qu'il souhaite visualiser grâce au cadre 1 en haut à droite sur la carte. Celle-ci déterminera la taille des arrêts. Les couches top1, top2, top3, top4, top5 coloreront les arrêts en vert ou rouge selon que les n plus proches écoles et les n plus accessibles soient les mêmes ou non (ceci est donc l'équivalent de la carte binaire présentée à la figure 4.3). De plus, la couche « Perte de temps » colore les arrêts en fonction de la différence de

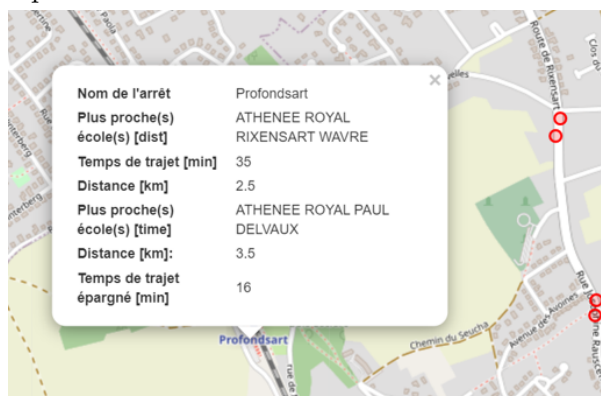
2. <https://python-visualization.github.io/folium/>

temps de trajet entre l'école la plus proche et l'école la plus accessible. En plus de ce choix d'affichage, l'utilisateur peut zoomer ou dézoomer à sa convenance (cadre 2 de la figure 4.12a, en haut à gauche) ainsi que chercher une école ou un arrêt particulier (cadre 3 de la figure 4.12a, en haut à gauche). Lors d'une recherche, la carte se centrera sur le point recherché et ajustera le zoom. Par ailleurs, plusieurs informations sont disponibles sur chacun des arrêts. La bulle de la figure 4.12b s'affiche lorsque l'utilisateur clique sur l'arrêt qui l'intéresse. Elle renseigne sur le nom de l'arrêt sélectionné, le nom de l'école qui en est la plus proche, le temps de trajet nécessaire pour rentrer depuis cette école ainsi que la distance en kilomètres entre l'arrêt et cette école. De plus, la ou les écoles les plus accessibles sont également listées, ainsi que les distances à vol d'oiseau avec l'arrêt, et la différence de temps de trajet pour rentrer depuis cette école plutôt que depuis l'école la plus proche.

Malheureusement, la carte ainsi obtenue était difficilement utilisable étant donné son volume. Folium intégrant toutes les données à visualiser dans le code de la page générant la carte, le temps de chargement était très élevé et les manipulations prenaient également



(a) Organisation de la carte : 1) Couleur des arrêts, 2) Zoom sur la carte, 3) Rechercher un arrêt ou une école spécifique



(b) Informations pour chaque station

FIGURE 4.12 – Outil de visualisation : carte interactive (Folium)

trop de temps. Le but de ce travail n'étant pas de travailler sur l'allègement d'une telle carte, mais bien de présenter nos résultats, nous avons décidé d'opter pour une autre méthode.

Ainsi, notre outil final est un tableau de bord, généré en utilisant l'outil Streamlit³. Streamlit est une librairie qui permet de créer aisément de petits sites internet en Python, le script créant la page étant rechargé chaque fois qu'un élément change sur celle-ci. La méthode s'est avérée effectivement utilisable en pratique et interactive. Sa description est présentée dans la sous-section 4.5.1.

4.5.1 Application avec Streamlit

Le tableau de bord que nous avons créé avec Streamlit permet à un utilisateur de calculer l'indice composite prévu par le décret Inscription pour l'établissement secondaire de son choix. Une comparaison avec un indice composite basé sur les temps de trajet est également fournie.

L'utilisateur indique les coordonnées géographiques de son domicile. Sur base de cela, notre outil affichera sur la carte l'emplacement du domicile, la liste des 5 écoles les plus proches à vol d'oiseau ainsi que les 5 écoles les plus accessibles. Ces écoles sont affichées à titre indicatif, pour éventuellement aider l'utilisateur à faire son choix sur l'établissement qui l'intéresse. Pour le calcul de l'indice composite, l'utilisateur doit ensuite entrer les noms de l'établissement secondaire qui l'intéresse, sa position dans son classement de préférence, ainsi que le nom de l'école primaire fréquentée par l'enfant et si une immersion est souhaitée. Lorsqu'il cliquera sur « calculer », notre outil calculera la valeur attribuée pour chacun des coefficients de l'indice composite selon les deux méthodes de calcul. Tous les coefficients sont calculés sauf le quatrième (le partenariat pédagogique), puisque les données sur ces partenariats ne sont pas disponibles en *open data* dans un format exploitable.

Ainsi, les valeurs d'indice composite calculées selon les critères du décret Inscription et selon notre approche sont affichées côte à côte afin de permettre une comparaison aisée de l'impact de notre changement de critère dans le cas de l'utilisateur. Par ailleurs, la distance à vol d'oiseau entre le domicile et l'école secondaire ainsi que le temps de trajet en transports en commun sont également indiqués. Un tracé du trajet est également affiché sur la carte.

Un tutoriel complet de l'utilisation de cet outil est fourni en annexe (annexe C).

3. <https://streamlit.io/>

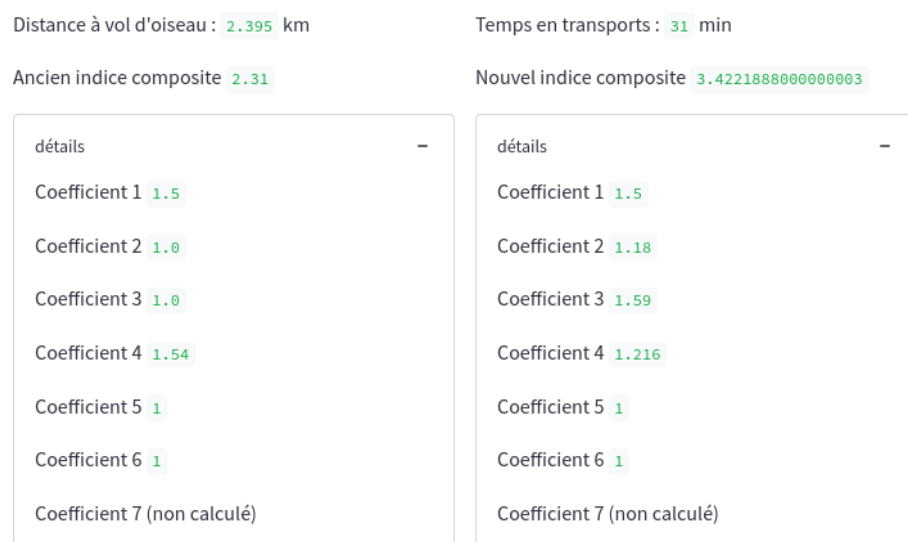
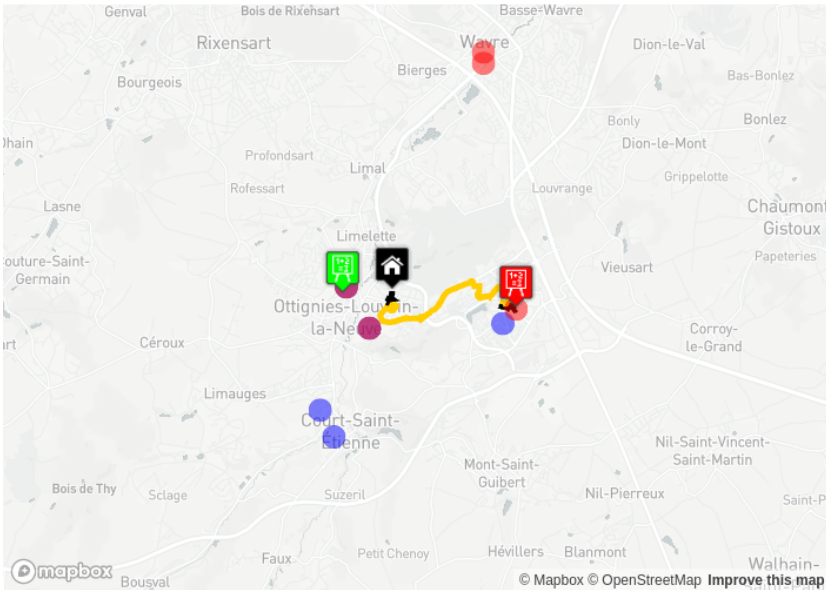


FIGURE 4.13 – Outil de visualisation : section résultats

Calcul de l'indice composite



 Domicile
  Écoles secondaires les plus accessibles

 École secondaire sélectionnée
  Écoles secondaires les plus proches

 École primaire sélectionnée
  Écoles secondaires proches et accessibles

Domicile de l'enfant

Longitude: - +
 Latitude: - +

École secondaire

École secondaire sélectionnée

Préférence

☐ Immersion secondaire

École primaire

École primaire sélectionnée

☐ Immersion primaire (depuis la 3e primaire au moins)

FIGURE 4.14 – Outil de visualisation : carte et données de l'élève

Chapitre 5

Discussion et Conclusion

Ce dernier chapitre sera l'occasion de prendre un peu de recul par rapport à notre travail et de mettre en évidence les aspects qui pourraient être améliorés ou les idées qui mériteraient d'être explorées dans de futurs projets de recherche.

5.1 Résumé de nos principaux résultats

Le premier objectif de ce mémoire était de mettre à profit les différentes sources de données à notre disposition afin d'étudier l'accessibilité (en transports en commun) des établissements secondaires ainsi que la pertinence des critères prévus par le décret Inscription pour calculer l'indice composite d'un élève. Ce travail a donc été l'occasion de mettre en lumière l'étendue de la discordance entre une mesure de proximité basée sur des distances à vol d'oiseau et les temps de trajet parcourus. Les conclusions que nous avons pu tirer sont les suivantes :

1. Il y a bien une distinction entre la proximité d'un établissement et son accessibilité : pour un peu plus de 30 % de la population, soit 1 personne sur 3, l'école la plus proche à vol d'oiseau n'est pas la plus accessible. Plus précisément, 34.4 % des personnes en région wallonne sont concernées contre 26.7 % en région de Bruxelles-Capitale.
2. Cette distinction entre les deux critères reste observable même lorsque l'on compare des listes d'écoles proches et accessibles de plus grande taille (nous avons effectué les analyses pour des listes allant jusqu'à 10 écoles proches).
3. En moyenne, la durée du trajet pour revenir de l'école la plus proche du domicile représente 22 minutes supplémentaires en Wallonie par rapport au trajet depuis l'école la plus accessible (ce qui équivaut en moyenne à une augmentation de la durée du trajet de 56 %). À Bruxelles, la différence est en moyenne de 4 minutes, ce qui représente une augmentation moyenne de 33 % par rapport au trajet le plus court.
4. De plus, pour un peu plus de 1 % de la population, il est impossible de revenir de l'école la plus proche en utilisant les transports en commun (sous les hypothèses que nous avons considérées). Le critère de distance à vol d'oiseau utilisé par le décret Inscription est donc particulièrement pénalisant pour ces personnes.
5. Nos analyses ont souligné le fait que les indices composites attribués à une école accessible mais qui n'est pas la plus proche sont significativement inférieurs aux autres

indices composites. Ces valeurs réduites d'indices composites sont problématiques puisqu'elles diminuent les chances d'obtenir une place dans l'établissement souhaité.

6. Lors de l'analyse des communes les plus touchées, nous avons pu observer certaines récurrences. Par exemple, à Bruxelles, Woluwe-Saint-Lambert est à la fois une des communes les plus impactées (avec environ 42 % de ses habitants pour lesquels il y a une discordance entre les critères), l'une des communes avec la plus grande différence de temps de trajet entre écoles proches et accessibles (6 minutes en moyenne) mais aussi l'une des communes dont le trajet supplémentaire représente la plus grande augmentation (52 % en moyenne).

De plus, au cours de nos analyses, nous avons pu observer des différences notables dans nos résultats d'une région à l'autre. Ceci s'explique entre autres par la différence de densité d'écoles dans ces régions mais aussi par la différence en termes de transports publics disponibles.

5.2 Limites de notre travail et voies d'amélioration

La validité de nos résultats est soumise à une série d'hypothèses listées et détaillées en section 4.1. Certaines de ces suppositions sont plutôt fortes et moyennement représentatives de la réalité. Ainsi, plusieurs aspects de notre approche pourraient être modifiés afin d'améliorer nos résultats.

Les navettes Certaines écoles organisent des navettes pour aider les élèves à se déplacer. Elles n'ont ici pas été prises en compte au vu du manque de données sur le sujet. Cependant, si de telles données pouvaient être obtenues ou approximées, l'idéal serait de les inclure dans les calculs afin de mieux évaluer l'accessibilité d'une école.

Prise en compte des différentes implantations d'un même établissement L'une des principales voies d'amélioration pour notre travail est d'ajouter une prise en compte des différentes implantations appartenant à un même établissement secondaire. En effet, le fichier FASE que nous utilisons afin d'identifier les écoles utilisées contient toutes les implantations d'une même école. Cependant, le fichier ne permet pas de différencier le cas d'une école secondaire possédant plusieurs adresses du cas de deux établissements qui ne sont aucunement liés mais qui possèdent des noms similaires. Ainsi, nous avons considéré chaque implantation d'école comme un établissement à part entière. Pour des travaux ultérieurs, il pourrait être intéressant de prendre cela en compte en générant une liste des implantations qui ne sont qu'un second site, ainsi que l'adresse de référence utilisée par le décret Inscription.

Adapter l'heure de départ aux horaires des écoles Ensuite, nous avons choisi une heure de départ de manière arbitraire (tous les élèves quittent leur école à 16 h). Même si cette hypothèse coïncide avec l'heure de pointe effectivement observée lorsque l'on considère les trajets école-domicile, imposer un départ à une heure fixe peut induire des temps d'attente aux arrêts parfois longs et inutiles. De plus, l'heure exacte de fin des cours varie d'une école à une autre, imposer une même heure de départ à tous les établissements peut donc en désavantager certains. Il serait alors souhaitable de disposer de données sur

les horaires de cours dispensés par chaque école. Il serait ainsi possible d'adapter l'heure de départ en fonction de l'école et par conséquent, d'obtenir des résultats plus pertinents. En l'absence de telles données, il pourrait également être envisagé d'autoriser un départ sur une certaine plage horaire (entre 15 h 30 et 17 h par exemple) et de sélectionner le trajet le plus optimal. Toutefois, cette approche a également ses limites. En effet, elle suppose que chaque école calque ses horaires sur ceux des transports en commun afin d'être aussi accessible que possible pour les élèves, ce qui est clairement loin de la réalité. Un tel modèle peut également s'avérer problématique dans le choix de l'heure à garder puisque l'accessibilité de l'école dépend aussi de la cellule où habite l'élève considéré. L'heure de fin des cours étant, à priori, la même pour tout le monde, une hypothèse doit être posée sur l'heure retenue : celle qui rend l'école accessible pour un maximum de monde ? Ou celle qui, en moyenne, minimise les temps de trajet de chacun ? Par conséquent, en l'absence de données précises par école, imposer une heure fixe de départ semble être la meilleure approche. Comparer les résultats pour différentes heures de départ serait toutefois intéressant.

Correction de l'hypothèse sur la population concernée Un autre point à discuter concerne le calcul de la population concernée par observation. En effet, nous avons utilisé les données sur l'entièreté de la population, sans distinction d'âge. Étant donné que, pour la plupart de nos analyses, nous nous sommes intéressés à la fraction de population concernée plutôt qu'à un nombre de personnes précis, nos résultats restent pertinents si la répartition des enfants fréquentant l'enseignement secondaire est similaire à la répartition de la population totale. Ceci n'est toutefois pas forcément le cas. Si de telles données peuvent être obtenues, il serait donc très intéressant de générer de nouveaux résultats ne prenant en compte que la population d'enfants de fin primaire/début secondaire sur le territoire. De plus, nous avons également supposé que l'entièreté de la population étudiée s'intéressait à l'enseignement francophone, ce qui n'est pas exactement le cas. Par exemple, il s'avère qu'à Bruxelles, 18 % des élèves étaient scolarisés dans l'enseignement néerlandophone pendant l'année scolaire 2016-2017 [43]. Nous n'avons pas non plus tenu compte des personnes habitant en Flandre et souhaitant s'inscrire dans l'enseignement francophone.

Prise en compte de moyens de transports alternatifs Pour ce travail, nous n'avons autorisé que des déplacements effectués à pied ou en transports en commun. Générer de nouveaux résultats en prenant en compte la possibilité d'utiliser le vélo ou la voiture pourrait donc être pertinent. Par exemple, un modèle plus complet pourrait laisser la possibilité d'un déplacement en voiture lorsque celui-ci représente un gain de temps significatif. Des logiciels tels que OSMR (Open Source Routing Machine) permettent de calculer le temps de trajet en voiture entre deux points assez efficacement et pourraient donc être mis à profit. Notons tout de même qu'étant donné l'heure de fin des cours, les possibilités de déplacement en voiture restent limitées puisque les parents travaillent généralement à cette heure, et les enfants de secondaire ne sont pas encore en âge de conduire. Un pourcentage maximum de personnes utilisant la voiture pourrait par exemple être imposé.

Génération de nouvelles métriques Nous avons ici essentiellement exploré les temps de trajet entre les écoles et le domicile. Une étude ultérieure pourrait explorer d'autres métriques, comme par exemple les kilomètres parcourus pour se rendre à l'école la plus proche ou à l'école la plus accessible. Sur base de cela, une analyse de l'impact écologique pourrait être menée, et serait un bon complément à nos conclusions. De même, détailler le temps de trajet en faisant la différence entre temps d'attente et déplacement pourrait également être intéressant, et permettre d'identifier d'éventuels problèmes (temps d'attente représentant la majorité du trajet par exemple). L'utilisation de la partie dynamique des données GTFS pourrait également être envisagée afin, par exemple, d'identifier les éventuelles récurrences dans les retards de certaines lignes, et d'en tenir compte dans le calcul d'accessibilité des écoles.

Exactitude des données Notons également que les données sont susceptibles de contenir des erreurs telles que des fautes dans les horaires ou de mauvaises coordonnées géographiques par exemple. Par ailleurs, il est probable que la répartition de la population ait varié depuis le recensement de 2018 sur lequel se basent nos analyses.

5.3 Conclusion et perspectives

En conclusion, ce mémoire a été l'occasion de mettre en évidence les faiblesses d'un critère de distance qui ne reflète pas correctement l'accessibilité d'un établissement scolaire. Ce critère est cependant utilisé pour calculer 3 des 7 coefficients de l'indice composite. Nous avons proposé un remplacement de ceux-ci au profit d'une mise en valeur de l'accessibilité des écoles, et donné un aperçu des changements d'indices qui pourraient ainsi être obtenus. Cependant, de plus amples échanges avec un expert politique ont permis de mettre en évidence que certains problèmes d'ordre légaux devaient également être pris en compte. Une manière plus équitable de valoriser l'accessibilité pourrait donc consister à la mise en place d'un nouveau coefficient bonus, similaire au coefficient 6, et accordant un bonus dans le cas où l'élève souhaite s'inscrire en immersion. Ce nouveau coefficient serait ainsi accordé uniquement aux personnes n'ayant pas de bonnes valeurs pour les coefficients 2 et 3. En effet, cela permettrait de rehausser l'indice composite d'un élève souhaitant s'inscrire dans un établissement accessible (mais pas proche du domicile) sans pour autant désavantager les écoles moins bien desservies.

Notons également que le décret Inscription a été promulgué pour tenter de réguler les inscriptions, mais le véritable problème réside dans les inégalités entre les écoles (la qualité d'enseignement qui y est dispensé, notamment). Il en résulte des inégalités dans les demandes d'inscription ainsi qu'un manque de place dans les établissements les plus prisés. Ainsi, des efforts de remise à niveau de certains établissements devraient également être mis en place.

En attendant, améliorer le décret Inscription afin qu'il soit plus en harmonie avec les besoins des familles nous semble urgent et essentiel. Nos recherches ont mis en évidence qu'en l'état actuel, les coefficients utilisés pour le calcul de l'indice composite pénalisent une part non négligeable des personnes souhaitant s'inscrire dans un établissement accessible pour eux. Même si un changement du décret n'est pas envisagé dans l'immédiat, une révision est prévue d'ici 2026. Il nous semble donc important d'engager d'ici là des discussions dans l'espoir de parvenir à une solution satisfaisante.

Bibliographie

- [1] Communauté FRANÇAISE. *Décret définissant les missions prioritaires de l'enseignement fondamental et de l'enseignement secondaire et organisant les structures propres à les atteindre*. Article 1. URL : https://www.gallilex.cfwb.be/fr/leg_res_01.php?ncda=21557&referant=102&bck_ncda=21557&bck_referant=100.
- [2] FÉDÉRATION WALLONIE-BRUXELLES. *Le décret actuel*. URL : <https://inscription.cfwb.be/le-decret-actuel/>. [Consulté le 30 avril 2022].
- [3] Nathalie RYELANDT. *Les décrets « inscriptions » et « mixité sociale » de la Communauté française*. URL : <https://www.cairn.info/revue-courrier-hebdomadaire-du-crisp-2013-23-page-5.htm>. [Consulté le 26 octobre 2021].
- [4] FÉDÉRATION WALLONIE-BRUXELLES. *L'indice composite*. URL : <https://inscription.cfwb.be/lindice-composite/>. [Consulté le 25 avril 2022].
- [5] BELGA. *Inscriptions dans le secondaire : le nouveau décret est définitivement validé*. URL : <https://www.lesoir.be/417638/article/2022-01-12/inscriptions-dans-le-secondaire-le-nouveau-decret-est-definitivement-valide>. [Consulté le 30 avril 2022].
- [6] FÉDÉRATION WALLONIE-BRUXELLES. *Les modifications du décret*. URL : <https://inscription.cfwb.be/le-decret/les-modifications-du-decret/>. [Consulté le 30 avril 2022].
- [7] Daphné V. OSSEL. *Décret inscription : « Dans 60% des écoles, on pourra être inscrit directement, selon Caroline Désir »*. URL : <https://www.rtb.be/article/decret-inscription-dans-60-des-ecoles-on-pourra-etre-inscrit-directement-selon-caroline-desir-10913087>. [Consulté le 3 février 2022].
- [8] Yvan DE BEAUFORT. *Décret inscription : manifestation ce dimanche...* URL : <http://www.yvandebeaufort.be/decret-inscription--manifestation-ce-dimanche-->. [Consulté le 30 avril 2022].
- [9] RTL INFO. *Manifestation de parents contre le décret Inscriptions à Bruxelles*. URL : <https://www.rtl.be/info/video/466740.aspx>. [Consulté le 30 avril 2022].
- [10] Service public fédéral MOBILITÉ ET TRANSPORTS. « Enquête Monitor sur la mobilité des belges ». In : (). Rapport de décembre 2019, Tiré de https://mobilit.belgium.be/fr/mobilite/mobilite_en_chiffres/enquetes_sur_la_mobilite_des_belges/monitor.
- [11] Opérateur de Transport de Wallonie. *TEC GTFS*. URL : <https://www.transportdata.be/fr/dataset/tec-gtfs>. [Téléchargé le 25 octobre 2021].

- [12] Société nationale des chemins de fer belges. *Public Data*. URL : <https://www.belgiantrain.be/fr/3rd-party-services/mobility-service-providers/public-data>. [Téléchargé le 25 octobre 2021].
- [13] Société des transports intercommunaux de Bruxelles. *GTFS*. URL : <https://opendata.stib-mivb.be/store/data>. [Téléchargé le 25 octobre 2021].
- [14] Vlaamse Vervoermaatschappij De Lijn. *De Lijn GTFS*. URL : <https://data.delijn.be/products/de-lijn-gtfs>. [Téléchargé le 25 octobre 2021].
- [15] Ministère de la Fédération Wallonie-Bruxelles. *Fichier signalétique des établissements d'enseignement de la Fédération Wallonie-Bruxelles (FASE)*. URL : <https://www.odwb.be/explore/dataset/signaletique-fase/>. [Téléchargé le 28 octobre 2021].
- [16] SPF Finances - Administration Générale de la Documentation Patrimoniale (AGDP). *Unités administratives*. URL : <https://www.geo.be/catalog/details/0e0b1866-7437-4b5a-9828-c5db2f47ad49>. [Téléchargé le 8 novembre 2021].
- [17] SPF Economie P.M.E. Classes moyennes et Energie - DG Statistique. *Répartition de la population sur une grille de cellules de 1 km² - Belgique*. URL : <https://www.geo.be/catalog/details/ea430cb8-2805-4c67-8392-9e2a47c1ef55>. [Téléchargé le 8 novembre 2021].
- [18] GOOGLE. *Présentation de GTFS statique*. URL : <https://developers.google.com/transit/gtfs>. [Consulté le 30 avril 2022].
- [19] MOBILITYDATA. *GTFS*. URL : <https://gtfs.org/>. [Consulté le 20 septembre 2021].
- [20] P. FORTIN. « Méthodologie de caractérisation d'un réseau de transport en commun ». Mém. de mast. (Mémoire de maîtrise, École Polytechnique de Montréal), 2016. Tiré de <https://publications.polymtl.ca/2202/>.
- [21] Frank SCHULZ. « Timetable information and shortest paths ». In : (jan. 2005).
- [22] Matthias MÜLLER-HANNEMANN et al. « Timetable Information : Models and Algorithms ». In : t. 4359. Jan. 2004, p. 67-90. ISBN : 978-3-540-74245-6. DOI : 10.1007/978-3-540-74247-0_3.
- [23] Edsger W DIJKSTRA et al. « A note on two problems in connexion with graphs ». In : *Numerische mathematik* 1.1 (1959), p. 269-271.
- [24] Robert SEDGEWICK et Kevin WAYNE. *Algorithms, 4th Edition*. Addison-Wesley, 2011, p. I-XII, 1-955. ISBN : 978-0-321-57351-3.
- [25] Richard BELLMAN. « On a routing problem ». In : *Quarterly of applied mathematics* 16.1 (1958), p. 87-90.
- [26] Lester R FORD JR. *Network flow theory*. Rapp. tech. Rand Corp Santa Monica Ca, 1956.
- [27] Edward F MOORE. « The shortest path through a maze ». In : *Proc. Int. Symp. Switching Theory, 1959*. 1959, p. 285-292.
- [28] Bernard ROY. « Transitivité et connexité ». In : *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences* 249.2 (1959), p. 216-218.

- [29] Robert W FLOYD. « Algorithm 97 : shortest path ». In : *Communications of the ACM* 5.6 (1962), p. 345.
- [30] Stephen WARSHALL. « A theorem on boolean matrices ». In : *Journal of the ACM (JACM)* 9.1 (1962), p. 11-12.
- [31] Donald B JOHNSON. « Efficient algorithms for shortest paths in sparse networks ». In : *Journal of the ACM (JACM)* 24.1 (1977), p. 1-13.
- [32] *Diagramme de Voronoï*. URL : https://fr.wikipedia.org/wiki/Diagramme_de_Vorono%C3%AF. [Consulté le 8 février 2022].
- [33] Markus MATERN. *Voronoi-Diagramm*. URL : https://commons.wikimedia.org/wiki/File:Voronoi_diagram.svg.
- [34] *Triangulation de Delaunay*. URL : https://fr.wikipedia.org/wiki/Triangulation_de_Delaunay. [Consulté le 27 mai 2022].
- [35] Peter J GREEN et Robin SIBSON. « Computing Dirichlet tessellations in the plane ». In : *The computer journal* 21.2 (1978), p. 168-173.
- [36] Michael Ian SHAMOS et Dan HOEY. « Closest-point problems ». In : *16th Annual Symposium on Foundations of Computer Science (sfcs 1975)*. IEEE. 1975, p. 151-162.
- [37] Steven FORTUNE. « A sweepline algorithm for Voronoi diagrams ». In : *Algorithmica* 2.1 (1987), p. 153-174.
- [38] Frédéric HÉRAN. « Des distances à vol d’oiseau aux distances réelles ou de l’origine des détours ». In : *Flux* 2 (2009), p. 110-121.
- [39] *Formule de haversine*. URL : https://fr.wikipedia.org/wiki/Formule_de_haversine. [Consulté le 25 mai 2022].
- [40] Sophie SCHOROCHOFF. « Évaluation et amélioration d’un réseau de transport en commun en terme de temps de trajet résidence-travail. » Mém. de mast. Ecole polytechnique de Louvain, Université catholique de Louvain, 2020. URL : <http://hdl.handle.net/2078.1/thesis:25243>. Prom. : Nijssen, Siegfried.
- [41] Guillaume DERVAL. *publictransportmap*. URL : <https://github.com/GuillaumeDerval/publictransportmap>. [Consulté le 26 octobre 2021].
- [42] *Carte choroplèthe*. URL : https://fr.wikipedia.org/wiki/Carte_choropl%C3%5C%A8the. [Consulté le 2 juin 2022].
- [43] perspective.brussels. *Données chiffrées*. URL : <https://beecole.brussels/fr/beecole/panorama-de-lecole-bruxelles/levolution-de-la-population-scolaire-bruxelles/donnees-chiffrees>. [Consulté le 3 février 2022].

Annexe A

Lecture et interprétations de nos graphes

Cette section a pour vocation d'aider le lecteur à interpréter les différentes figures qui sont présentées dans nos résultats. Chaque section décrira un type de graphe qui a été utilisé et fournira des explications sur la façon de l'interpréter.

A.1 Boxplot

Un graphe de type *boxplot* (également connu sous le nom « boîte à moustaches ») permet de représenter des statistiques descriptives sur des ensembles de données. Chaque ensemble de données est représenté par une boîte avec un segment au-dessus et en dessous. Cette boîte s'étend du premier quartile¹ des données au troisième quartile². Elle contient une barre horizontale colorée représentant la médiane³. Les segments au-dessus et en dessous de la boîte s'étendent jusqu'aux valeurs maximale et minimale observées, exception faite des *outliers*⁴ qui sont représentées par des points.

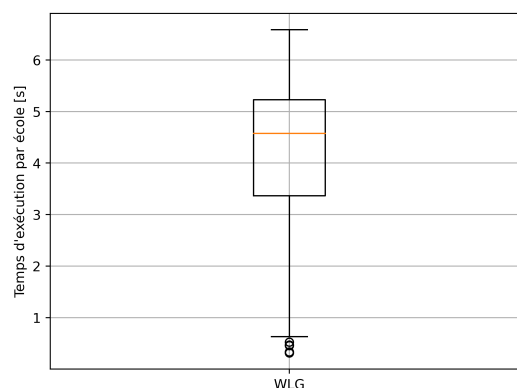


FIGURE A.1 – Exemple de boxplot

1. Valeur telle que 25 % des données lui sont inférieures.
2. Valeur telle que 75 % des données lui sont inférieures.
3. Valeur telle que la moitié des données se trouvent au-dessus, et l'autre moitié en dessous.
4. Valeurs exceptionnellement grandes ou petites comparées au reste des données.

Dans l'exemple de la figure A.1, on observe donc que la médiane vaut 4.6. La valeur maximale est de 6.6 et la valeur minimale est de 0.6, exception faite de quelques points inférieurs, considérés comme *outliers* (le point le plus bas est alors 0.3). Le premier quartile vaut 3.4, et le troisième quartile vaut 5.2.

A.2 Histogramme

Un histogramme est un type de graphe permettant de séparer les données en différentes classes, chaque classe étant définie comme l'ensemble des valeurs comprises entre le côté gauche de la colonne et le côté droit. La hauteur d'une colonne est d'autant plus grande que le nombre de données appartenant à la classe qu'elle représente est élevée.

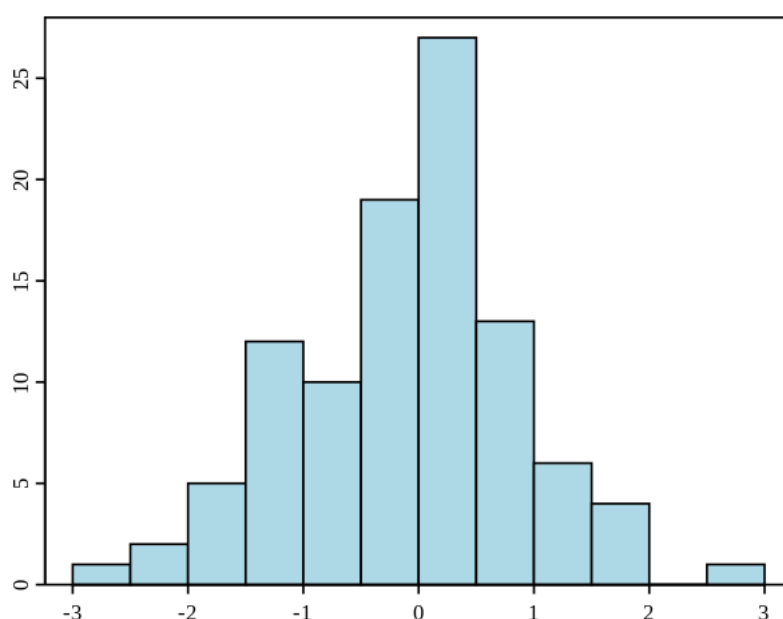


FIGURE A.2 – Exemple d'histogramme

Dans l'exemple de la figure A.2 (image tirée de⁵), on observe que 27 données ont une valeur comprise entre 0 et 0.5, alors que seules 5 données ont une valeur comprise entre -1.5 et -2.

A.2.1 Variante : Histogramme cumulatif

L'histogramme cumulatif est un cas particulier d'histogramme dans lequel chaque colonne contient le nombre d'éléments appartenant à cette colonne, ou à une des colonnes présentes à sa gauche. Ainsi, une colonne sera toujours plus grande ou égale à une colonne située à sa gauche sur le graphique, et plus petite que les colonnes se trouvant plus à la droite du graphe.

Nous avons utilisé des histogrammes cumulatifs inversés, le principe est donc inversé : la colonne la plus grande est celle qui se trouve le plus à gauche, et chaque colonne

5. https://commons.wikimedia.org/wiki/File:Histogram_example.svg

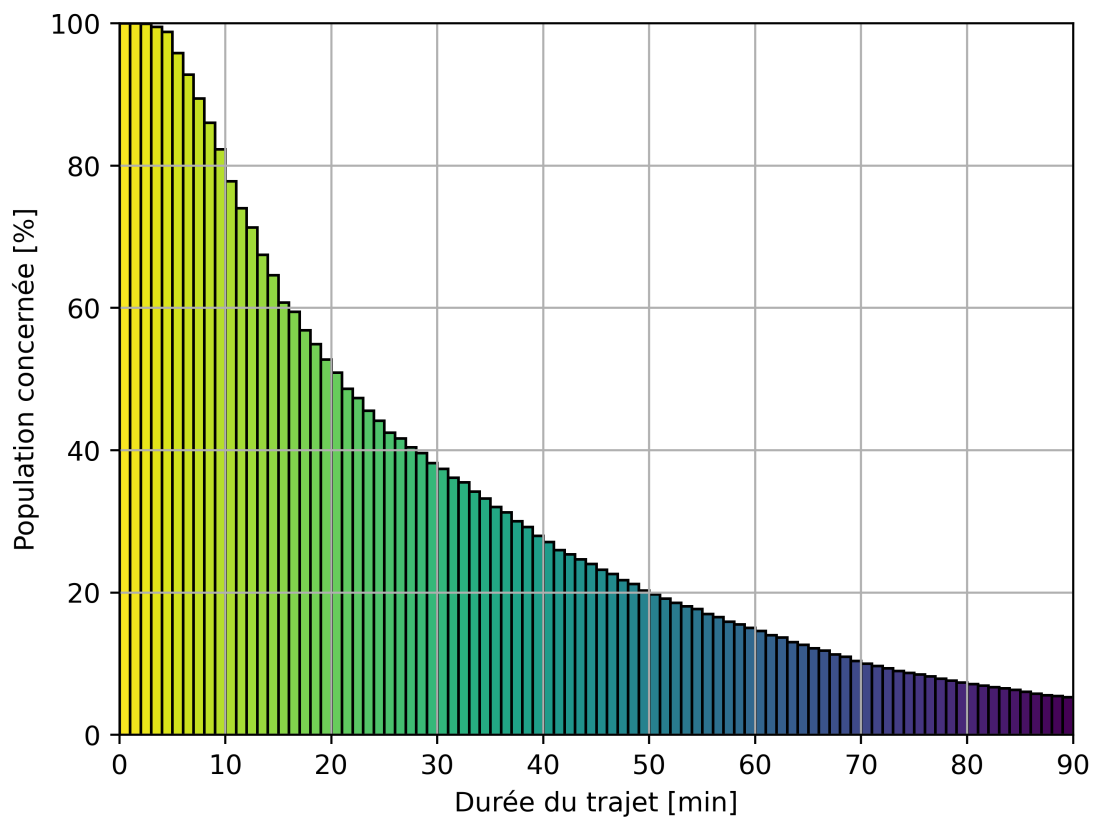


FIGURE A.3 – Exemple d'histogramme cumulatif inversé

représente le nombre d'éléments appartenant à cette colonne, ou à une des colonnes présentes à sa droite.

En termes d'interprétation, si nous prenons l'exemple de la figure A.3 nous pouvons voir que 100 % de la population a un temps de trajet de 0 minute *ou plus*. Ceci inclut donc les personnes ayant un trajet de 0 minute, mais aussi toutes celles qui ont un temps de trajet plus long, comme par exemple 15 minutes de trajet. De même, on peut voir grâce à la dernière colonne qu'environ 5 % de la population a un temps de trajet de minimum 90 minutes.

A.3 Carte choroplèthes

Une carte choroplèthe permet de quantifier la variation de la métrique d'intérêt sur un territoire donné. En l'occurrence, les cartes que nous avons présentées illustrent la variation des métriques d'une commune à l'autre, dans les régions wallonne et bruxelloise. La valeur de la métrique pour un territoire donné est indiquée par la couleur attribuée à ce territoire.

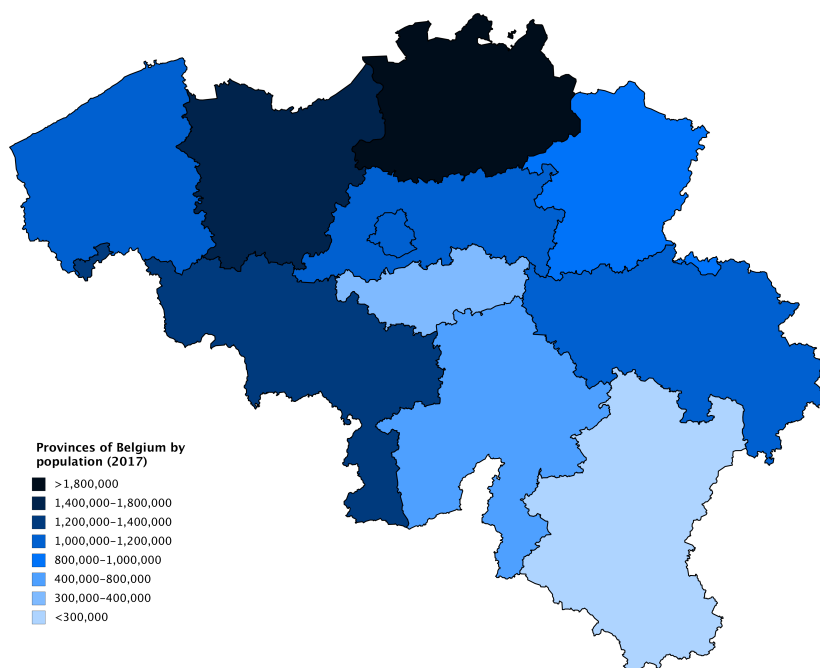


FIGURE A.4 – Exemple de carte choroplèthe

Sur l'exemple de la figure A.4 (image tirée de⁶), on observe que la province la plus peuplée de Belgique est la province d'Anvers, avec plus de 1 800 000 habitants. Pour les autres provinces, le dégradé de couleur varie selon l'importance de la population, jusqu'à la province de Luxembourg qui est la moins peuplée de Belgique. Elle apparaît en effet comme la plus claire des provinces.

A.3.1 Variante : Carte 1 km²

La carte de la Belgique en cellule de 1 km² est un cas particulier de carte choroplèthe dans laquelle les territoires sont les cellules pour lesquelles nous avons des données sur la population. Le principe est inchangé par rapport aux autres cartes choroplèthes : la couleur de chaque cellule reflète la valeur de la métrique qui lui est attribuée.

6. [https://commons.wikimedia.org/wiki/File:Provinces_of_Belgium_by_population_\(2017\).png](https://commons.wikimedia.org/wiki/File:Provinces_of_Belgium_by_population_(2017).png)

A.4 Diagramme circulaire

Un diagramme circulaire, ou *pie chart*, permet de représenter la répartition d'un ensemble de données en un nombre réduit de classes. La part qu'une classe particulière occupe dans le diagramme est proportionnelle au nombre de données appartenant à celle-ci. La somme de la part occupée par chaque classe vaut 100 %.

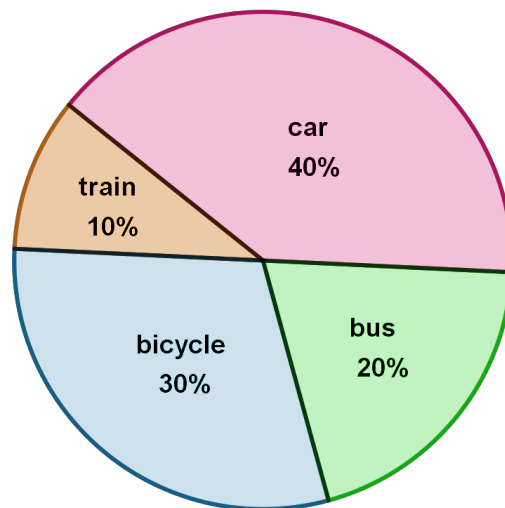


FIGURE A.5 – Exemple de diagramme circulaire

Dans l'exemple de la figure A.5 (image tirée de⁷), on observe que 40 % de la population se déplacent en voiture, tandis que 20 % préfèrent se déplacer en bus.

7. <https://mathemania.com/lesson/pie-chart/>

Annexe B

Rapport remis aux instances politique (02/2022)

Les pages qui suivent contiennent le rapport que nous avons envoyé en février 2022 au cabinet du Ministre-Président de la Fédération Wallonie-Bruxelles et au cabinet de la Ministre de l'Éducation. Nous nous sommes entretenus avec un conseiller du cabinet du Ministre-Président lors une réunion le 15 mars pour lui présenter nos résultats.

Impact du réseau de transport en commun sur les critères du décret Inscription

Auteurs / Contact

Cédric DE BONTRIDDER

Céline EVERAERT

Guillaume DERVAL

Pierre SCHAUS

cedric.debontridder@student.uclouvain.be

celine.everaert@student.uclouvain.be

gderval@uliege.be

pierre.schaus@uclouvain.be

Février 2022

UCLouvain
École polytechnique de Louvain



Résumé opérationnel

L'indice composite défini dans le décret Inscription se base notamment sur des critères de distance par rapports aux établissements scolaires. Ces critères utilisent la **distance à vol d'oiseau** entre l'habitation et l'école. **Ce document explore la discordance entre distance à vol d'oiseau et temps de parcours en transport en commun.** Les conclusions principales (soumises néanmoins à plusieurs hypothèses décrites dans le document complet) sont les suivantes :

Distance à vol d'oiseau versus temps de trajet Pour un habitant sur quatre en régions wallonne et bruxelloise, l'école la plus proche à vol d'oiseau n'est pas la plus accessible en transport en commun. Autrement dit, pour ces personnes, il existe au moins une école plus rapidement accessible en utilisant les transports en commun.

Augmentation relative du temps de trajet (figure 1) Pour 40% des personnes concernées, se rendre à l'école la plus proche en utilisant les transports en commun augmente de moitié le temps de trajet par rapport au trajet le plus rapide vers une école.

Temps de trajet supplémentaire (figure 2) Une personne sur 25 mettrait au moins 30 minutes supplémentaires pour rejoindre l'école la plus proche en distance plutôt que l'école la plus rapide à rejoindre.

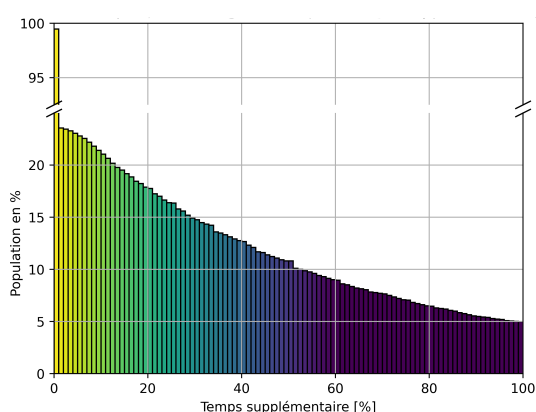


FIGURE 1 – Pourcentage de la population (des régions wallonne et bruxelloise) affectée par temps relatif de trajet supplémentaire (en % par rapport au trajet vers l'école la plus rapide à atteindre)

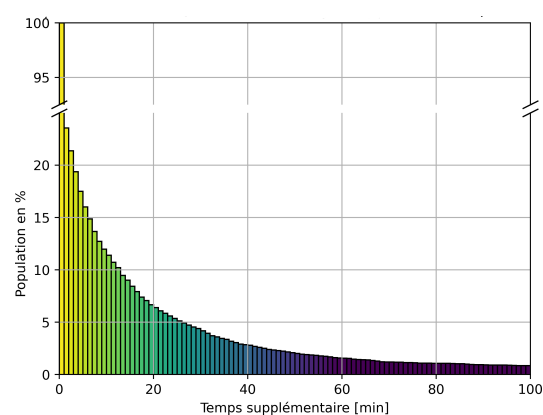


FIGURE 2 – Pourcentage de population (des régions wallonne et bruxelloise) affectée par minute de trajet supplémentaire pour se rendre à l'école la plus proche en distance par rapport à la plus rapide à atteindre

Ces différences peuvent donc amener des écoles pourtant plus accessibles à avoir des indices composites plus faibles.

Présentation

Nous sommes deux étudiants à l'École polytechnique de Louvain, respectivement en « Ingénieur civil en informatique » et « Ingénieur civil en sciences des données ». Dans le cadre de notre travail de fin d'études, nous analysons le réseau de transports en commun belge. Nous nous intéressons plus particulièrement à l'accessibilité des écoles en transports en commun ainsi qu'à une possible amélioration concernant le décret Inscription.

1 Introduction

La réforme du décret Inscription devrait entrer en vigueur pour la rentrée 2023. Il a notamment été annoncé que, dans cette nouvelle version, le critère de distance entre le domicile et l'établissement secondaire prendrait plus d'importance pour le calcul de l'indice composite attribué aux élèves [1]. Actuellement, la proximité est estimée sur base de la distance à vol d'oiseau entre le domicile et l'école [2]. Cependant, notre travail nous a permis de souligner qu'une telle mesure ne reflète pas le temps de trajet nécessaire pour se rendre à l'école en question. Nous pensons qu'un indice composite prenant en compte le temps de trajet serait plus pertinent pour départager les inscriptions.

Dans ce rapport, nous comparons les critères de la distance à vol d'oiseau et du temps de trajet. Nos résultats sont présentés dans la section 2 tandis que notre démarche pour obtenir ces résultats ainsi que les données qui ont été utilisées sont décrites dans la section 3.

Résumé des hypothèses de travail

Les résultats présentés dans ce rapport ne concernent que les écoles de l'enseignement secondaire ordinaire de la Fédération Wallonie-Bruxelles. De plus, seuls les déplacements à pied et en transports en commun ont été étudiés. Les temps de trajet ont été calculés pour un départ de l'école à 16h et comprennent le temps de marche, le temps d'attente ainsi que le temps passé dans les transports en commun. Nous avons considéré que les élèves cherchent à rejoindre l'arrêt qui est le plus proche de chez eux. De plus, le temps de marche entre l'école et un arrêt a été limité à 20 minutes.

2 La distance à vol d'oiseau n'est pas représentative du temps de trajet entre l'école et le domicile

Pour chaque arrêt dans les régions wallonne et bruxelloise, nous avons calculé le temps de trajet en transports en commun depuis l'école la plus proche (en terme de distance à vol d'oiseau). Celui-ci a été comparé au temps de trajet minimal pour rejoindre l'arrêt considéré depuis n'importe quelle école. Nous avons ainsi pu mettre en évidence que, pour plus d'une personne sur quatre, l'école pour laquelle le temps de trajet est minimum ne correspond pas à l'école la plus proche à vol d'oiseau. Pour ces personnes, la version actuelle du calcul de l'indice composite favorise donc des trajets plus longs. La figure 3 représente le pourcentage de population (la population de la région flamande n'est évidemment pas prise en compte) affecté par temps de trajet supplémentaire (exprimé en minutes, calculé pour un retour d'école à partir de 16h). Ainsi, environ une personne sur quinze voit son temps de trajet augmenté d'au moins 20 minutes.

Afin de mieux évaluer l'impact réel sur les temps de trajet de chacun, la même analyse peut être effectuée en terme de temps de trajet relatifs (trajet supplémentaire exprimé en pourcentage du trajet le plus rapide pour se rendre à une école). La figure 4 représente donc le pourcentage de population affecté par pourcentage de temps de trajet supplémentaire. Pour une personne sur dix, revenir de l'école la plus proche (à vol d'oiseau) augmente au moins de moitié le trajet le plus rapide vers une école! De plus, une personne sur vingt voit son temps de trajet au moins doublé.

Notons également que pour une petite partie de la population (0.5%), il n'est pas possible de rentrer en transports en commun en quittant l'école la plus proche à 16h. Ceci peut être dû à un temps de trajet à pied trop long pour rejoindre un arrêt (nos calculs n'en tiennent donc pas compte) ou à une absence de possible trajet en transports en commun pour ces personnes. Afin de s'assurer qu'il n'y avait pas de biais dû au réseau considéré, nous avons également effectué les calculs séparément pour les réseaux libre et officiel. Aucun changement significatif n'a été observé dans les résultats.

3 Démarche de calculs

3.1 Données utilisées

- **Liste des écoles** : Pour connaître le nom et l'emplacement des écoles de la Fédération Wallonie-Bruxelles, nous avons utilisé le fichier FASE [3]. Nous avons extrait de ce tableau les données relatives aux écoles dont le champ type_d_enseignement est indiqué comme Secondaire ordinaire.
- **Limites administratives** : Pour déterminer la région dans laquelle se trouve chaque arrêt de transports en commun, nous avons utilisé les unités administratives belges [4]. Ces cartes contiennent les différentes limites administratives belges : pays, régions, provinces, arrondissements et communes.
- **Horaires des transports en commun et emplacements des arrêts** : Nous avons utilisé les données GTFS mises à disposition du public par les différentes compagnies belges de transports en commun :
 - De Lijn : [5]
 - SNCB : [6]
 - STIB : [7]
 - TEC : [8]Ces données contiennent entre autres le nom et l'emplacement des différents arrêts ainsi que les horaires d'arrivée et de départ des transports en commun à ces arrêts. Tous les calculs sur ces données ont été effectués en supposant un départ de l'école le 26 octobre 2021 à 16h.
- **Population par zone** : Pour estimer le pourcentage de population concerné par un certain temps de trajet supplémentaire, nous avons utilisé la grille de répartition de la population belge [9]. Ce fichier comprend les données de population pour chaque cellule de la Belgique découpée en cellules carrées de 1 km².

3.2 Hypothèses de travail

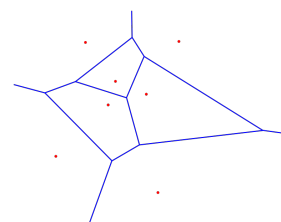
Afin de réaliser nos calculs, nous avons dû définir quelques hypothèses :

- Seuls les établissements indiqués comme « écoles secondaires ordinaires » ont été considérés dans nos calculs.
- Les temps de trajets concernent ici un retour au domicile en quittant l'école à 16h.
- Seuls les déplacements à pied et en transports en commun ont été pris en compte (ainsi que le temps d'attente à l'arrêt si nécessaire).
- Le temps de trajet à pied entre deux points peut être approximé proportionnellement à la distance à vol d'oiseau qui les sépare [10]. Dans le cadre de nos calculs, nous avons considéré une vitesse réduite de 3 km/h.
- Le temps trajet à pied pour rejoindre un arrêt depuis l'école est limité à 20 minutes. Ceci correspond donc à un rayon de 1 km autour de l'école.
- En quittant l'école, les élèves cherchent à se rendre à l'arrêt le plus proche de chez eux (en terme de distance à vol d'oiseau). Le trajet à pied entre le domicile et l'arrêt est négligé.
- Lors de l'estimation de la population impactée, l'entièreté de la population des régions wallonne et bruxelloise a été prise en compte. Nos observations ne se concentrent donc pas uniquement sur le nombre d'enfants concernés. De plus, nous avons supposé que la population est répartie de manière uniforme au sein de chaque cellule de 1 km² de la grille de répartition de la population belge.

3.3 Méthodologie

3.3.1 Calcul de la population impactée

Diagramme de Voronoï « En mathématiques, un diagramme de Voronoï est un pavage (découpage) du plan en cellules (régions adjacentes) à partir d'un ensemble discret de points appelés "germes". Chaque cellule enferme un seul germe, et forme l'ensemble des points du plan plus proches de ce



germe que d'aucun autre. » [12] Ainsi, sur la figure 5, chaque région de Voronoï, délimitée par les lignes bleues, correspond à l'ensemble des points du plan qui sont les plus proches des centres rouges.

Afin d'estimer le pourcentage de population concerné par un certain temps de trajet supplémentaire, nous avons considéré que chaque personne « habite » à l'arrêt le plus proche de chez elle. Il s'agit bien sûr d'une première approximation qui sera raffinée ultérieurement. Nous avons donc calculé les zones de Voronoï autour des arrêts, et avons estimé la population dans chaque région de Voronoï à partir de la grille de répartition de la population décrite précédemment. Lorsqu'une cellule de la grille de répartition appartient à plusieurs zones de Voronoï, la contribution de cette cellule à la population de chaque zone a été estimée en fonction de la proportion de la cellule présente dans chaque zone. Prenons l'exemple d'une cellule de la grille de répartition qui contiendrait 100 habitants et qui serait majoritairement présente sur la région 1, mais dont 10 % de la surface dépasserait sur la région 2. Dans ce cas, nous considérerions que 90 personnes se rendent à l'arrêt de la région 1, tandis que les 10 restantes iraient à l'arrêt de la région 2.

De cette façon, nous avons pu associer à chaque arrêt un nombre de personnes concernées, qui est plus tard utilisé dans la création de l'histogramme. Ensuite, le temps de trajet jusqu'à chaque arrêt a été calculé.

3.3.2 Calcul des temps de trajets

Les temps de trajet depuis les écoles jusqu'à chaque arrêt ont été calculés sur base des données GTFS décrites précédemment. L'algorithme calcule les plus courts chemins entre les écoles et les arrêts considérés. Il en résulte une grande matrice contenant tous ces temps de trajets, chaque ligne représentant les trajets vers un arrêt particulier. Il suffit donc de sélectionner le temps de trajet minimum pour trouver l'école la plus proche en transports en commun.

La distance à vol d'oiseau entre un arrêt et une école est calculée à partir de leurs coordonnées. Le temps de trajet entre cette école et l'arrêt est récupéré depuis la matrice calculée décrite précédemment.

4 Conclusion

Nos résultats ont souligné la faiblesse d'un critère simplement basé sur la distance à vol d'oiseau entre le domicile et une école. En effet, la formule actuelle est susceptible de défavoriser l'inscription des enfants dans une école plus accessible en transports en commun. Les différences de temps de trajets ne sont d'ailleurs pas négligeables puisqu'une personne sur vingt met le double du temps pour revenir de l'école considérée comme la plus proche, comme vu dans la section 2. Ceci ayant été souligné, nous comptons désormais confirmer la robustesse de ces résultats en prenant en compte les cinq écoles les plus proches. Notons également que, pour le moment, nos calculs supposent que chaque personne se rend à l'arrêt le plus proche de chez elle. Il s'agit bien sûr d'une approximation pour nos premiers résultats, par la suite nous voulons prendre en compte plusieurs arrêts possibles à proximité du domicile afin d'avoir des estimations des temps de trajet plus précises.

Finissons par quelques remarques qui nous paraissent pertinentes à souligner.

D'abord, nous nous sommes concentrés sur le troisième critère de l'indice composite, c'est-à-dire sur la distance entre le domicile et l'établissement secondaire. Il est évident que la même démarche peut également être appliquée au critère de la distance avec l'école primaire.

Ensuite, nous avons utilisé les données de population des régions wallonne et bruxelloise. Celles-ci incluent donc celle de la Communauté germanophone, moins susceptible de s'inscrire dans des écoles de la Fédération Wallonie-Bruxelles. À l'inverse, nous n'avons pas tenu compte des personnes résidant en Flandre et souhaitant s'inscrire dans l'enseignement francophone. De plus, à Bruxelles, nous avons considéré que la totalité de la population souhaite s'inscrire dans l'enseignement francophone. Or, environ 18 % des élèves étudiant en région bruxelloise étaient scolarisés dans l'enseignement néerlandophone pendant l'année scolaire 2016-2017 [13]. Nous rappelons également que nos chiffres ne se concentrent pas uniquement sur la proportion d'élèves impactés, mais bien sur l'entièreté de la population.

Références

- [1] Daphné V. OSSEL. *Décret inscription : « Dans 60% des écoles, on pourra être inscrit directement, selon Caroline Désir »*. URL : <https://www.rtb.be/article/decret-inscription-dans-60-des-ecoles-on-pourra-etre-inscrit-directement-selon-caroline-desir-10913087>. [Consulté le 3 février 2022].
- [2] Communauté française. *Décret définissant les missions prioritaires de l'enseignement fondamental et de l'enseignement secondaire et organisant les structures propres à les atteindre*. Article 79/2, alinéa 1er, 2°. URL : https://www.gallilex.cfwb.be/fr/leg_res_01.php?ncda=21557&referant=102&bck_ncda=21557&bck_referant=100.
- [3] Ministère de la Fédération Wallonie-Bruxelles. *Fichier signalétique des établissements d'enseignement de la Fédération Wallonie-Bruxelles (FASE)*. URL : <https://www.odwb.be/explore/dataset/signaletique-fase/>. [Téléchargé le 28 octobre 2021].
- [4] SPF Finances - Administration Générale de la Documentation Patrimoniale (AGDP). *Unités administratives*. URL : <https://www.geo.be/catalog/details/0e0b1866-7437-4b5a-9828-c5db2f47ad49>. [Téléchargé le 8 novembre 2021].
- [5] Vlaamse Vervoermaatschappij De Lijn. *De Lijn GTFS*. URL : <https://data.delijn.be/products/de-lijn-gtfs>. [Téléchargé le 25 octobre 2021].
- [6] Société nationale des chemins de fer belges. *Public Data*. URL : <https://www.belgiantrain.be/fr/3rd-party-services/mobility-service-providers/public-data>. [Téléchargé le 25 octobre 2021].
- [7] Société des transports intercommunaux de Bruxelles. *GTFS*. URL : <https://opendata.stib-mivb.be/store/data>. [Téléchargé le 25 octobre 2021].
- [8] Opérateur de Transport de Wallonie. *TEC GTFS*. URL : <https://www.transportdata.be/fr/dataset/tec-gtfs>. [Téléchargé le 25 octobre 2021].
- [9] SPF Economie P.M.E. Classes moyennes et Energie - DG Statistique. *Répartition de la population sur une grille de cellules de 1 km² - Belgique*. URL : <https://www.geo.be/catalog/details/ea430cb8-2805-4c67-8392-9e2a47c1ef55>. [Téléchargé le 8 novembre 2021].
- [10] Frédéric HÉRAN. « Des distances à vol d'oiseau aux distances réelles ou de l'origine des détours ». In : *Flux 2* (2009), p. 110-121.
- [11] Markus MATERN. *Voronoi-Diagramm*. URL : https://commons.wikimedia.org/wiki/File:Voronoi_diagram.svg.
- [12] *Diagramme de Voronoï*. URL : https://fr.wikipedia.org/wiki/Diagramme_de_Vorono%C3%AF. [Consulté le 8 février 2022].
- [13] perspective.brussels. *Données chiffrées*. URL : <https://beecole.brussels/fr/be-ecole/panorama-de-lecole-bruxelles/levolution-de-la-population-scolaire-bruxelles/donnees-chiffrees>. [Consulté le 3 février 2022].

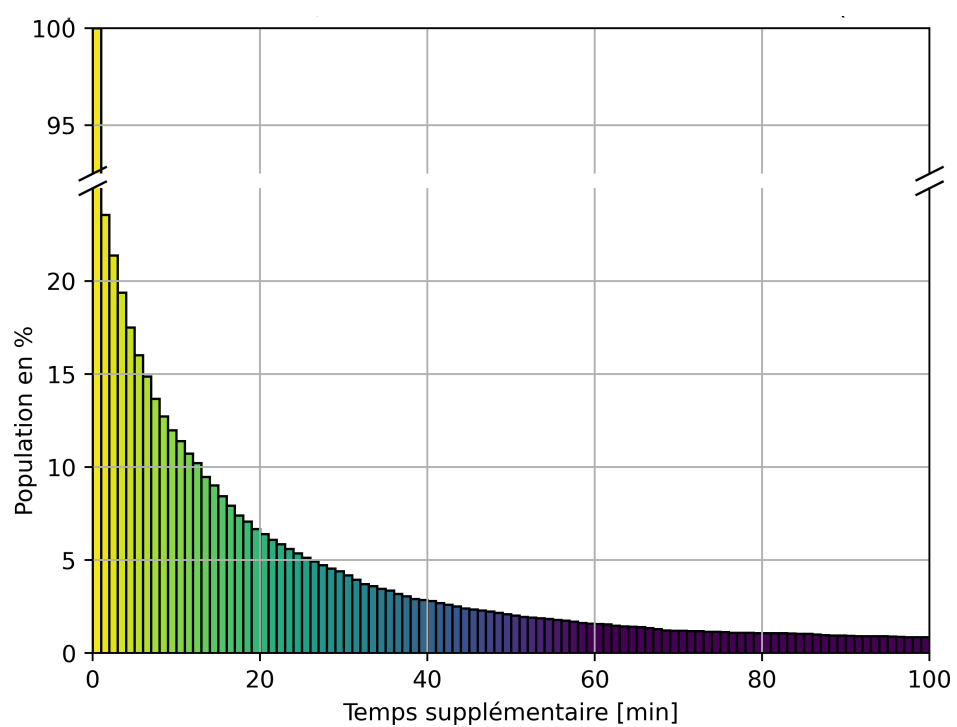


FIGURE 3 – Histogramme : population (en %) affectée par minute supplémentaire nécessaire pour se rendre à l'école la plus proche en distance (comparaison avec le temps de trajet le plus court pour se rendre à une école)

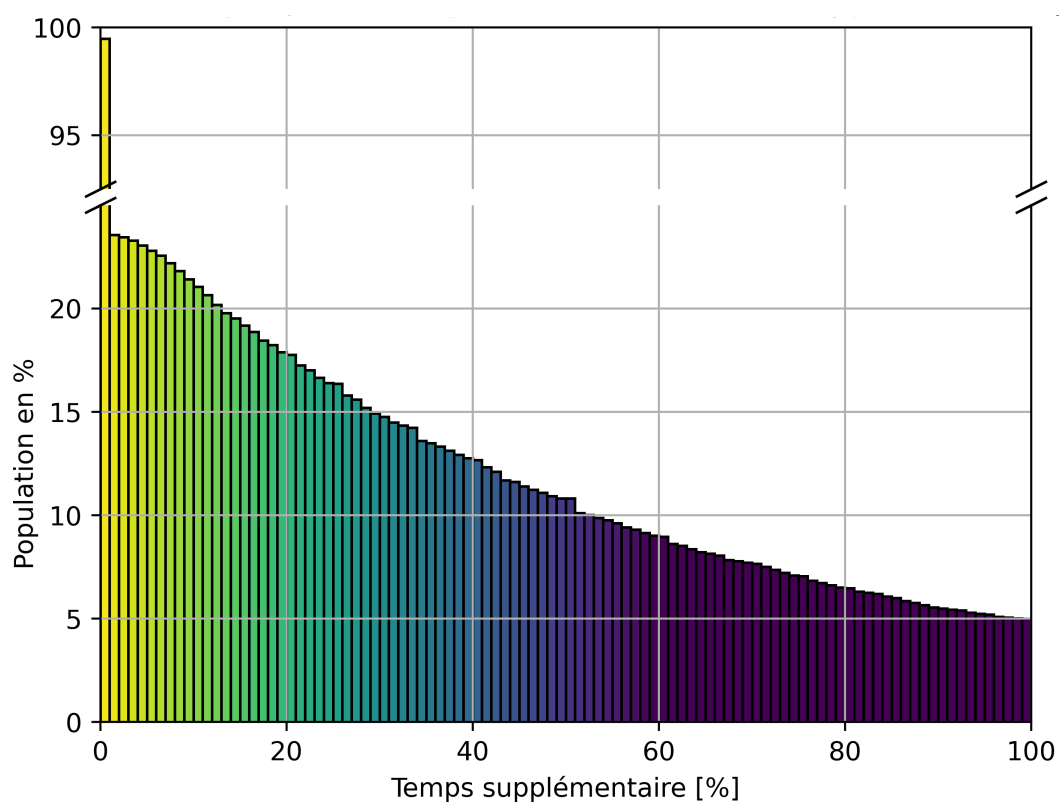


FIGURE 4 – Histogramme : population (en %) affectée par pourcentage de temps de trajet supplémentaire nécessaire pour se rendre à l'école la plus proche en distance (comparaison avec le temps de trajet le plus court pour se rendre à une école)

Annexe C

Guide d'utilisation de l'outil (Streamlit)

Notre outil de visualisation est généré via le script `6_tool.py`. Notons qu'il se base sur des résultats calculés par les scripts précédents : avant d'utiliser l'outil, il est nécessaire de faire tourner notre code jusqu'aux scripts de la partie 3 (`3a_school_computations.py` et `3b_computation_grid.py`). Veuillez également à installer les packages requis afin de pouvoir l'utiliser.

C.1 Lancement l'application

Suivez la procédure suivante pour accéder à notre outil de visualisation :

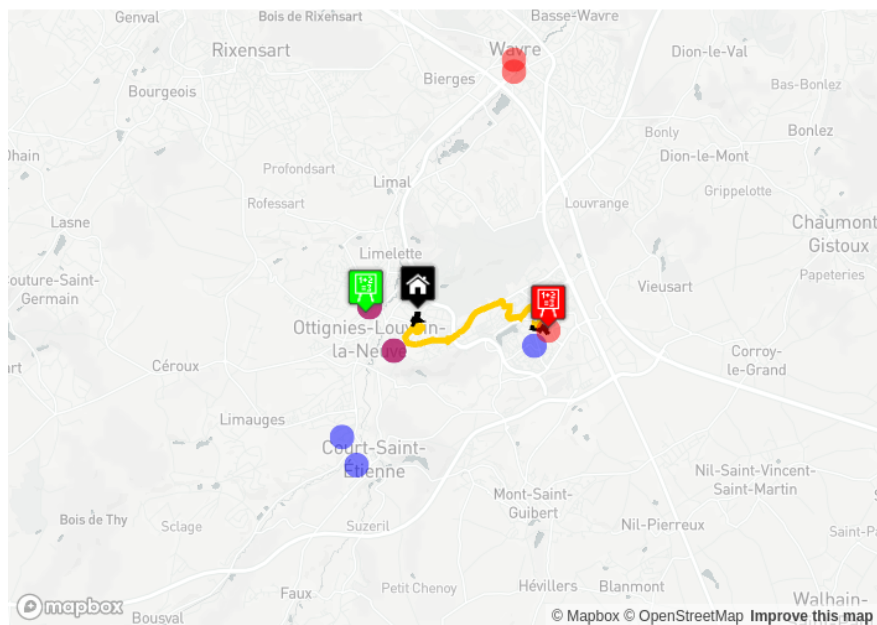
1. Ouvrir un terminal et se placer dans le dossier où se trouve le script `6_tool.py`
2. Exécuter la commande suivante : `streamlit run 6_tool.py`. Cela démarrera le serveur Streamlit.
3. Récupérer l'adresse IP qui s'affiche dans le terminal.

Une fois l'adresse récupérée, l'ouvrir dans un navigateur (au choix). Une fenêtre de chargement devrait apparaître (les données utiles sont chargées, cela peut nécessiter quelques secondes).

C.2 L'outil

Une fois la page chargée, une fenêtre similaire à la figure C.1 devrait apparaître. La page est divisée en deux sections : une carte de la zone et les données de l'utilisateur. Par défaut, les coordonnées du domicile sont centrées sur Louvain-la-Neuve, et les champs à compléter sont remplis par le premier choix disponible dans les données.

Calcul de l'indice composite



- | | | | |
|---|-------------------------------|---|---|
|  | Domicile |  | Écoles secondaires les plus accessibles |
|  | École secondaire sélectionnée |  | Écoles secondaires les plus proches |
|  | École primaire sélectionnée |  | Écoles secondaires proches et accessibles |

Domicile de l'enfant

Longitude

4.58078

- +

Latitude

50.67033

- +

École secondaire

École secondaire sélectionnée

LYCEE MARTIN V

Préférence

1re

☐ Immersion secondaire

École primaire

École primaire sélectionnée

Ecole fondamentale annexée Paul Delvaux - Ottignies

☐ Immersion primaire (depuis la 3e primaire au moins)

Calculer

FIGURE C.1 – Affichage par défaut

C.3 Calcul de l'indice composite

Pour calculer l'indice composite attribué pour l'établissement secondaire qui l'intéresse, l'utilisateur doit renseigner différentes informations dans l'application :

1. Les coordonnées de son domicile. Les deux champs (longitude et latitude) peuvent directement être modifiés par l'utilisateur. Les coordonnées exactes de son domicile peuvent facilement être trouvées en utilisant un outil tel que Google Maps, par exemple.
2. L'école secondaire pour laquelle il souhaite un calcul de l'indice composite. L'utilisateur peut sélectionner l'école de son choix en utilisant le menu déroulant, ou écrire directement le nom de l'établissement.
3. La place qu'occupe cet établissement dans la liste de préférence de l'utilisateur (sélectionner la place dans le menu déroulant).
4. Si une inscription en immersion est souhaitée, la case « Immersion secondaire » doit être cochée.
5. L'école primaire fréquentée par l'enfant. Comme pour l'établissement secondaire, l'utilisateur peut sélectionner l'école de son choix en utilisant le menu déroulant, ou directement écrire le nom de l'établissement.
6. Si une immersion a été suivie en primaire. La case « Immersion en primaire (depuis la 3e primaire au moins) » devrait être cochée uniquement si la langue d'immersion en primaire est identique à la langue d'immersion souhaitée pour le secondaire. De plus, l'élève est supposé avoir suivi un enseignement en immersion depuis la troisième année primaire au moins.

Une fois que toutes ces informations sont correctement complétées, il suffit de cliquer sur le bouton « Calculer » pour afficher les résultats. Une section supplémentaire devrait alors apparaître au bas de la page (figure C.2).

Distance à vol d'oiseau : 2.395 km	Temps en transports : 31 min
Ancien indice composite 2.31	Nouvel indice composite 3.4221888000000003
détails — Coefficient 1 1.5 Coefficient 2 1.0 Coefficient 3 1.0 Coefficient 4 1.54 Coefficient 5 1 Coefficient 6 1 Coefficient 7 (non calculé)	détails — Coefficient 1 1.5 Coefficient 2 1.18 Coefficient 3 1.59 Coefficient 4 1.216 Coefficient 5 1 Coefficient 6 1 Coefficient 7 (non calculé)

FIGURE C.2 – Section résultats de l'outil

Cette section indique la distance à vol d’oiseau qui sépare le domicile de l’établissement secondaire sélectionné ainsi que l’estimation du temps de trajet nécessaire pour revenir à son domicile en utilisant les transports en commun (départ à 16 h de l’école). Notez que cette valeur de temps est calculée pour le jour indiqué dans la variable DATE contenue dans le script `params.py` (dans notre cas, il s’agit donc du 26 octobre 2021). La valeur de l’indice composite est ensuite affichée. Deux valeurs sont données afin de comparer l’indice obtenu selon la méthode de calcul utilisée :

- La première colonne (Ancien indice composite) affiche la valeur de l’indice composite selon les coefficients du décret Inscription actuel. Le critère de distance est donc basé sur les distances à vol d’oiseau.
- La seconde colonne (Nouvel indice composite) affiche la valeur de l’indice composite calculée à partir de coefficients prenant en compte l’accessibilité de l’établissement plutôt que sa proximité.

Il s’agit de boîtes extensibles, cliquer sur « détails » permet d’accéder au détail de la valeur attribuée à chacun des coefficients qui composent l’indice composite.

C.4 Affichage de la carte

La carte présente au début de la page (figure C.3) est mise à jour à chaque fois qu’une donnée est modifiée par l’utilisateur. Elle permet d’afficher plusieurs éléments qui nous semblent intéressants pour l’utilisateur :

- L’emplacement du domicile renseigné.
- La localisation de l’école primaire fréquentée ainsi que celle de l’établissement secondaire sélectionné.
- De plus, l’itinéraire à utiliser pour rentrer de l’école à 16 h est affiché sur la carte. Le chemin calculé est généré sur base de données actualisées chaque jour, et ne correspond donc pas forcément aux temps de trajet calculés pour la journée DATE paramétrée dans `params.py`.
- Par ailleurs, nous avons également affiché les 5 écoles les plus proches (en bleu), ainsi que les 5 plus accessibles (en rouge) pour le domicile indiqué. Lorsqu’une école fait partie des deux listes, une couleur intermédiaire (mauve) est utilisée pour l’indiquer.

Enfin, cette carte affiche des données supplémentaires lorsque l’utilisateur sélectionne un élément avec sa souris :

- **Domicile** : Les coordonnées sont affichées.
- **École primaire fréquentée** : Le nom complet de l’école est affiché.
- **Écoles secondaires** : Le nom complet de l’établissement est affiché. Le temps de trajet nécessaire pour revenir en transports en communs depuis celui-ci est également affiché.
- **Le trajet** : La ou les lignes à utiliser sont affichées. Les parties de trajet parcourues à pied sont également renseignées.

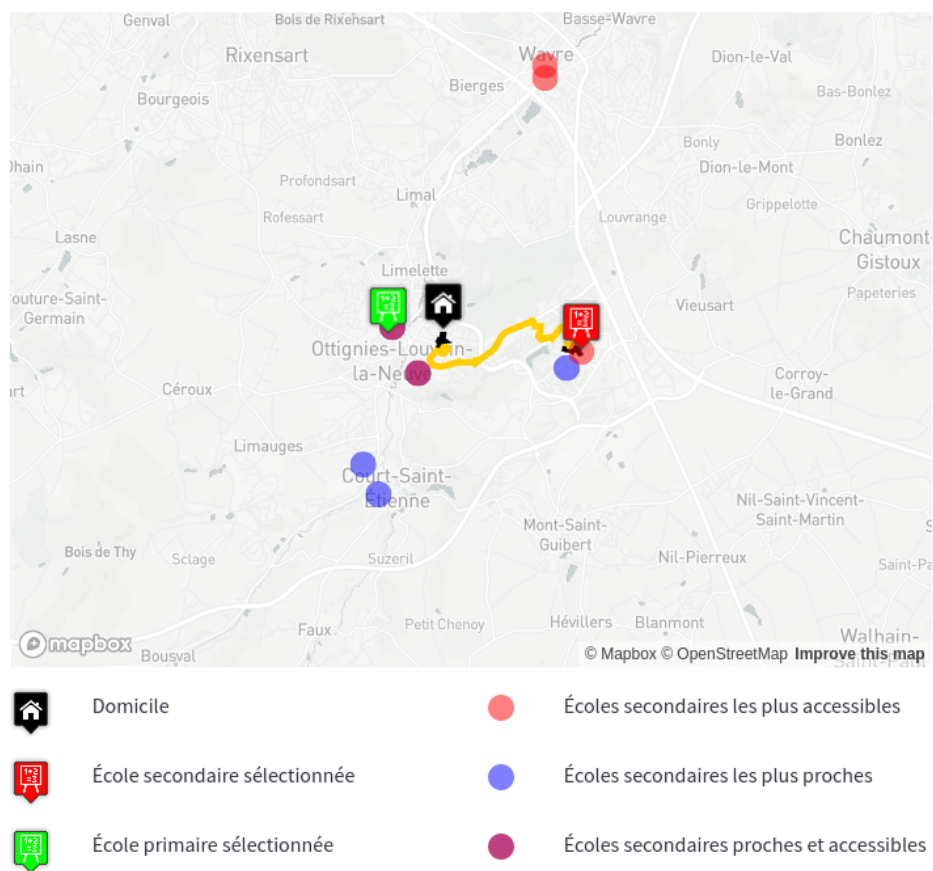


FIGURE C.3 – La carte affichée

Annexe D

Description des scripts / organisation du code

Pour plus de lisibilité, notre programme est divisé en plusieurs scripts, chacun étant chargé d'une partie des calculs. Une vue d'ensemble de l'organisation du code se trouve à la sous-section 3.5.1, avec, entre autres, un flowchart représentant la structure (figure 3.6) de l'ensemble. Cette annexe détaille le contenu de chacun des scripts.

D.1 `params.py`

Ce script permet de centraliser les paramètres utilisés lors de nos calculs. Il permet donc une modification aisée de ces derniers puisque l'intégralité des autres scripts se base sur les valeurs de paramètres listées dans ce fichier.

Cela inclut les différents noms de dossiers utilisés (où se trouvent les données, où placer les résultats) ainsi que des paramètres tels que la vitesse de marche, la distance maximale de marche autorisée, etc. Ces paramètres sont listés de manière exhaustive ci-dessous :

- Les dossiers dans lesquels les résultats seront stockés : `PRODUCE_DIR`, `FIGS_DIR`.
- Les dossiers dans lesquels les données sont cherchées : `PUBLIC_DATA_DIR`, `GTFS_DIR`, `MAPS_DIR`, `SCHOOLS_DIR`.
- Le nombre de cœurs à utiliser pour les parties de code effectuées en parallèle : `POOL_SIZE`.
- Les paramètres de la marche (voir nos choix de valeurs en section 3.1.2 et sous-section 3.1.3) :
 - `WALKING_SPEED`, la vitesse de marche considérée, en km/h (une vitesse réduite est conseillée afin de prendre en compte le détour du trajet effectué par rapport à la distance à vol d'oiseau entre deux points).
 - `MAX_WALKING_TIME_ST`, le temps de marche maximal entre deux arrêts. Ce paramètre permet d'autoriser les correspondances, en secondes.
 - `MAX_WALKING_DIST_ST`, la distance de marche maximale entre deux arrêts, lors d'une correspondance, en km. Ce paramètre est actuellement basé sur le précédent afin d'éviter les incohérences.

- MAX_RADIUS_ST, le rayon de marche maximal entre deux arrêts lors d’une correspondance, en radians. Ce paramètre est actuellement basé sur le précédent afin d’éviter les incohérences.
- MAX_WALKING_TIME_SCH, le temps de marche maximal entre une école et l’arrêt de départ utilisé, en secondes.
- MAX_WALKING_DIST_SCH, la distance de marche maximale entre l’école et l’arrêt utilisé, en km. Ce paramètre est actuellement calculé sur le précédent afin d’éviter les incohérences.
- MAX_RADIUS_SCH, le rayon de marche maximal entre l’école et la station utilisée, en radians.
- MAX_NB_NEIS, le nombre maximal de nœuds voisins autorisés pour une cellule.
- La zone géographique évaluée : GEO_ZONE. Il s’agit d’une liste contenant les provinces ou les régions à considérer, ou BEL pour effectuer les calculs sur toute la Belgique.
- La date des calculs : DATE. Attention, il faut s’assurer que les données GTFS utilisées couvrent bien cette date afin d’éviter tout problème.
- L’heure de départ de l’école : DEPARTURE_TIME. Cette valeur est exprimée en dizaines de secondes.
- La taille des listes d’écoles « proches » et « accessibles » que l’on souhaite générer : TOP. En cas d’égalité de distance (pour les écoles proches) ou de temps de trajet (pour les écoles accessibles), toutes les écoles possédant la même valeur sont ajoutées, la taille de certaines listes peut donc être, en pratique, plus grande que la valeur assignée à TOP.

D.2 utils.py

Dans ce script sont stockées les fonctions utilisées par plusieurs scripts différents. Nous les avons centralisées pour éviter la duplication inutile de code, et ainsi augmenter la lisibilité de nos fichiers. `utils.py` contient également quelques variables utilisées par les autres scripts qui ne sont pas censées être changées par les utilisateurs (par exemple les codes des régions et des provinces).

Les fonctions présentes dans ce script incluent des fonctions utilisées pour le calcul des distances (`distance_to_walking_time`, `haversine`), pour la création des histogrammes (`plot_hist`, `plot_broken_hist`) ou d’analyse de certains paramètres fournis dans `params.py` (`geo_zones`).

D.3 0_parse.py

Dans ce script, nous effectuons le *parsing* des données GTFS et du fichier FASE.

Pour les données GTFS, nous chargeons les fichiers, puis les analysons pour en extraire le nom et la position de chaque station, ainsi que ses voisins et les horaires de déplacement permettant de les rejoindre.

Pour le fichier FASE, nous chargeons le fichier, puis effectuons un filtrage sur les établissements. Seules les écoles primaires et secondaires ordinaires sont conservées. Pour

chaque école, nous gardons son id, son nom, sa latitude et sa longitude. Nous ajoutons son type (P ou S selon qu'elle est primaire ou secondaire, respectivement) ainsi que le réseau auquel elle appartient (confessionnel (C) ou non confessionnel (NC)). Enfin, nous calculons les arrêts de transports en commun qui lui sont voisins. Cela prend en compte tous les arrêts qui se trouvent dans un rayon `MAX_RADIUS_SCH`, avec un maximum de `MAX_NB_NEIS`.

D.4 `1_precomputation.py`

Dans ce script, nous effectuons quelques calculs destinés à simplifier nos démarches ultérieures.

Nous commençons par calculer la population totale assignée à chaque arrêt. Pour cela, nous calculons les zones de Voronoï autour des arrêts en Belgique. Nous calculons ensuite les intersections entre ces zones de Voronoï et les cellules de la grille de répartition de la population belge. Nous supposons que la population est uniformément répartie au sein de chaque cellule. Le calcul des intersections est effectué en parallèle dans un `Pool` contenant `POOL_SIZE` threads.

Nous calculons ensuite le temps de trajet à pied entre chaque arrêt ou école et tous les autres arrêts ou écoles situées dans un rayon de `MAX_RADIUS_ST`, avec un maximum de `MAX_NB_NEIS`.

Enfin, nous calculons le graphe comprenant tous les arrêts et écoles. Sur base des données parsées par le script précédent et des temps de trajet à pied, un nœud est créé pour chaque arrêt/école, à chaque temps possible. Une arête est créée pour chaque déplacement possible entre deux arrêts/écoles, et une supplémentaire pour représenter l'attente à un arrêt.

D.5 `2_resolve.py`

Dans ce script, nous effectuons le calcul des temps de parcours par les plus courts chemins. Une fonction effectue le calcul du temps de trajet d'un point de départ vers tous les arrêts en Belgique. Cette fonction renvoie un vecteur contenant tous les temps de trajet. Elle est appelée avec comme point de départ chaque école sélectionnée dans le script de passage, et le vecteur renvoyé est sauvegardé. Le script utilise un `Pool` contenant `POOL_SIZE` threads, afin d'accélérer le temps d'exécution en exécutant les calculs en parallèle.

À la fin de ce script, nous réassemblons les vecteurs sauvegardés en un seul tableau, et effectuons un filtrage pour ne garder que les arrêts situés en Fédération Wallonie-Bruxelles. Un tableau similaire est créé avec les distances à vol d'oiseau.

D.6 `3a_schools_computations.py`

Dans ce script, nous effectuons différents calculs sur base des résultats générés par le script précédent.

Tout d’abord, nous calculons la liste des TOP écoles les plus proches et les TOP écoles les plus accessibles pour chaque station. Ceci est effectué aussi bien pour les établissements secondaires que pour les écoles primaires.

Pour cela nous avons écrit une nouvelle fonction, afin de récupérer les n minima pour chaque école : `find_n_min`. Cette fonction trie le vecteur qu’elle reçoit en argument, et en extrait la n ième valeur (par exemple, si une liste de taille 3 est souhaitée, et que 3 données correspondent à la plus petite valeur observée, le but de notre fonction est de renvoyer cette plus petite valeur, et non la troisième valeur la plus petite dans l’ensemble du vecteur). Une fois ces valeurs extraites, une autre a pour objectif de récupérer les indices de toutes les valeurs plus petites ou égales à cette valeur. Ceci nous permettra effectivement de retrouver *toutes* les écoles accessibles ou proches. Ces indices seront ensuite utilisés pour récupérer les noms des écoles, les temps de trajet correspondants ainsi que les distances à vol d’oiseau associées.

L’étape suivante est de calculer l’intersection entre la liste des index des écoles les plus proches, et la même liste générée pour les écoles les plus accessibles. Sur base de cela, notre script calcule des cartes binaires, pour lesquelles chaque arrêt est coloré en vert ou en rouge selon que l’école la plus proche et la plus accessible sont identiques ou non (les arrêts sont donc verts s’ils appartiennent à l’intersection que nous avons calculée précédemment).

Ensuite, nous calculons le temps additionnel pour revenir de l’école la plus proche en transports en commun, par rapport au temps de trajet depuis l’école la plus accessible. Cette valeur permet également de calculer la différence de temps en termes de pourcentage de temps de trajet.

Ce script contient également une fonction permettant de calculer l’indice composite pour l’école la plus proche et pour l’école la plus accessible.

Enfin, nous exportons l’ensemble de ces données au format GeoJSON, avec plusieurs paramètres calculés supplémentaires. Ce fichier GeoJSON sera utilisé par la plupart des scripts suivants.

D.7 3b_computation_grid.py

Ce script est similaire au précédent, mais effectue cette fois-ci les calculs pour l’entièreté des cellules plutôt que pour les stations. Nous commençons par les stations MAX_NB_NEIS les plus proches du centre de la cellule. Ensuite, le temps de trajet à pied pour se rendre à ces arrêts est calculé. Ceci permet de générer une version étendue de la matrice des temps de trajet entre une école et les arrêts : une matrice des trajets entre les écoles et le centre des cellules. Pour cela, la procédure suivante est répétée pour chaque cellule :

1. Récupérer la matrice du temps de trajet depuis toutes les écoles, en ne sélectionnant que les colonnes correspondant aux arrêts proches de son centre.
2. Mettre à jour les valeurs de temps de trajet, en ajoutant à chaque colonne le temps nécessaire pour se rendre à l’arrêt qu’elle représente
3. Sélectionner pour chaque école la valeur de trajet minimale. Il s’agit donc d’un minimum effectué sur les lignes de la sous matrice extraite au point 1.

Une fois la matrice obtenue, un fichier GeoJSON similaire à celui du script 3a_schools_computations est généré. Il sera utilisé pour la plupart de nos résultats et contient des valeurs de temps

de trajet et de distance vers les écoles les plus proches et les plus accessibles, la liste de ces écoles, les différences en temps de trajet, etc.

Ce script n'a été prévu que pour une utilisation avec le TOP 1.

D.8 3c_computation_grid_bonus.py

Ce script additionnel a été utilisé pour calculer la robustesse de nos résultats. En effet, comme expliqué à la section précédente, le script 3b_computation_grid.py est plutôt lent et n'est pas adapté au calcul de listes de taille trop longue. Nous avons donc créé ce script dans le but de générer les pourcentages de similitude décrits en sous-section 4.2.4.

Ainsi, plutôt que de calculer une version étendue de la matrice des temps de trajet, ce script récupère les listes des écoles les plus proches de l'arrêt qui seront utilisées pour chaque cellule.

Lors de la génération du GeoJSON, nous générons ainsi plusieurs métriques, qui n'ont finalement pas toutes été utilisées. Ainsi, la seule métrique d'intérêt est la colonne `similarity_pc`, qui permettra la génération des diagrammes circulaires (présentés en sous-section 4.2.4 et en section E.5 de l'annexe E).

D.9 4a_histograms.py, 4b_choropleth.py, 4c_pixel_map.py, 4d_ic.py

Ces scripts ont pour fonction de générer les figures présentées dans ce mémoire. Ils chargent les données générées par les scripts précédents et appliquent différentes opérations : conversion en minutes des temps de trajet, *clip* des vecteurs si nécessaire (figure D.1), choix des couleurs et des échelles sélectionnées, etc.

Sans entrer dans les détails, ces scripts permettent de générer les images souhaitées et sont séparés en fonction du type de graphe généré. Les noms donnés à ces scripts parlent généralement d'eux-mêmes, notez tout de même que 4d_ic est celui qui génère les graphes circulaires décrits en section 4.4, son nom indique simplement qu'il concerne l'indice composite.

-1	0	5	17	35	87	89	93	99	125
↓ clip(0, 90)									
0	0	5	17	35	87	89	90	90	90

FIGURE D.1 – Exemple d'application de la fonction clip

D.10 6_tool.py

Dans ce script, nous créons un outil utilisable sur internet, permettant à toute personne d'évaluer les changements qui leur seraient apportés par une modification de la méthode du calcul de l'indice composite.

Nous avons pour cela utilisé le module Streamlit, qui permet de créer ce genre de petit outil en une page assez facilement.

D'abord est effectué le chargement des données que nous utilisons dans l'outil.

La carte est affichée en premier sur la page, accompagnée de sa légende.

Ensuite, le formulaire de la page est créé, avec les différents champs nécessaires pour le calcul de l'indice composite : emplacement du domicile, établissement secondaire et préférence de celui-ci, école primaire, immersion éventuelle.

Enfin, un bouton en fin de page permet d'effectuer réellement le calcul de l'indice composite par l'ancienne et la nouvelle méthode. Les résultats sont disposés côte à côte, afin de faciliter la comparaison. Le détail des différents coefficients est également disponible dans un module extensible réduit par défaut.

Un guide pour utiliser l'outil créé par ce script se trouve dans l'annexe C.

Annexe E

Détails sur les résultats

E.1 Comparaison des résultats par réseau

La figure E.1 illustre la différence des résultats obtenus en considérant uniquement les établissements du réseau confessionnel (graphe orange) ou uniquement les établissements du réseau non confessionnel (graphe bleu). Comme mentionné dans le texte, les tendances observées pour les deux réseaux sont identiques. Aucune différence significative n'est observée. Par conséquent, nous n'avons pas généré l'entièreté des figures présentées dans ce travail pour chacun des réseaux.

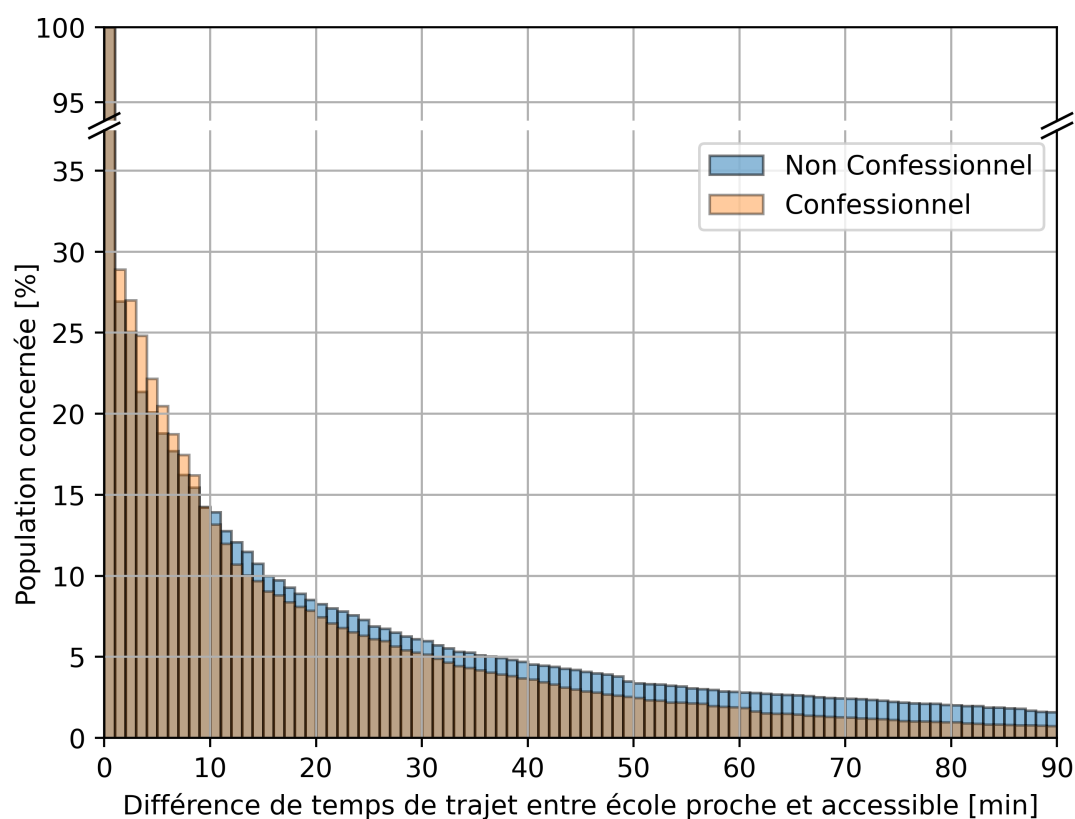


FIGURE E.1 – Population impactée par minute de trajet supplémentaire : réseau confessionnel (orange) VS réseau non confessionnel (bleu)

E.2 Distance moyenne avec l'école la plus proche

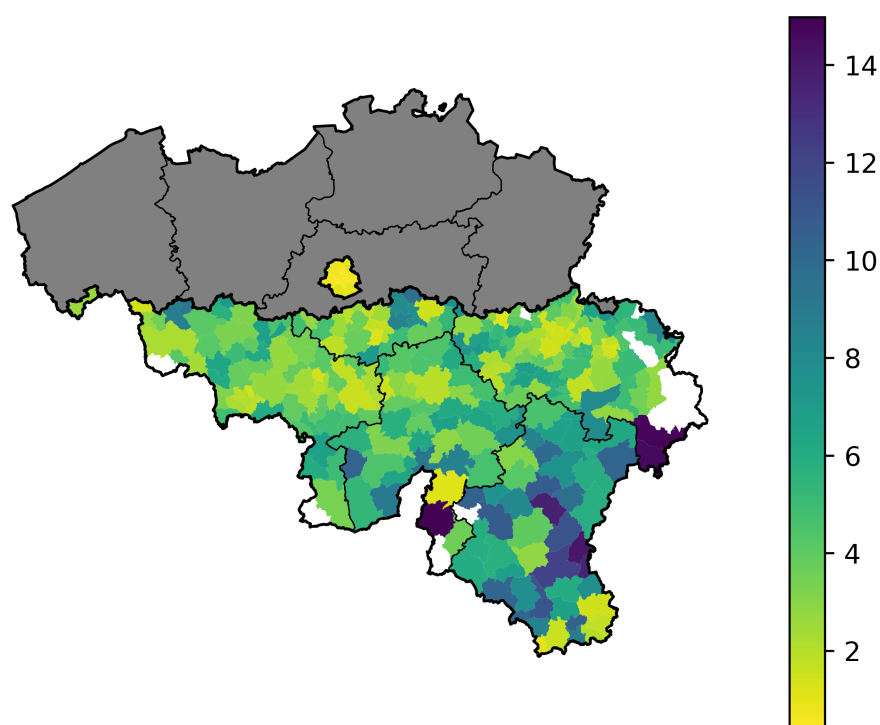


FIGURE E.2 – Distance moyenne par commune [km]

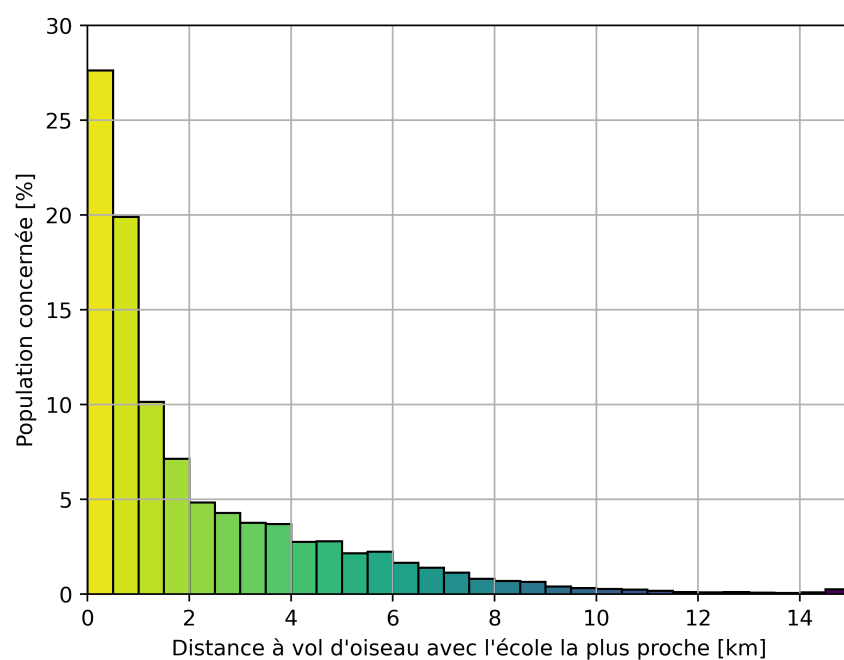


FIGURE E.3 – Population concernée par valeur de distance (réparties par tranche de 500m)

E.3 Temps de trajet moyen pour rentrer depuis l'école la plus proche

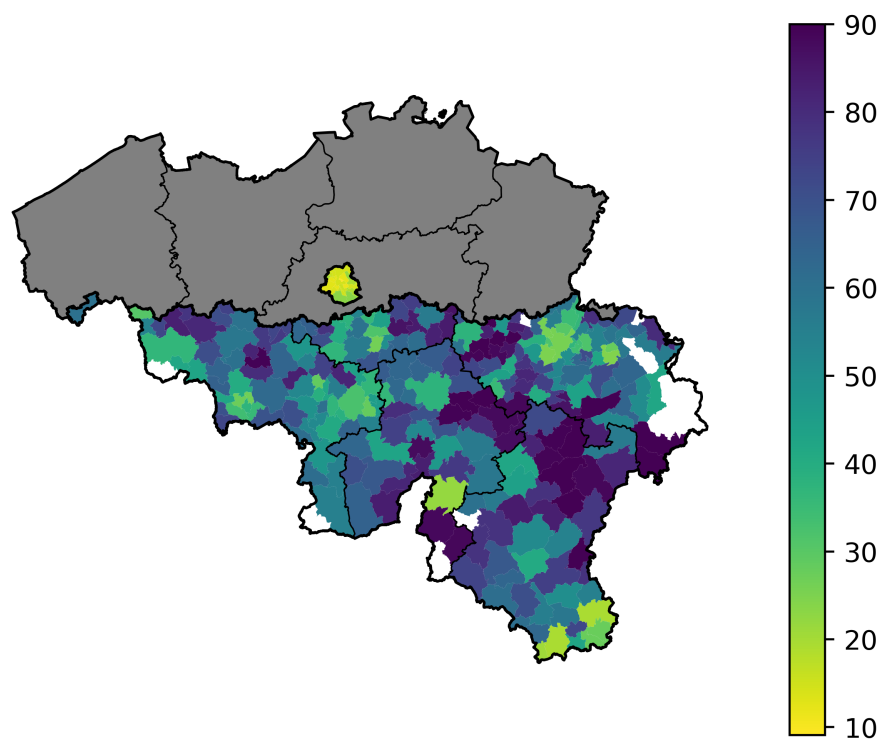


FIGURE E.4 – Temps de trajet [min]

E.4 Cartes (au km²) des temps de trajet pour revenir des écoles

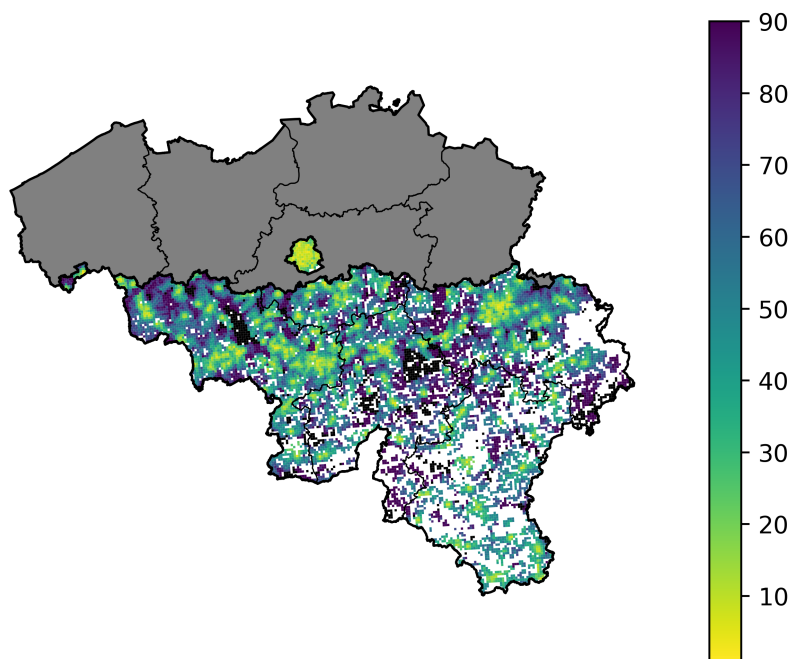


FIGURE E.5 – Temps de trajet depuis l'école la plus proche

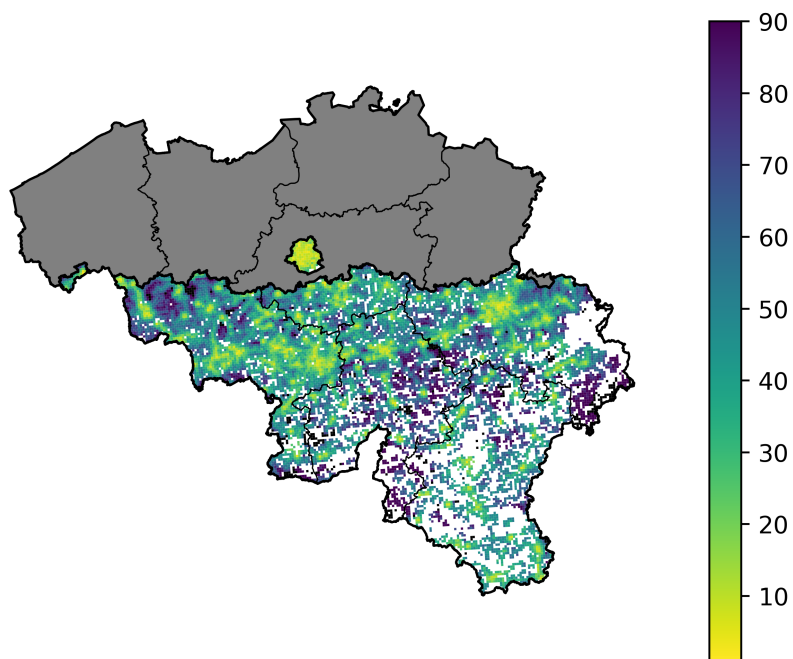


FIGURE E.6 – Temps de trajet depuis l'école la plus accessible

E.5 Robustesse des résultats : listes de tailles deux à 10

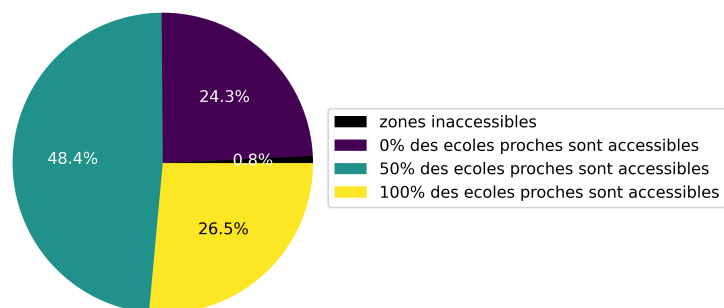


FIGURE E.7 – Liste de taille 2

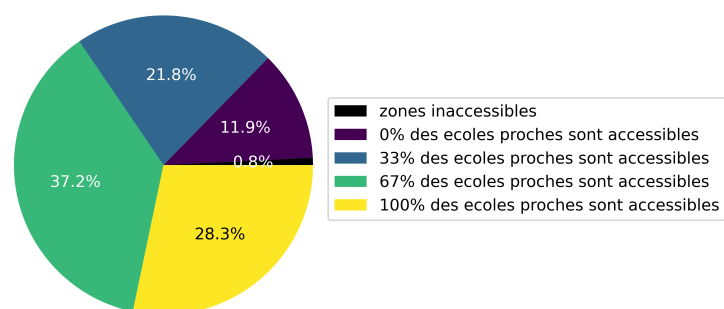


FIGURE E.8 – Liste de taille 3

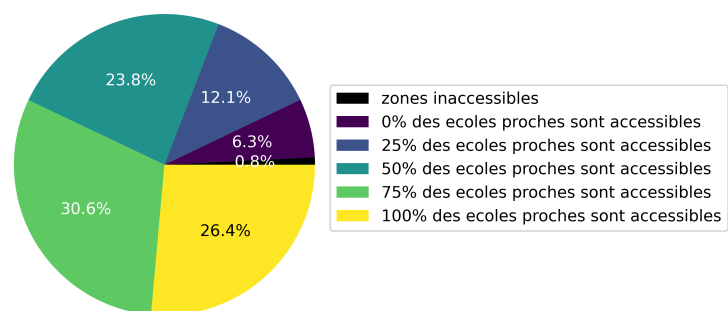


FIGURE E.9 – Liste de taille 4

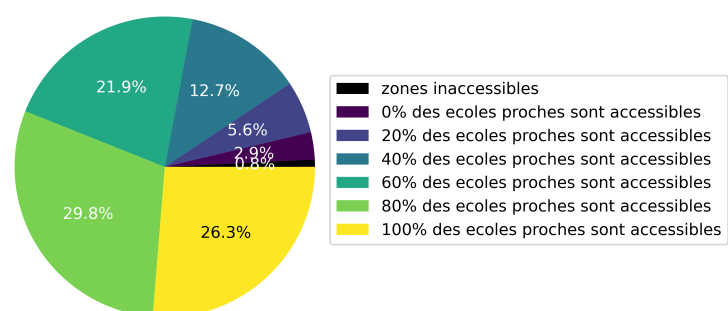


FIGURE E.10 – Liste de taille 5

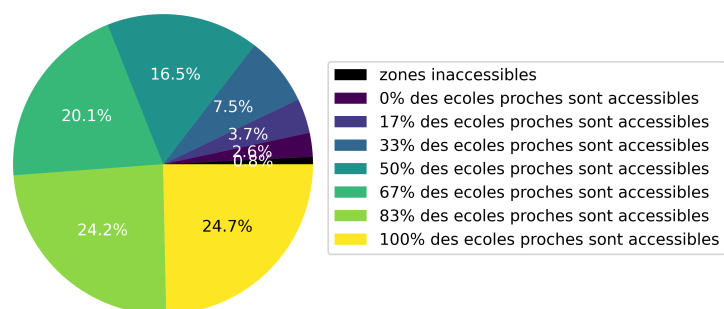


FIGURE E.11 – Liste de taille 6

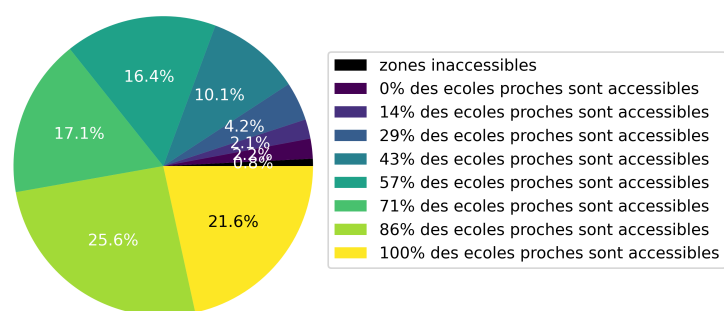


FIGURE E.12 – Liste de taille 7

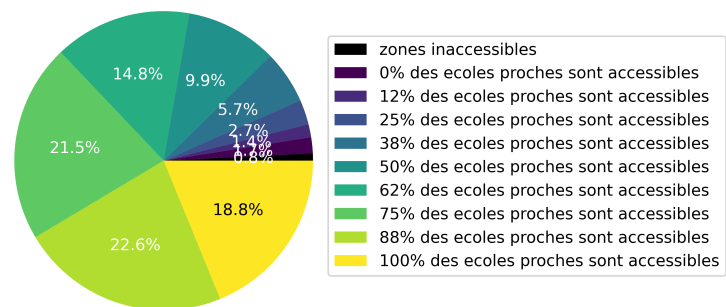


FIGURE E.13 – Liste de taille 8

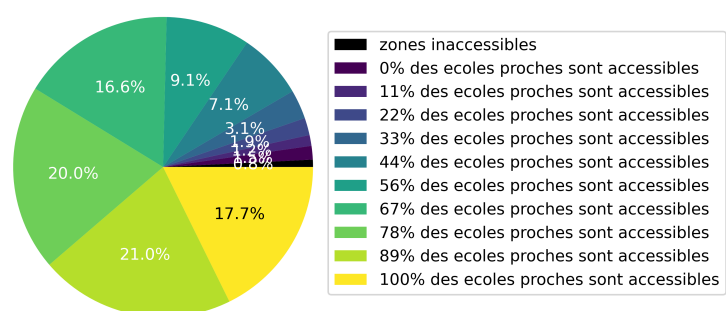


FIGURE E.14 – Liste de taille 9

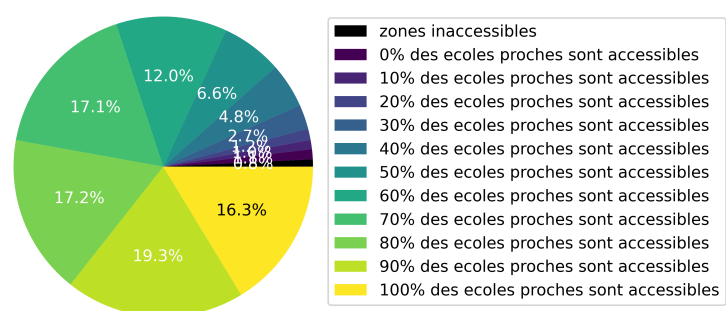


FIGURE E.15 – Liste de taille 10

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

École polytechnique de Louvain

Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve, Belgique | www.uclouvain.be/epl