

# Relations de voisinage entre les stations de base mobiles

Stage ČVUT Prague, Mai-Août 2025

Amandine BURÇON

Gautier VASSE

# Objectif

**Objectif principal :**

Construire un graphe de voisinage reflétant le recouvrement possible

**Contraintes :**

Base ARCEP, 108 838 stations (France, T4 2023)

---

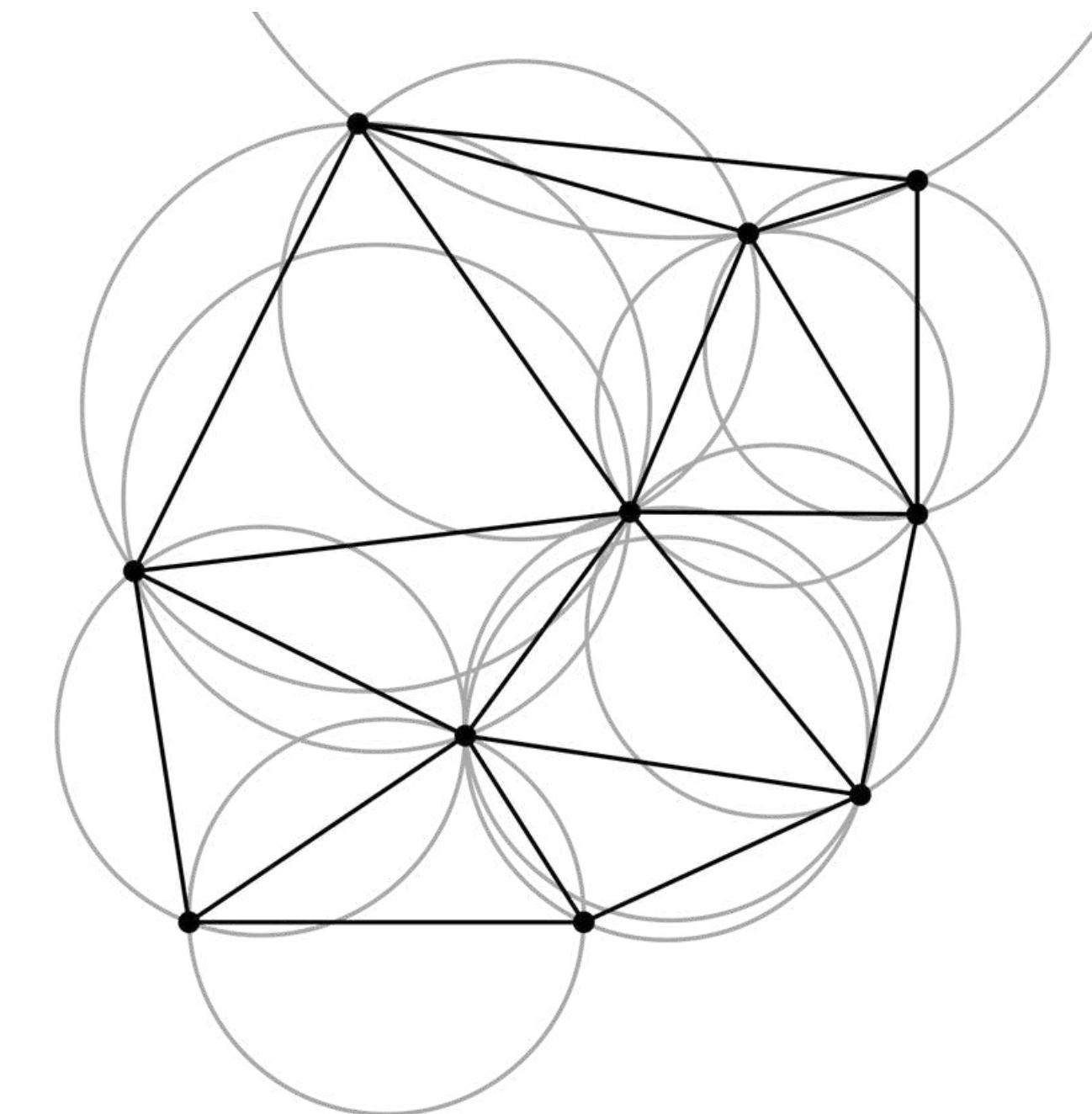
Column	Description
<i>id_station_anfr</i>	ANFR identifier of the station
<i>nom_op</i>	Provider name
<i>x, y, latitude, longitude</i>	Station coordinates
<i>nom_reg, nom_dep, nom_com</i>	Additional location information
<i>site_2g, 3g, 4g, 5g</i>	Technology equipping the station

---

# Méthode de base (Delaunay)

Génère un réseau de triangles sans cercle circonscrit contenant d'autres points

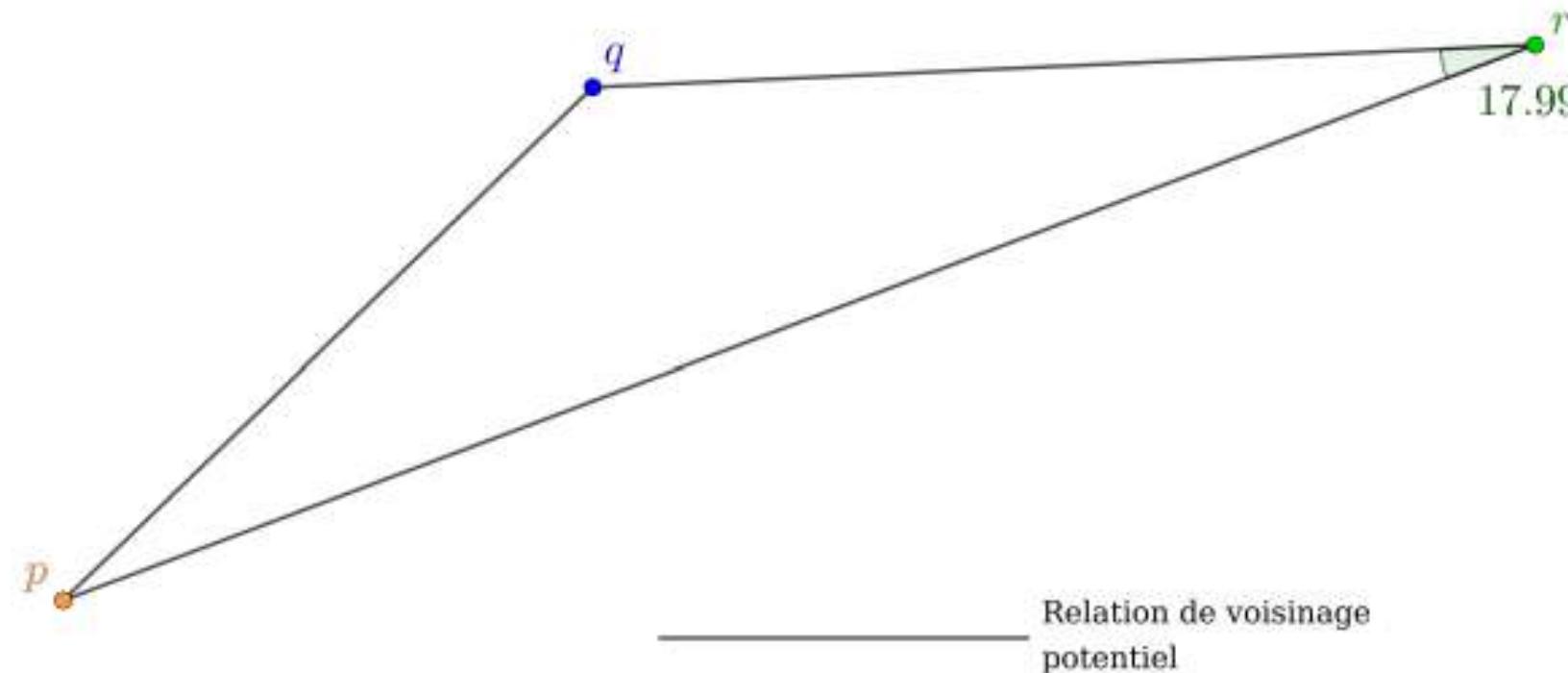
Liste brute de voisins pour chaque station



# Filtrage adaptatif

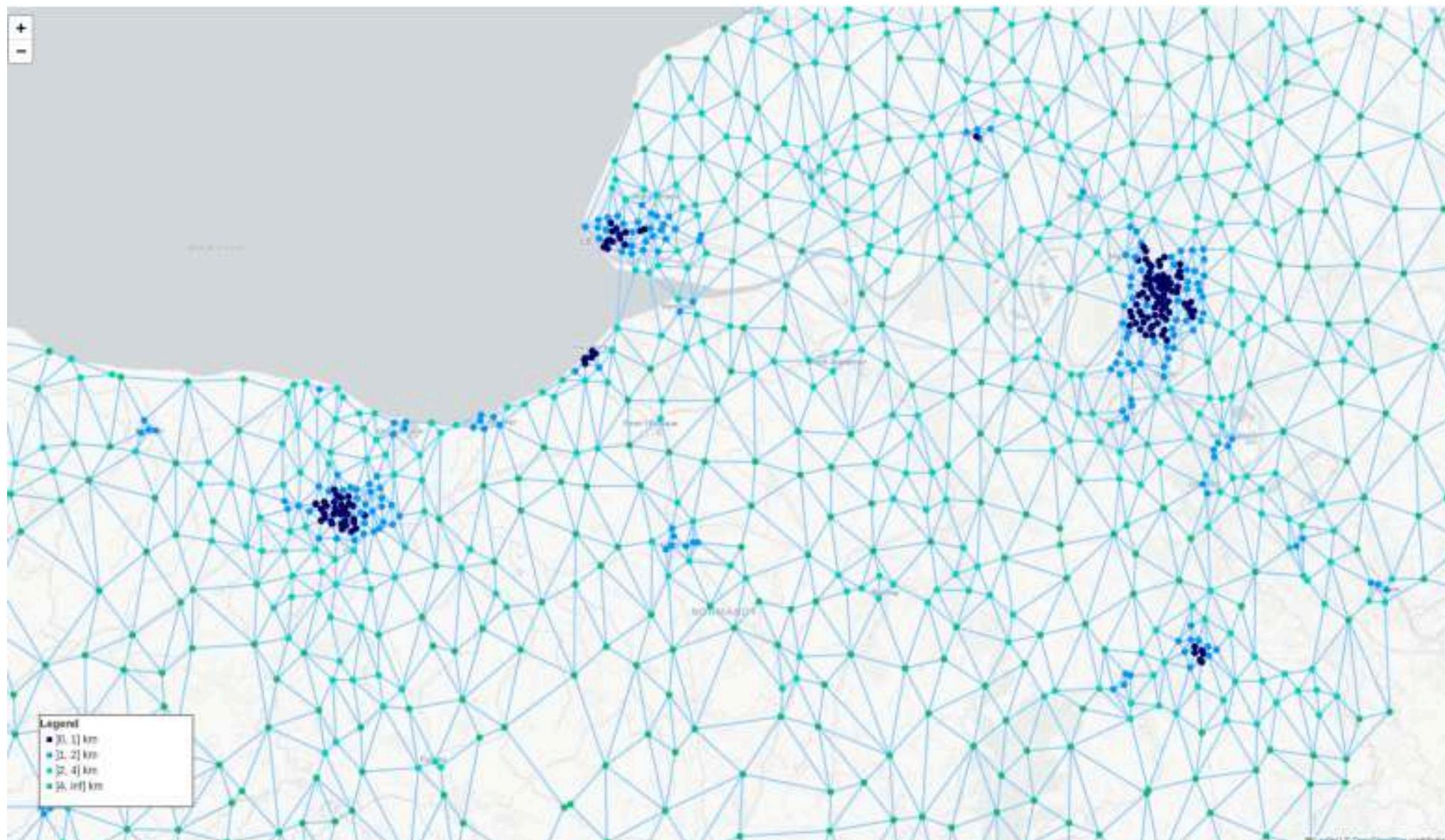
- Estimation de densité locale (3-NN)
- Catégorisation en 4 zones : centre-ville / urbain / périurbain / rural

Critère de filtrage selon angle et distance



# Résultat & Conclusion

- Graphe final : visualisation (Normandie)
- Limites : pas de vérité-terrain, géométrie pure, seuils manuels

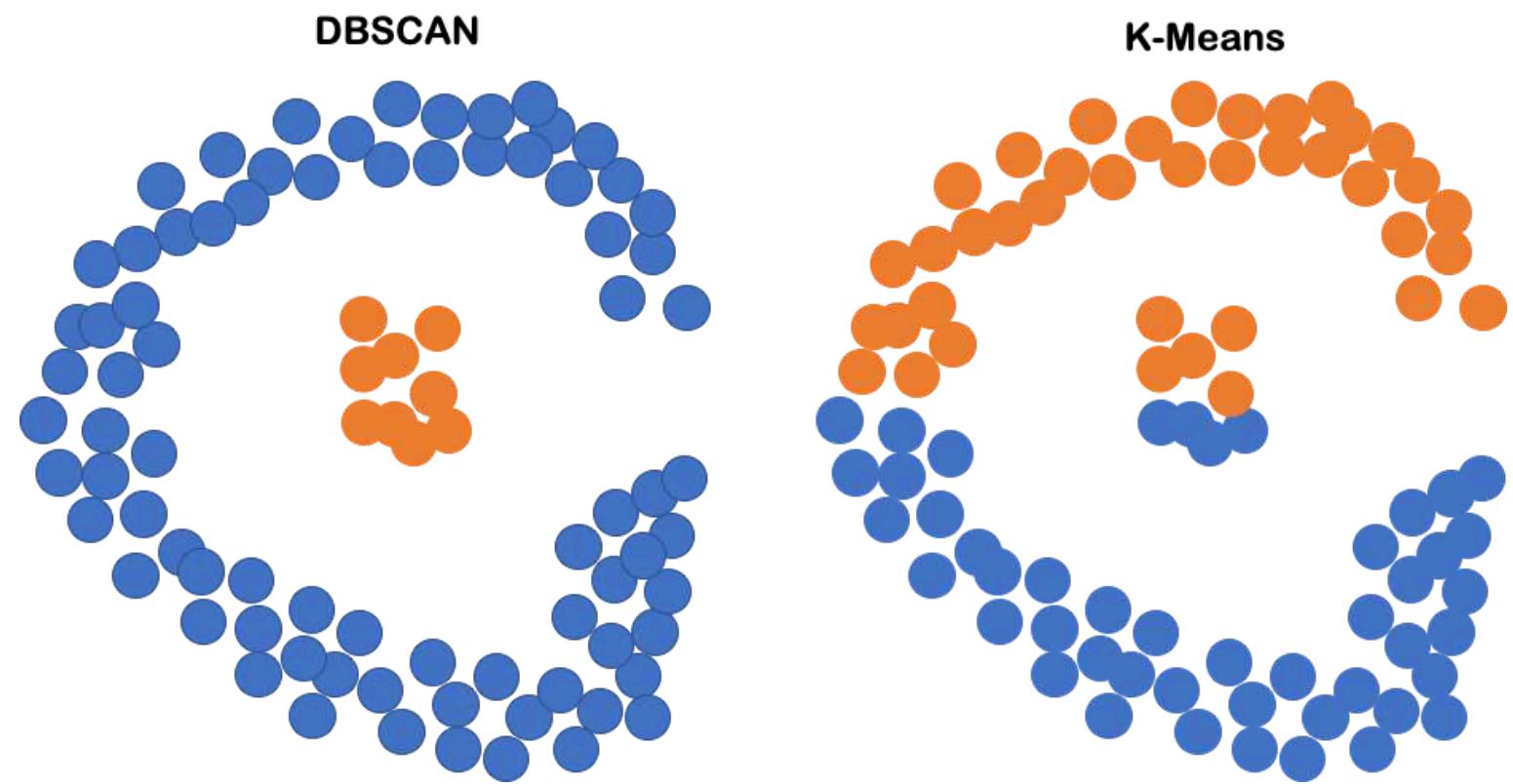


**26-30 Mai**

**Détection des routes et axes ferroviaires**

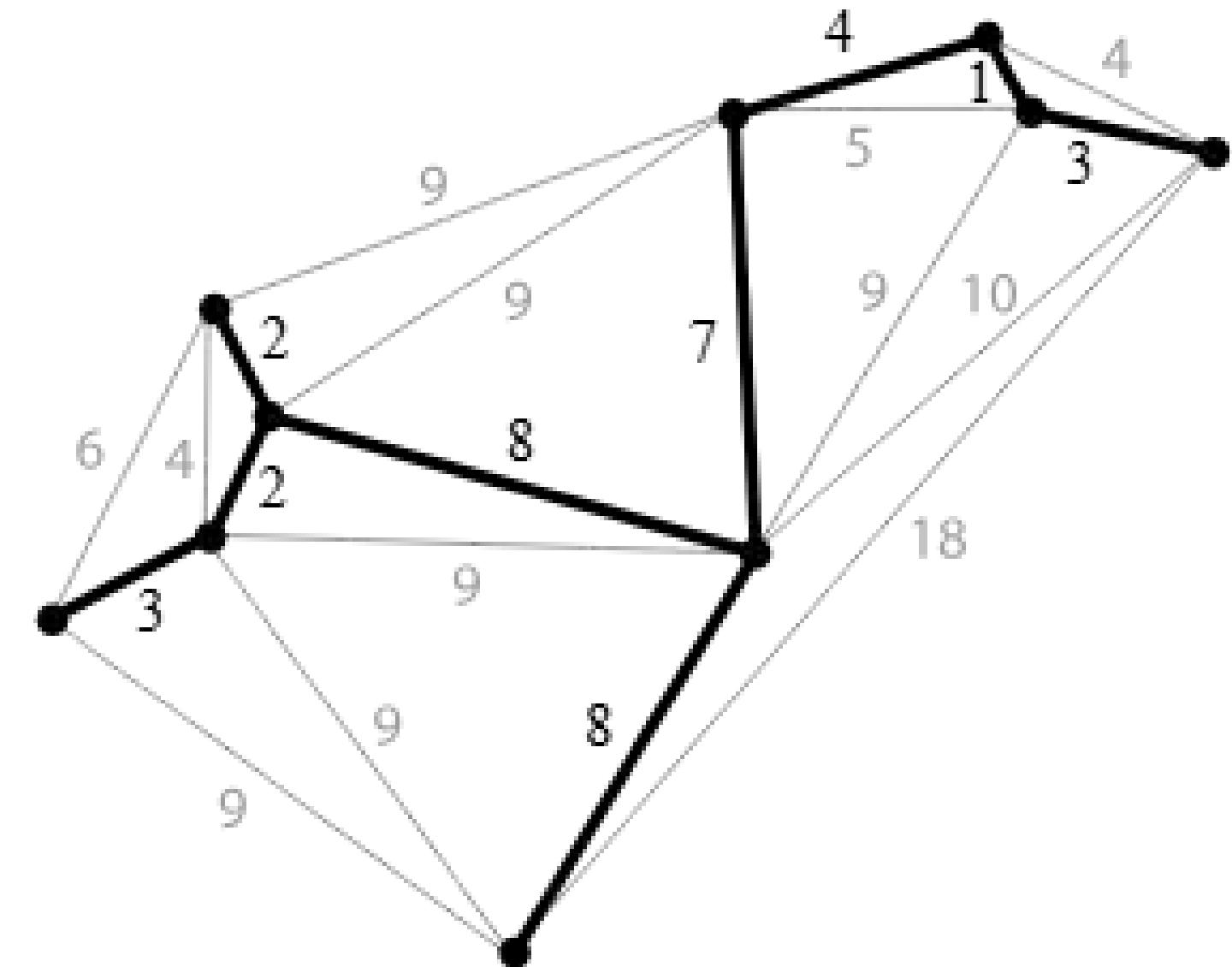
# DBSCAN

- Regarde ses voisins à distance  $\leq \varepsilon$ .
- Si  $\geq \text{MinPts}$  : marque-le comme core et crée un nouveau cluster.
- Explore récursivement tous les voisins core pour y ajouter leurs voisins.
- Sinon, marque-le comme noise (bruit).



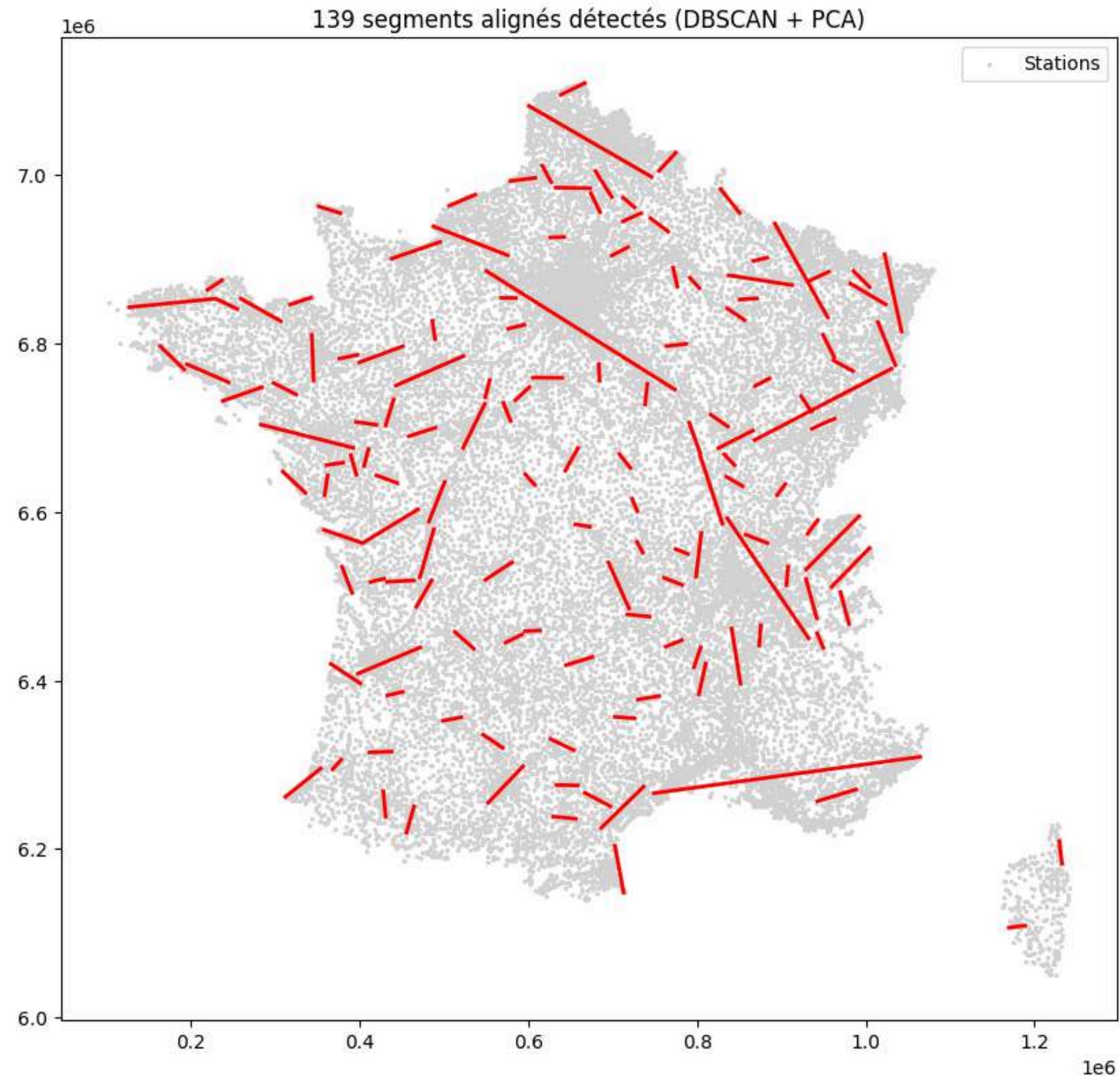
# Minimum Spanning Tree (MST)

- Obtenir la charpente minimale qui relie tous vos points (ici : stations) avec le coût (distance) cumulé le plus petit.
- Mettre en évidence les “lignes de force” ou couloirs principaux entre équipements.



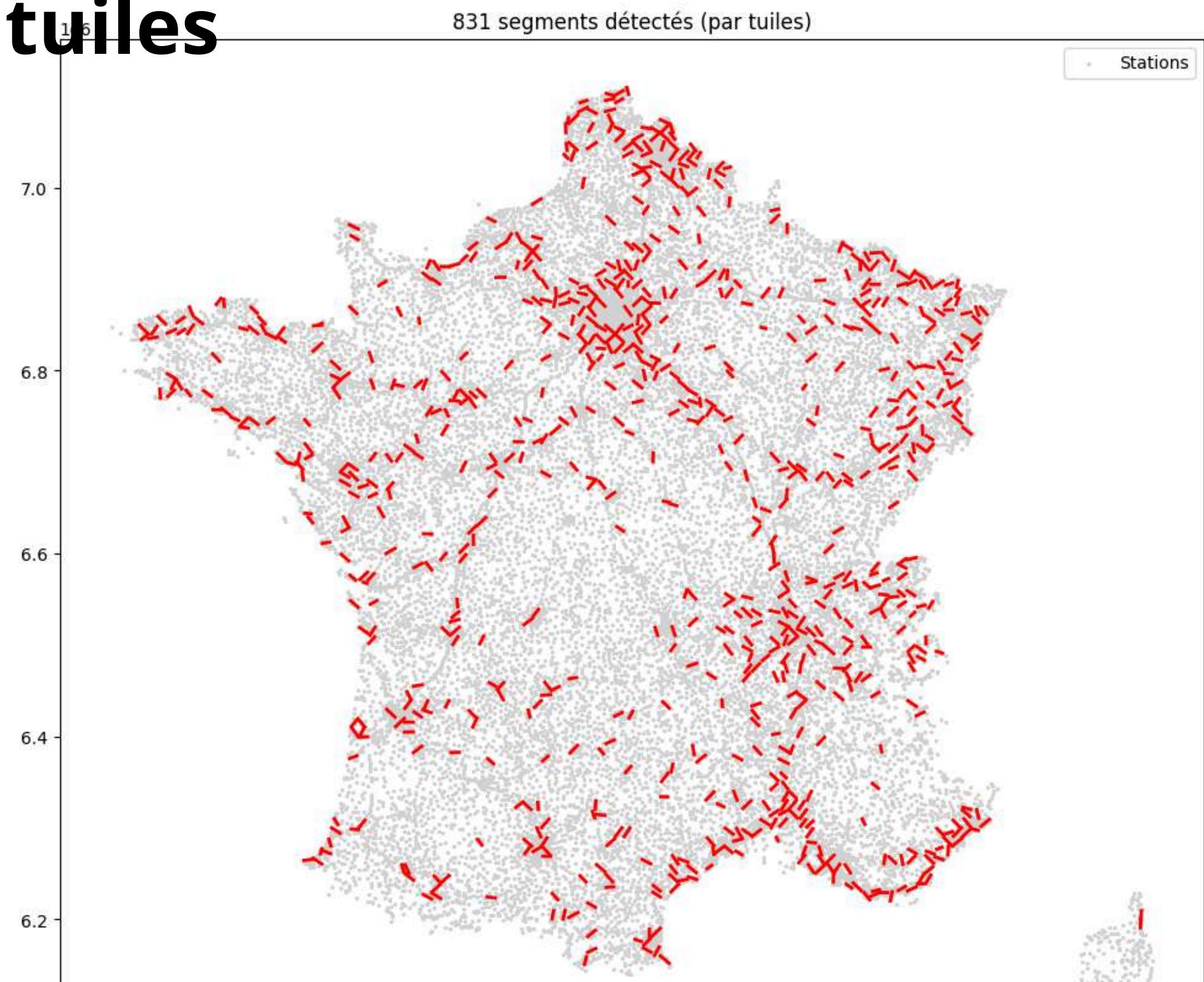
# DBSCAN + PCA

- DBSCAN pour commencer
- Pour chaque cluster, tester la linéarité
- Tracer segment dans la direction



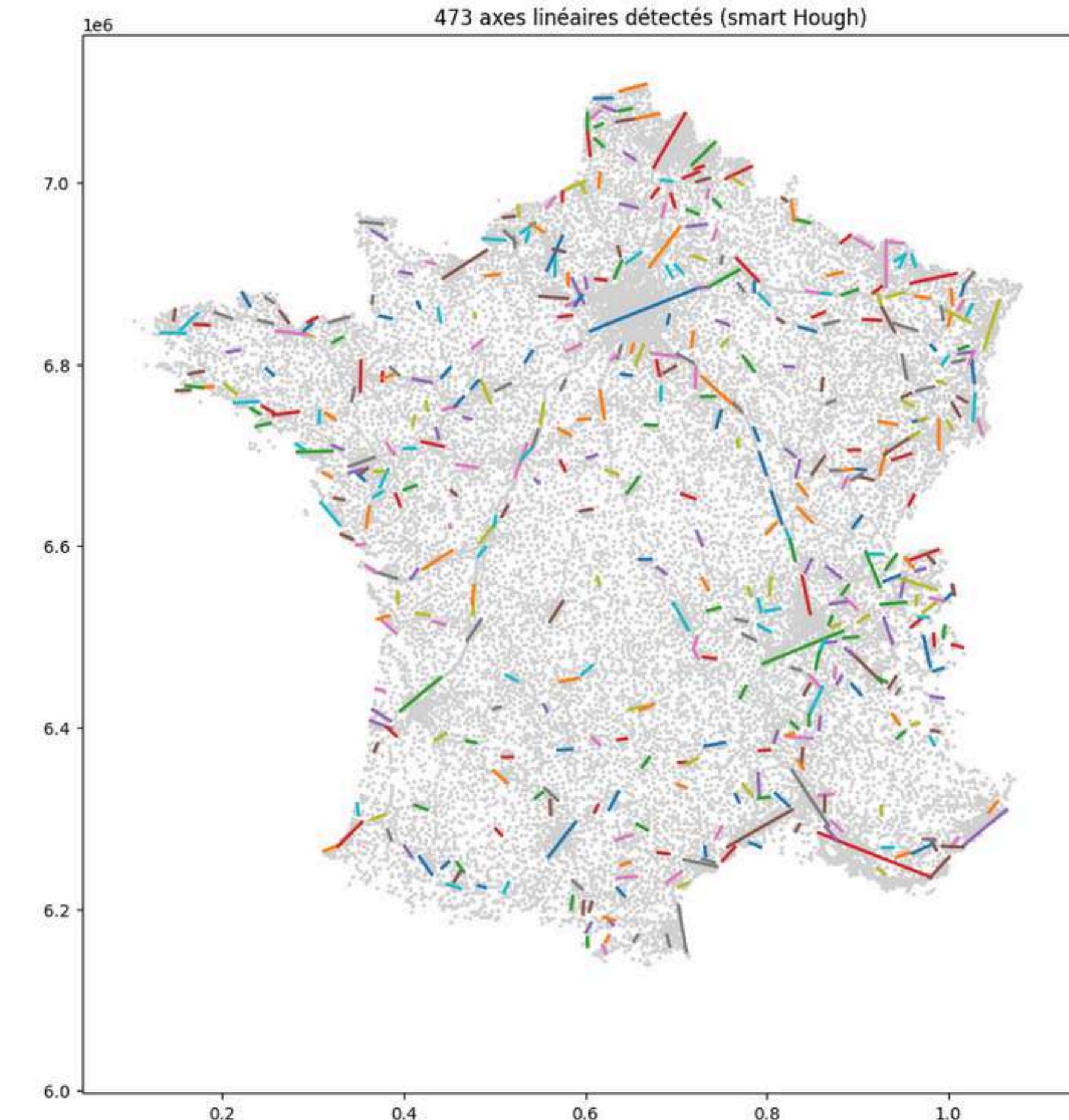
# DBSCAN + PCA par tuiles

- On sépare la carte en quadrillage
- Puis même méthode



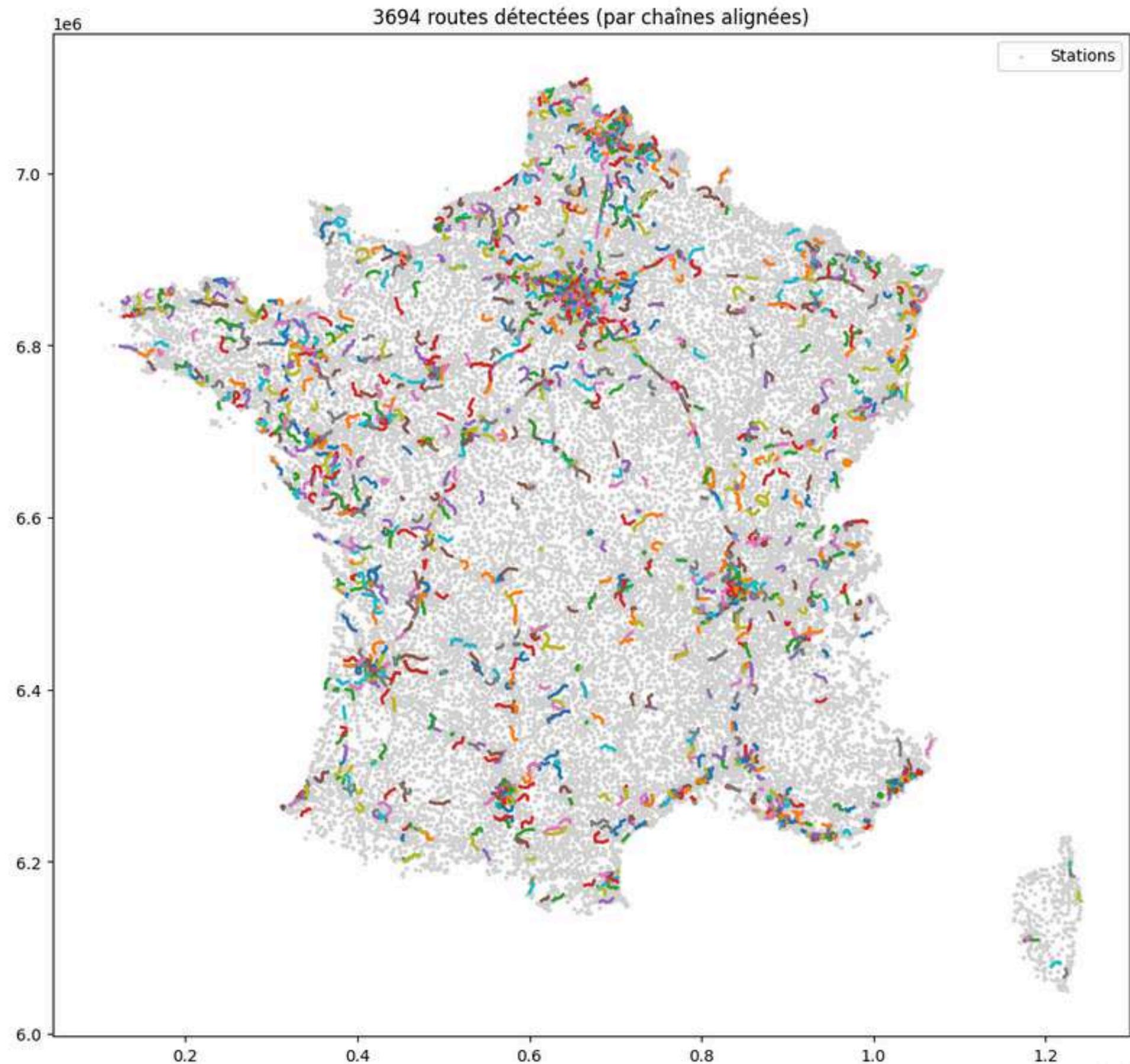
# Smart Hough Transform

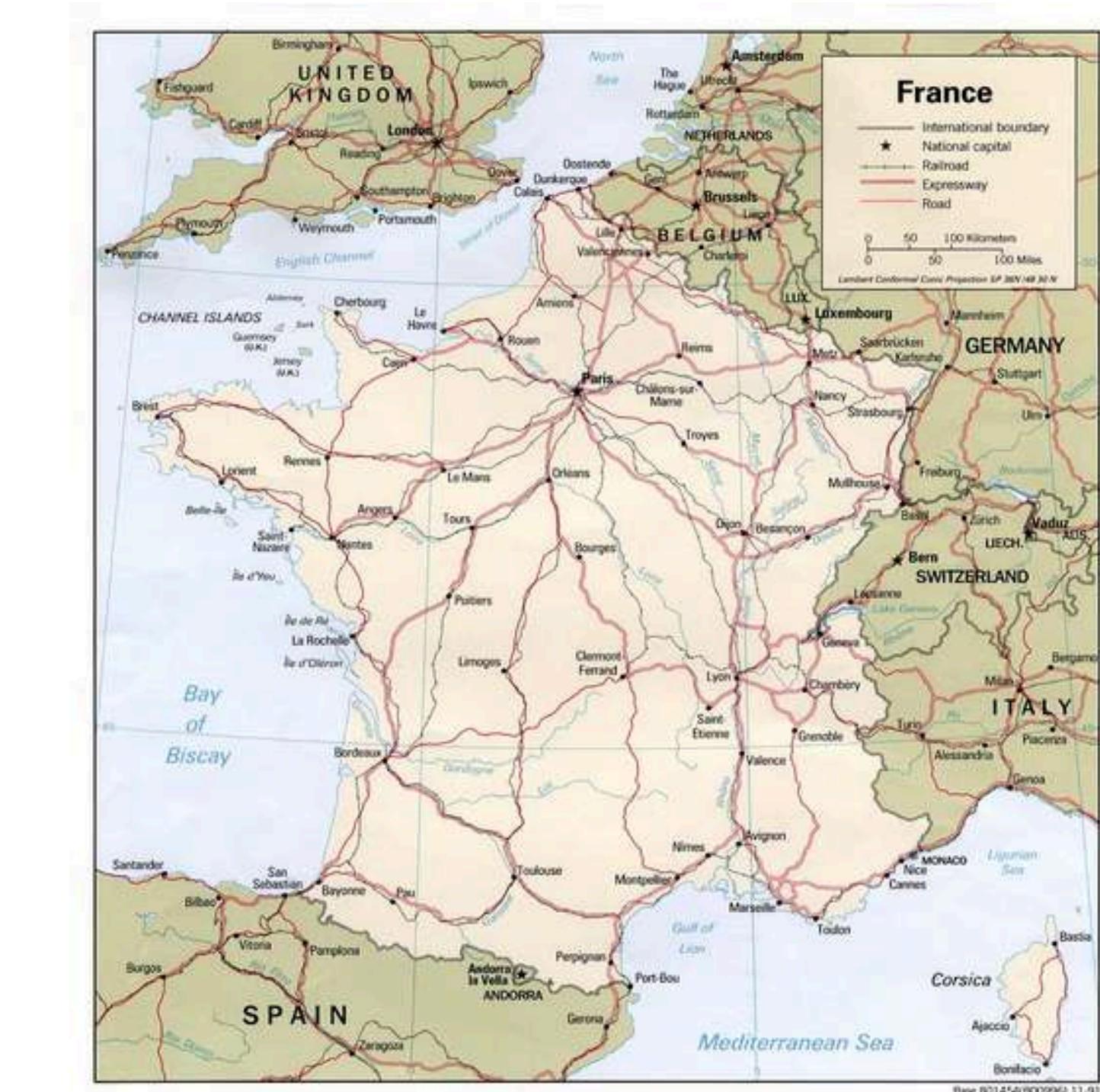
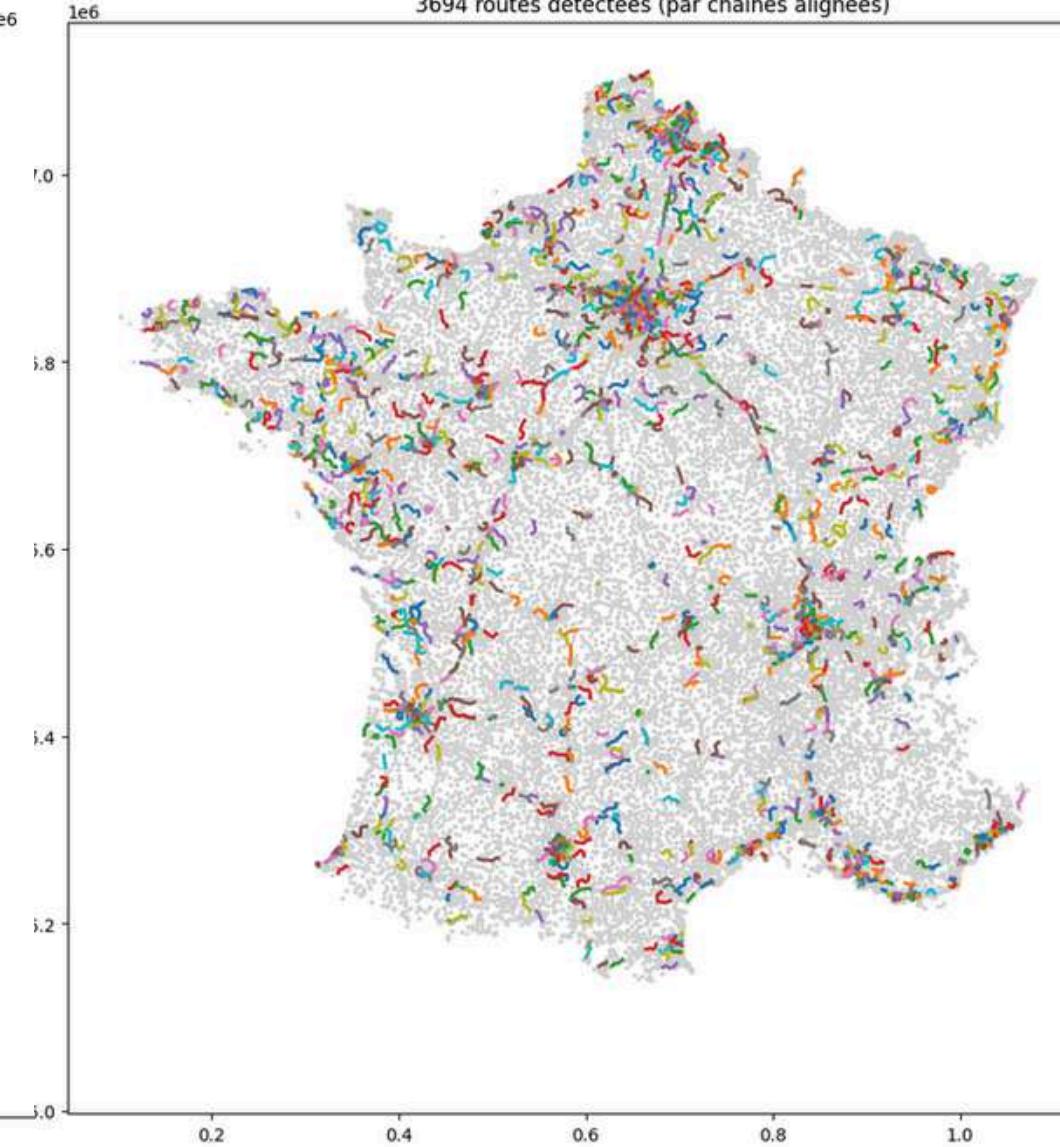
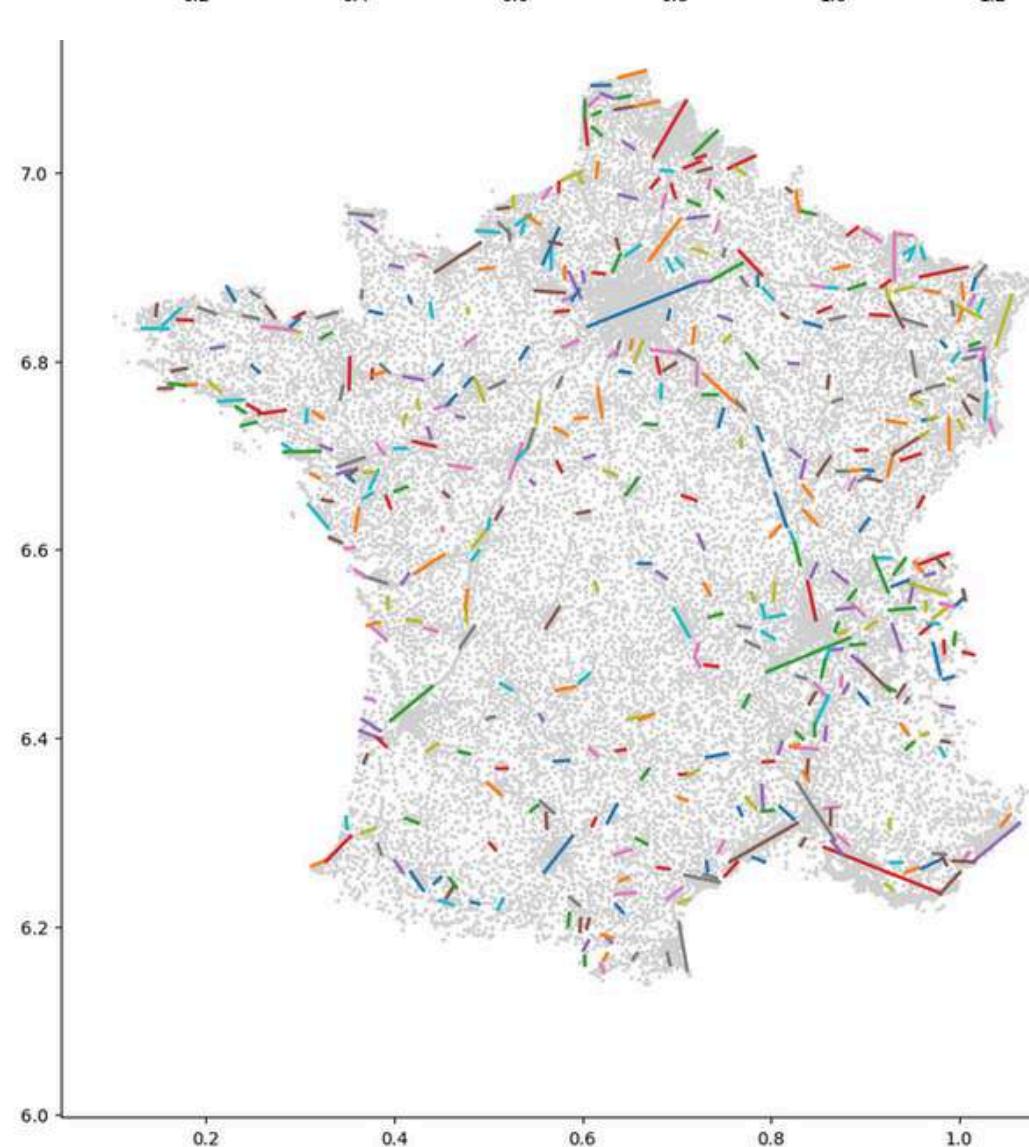
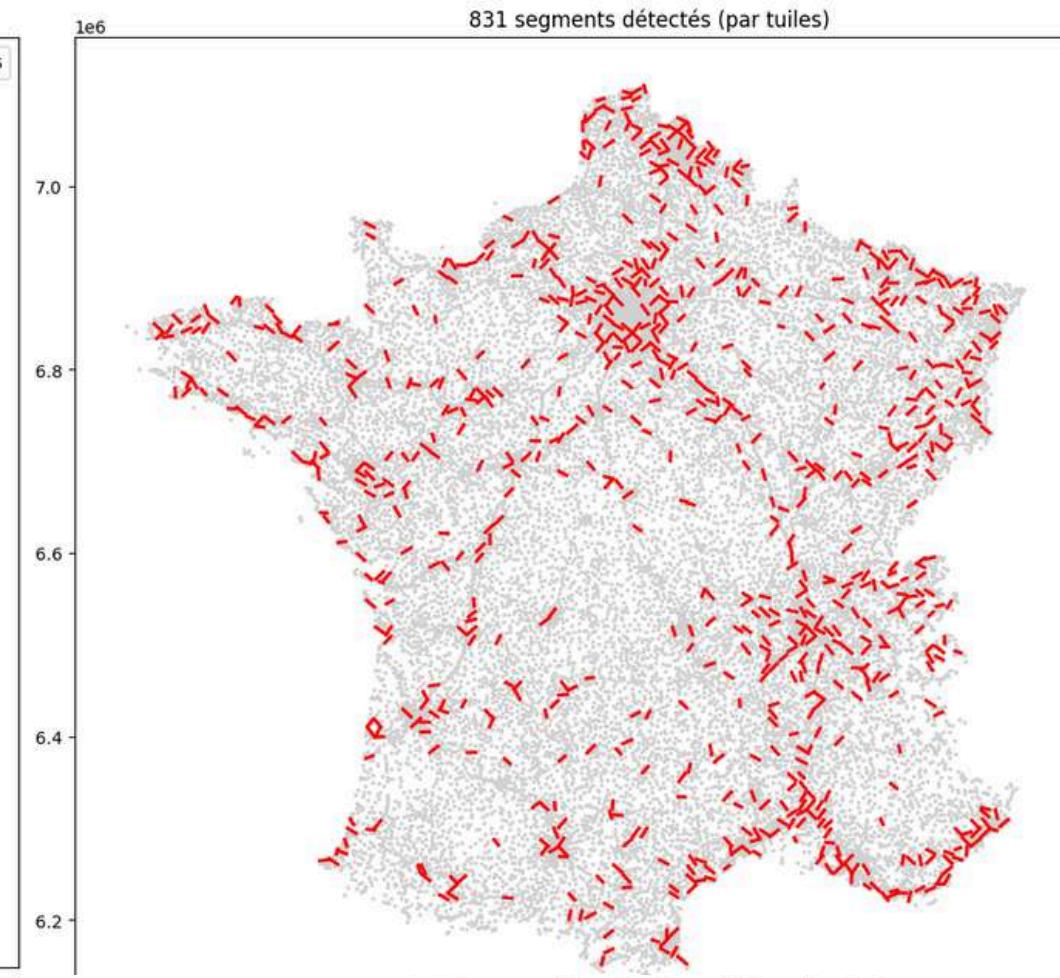
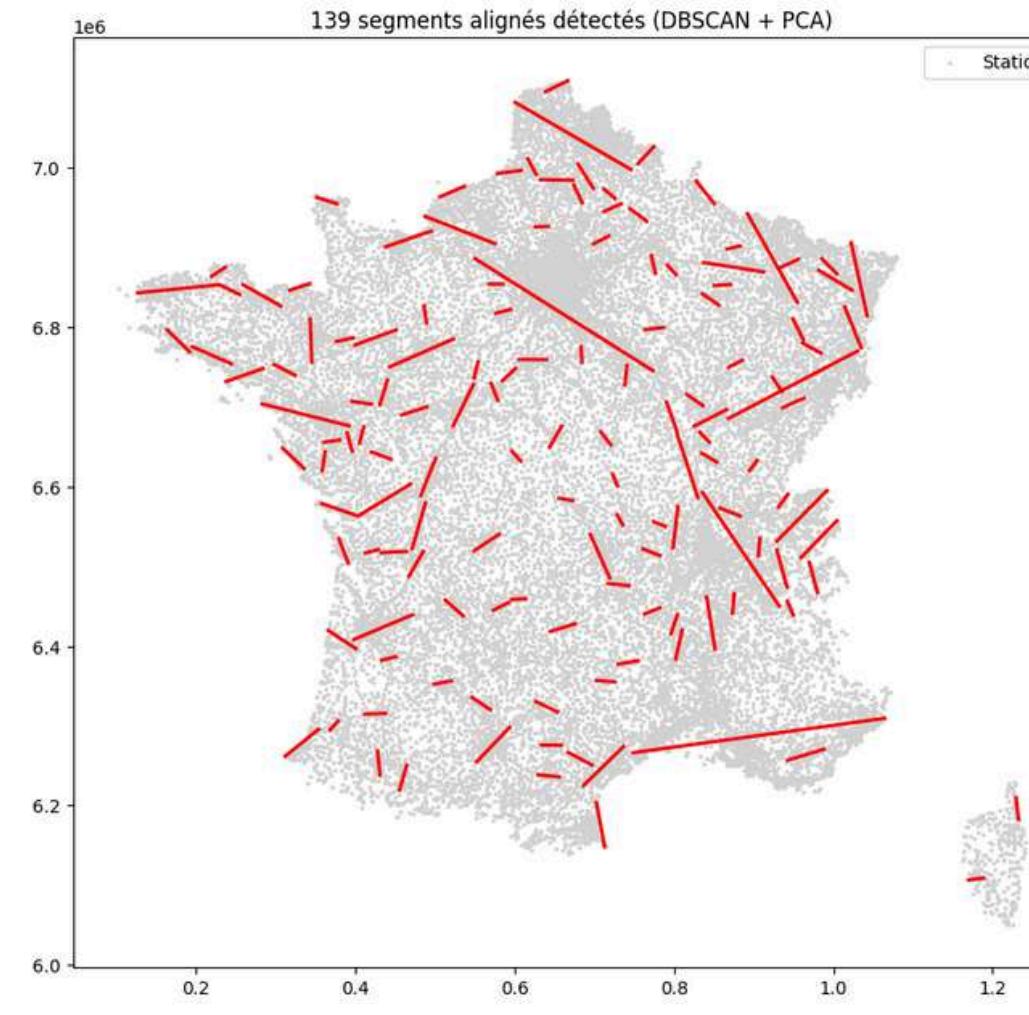
- Pour chaque station, cherche ses  $k$  plus proches voisins
- Si elles sont à peu près alignées, définit un segment local.
- Regroupe (clusterise) les segments par orientation + position
- Trace un segment dans la direction générale de chaque cluster



# Chaînes de stations

- On cherche des groupes de 3 stations “presque” alignées
- On étends les chaînes en les concaténant, si possible
- On affiche les chaînes de plus de *min\_stations*





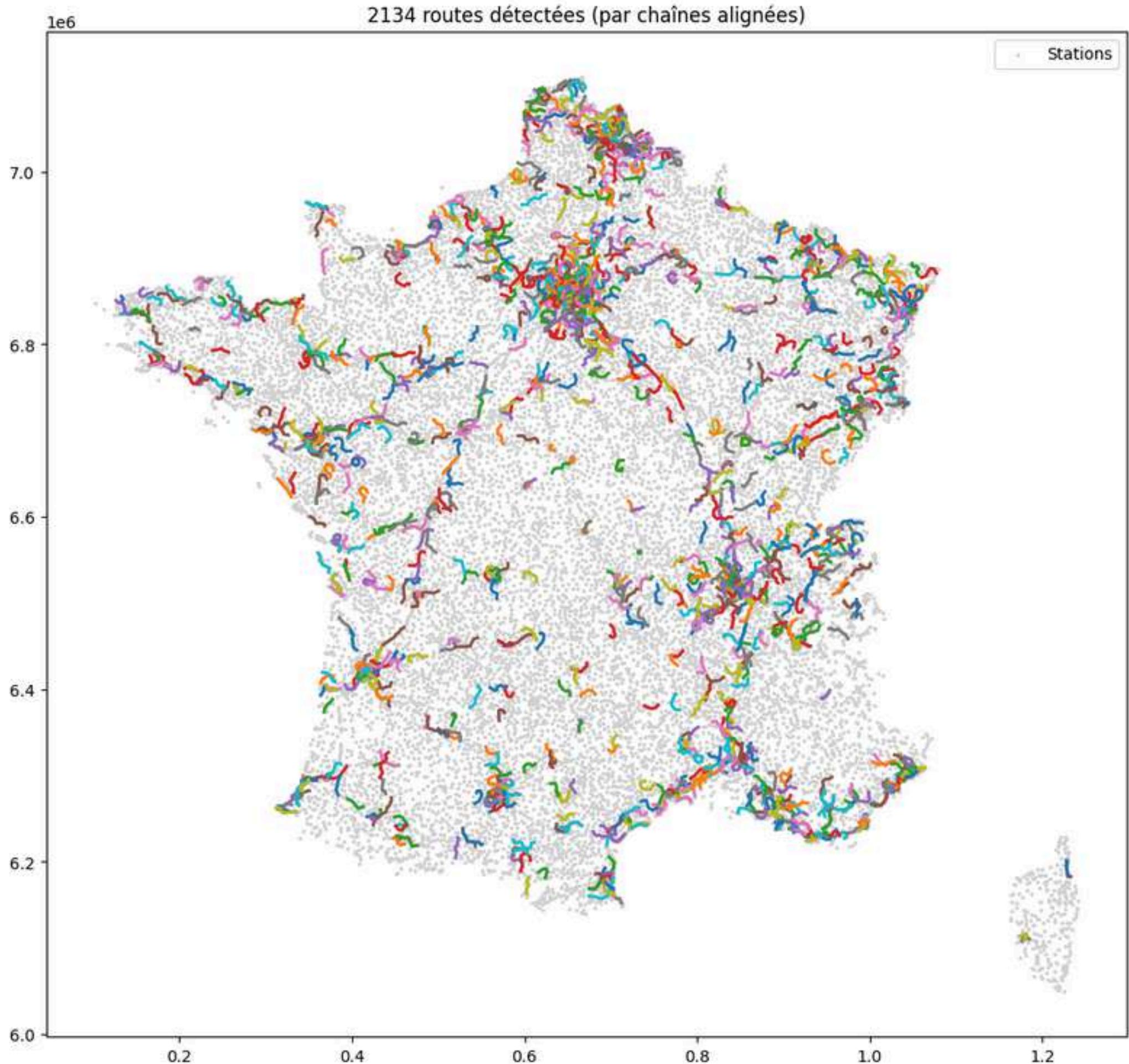
**2 - 6 Juin**

**Détection des axes routiers et ferroviaires  
Améliorations**

Rappel...

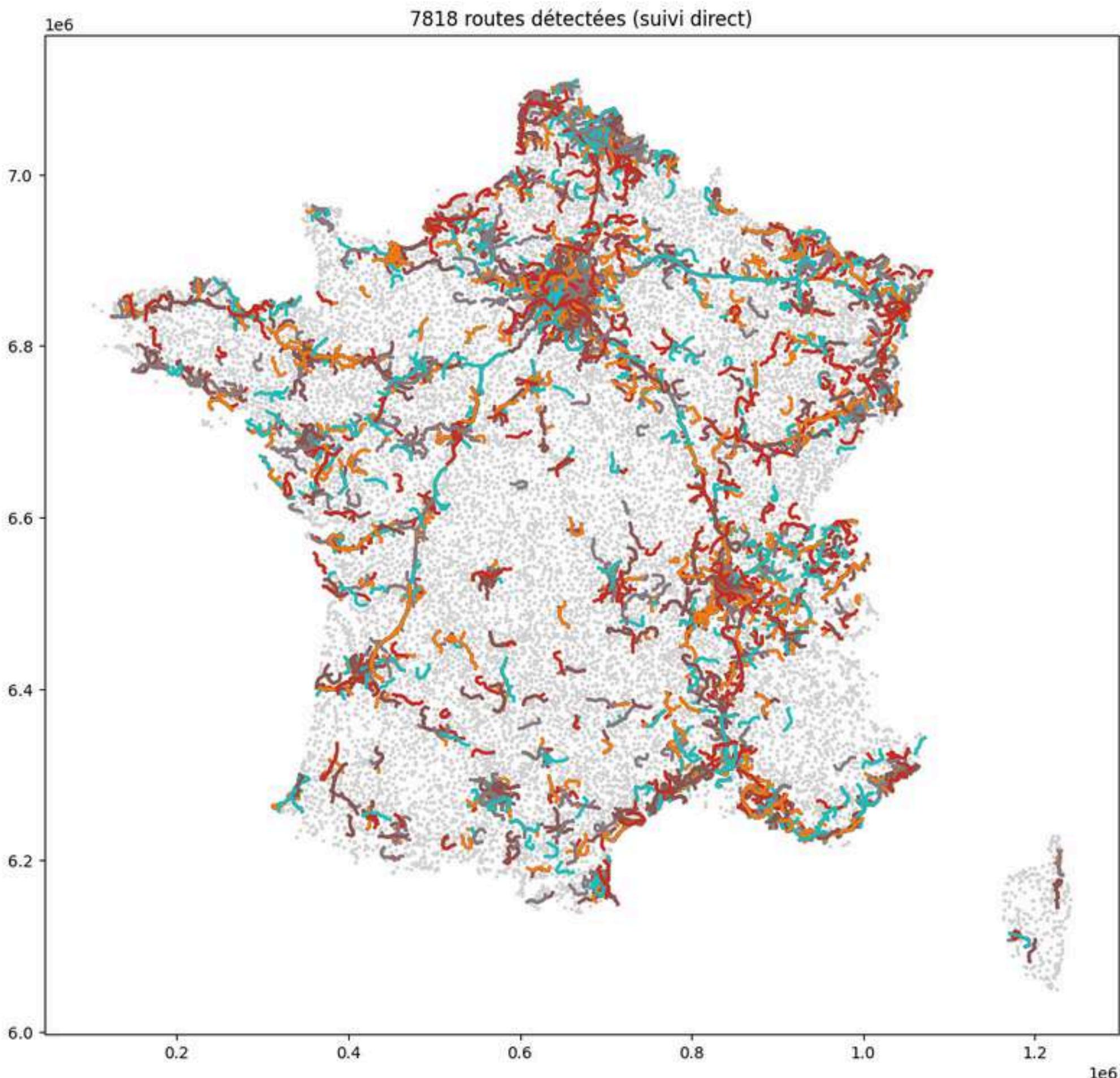
## Triplets de stations

- Pour chaque station, on cherche 2 voisins avec lesquels elle est suffisamment alignée
- On essaye de concaténer les triplets
- On affiche les chaînes de plus de *min\_stations*



# Suivi de chemin

- Pour chaque station, on suit une direction (celle du voisin le plus proche)
- A chaque étape, on considère la station la plus proche dans la même direction
- On affiche les chaînes de plus de *min\_stations*



## **Problème :**

- Certaines routes ne sont pas assez détectées
- Discontinuités sur les routes détectées

## **Améliorations :**

- Considérer plus que seulement une direction par station
- Fusionner/Concaténer les chaînes qui appartiennent sûrement à la même route

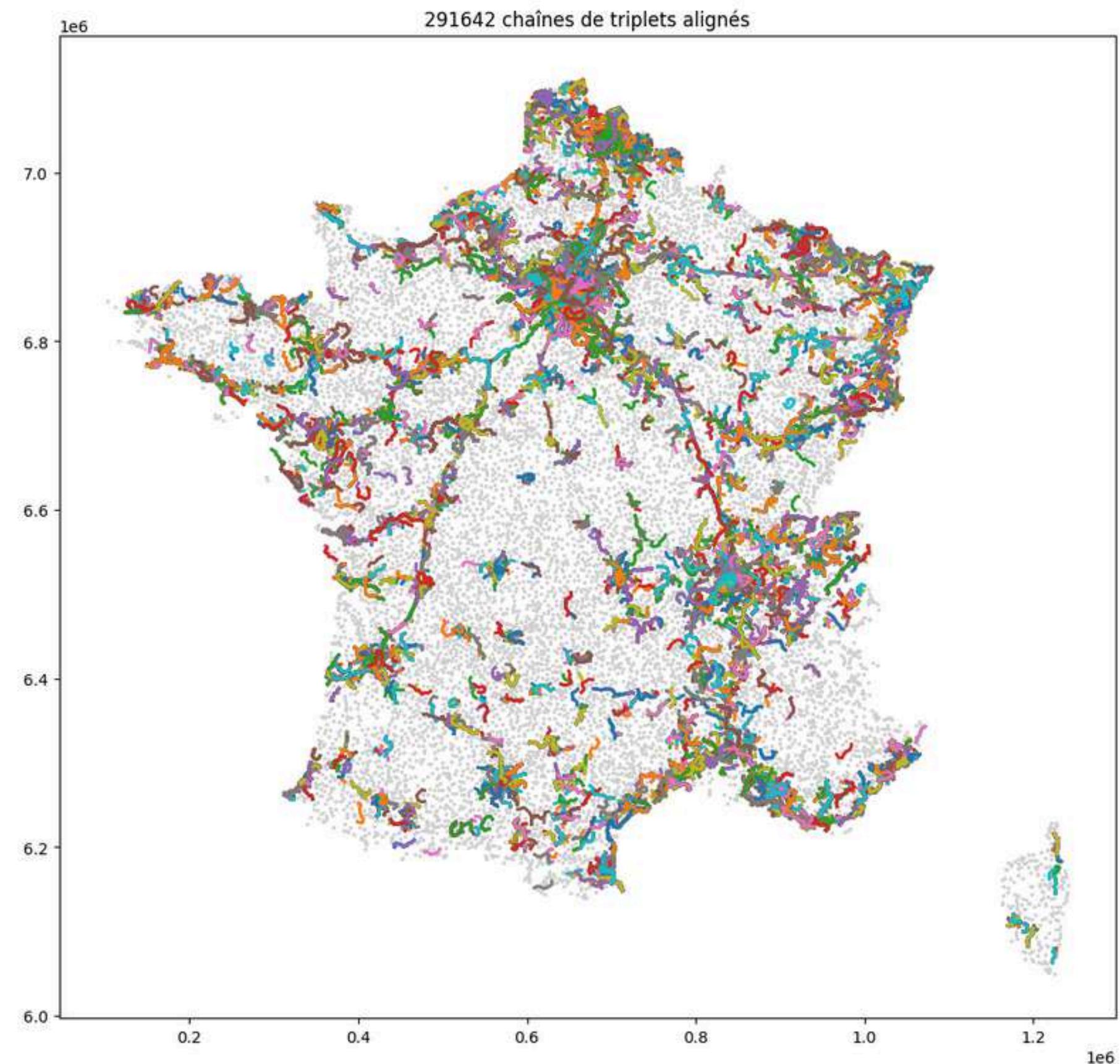
# Considérer plus de directions

- On autorise une station à appartenir à plusieurs routes pour prendre en compte les croisements de routes
- Méthode Triplets : On considère tous les triplets suffisamment alignés parmi les  $k$  plus proches voisins de la station (plutôt que le meilleur)
- Méthode Suivi de chemin : On part dans les directions des  $k$  plus proches voisins (plutôt que seulement le plus proche)

# Considérer plus de directions de départ

## Méthode Triplets

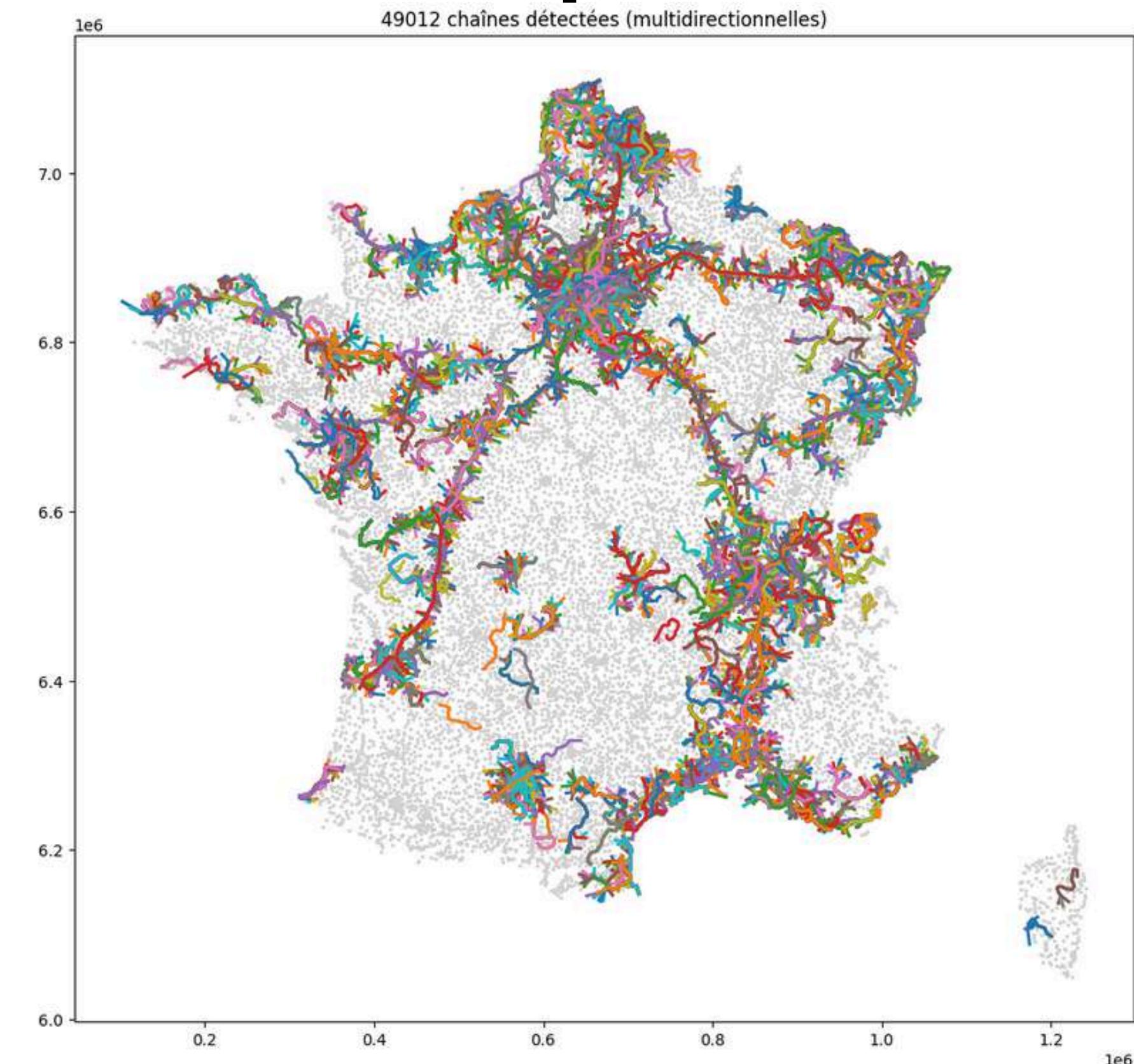
- On considère les 10 plus proches voisins
- Beaucoup moins de discontinuités des routes : considérer seulement le tripler le plus aligné par station faisait passer à côté de beaucoup de sections de routes



# Considérer plus de directions de départ

## Méthode suivi de chemin

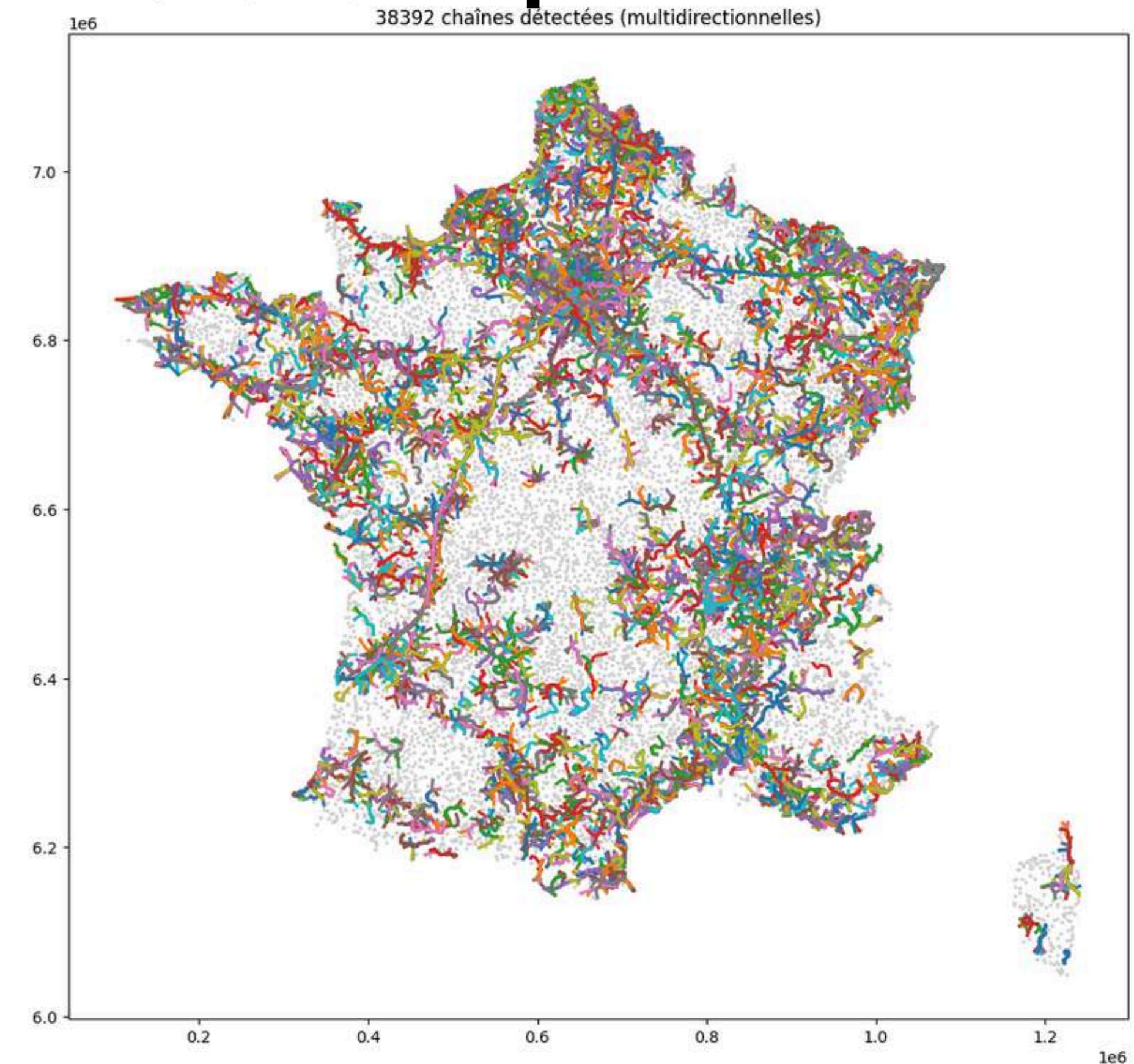
- Ici 10 directions par station
- Beaucoup de bruit autour des routes, beaucoup de chemins très courts détectés
- Affichage seulement des routes de plus de 20 stations ici



# Considérer plus de directions de départ

## Méthode suivi de chemin

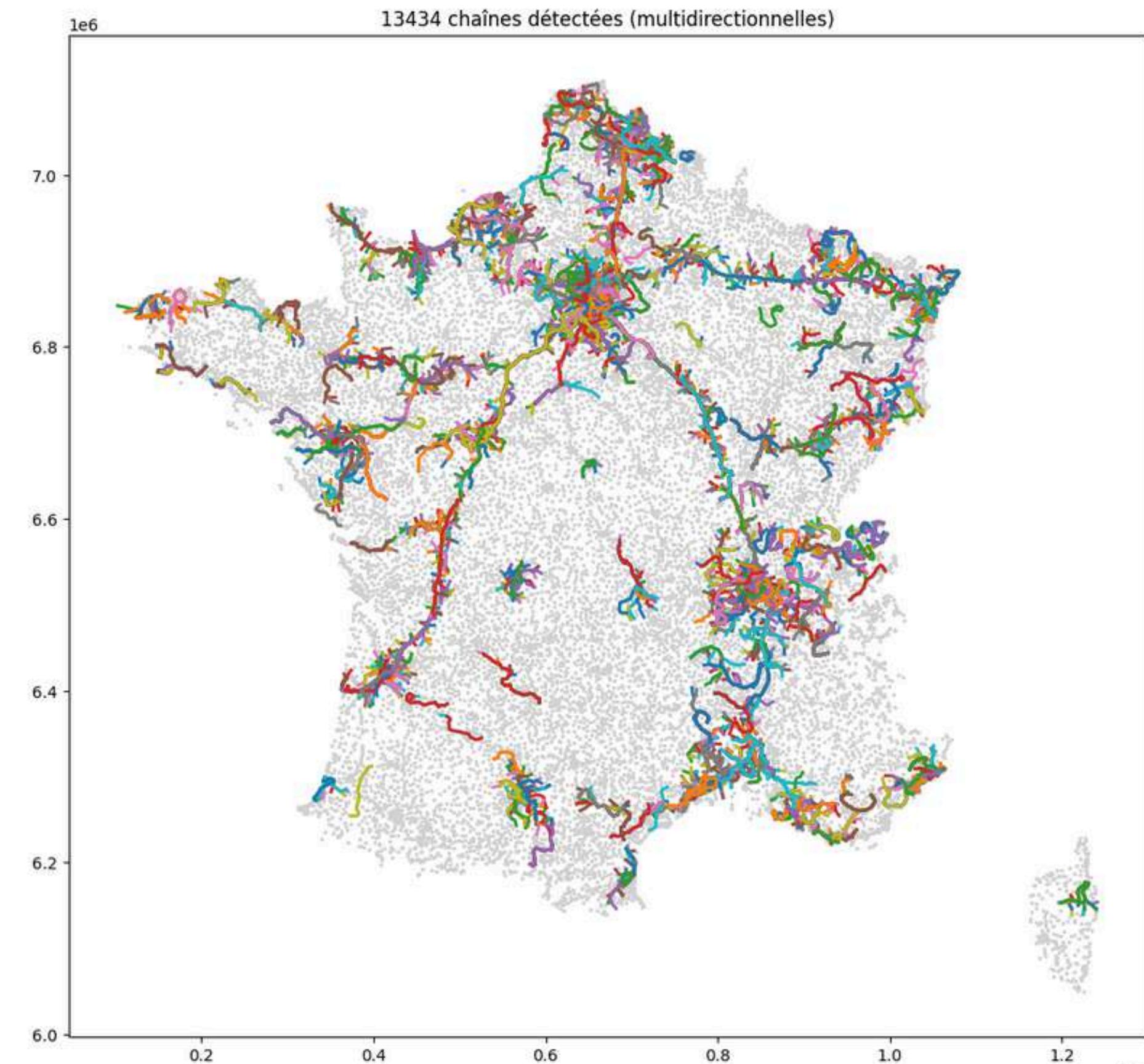
- Ici 5 directions par station
- Beaucoup de bruit autour des routes
- Nombre minimum de stations par route : 9



# Considérer plus de directions de départ

## Méthode suivi de chemin

- 5 directions par station
- Affichage ici seulement des routes de plus de 20 stations ici, au lieu de 9 avant



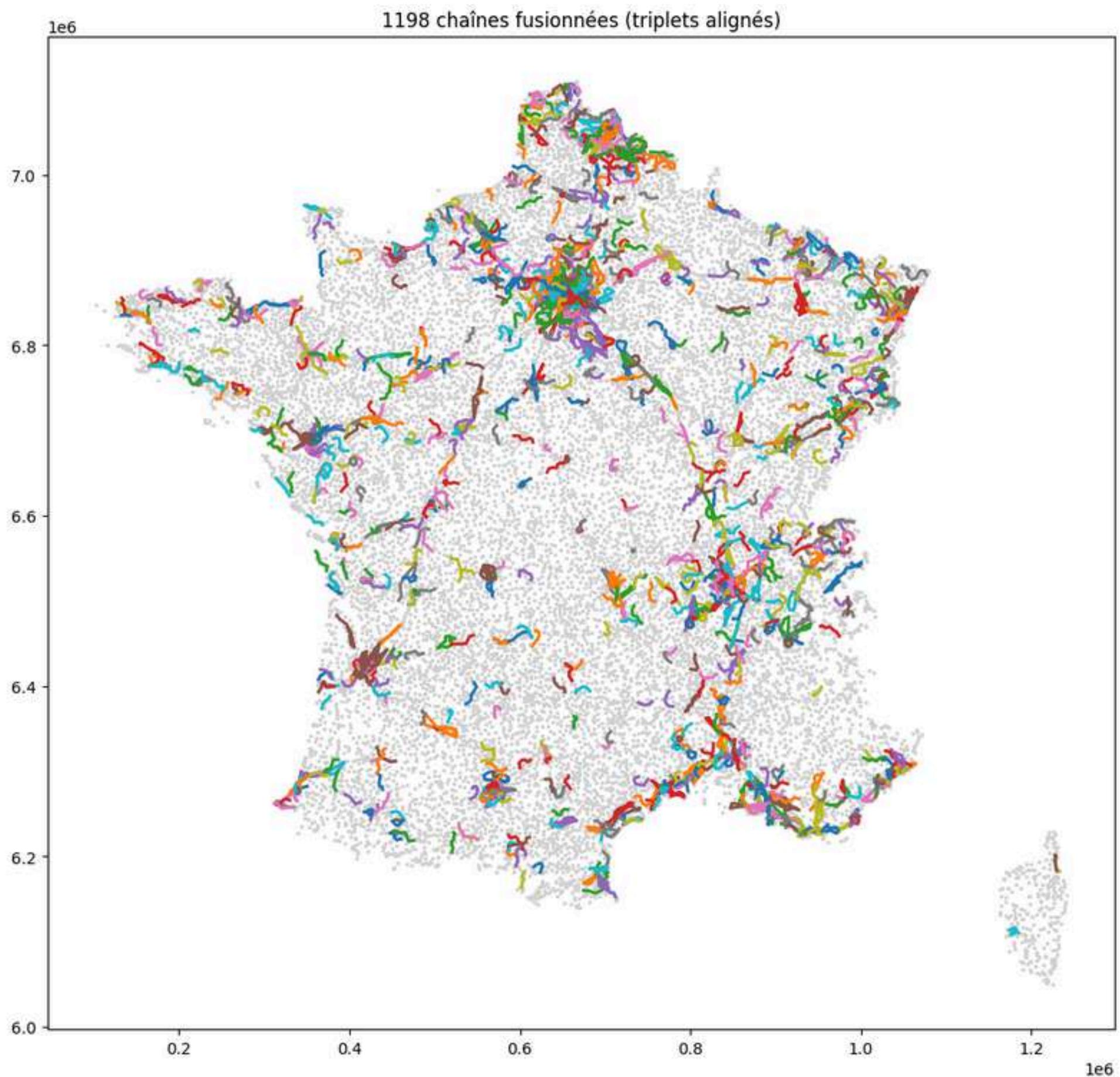
# Fusionner les routes

- On considère la distance entre les extrémités de 2 chemins, et leur direction générale
- 2 paramètres ajustables : la distance et l'angle entre les directions maximum pour accepter la fusion

# Fusionner les routes

## Méthode Triplets

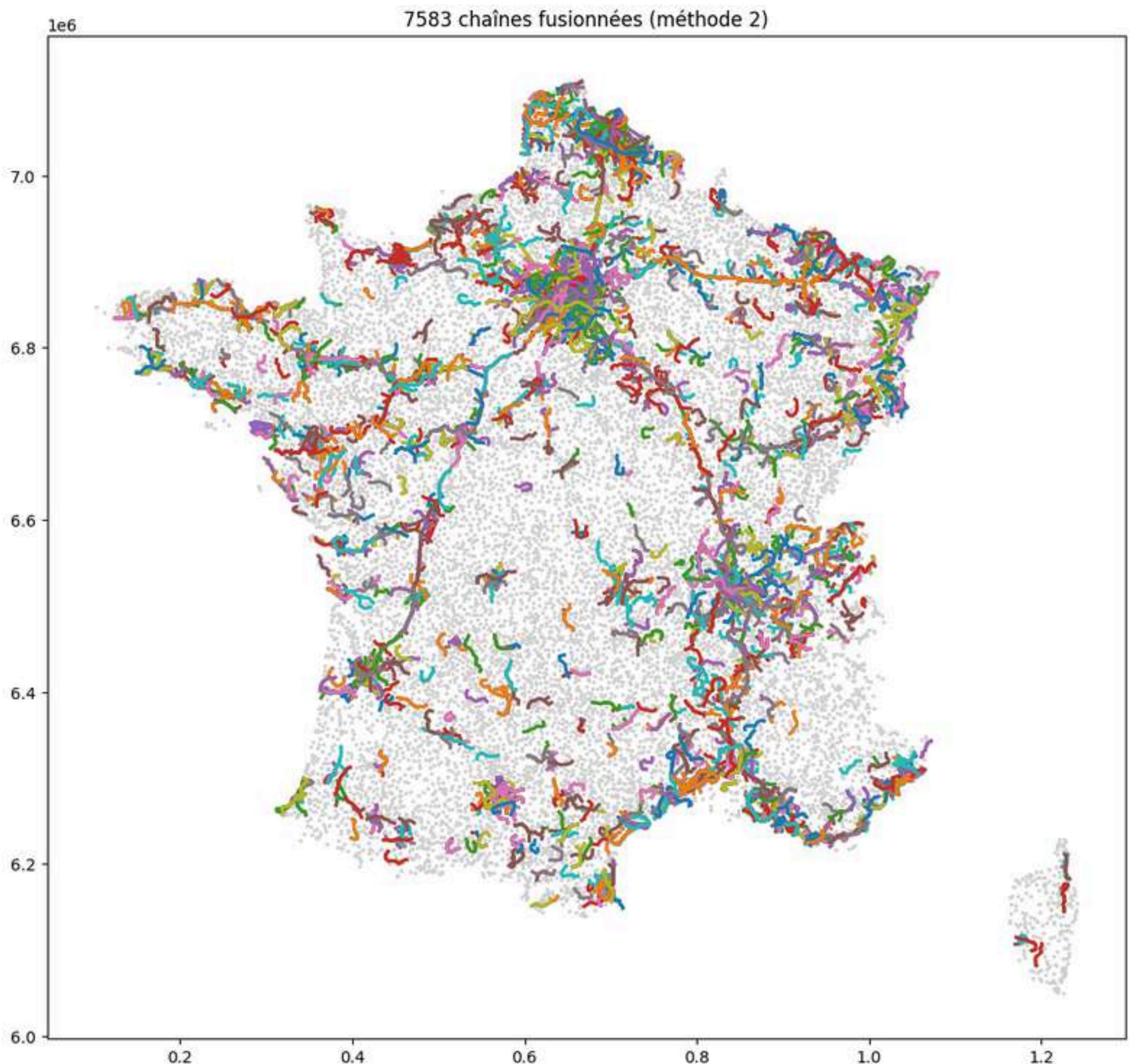
- $2134 \rightarrow 1198$  chaînes
- Pas de grande amélioration des routes détectées
- Sans doute améliorable en jouant avec les paramètres



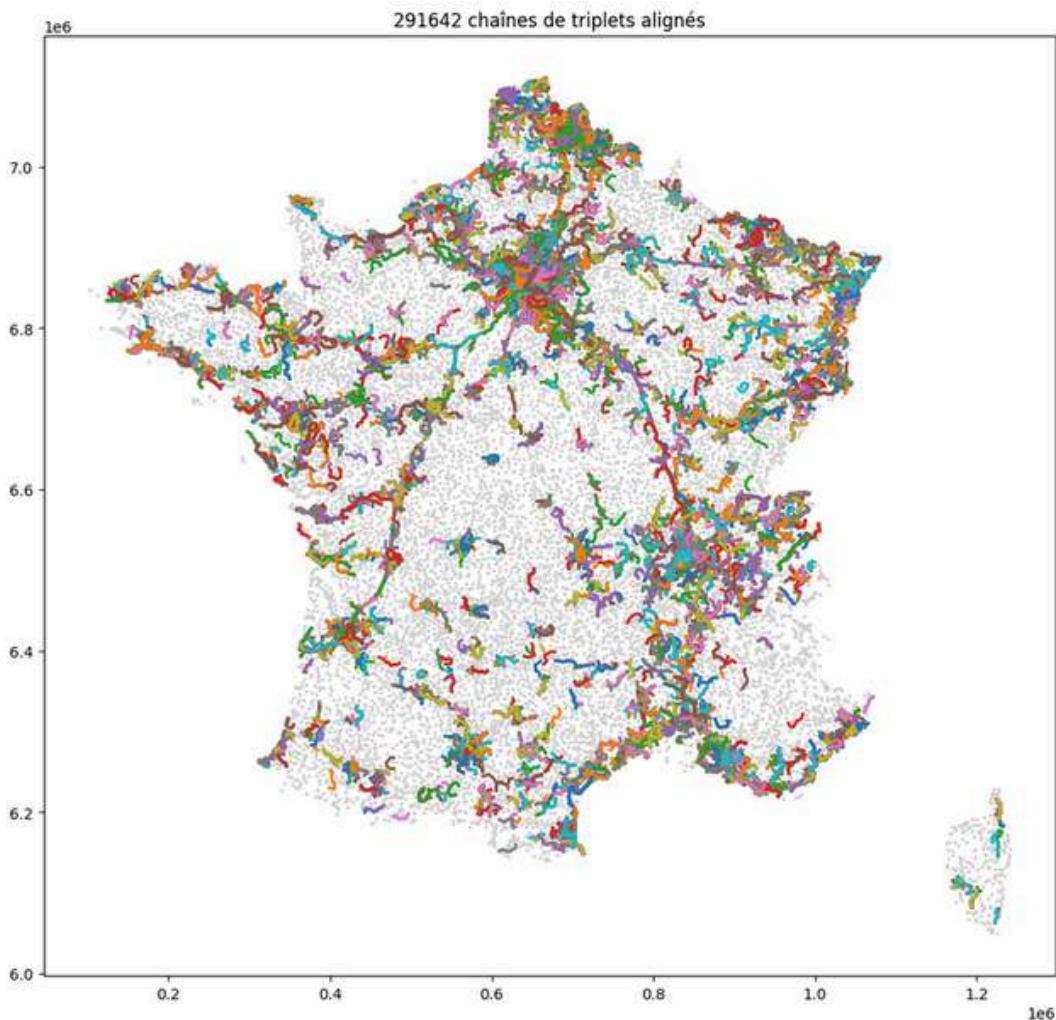
# Fusionner les routes

## Méthode suivi de chemin

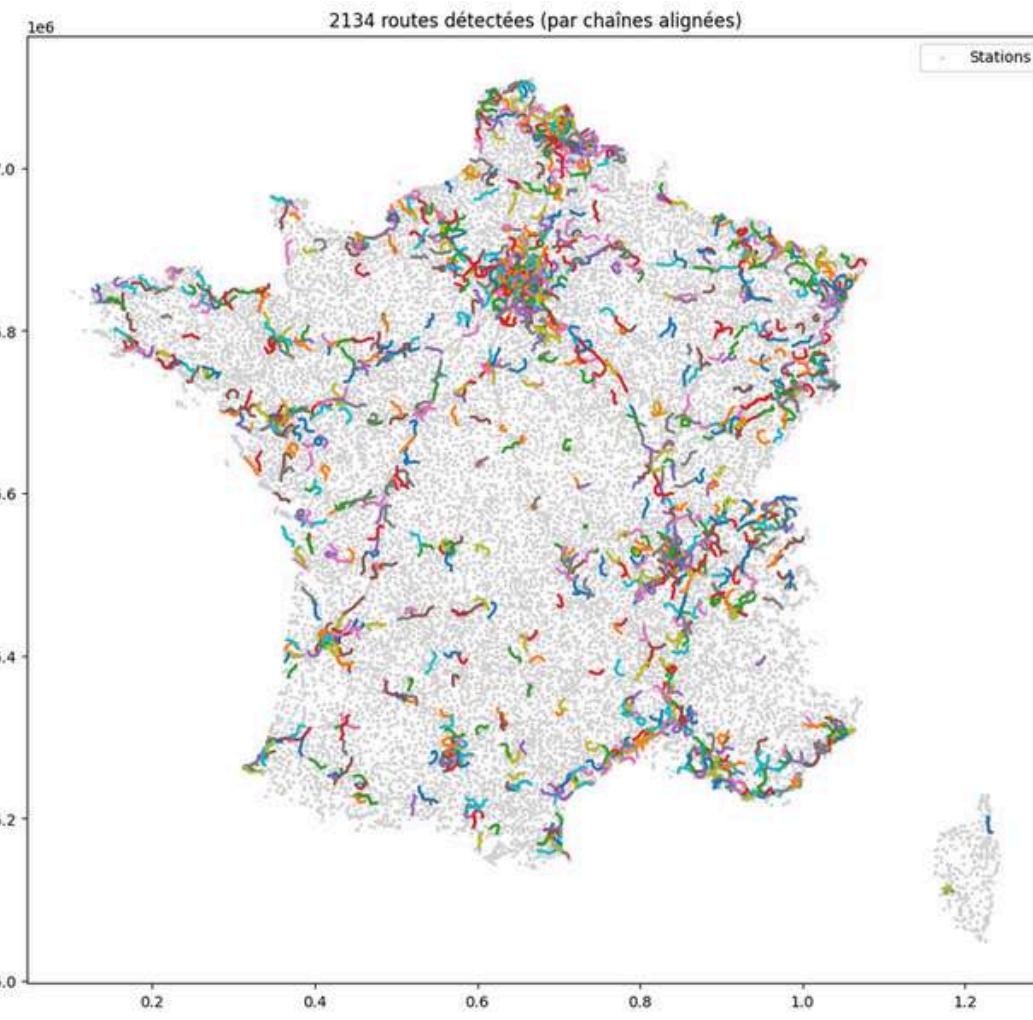
- $7818 \rightarrow 7583$  chaînes
- Sans doute encore améliorable en jouant avec les paramètres



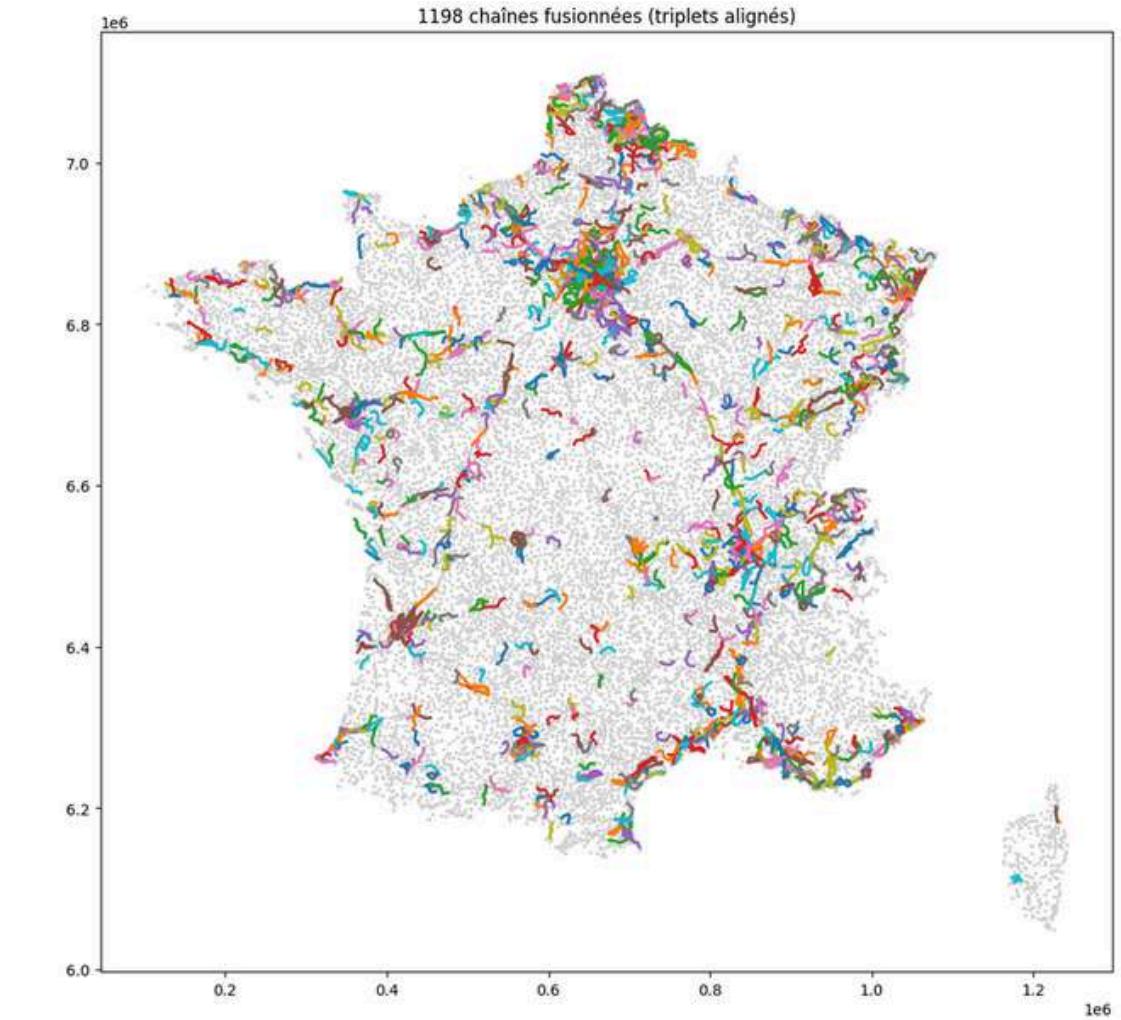
# Triplets de stations



Multi-directions

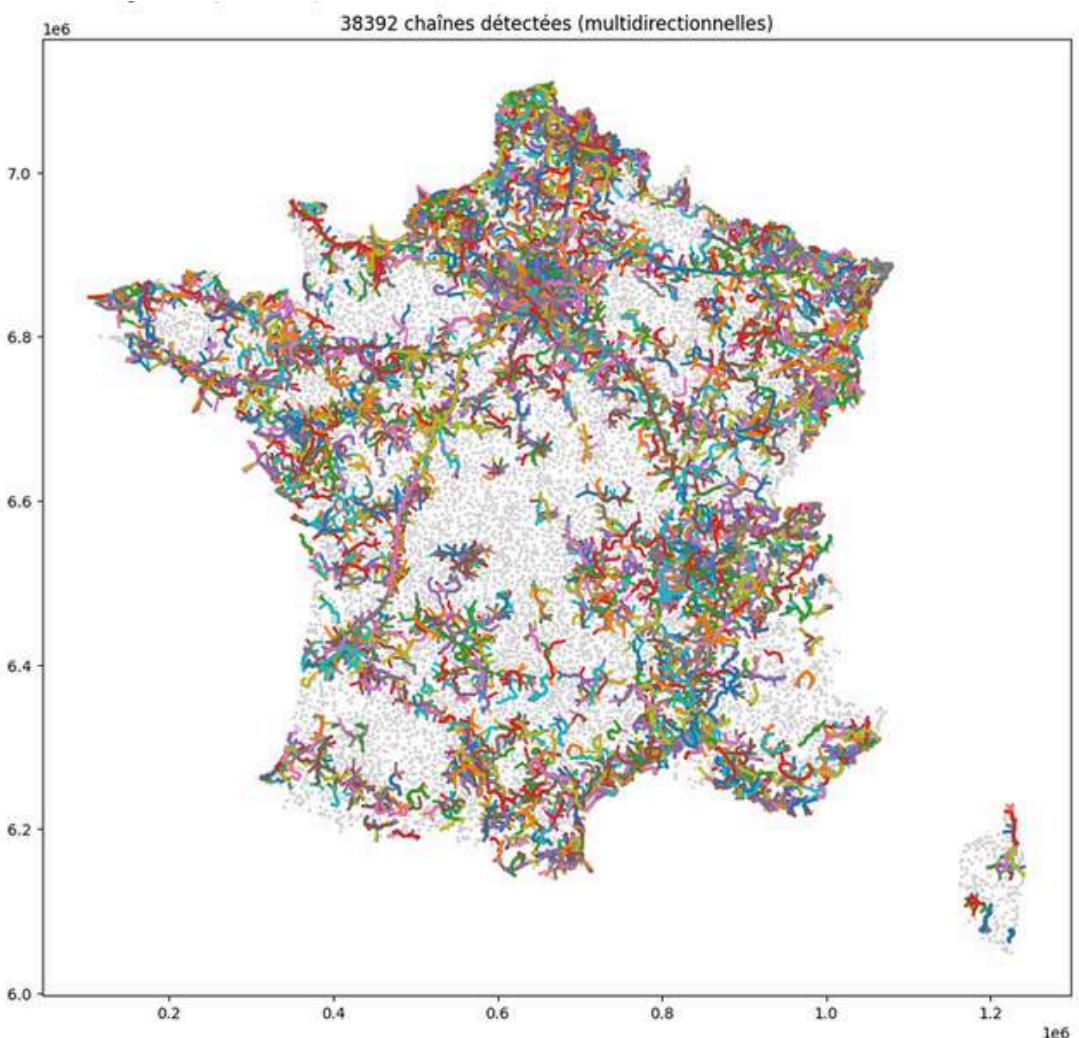


Base

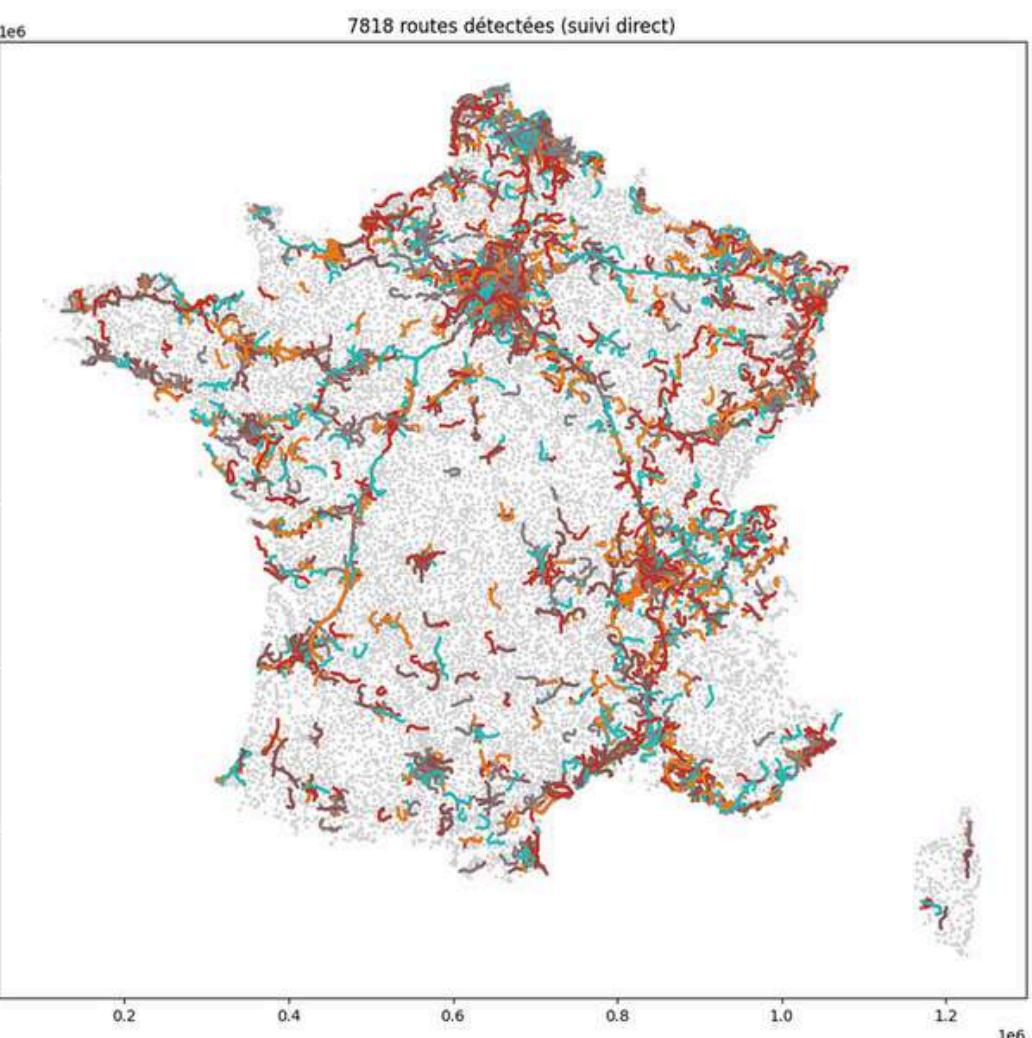


Fusion des routes

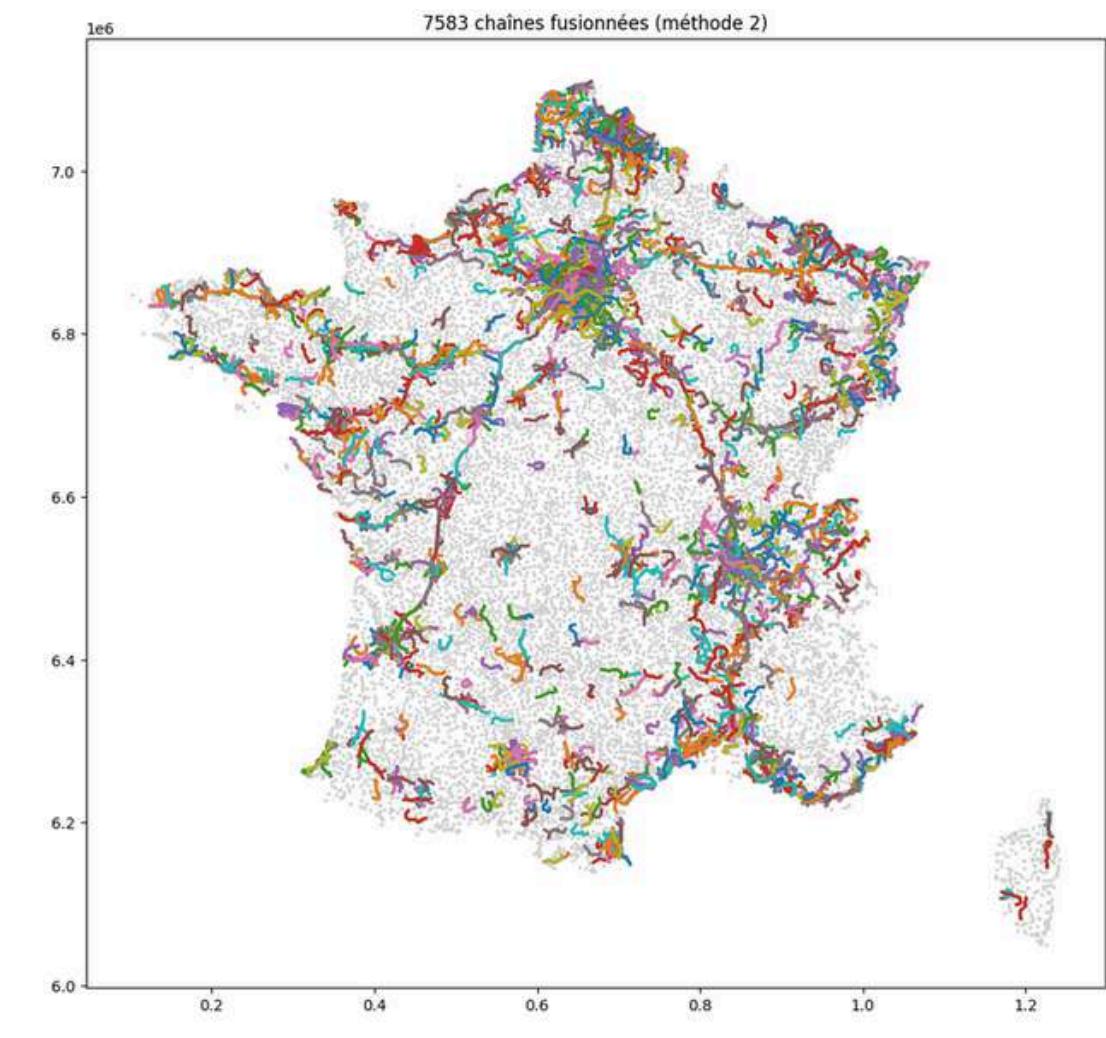
# Suivi de chemin



Multi-directions



Base



Fusion des routes

## Résultats :

- Améliorations coûteuses, ajustement des paramètres chronophage
- Résultats sur seulement une région possible mais moins facile d'interpréter les résultats, notamment ne permet pas de voir les disparités au niveau national entre les routes plus ou moins importantes

## Idées :

- Ajuster les paramètres par région, prendre en compte densités de stations et de population, et topologie
- Combiner les 2 améliorations (très couteux car beaucoup plus de chemins à fusionner)

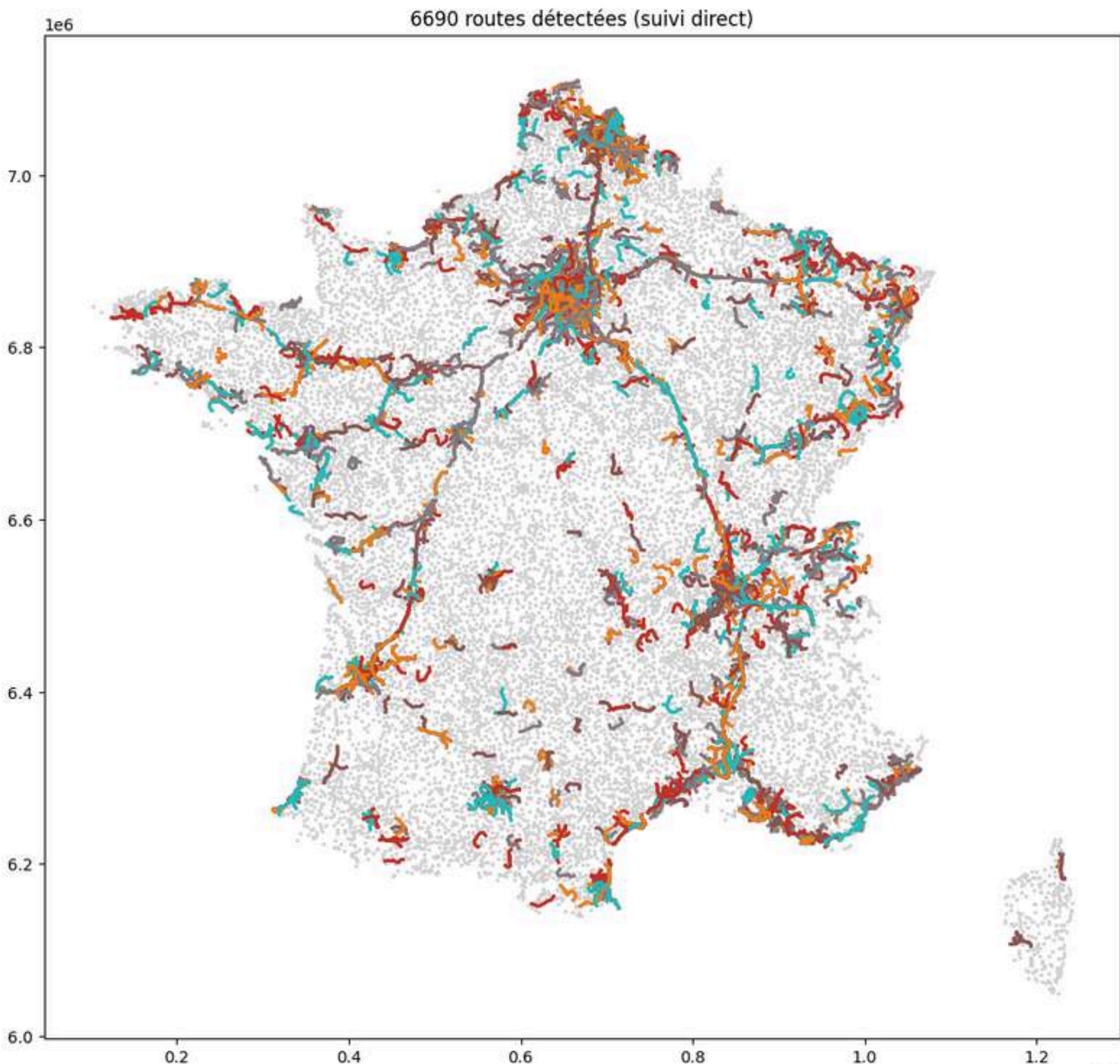
**9 - 13 Juin**

**Détection des axes routiers et ferroviaires  
Suite**

Rappel...

# Suivi de chemin

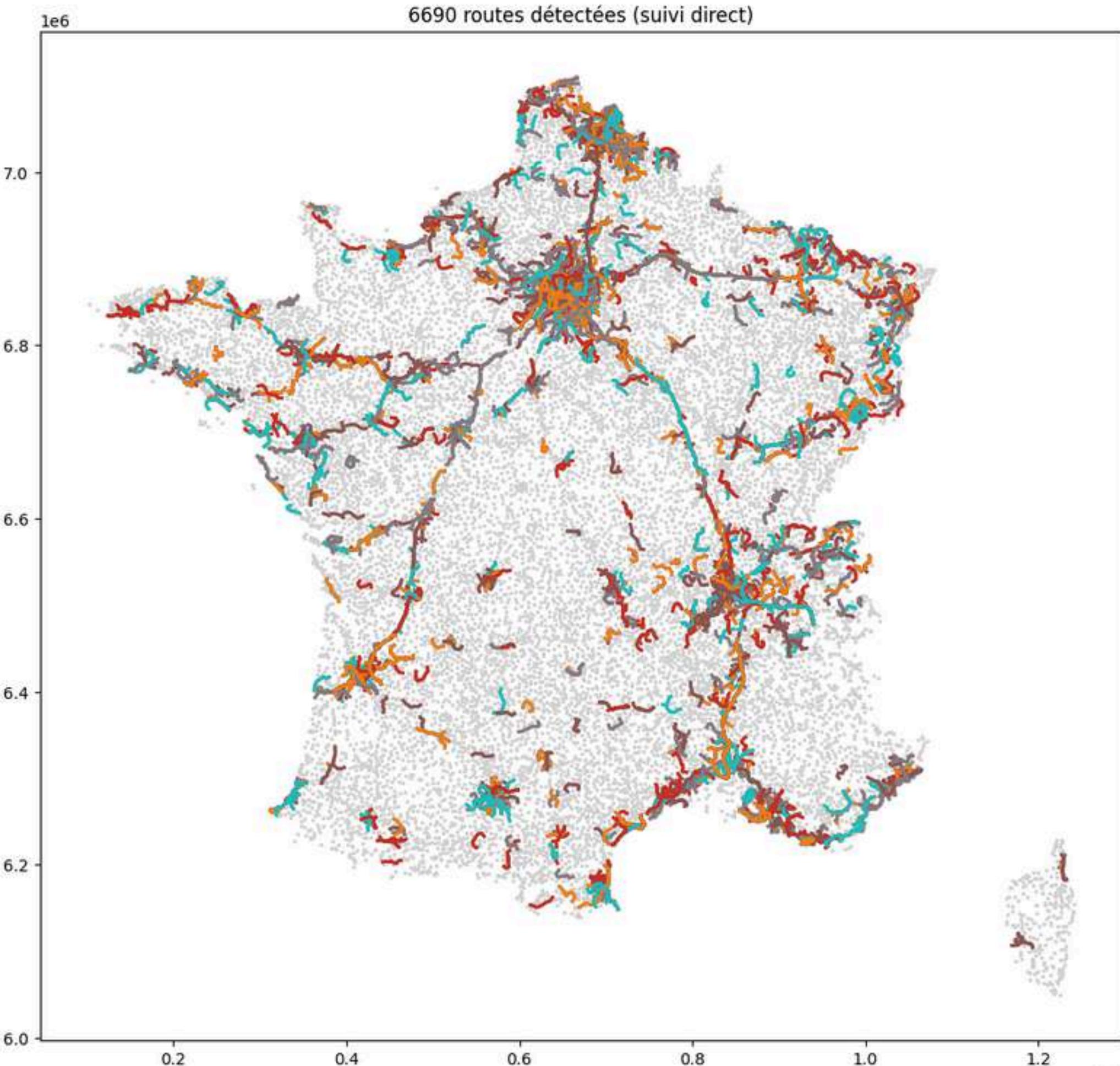
- Pour chaque station, on suit une direction (celle du voisin le plus proche)
- A chaque étape, on considère la station la plus proche dans la même direction
- On affiche les chaînes de plus de *min\_stations*



# Suivi de chemin

## Paramètres ici

- $\text{min\_chain\_len} = 9$
- A chaque étape on prends le meilleur voisin parmi les  $k = 10$  plus proches. Il doit être :
  - à moins de  $\text{dist\_max} = 4500m$
  - faire un angle de maximum  $\text{angle\_max} = 80^\circ$  avec la direction générale de la route (en valeur absolue)

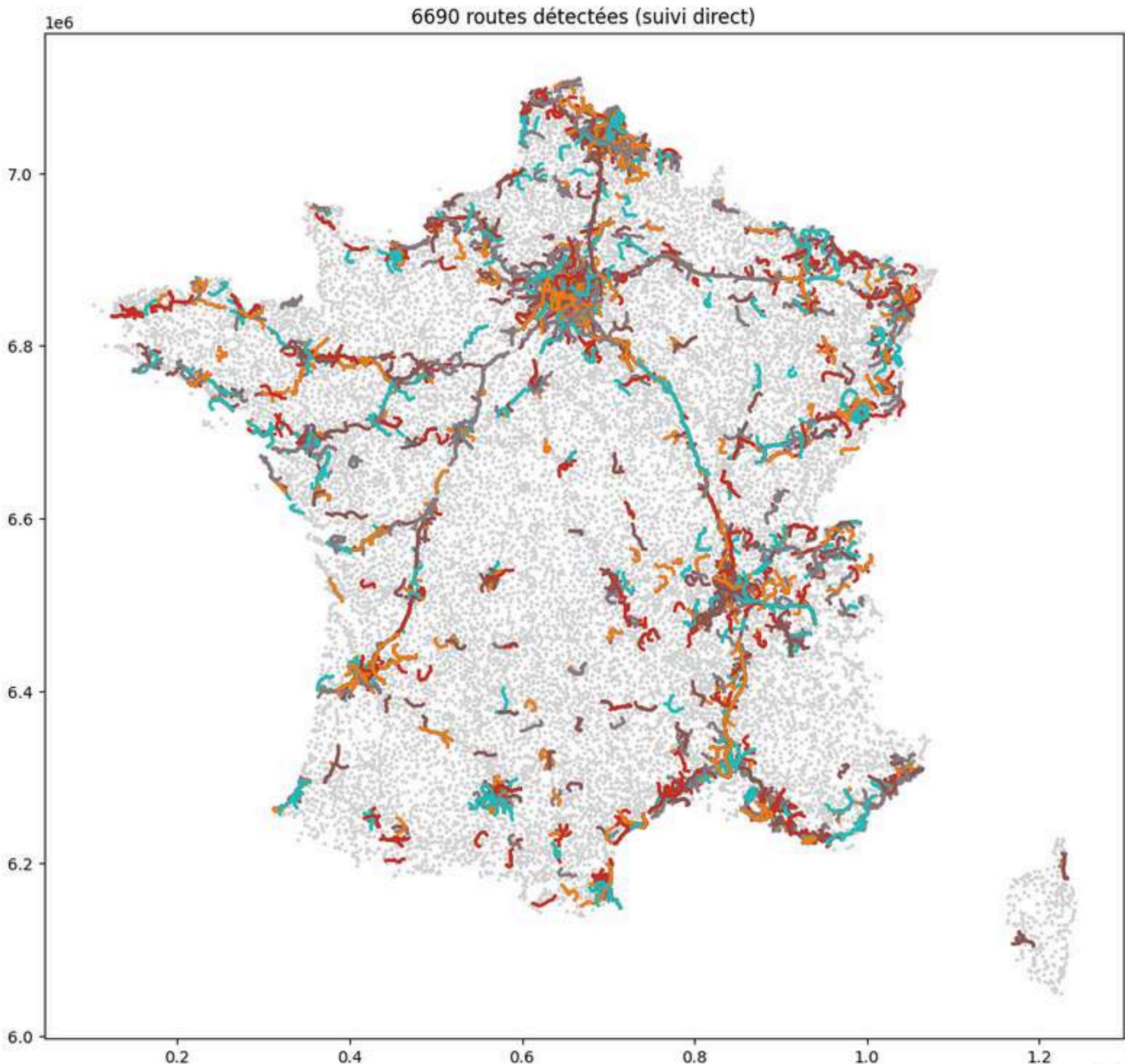


## Améliorations :

- On considère plus de stations de départ (donc plus de directions) :
  - Ajout d'un paramètre *n\_directions*
  - Et *angle\_tol\_deg* : on veut un certain angle entre 2 directions de départ, pour ne pas suivre 2 fois la même route
- On fusionne les chaînes trouvées :
  - Ajout de *fusion\_dist* : On fusionne 2 chaînes si leur extrémités sont assez proches
  - Ajout de *fusion\_angle* : On fusionne 2 chaînes si leur direction est presque la même

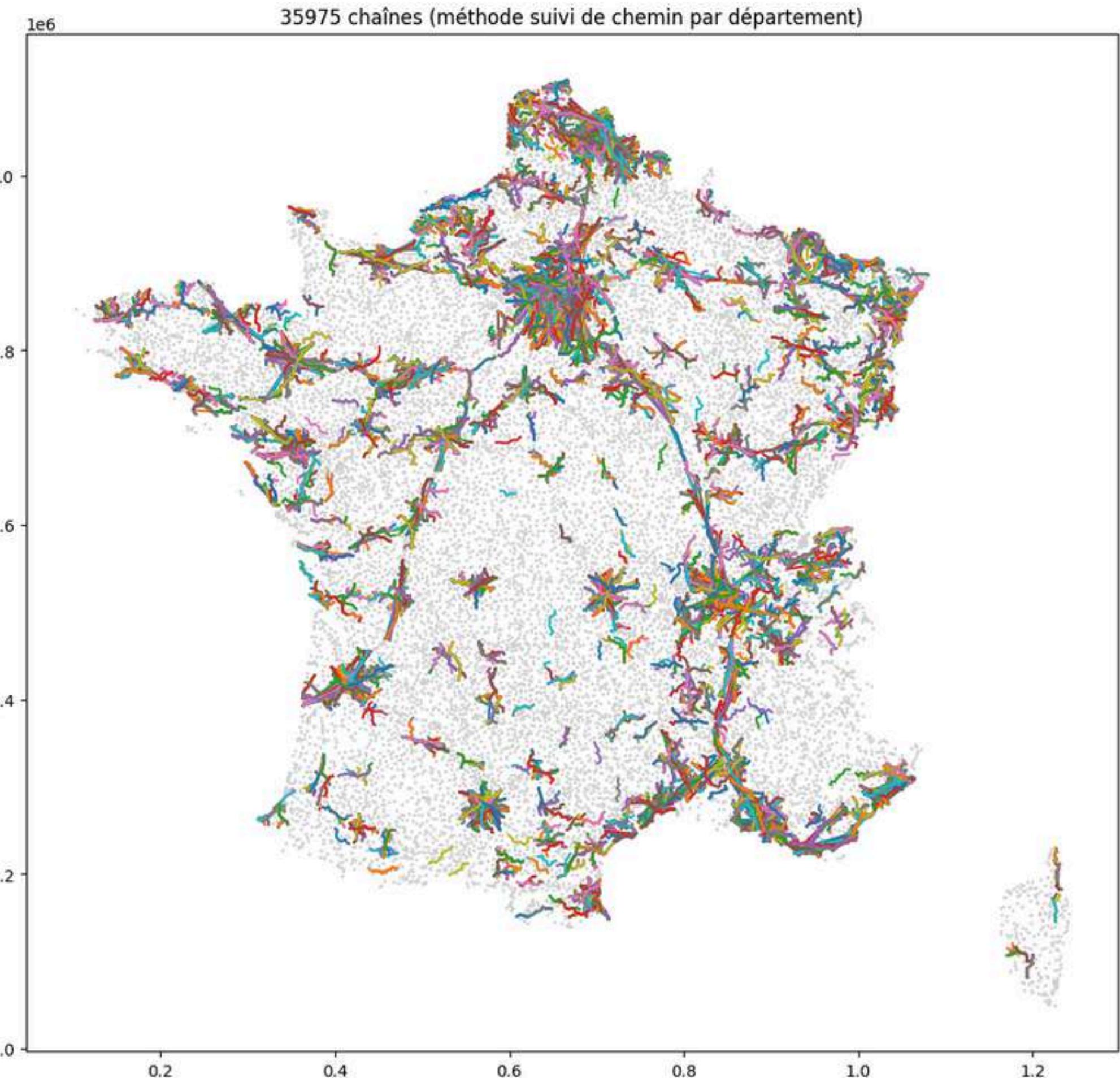
# Suivi de chemin Par département

- On applique la méthode du suivi de chemin pour chaque département avec des paramètres adaptés
- Concrètement : dictionnaire de paramètres en fonction des départements



# Suivi de chemin Par département

- Ici valeurs de paramètres par défaut :
  - `n_directions` = 5
  - `angle_tol_deg` = 2
  - `k` = 15
  - `dist_max` = 4500
  - `angle_max` = 80
  - `min_len` = 9
  - `fusion_dist` = 350
  - `fusion_angle` = 25



## **Améliorations :**

- Ajuster au mieux les paramètres pour chaque département
- Fusionner les routes entre elles après l'avoir fait par département (coûteux)

## **Idéalement :**

- Modifier les paramètres automatiquement en fonction de la densité de stations autour de la station

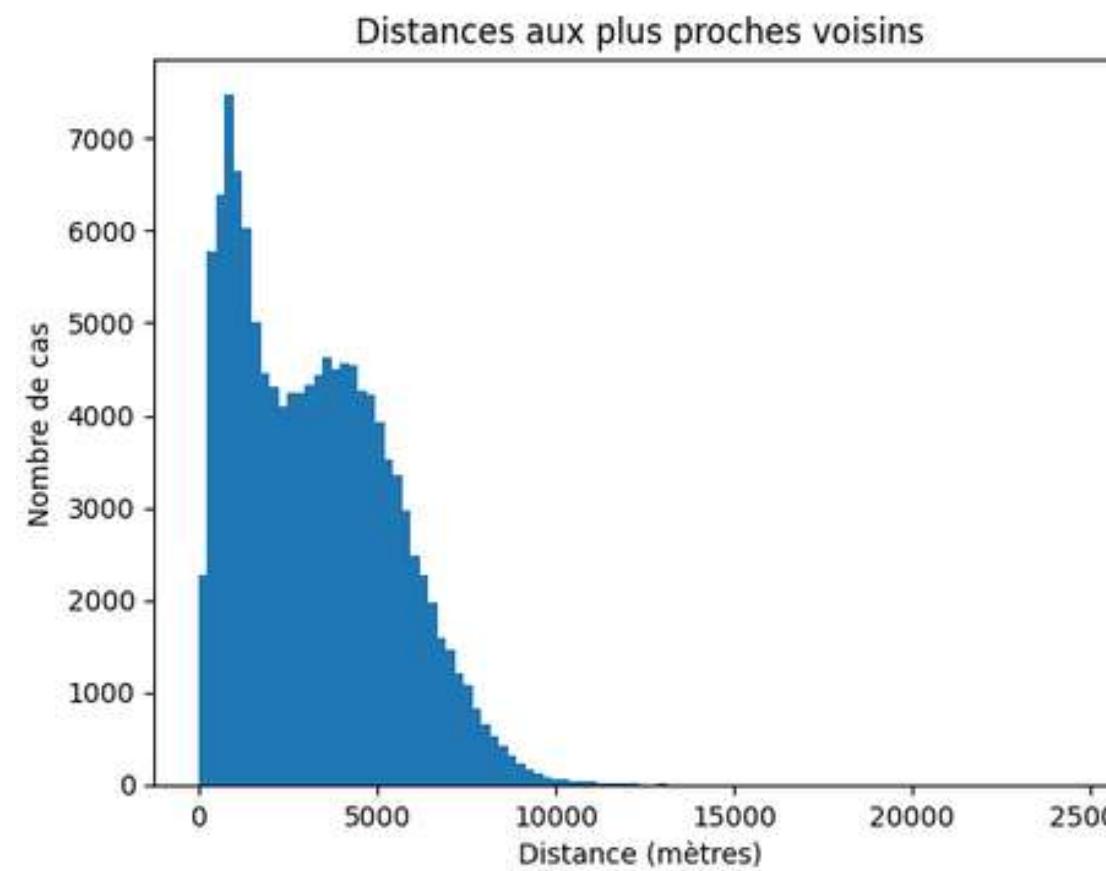
**13 - 18 Juin**

# **Améliorations de la méthode suivi de chemin**

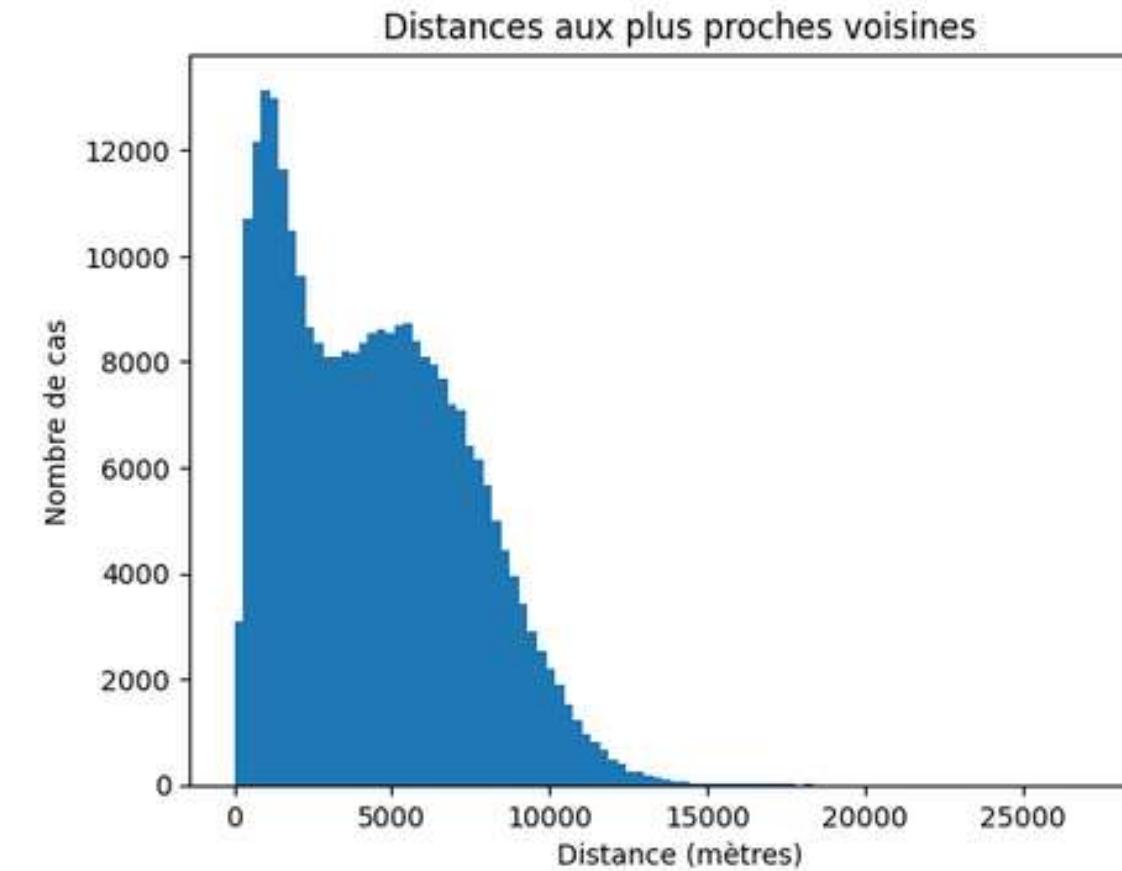
## Suite

# Paramètre *dist\_max*

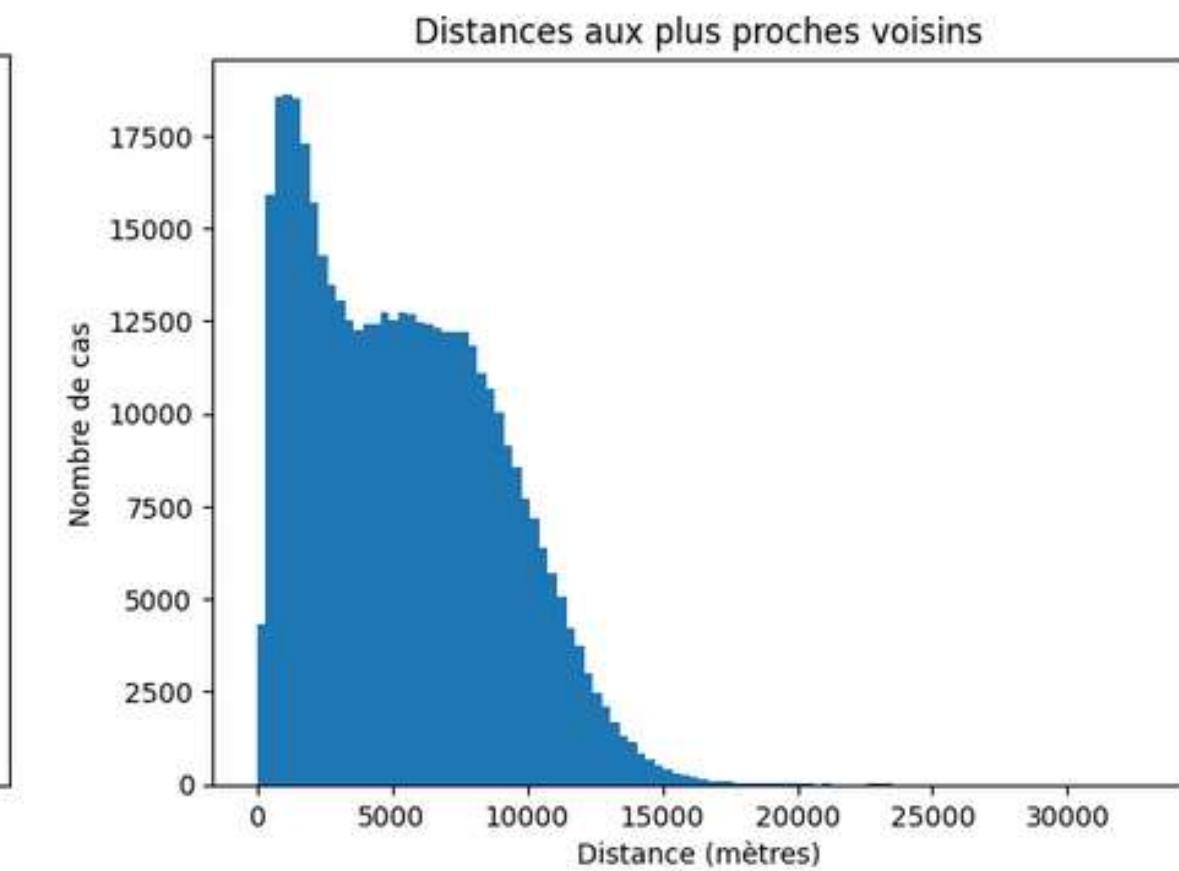
- Distance maximale entre 2 stations pour étendre une route
- Qualitativement : entre 4000 et 6000 mètres
- Nombre de stations par distance moyenne à ses  $k$  plus proches voisins :



$k = 5$



$k = 10$

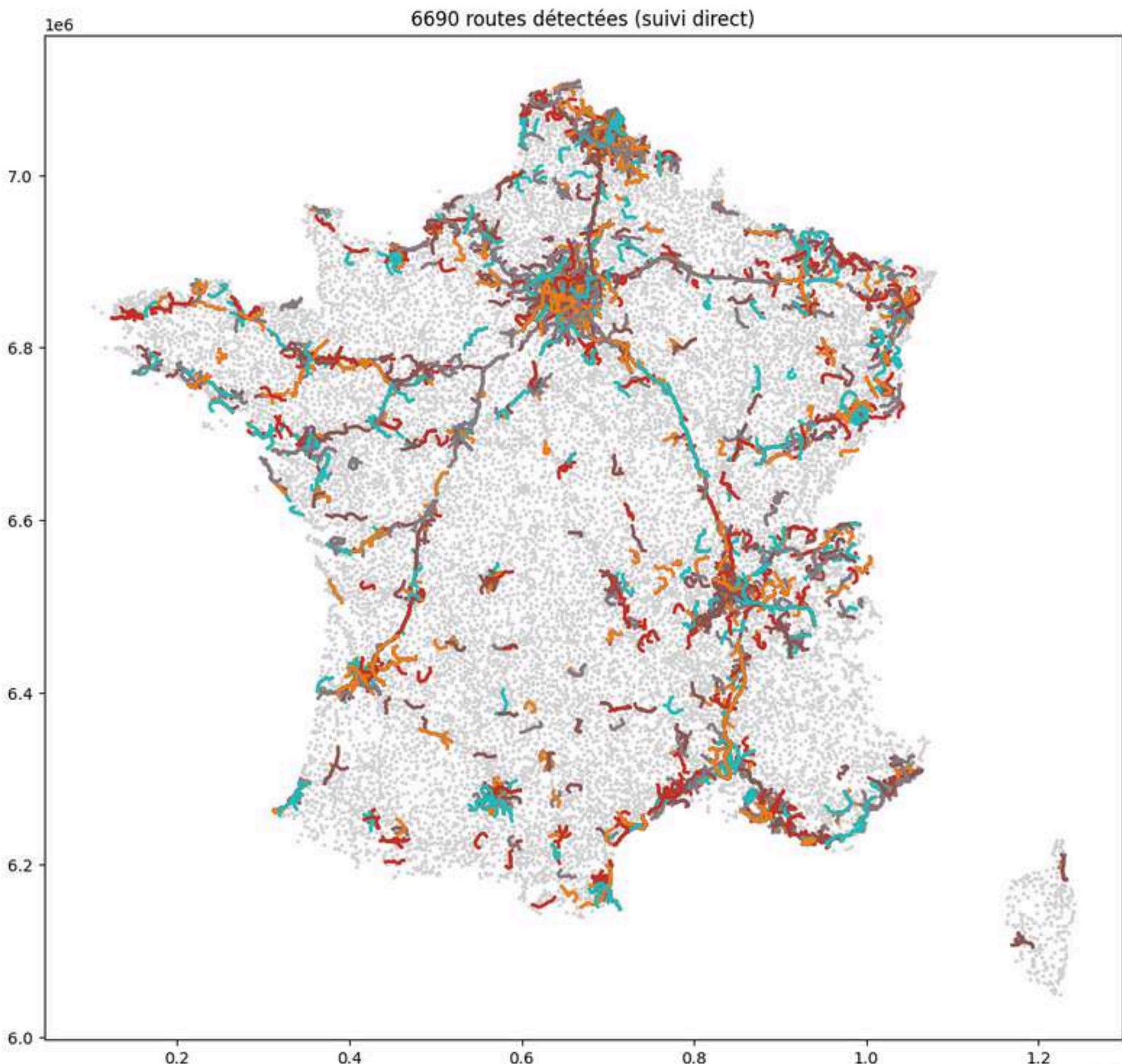


$k = 15$

Rappel...

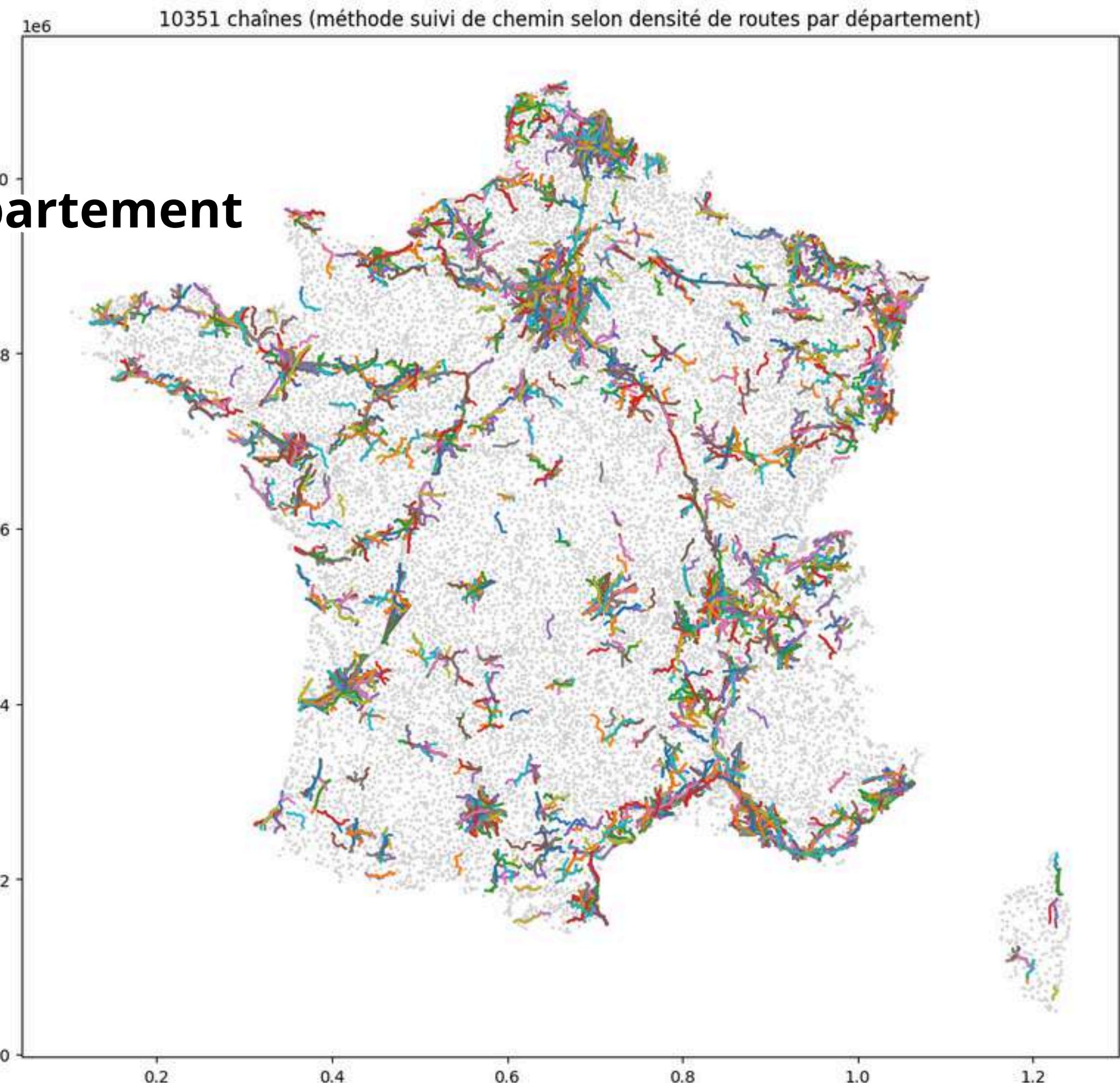
# Suivi de chemin

- Pour chaque station, on suit les directions des  $n$  voisins les plus proches
- A chaque étape, on considère la station la plus proche dans la même direction
- On affiche les chaînes de plus de  $min\_stations$



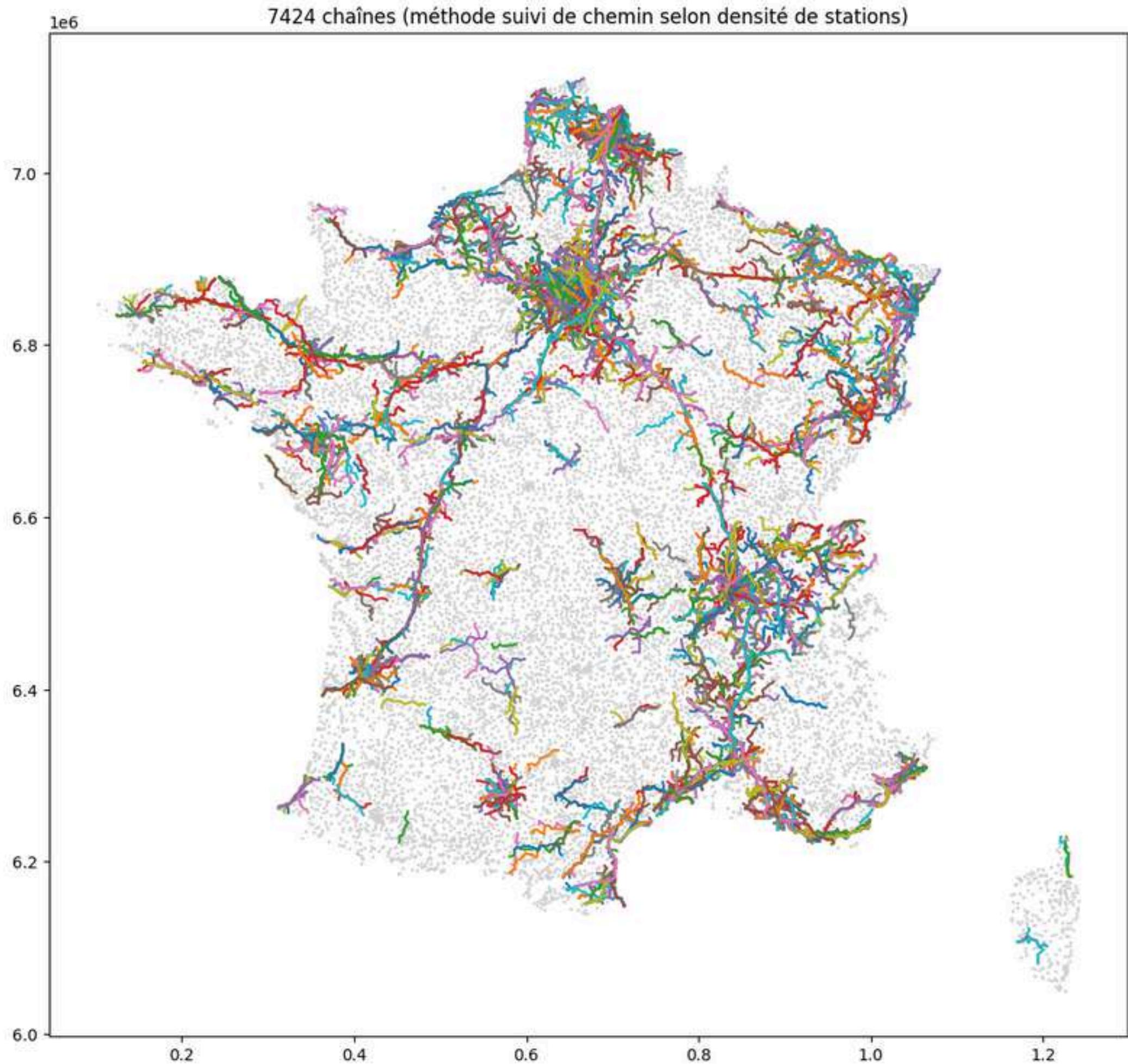
# Suivi de chemin Selon densité de routes par département

- Pour chaque département, on calcule la longueur de route+chemin de fer proportionnellement à sa superficie
- On normalise cette valeur
- On calcule en fonction le paramètre de distance à utiliser, qui sera dans un certain intervalle choisi.
- $min\_len = 9$



# Suivi de chemin Selon densité de stations

- Pour chaque station, on calcule la densité de stations dans un rayon de  $r_{density} = 20\ 000$  mètres
- On normalise cette densité
- On calcule en fonction les paramètres de distance et d'angle à utiliser, qui seront dans un certain intervalle choisi.
- $min\_len = 15$



# Conclusion

- Selon densité de routes par département : mieux que méthode de base, mais on a toujours des discontinuités sur les routes, des petits segments.
- Selon densité de stations : peu de discontinuités, chaînes plus longues, et certaines nouvelles routes détectées.

# Idée

- Comparer la longueur de route+chemin de fer réelle avec celle trouvée, pour avoir une mesure d'erreur de mes méthodes.

**19 - 26 Juin**

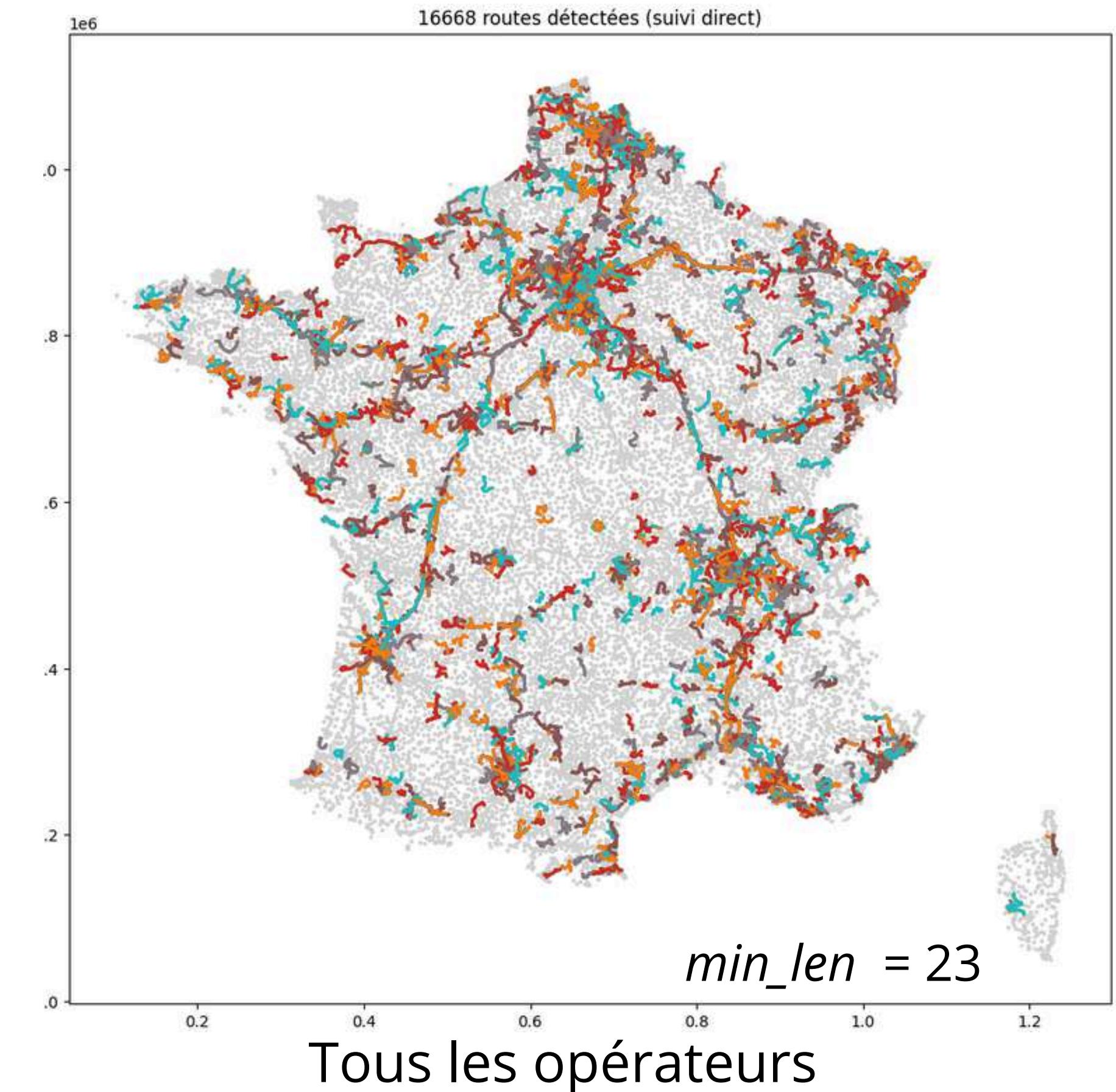
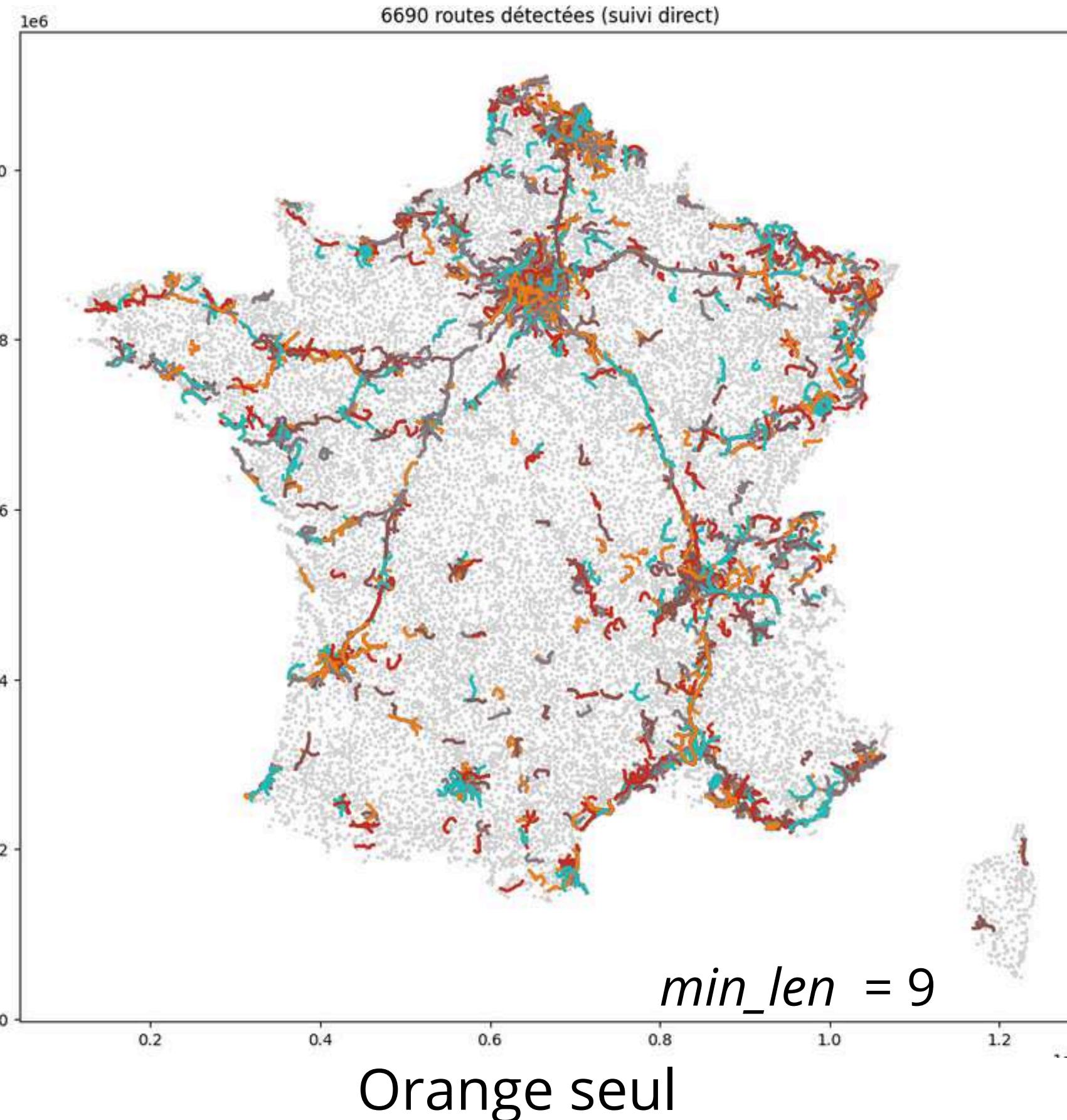
# **Fonction d'erreur**

# Fonction d'erreur

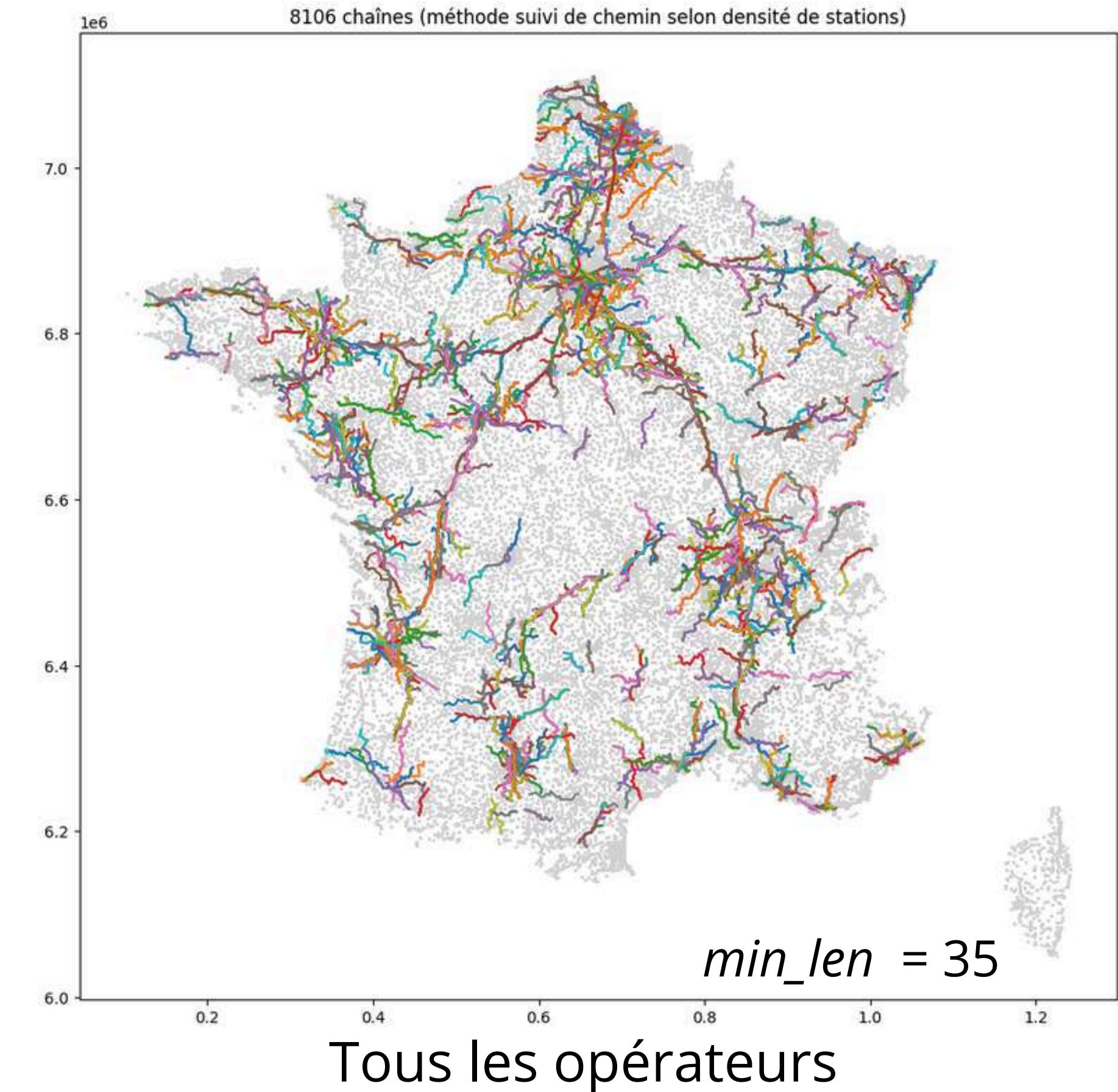
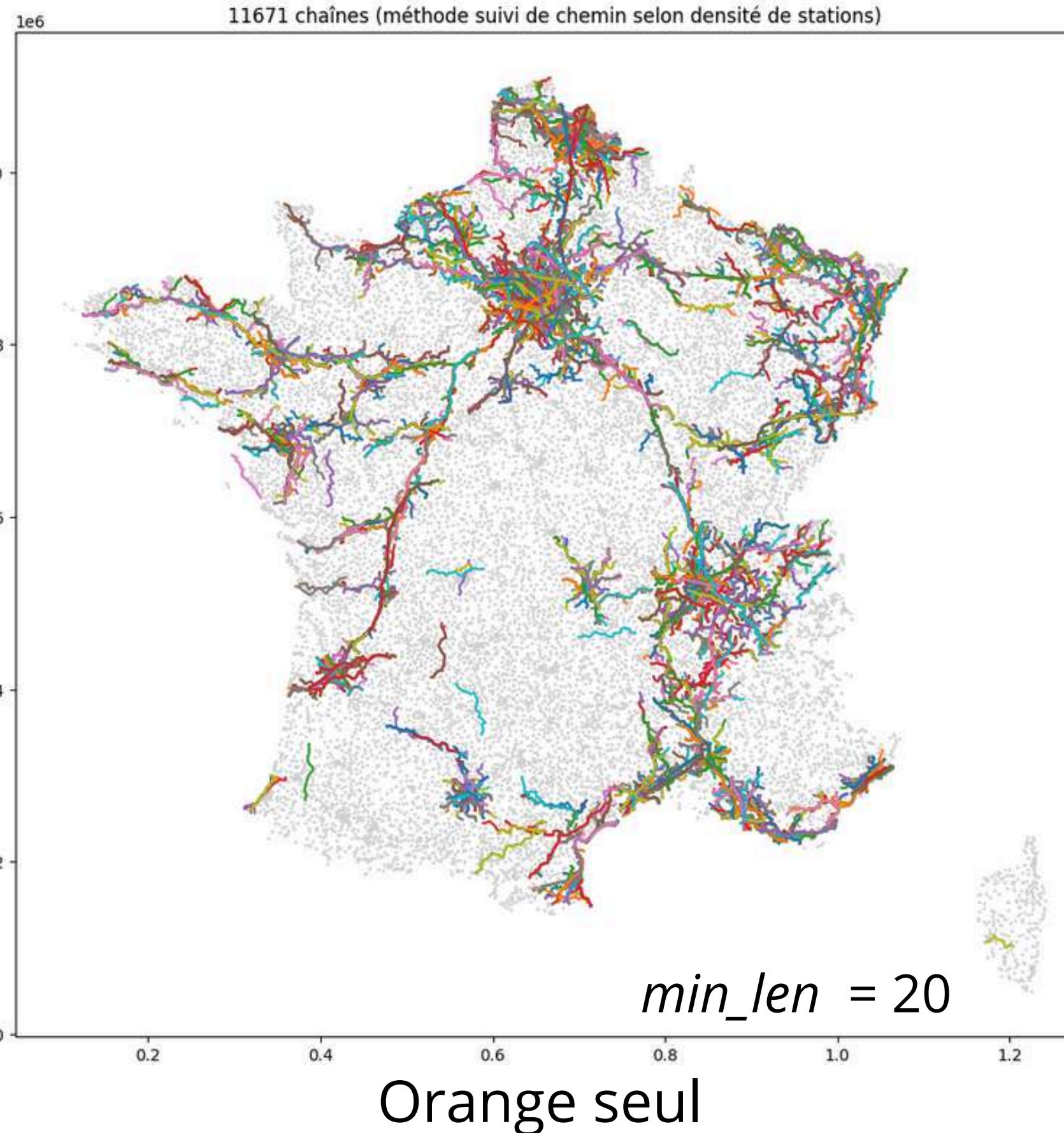
- Compare la longueur de route trouvée à la longueur réelle
- Une fonction qui compare globalement, une autre par département
- On ne compare que la longueur des routes, pas leur position -> pas idéal.
- Données : (1) chemins fer, (2) Autoroutes, (3) routes nationales, (4) routes départementales et communales
- **Problème** pour la longueur réelle : On détecte plus que (1)+(2)+(3), mais moins que (1)+(2)+(3)+(4)

**Tous les opérateurs / seulement Orange**

# Suivi de chemin (sans améliorations)



# Suivi de chemin Par densité de stations



**27 Juin - 2 Juillet**

# **Code propre**

# Données

## Problème :

- Longueur réelle de route qui ne sépare pas routes départementales et communales

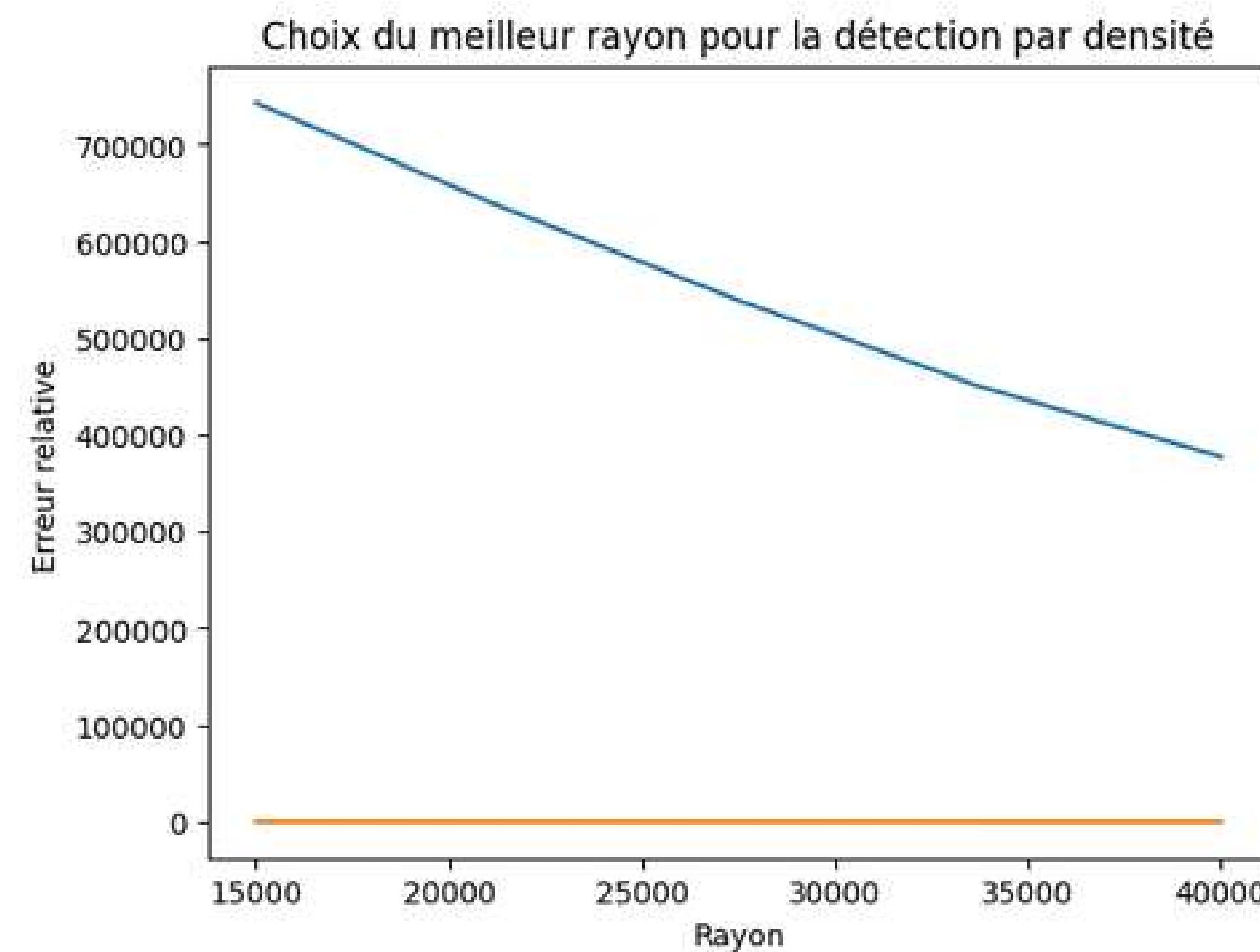
## Solution :

- D'après *statista* et autres sources, longueur en france :
    - Routes départementales : 380 000 km
    - Routes communales : 710 000km
- Utilisation de ces chiffres pour considérer les routes départementales mais pas celles communales

# **Rayon de densité idéal**

# Suivi de chemin par densité de stations

- Auparavant : 20 km, environ rayon de l'Ile de France
- Objectif : Trouver le rayon qui minimise l'erreur



→ A continuer pour avoir un résultat

**4 - 14 juillet**

# **Rayon de densité optimal**

Suivi de chemin par densité de stations

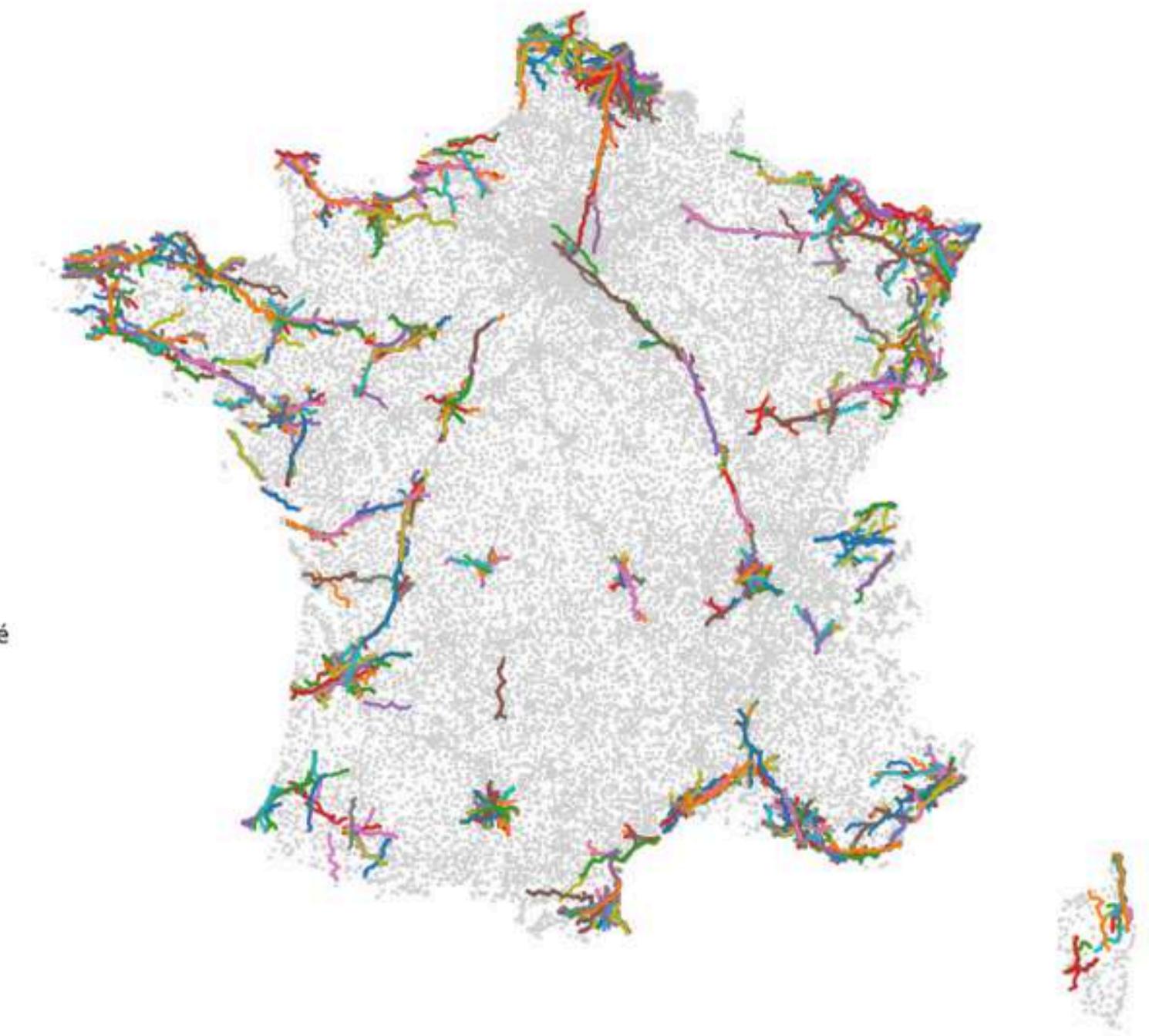
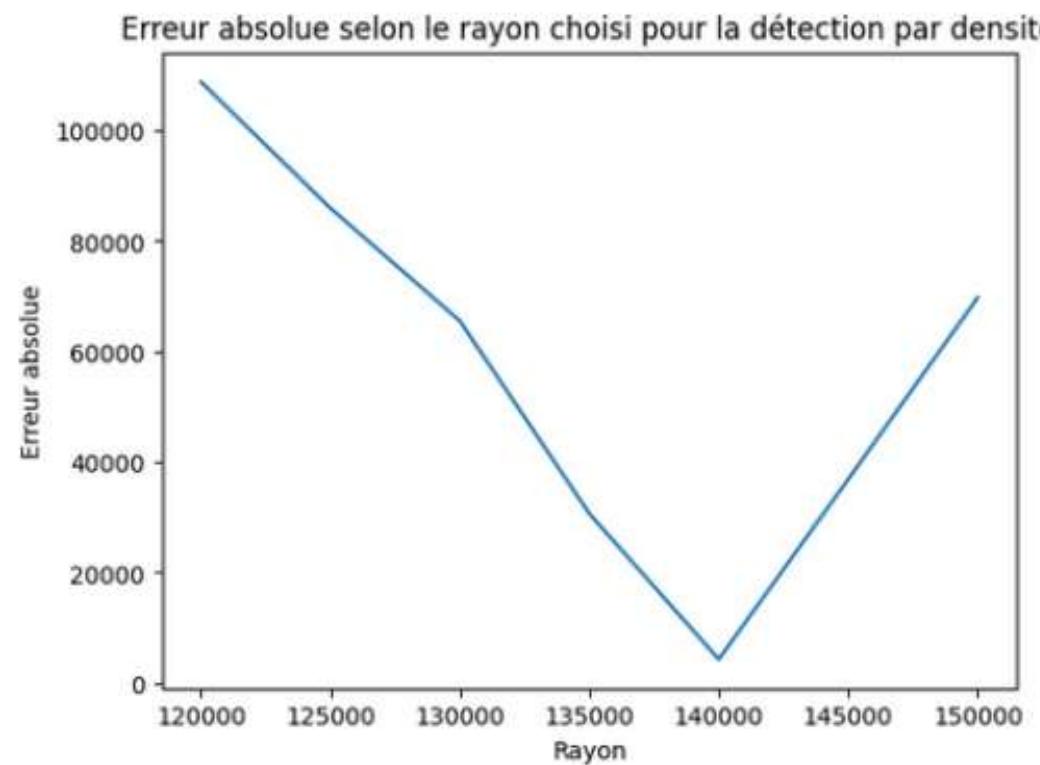
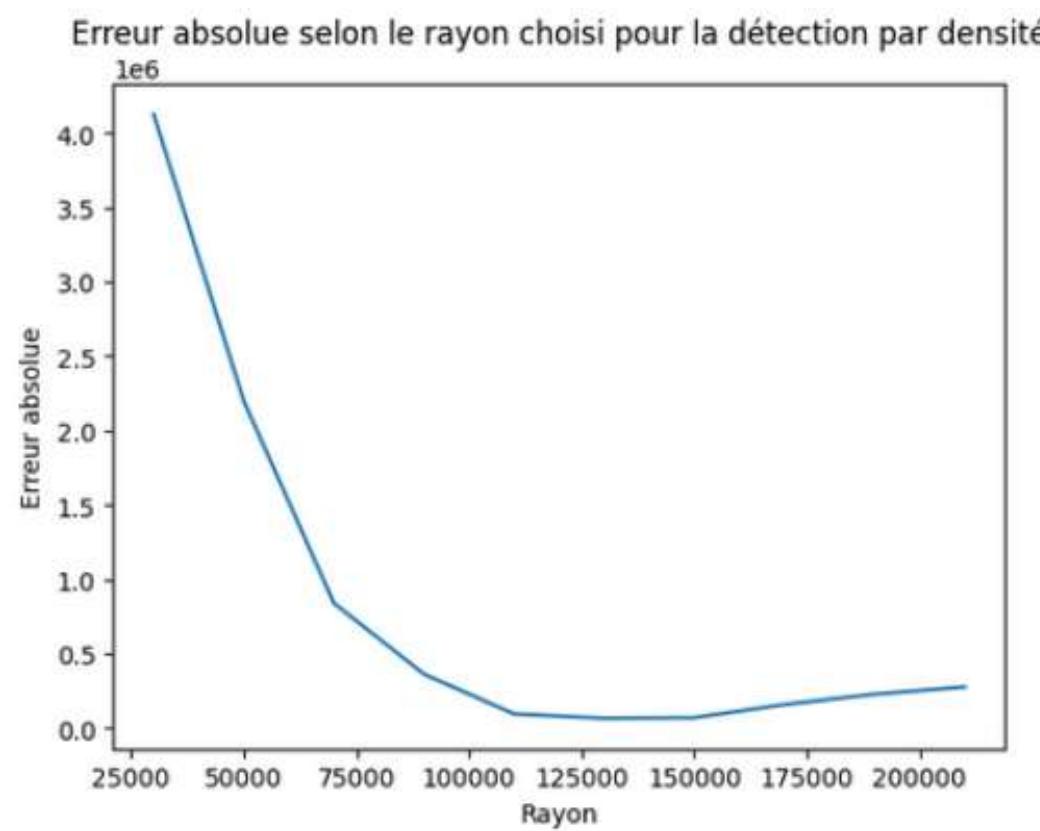
# Suivi de chemin par densité de stations

Objectif : Trouver le rayon qui minimise l'erreur

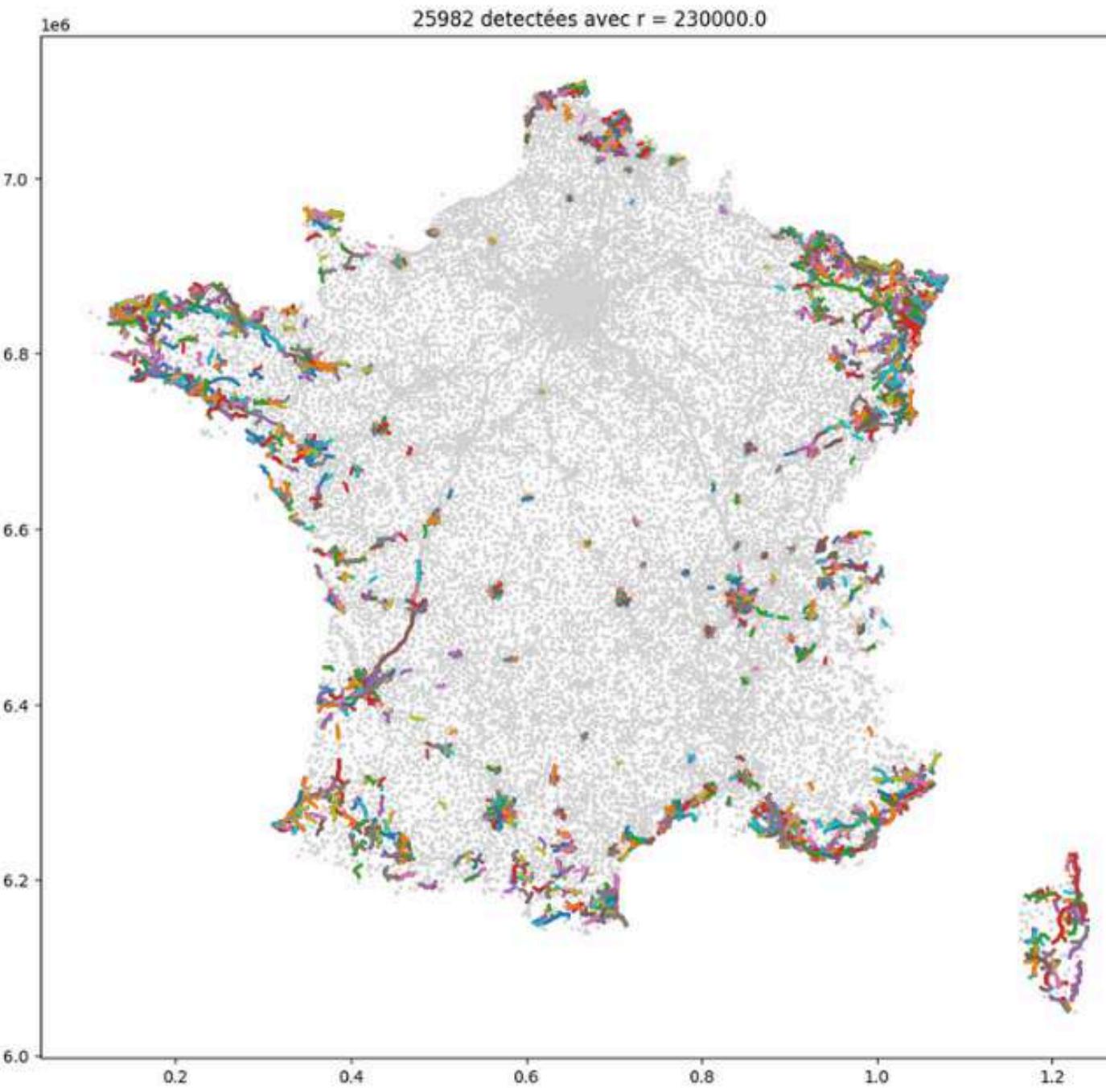
- Rappel : ajustement des paramètres selon la densité des stations dans un certain rayon autour de la station considérée.
- Par hypothèse, un rayon de 20 à 40 km, qui correspond environ à la superficie de l'Île de France donc zone la plus dense.

# Rayon optimal

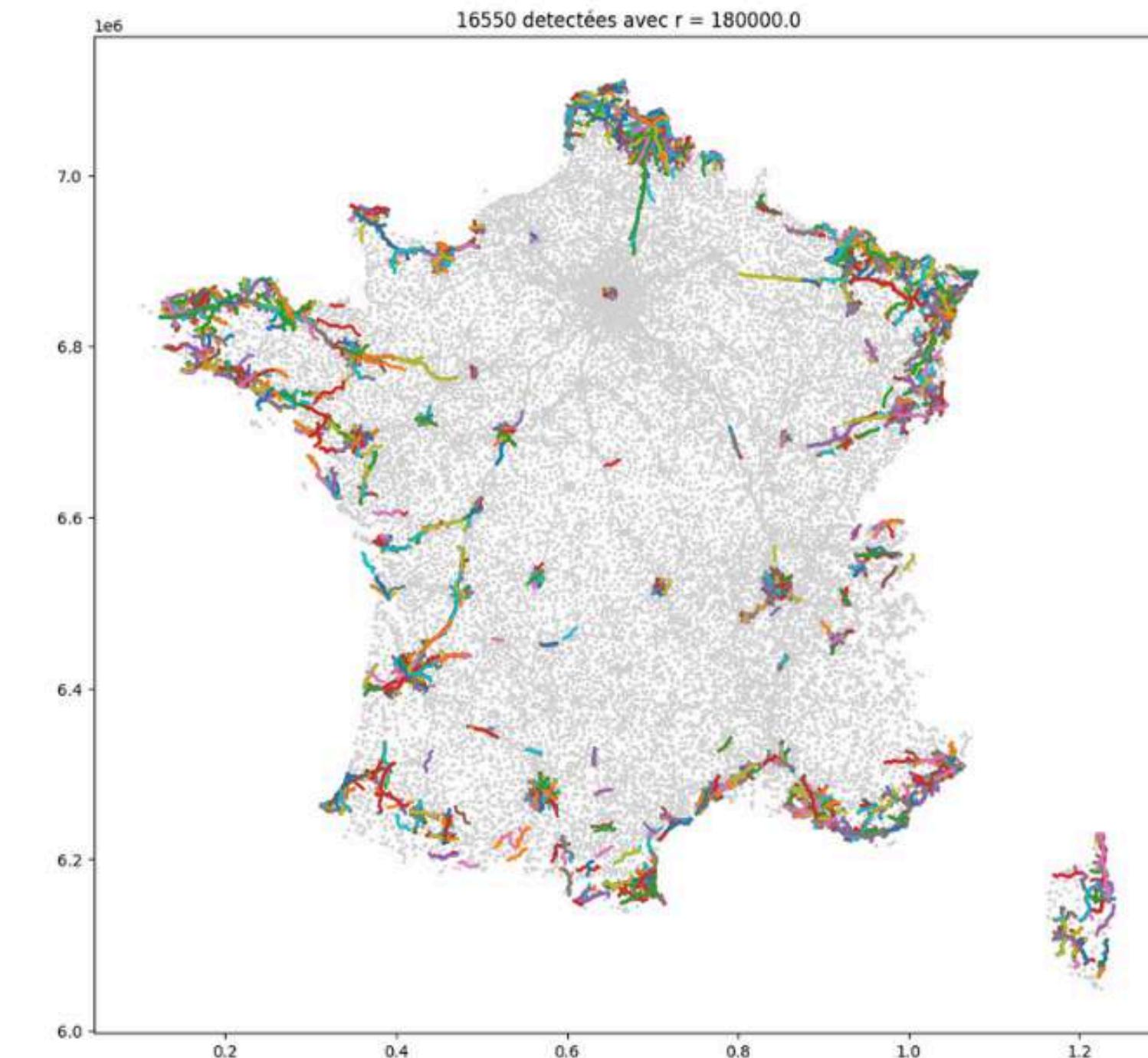
- Stations d'Orange + 4G
- Autres paramètres :
  - $n\_neighbors = 10$
  - $dist\_min = 5$
  - $dist\_max = 7000$
  - $angle\_min = 20$
  - $angle\_max = 90$
  - $n\_directions = 5$
  - $angle\_tol = 5$
  - $min\_len = 15$



# Rayon optimal

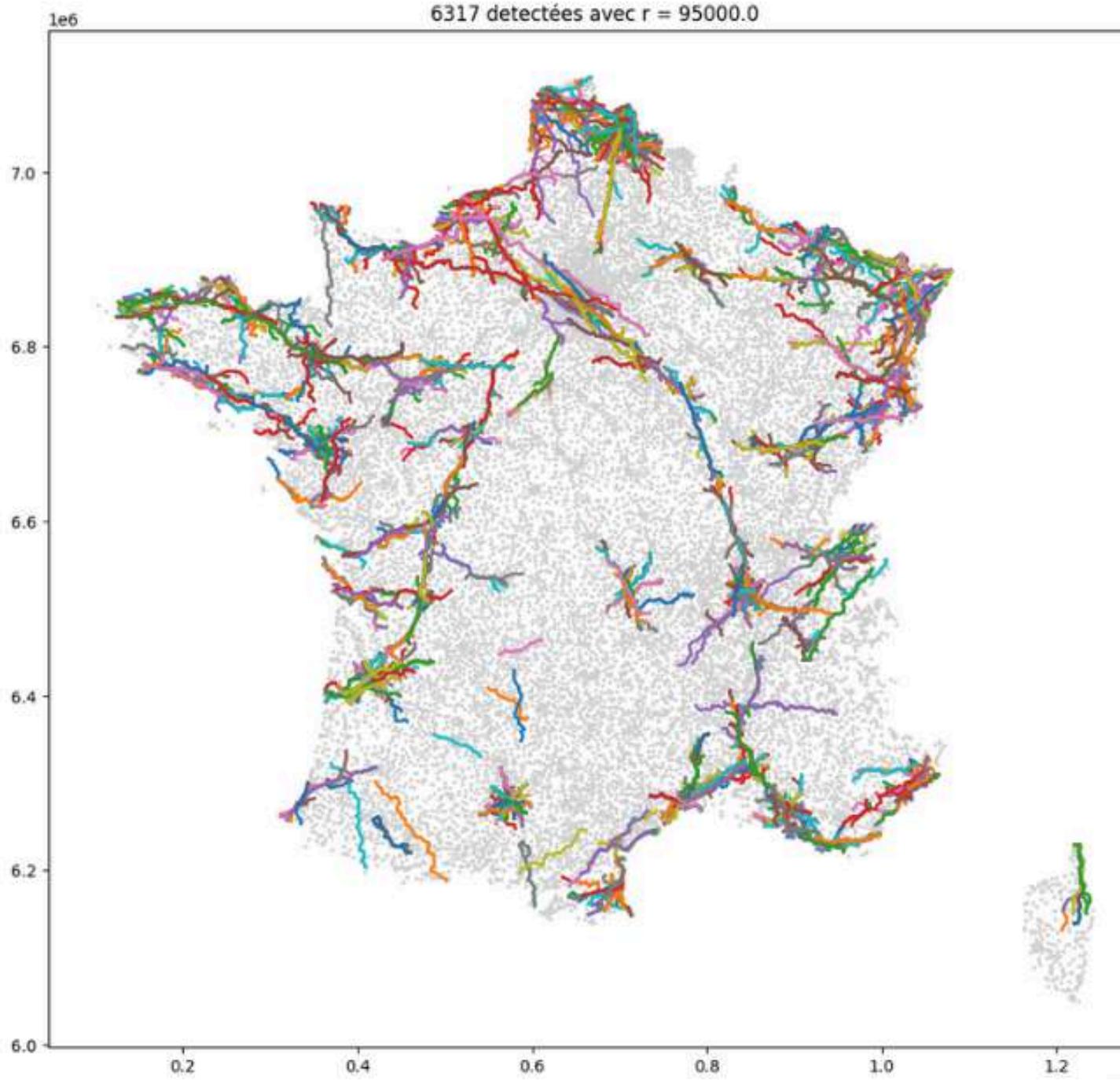


$min\_len = 5$

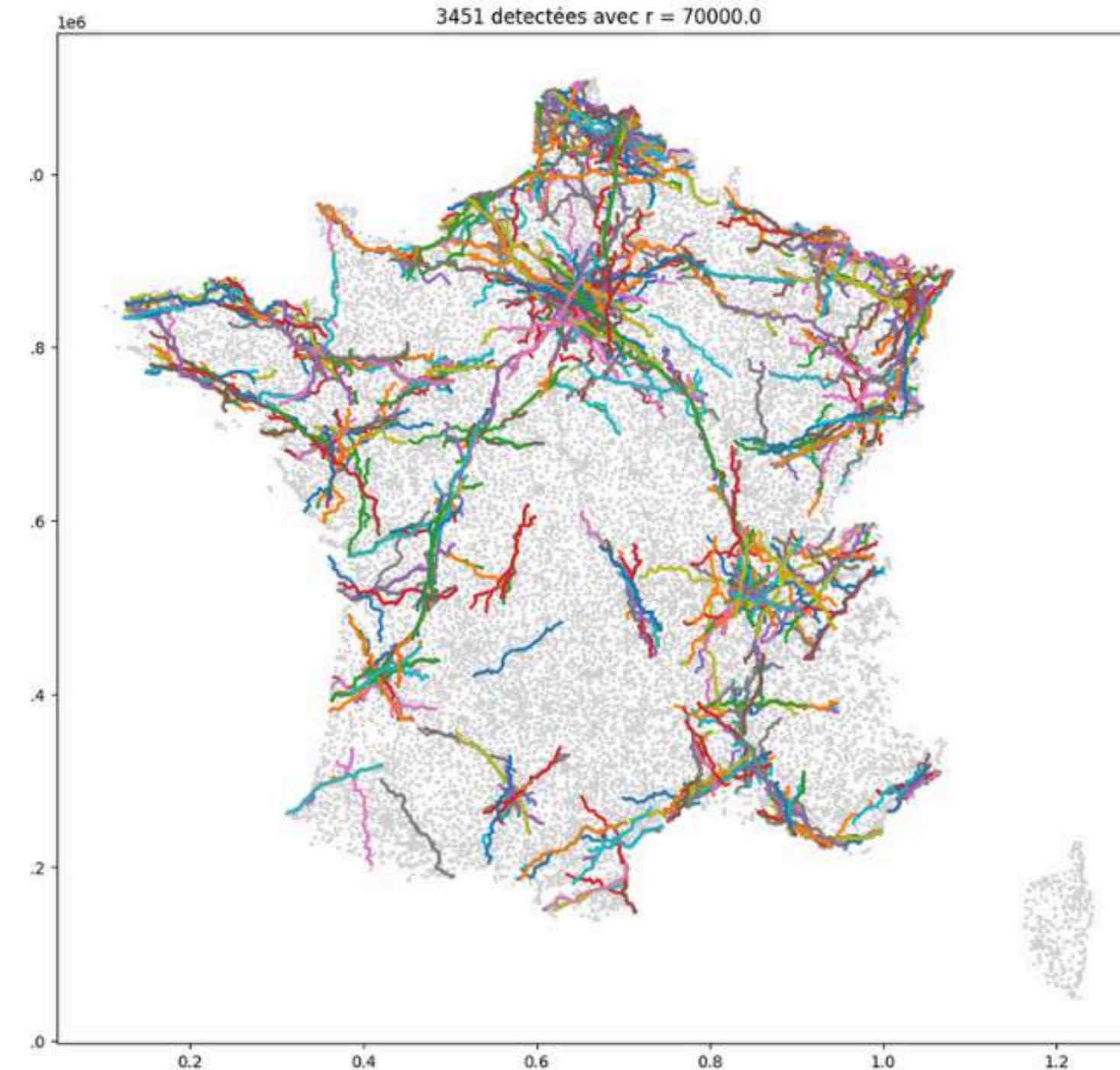


$min\_len = 10$

# Rayon optimal



*min\_len* = 20



*min\_len* = 30

# Rayon optimal

**Problème :** Tous les paramètres influencent le résultat

**Solution :** Librairie python *skopt* : Optimiser tous les paramètres à la fois.

Fonctionnement : Intervalle pour chaque paramètre.

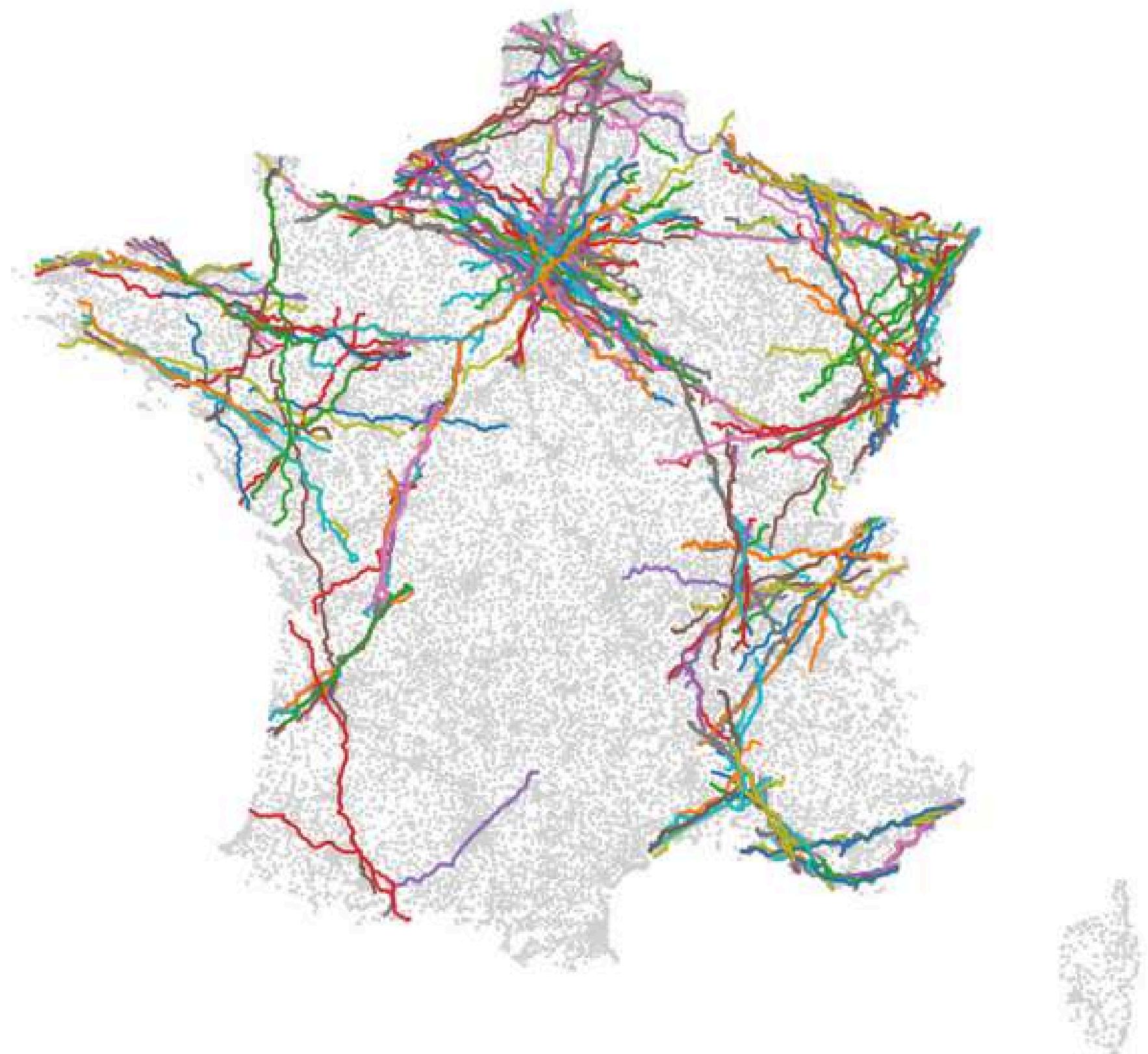
Réalise  $n\_calls$  tests intelligents de combinaisons de ces paramètres.

# **Paramètres optimaux**

Suivi de chemin par densité de stations

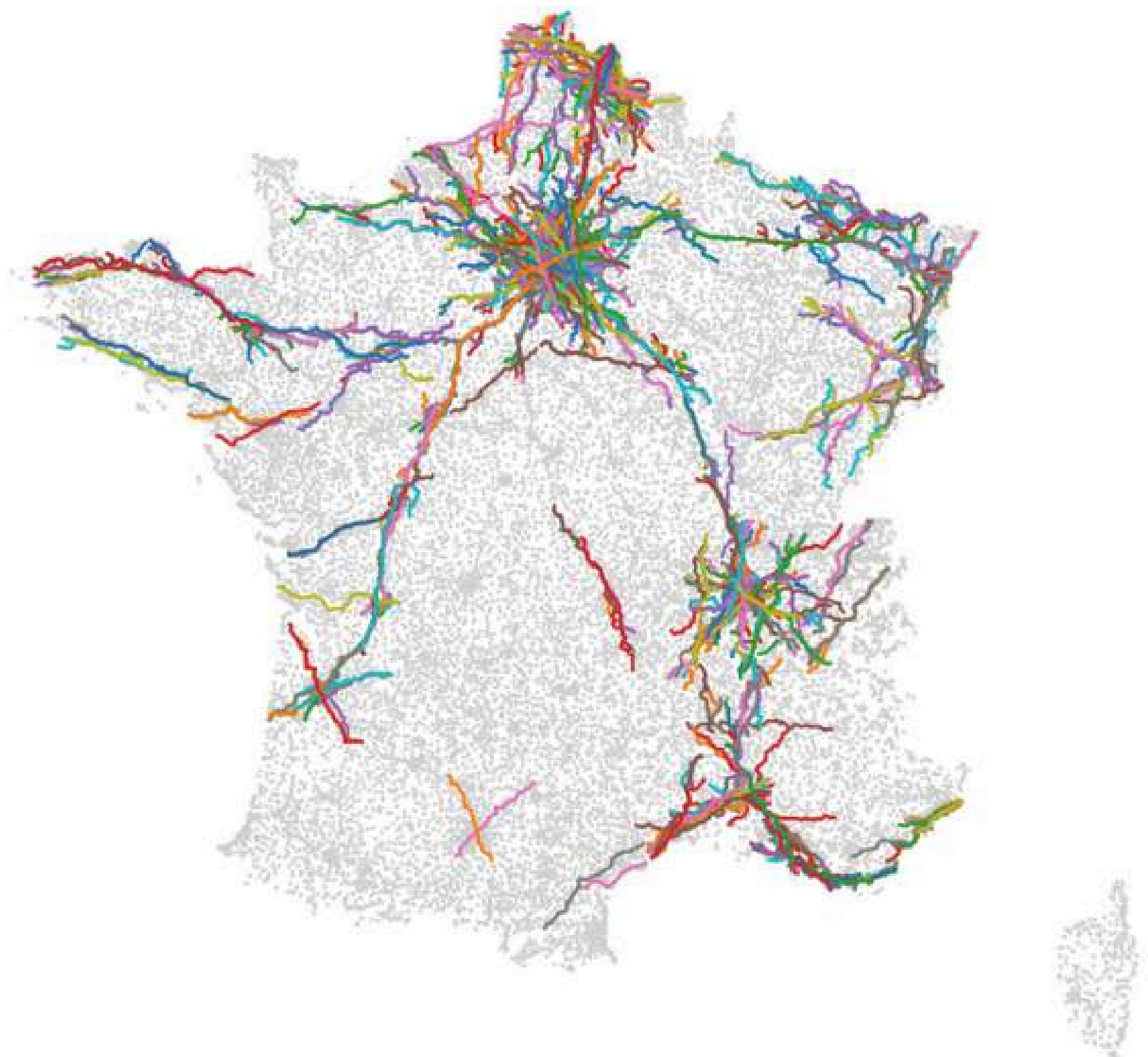
# Paramètres optimaux

- Stations Orange + 4G
- Intervalles des paramètres :
  - $rayon = 20\ 000 - 120\ 000$
  - $n\_neighbors = 5 - 30$
  - $dist\_min = 1 - 200$
  - $dist\_max = 2\ 000 - 10\ 000$
  - $angle\_min = 0 - 80$
  - $angle\_max = 40 - 90$
  - $n\_directions = 1 - 5$
  - $angle\_tol = 1 - 20$
  - $min\_len = 5 - 60$
- Paramètres trouvés :
  - $rayon = 100\ 039$
  - $n\_neighbors = 13$
  - $dist\_min = 140$
  - $dist\_max = 8032$
  - $angle\_min = 77$
  - $angle\_max = 82$
  - $n\_directions = 5$
  - $angle\_tol = 15$
  - $min\_len = 50$
- Erreur : 204 771 km



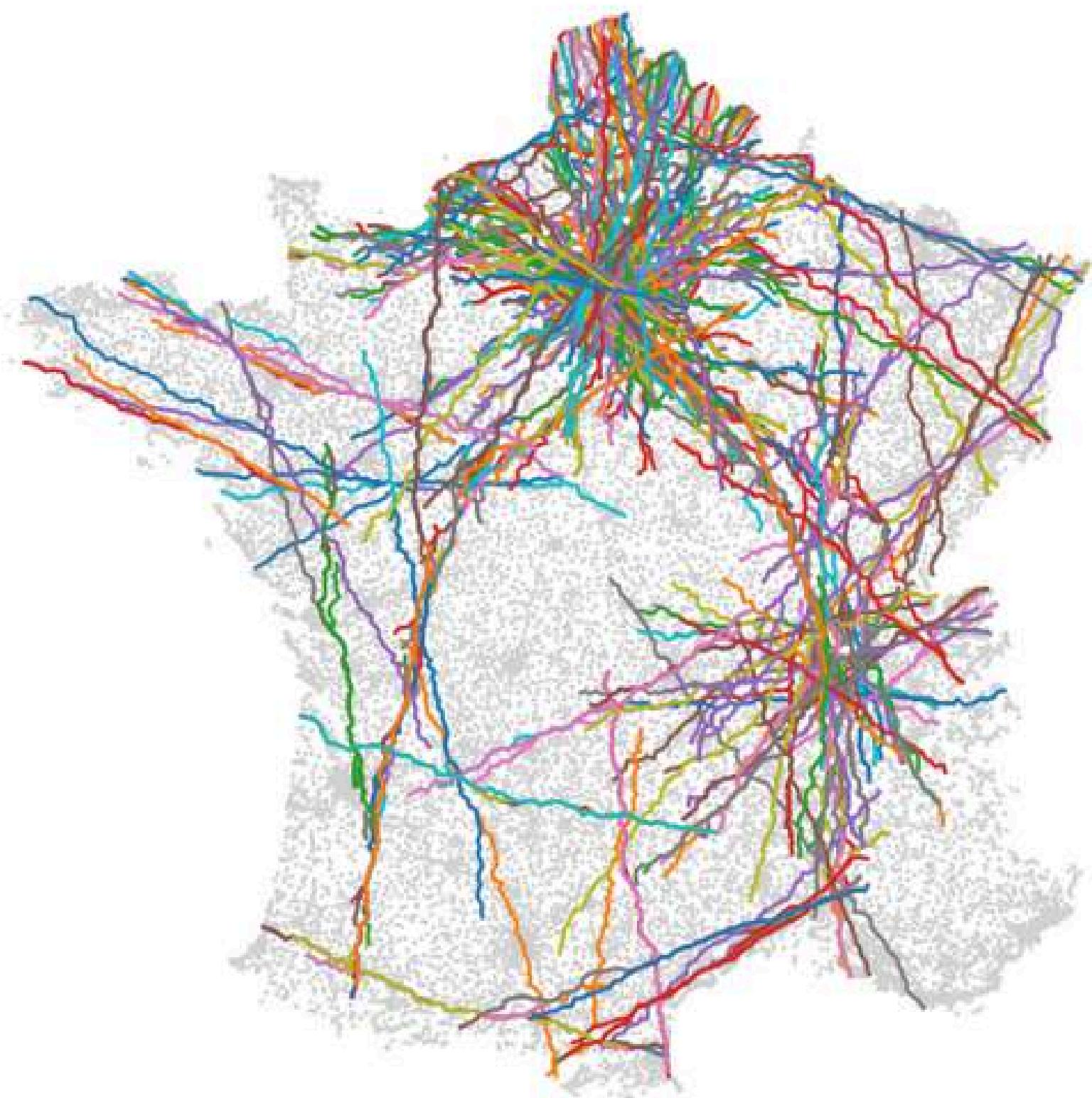
# Paramètres optimaux

- Stations Orange + 4G
- Intervalles des paramètres :
  - $rayon = 15\ 000 - 60\ 000$
  - $n\_neighbors = 5 - 30$
  - $dist\_min = 1 - 200$
  - $dist\_max = 2\ 000 - 10\ 000$
  - $angle\_min = 0 - 80$
  - $angle\_max = 40 - 90$
  - $n\_directions = 1 - 5$
  - $angle\_tol = 1 - 20$
  - $min\_len = 5 - 60$
- Paramètres trouvés :
  - $rayon = 47\ 727$
  - $n\_neighbors = 13$
  - $dist\_min = 115$
  - $dist\_max = 6\ 167$
  - $angle\_min = 77$
  - $angle\_max = 82$
  - $n\_directions = 4$
  - $angle\_tol = 11$
  - $min\_len = 37$
- Erreur : 114 278 km



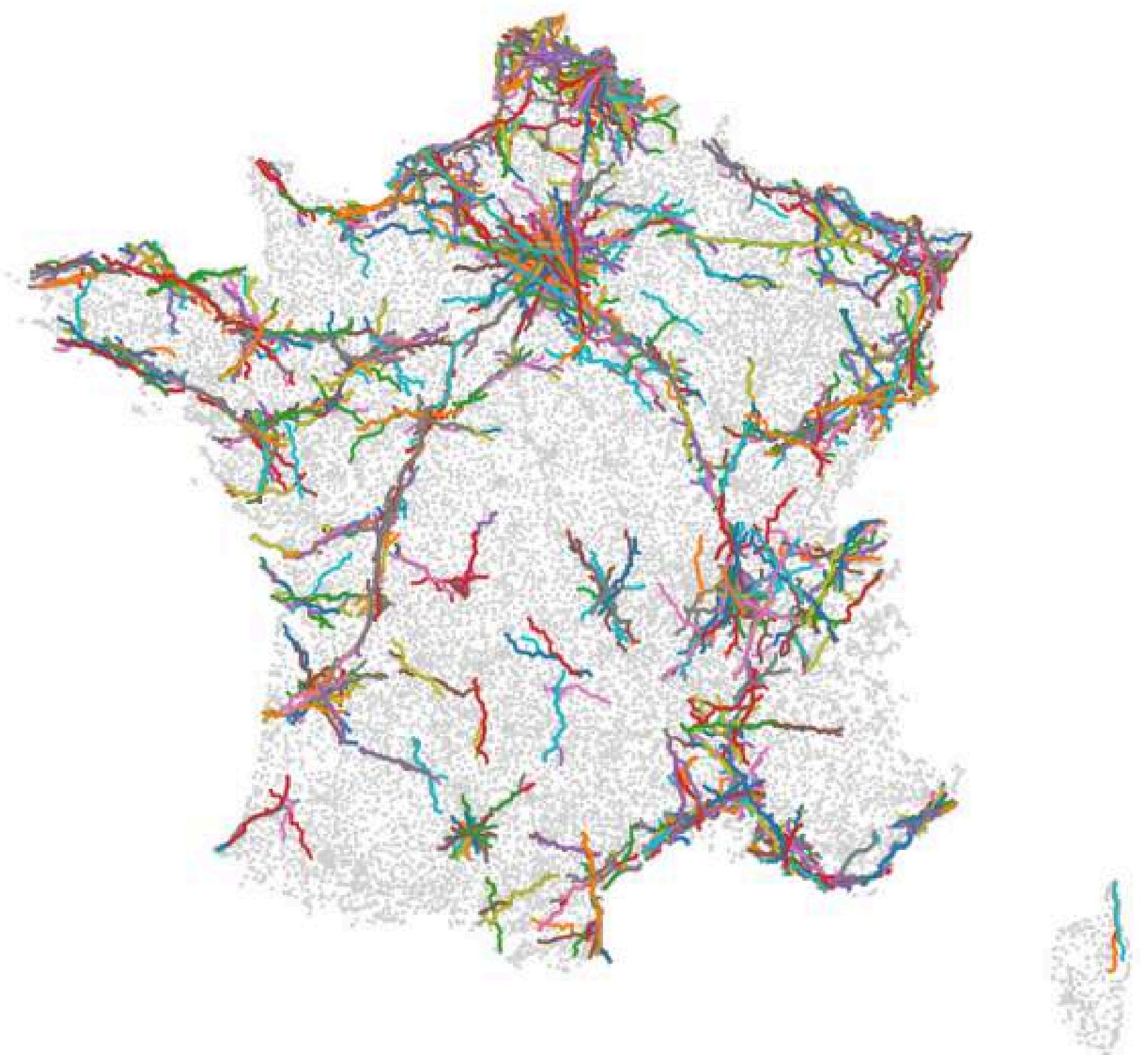
# Paramètres optimaux

- Stations Orange
- Intervalles des paramètres :
  - $rayon = 15\ 000 - 60\ 000$
  - $n\_neighbors = 5 - 30$
  - $dist\_min = 1 - 200$
  - $dist\_max = 2\ 000 - 10\ 000$
  - $angle\_min = 0 - 80$
  - $angle\_max = 40 - 90$
  - $n\_directions = 1 - 5$
  - $angle\_tol = 1 - 20$
  - $min\_len = 5 - 65$
- Paramètres trouvés :
  - $rayon = 48\ 000$
  - $n\_neighbors = 30$
  - $dist\_min = 1$
  - $dist\_max = 8500$
  - $angle\_min = 35$
  - $angle\_max = 55$
  - $n\_directions = 5$
  - $angle\_tol = 10$
  - $min\_len = 60$
- Erreur : 136 207 km



# Paramètres optimaux

- Stations Orange
- Intervalles des paramètres :
  - $rayon = 20\ 000 - 60\ 000$
  - $n\_neighbors = 10 - 30$
  - $dist\_min = 1 - 100$
  - $dist\_max = 4\ 000 - 10\ 000$
  - $angle\_min = 0 - 80$
  - $angle\_max = 60 - 90$
  - $n\_directions = 2 - 5$
  - $angle\_tol = 1 - 20$
  - $min\_len = 10 - 55$
- Paramètres trouvés :
  - $rayon = 58\ 000$
  - $n\_neighbors = 20$
  - $dist\_min = 5$
  - $dist\_max = 6500$
  - $angle\_min = 30$
  - $angle\_max = 70$
  - $n\_directions = 4$
  - $angle\_tol = 5$
  - $min\_len = 20$
- Erreur : 33 536 km



# Paramètres optimaux

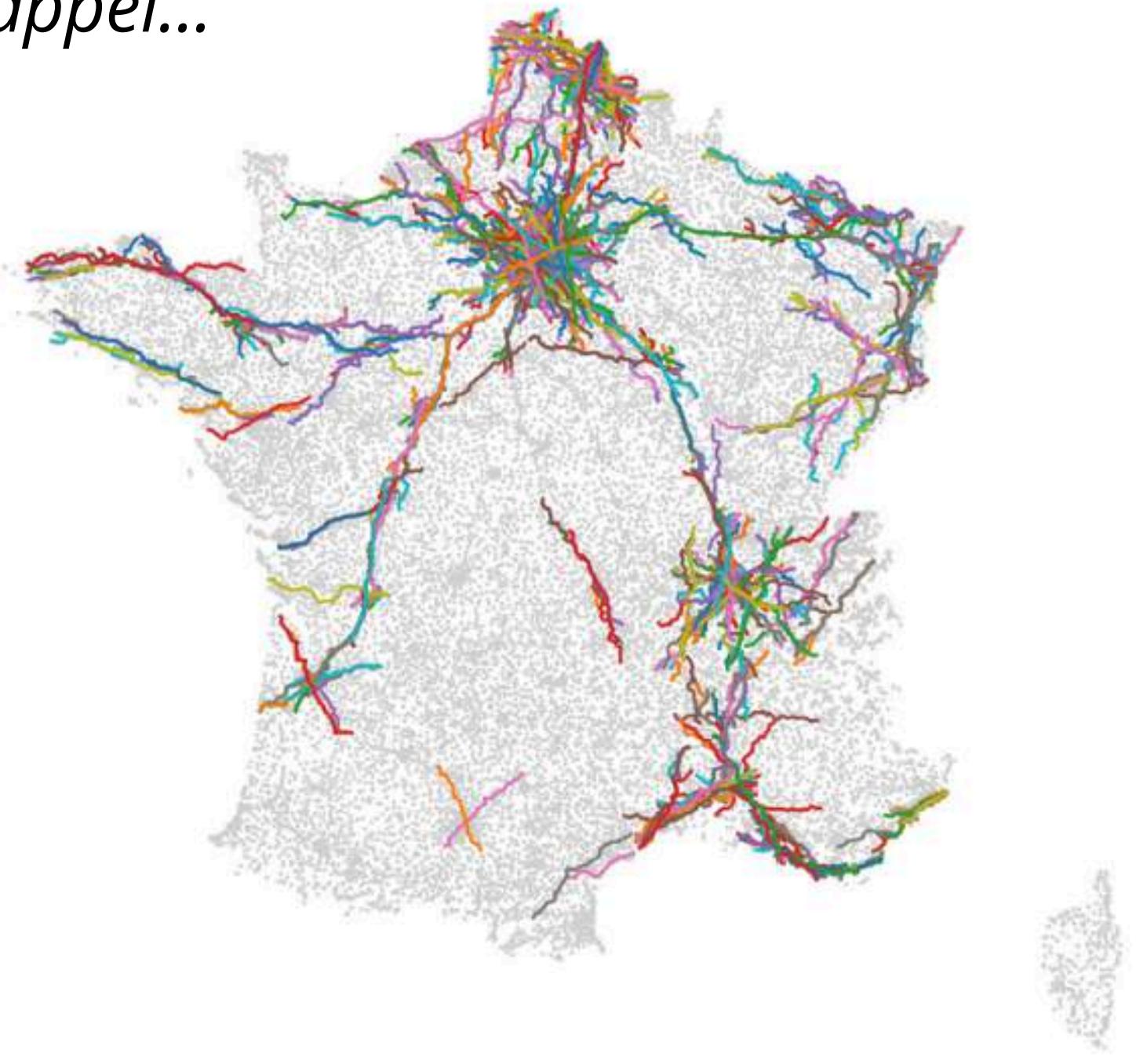
- Ajuster les intervalles des paramètres selon les résultats trouvés aide à réduire les possibles et donc donne de meilleurs résultats
- Fonction d'erreur globale → Résultats avec faible erreur mais incorrects géographiquement. Modification possible : minimiser la différence d'erreur entre les départements, pour s'assurer de détecter aussi bien (ou mal) tous les départements.

**15 - 21 Juillet**

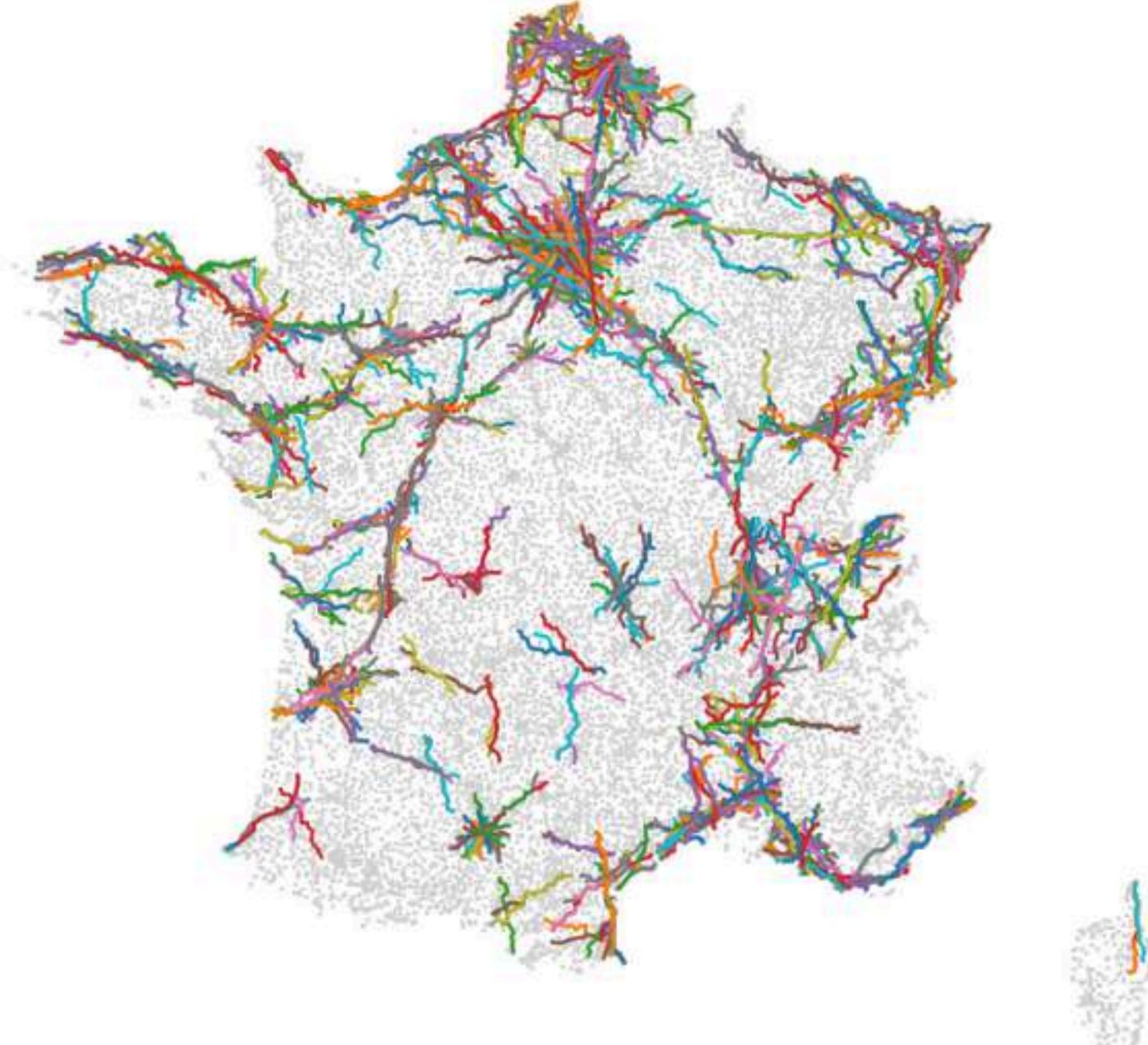
# **Paramètres optimaux**

Suivi de chemin par densité de stations

# Rappel...



- Paramètres trouvés :
  - $angle\_min = 77$
  - $angle\_max = 82$
  - $rayon = 47\ 727$
  - $n\_neighbors = 13$
  - $dist\_min = 115$
  - $dist\_max = 6\ 167$
  - $n\_directions = 4$
  - $angle\_tol = 11$
  - $min\_len = 37$
- Erreur : 114 278 km (26%)



- Paramètres trouvés :
  - $angle\_min = 30$
  - $angle\_max = 70$
  - $rayon = 58\ 000$
  - $n\_neighbors = 20$
  - $dist\_min = 5$
  - $dist\_max = 6500$
  - $n\_directions = 4$
  - $angle\_tol = 5$
  - $min\_len = 20$
- Erreur : 33 536 km (7%)

# **Longueur de référence**

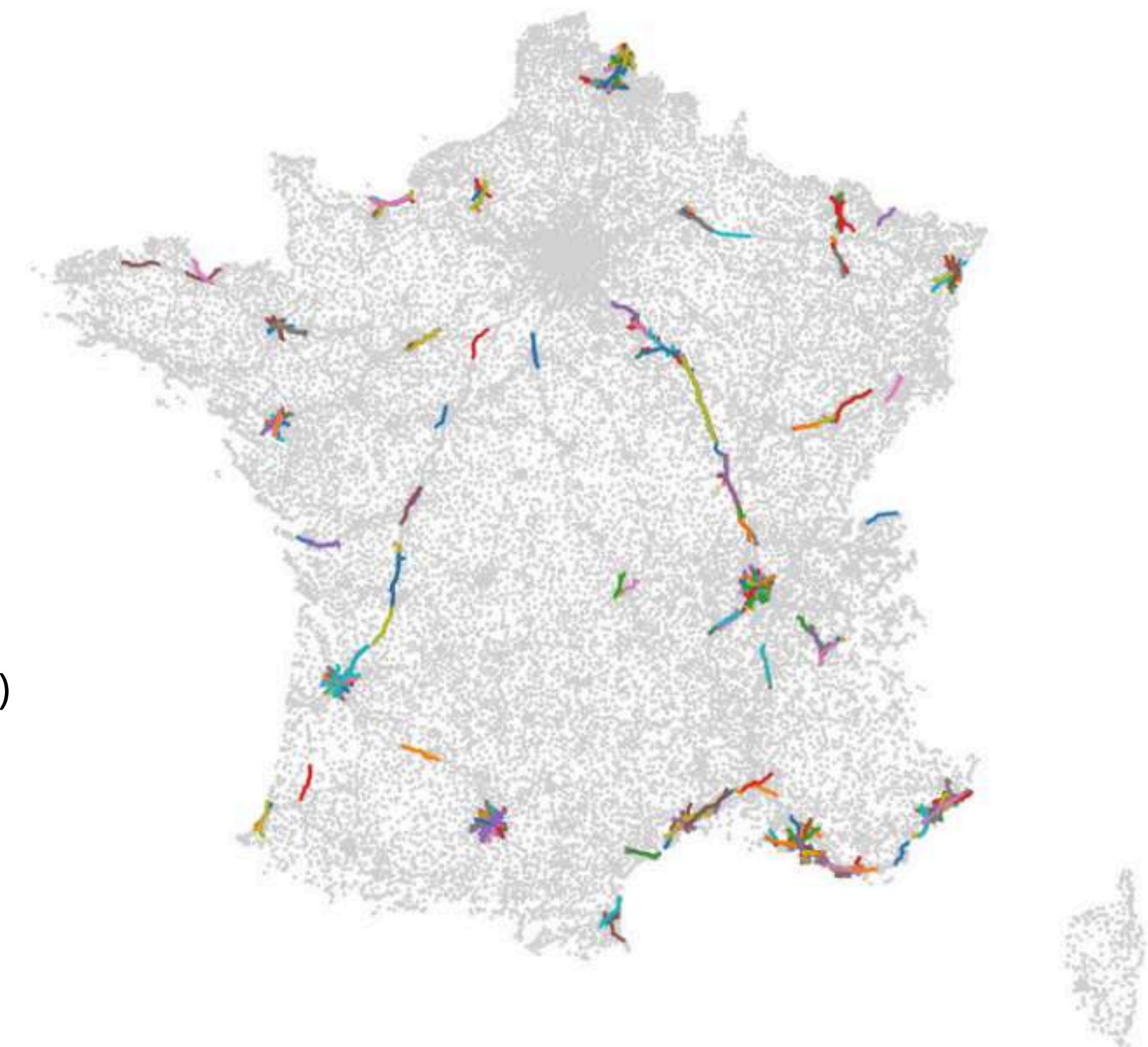
Fonction d'erreur

# Longueur de référence

- Longueur de référence utilisée : chemins de fer + Autoroutes + nationales + **départementales**
- On veut détecter les plus grands axes routiers/ferroviaires seulement → pas forcément les départementales.
- Suppression puis calcul des paramètres optimaux de nouveau.

# Paramètres optimaux Sans départementales

- Stations Orange
- Intervalles des paramètres :
  - $rayon = 20\ 000 - 60\ 000$
  - $n\_neighbors = 10 - 30$
  - $dist\_min = 1 - 70$
  - $dist\_max = 4\ 000 - 10\ 000$
  - $angle\_min = 0 - 80$
  - $angle\_max = 60 - 90$
  - $n\_directions = 2 - 5$
  - $angle\_tol = 1 - 20$
  - $min\_len = 10 - 55$
- Paramètres trouvés :
  - $rayon = 56\ 000$
  - $n\_neighbors = 15$
  - $dist\_min = 1$
  - $dist\_max = 4000$
  - $angle\_min = 5$
  - $angle\_max = 75$
  - $n\_directions = 2$
  - $angle\_tol = 10$
  - $min\_len = 15$
- Erreur : 6 131 km (12%)



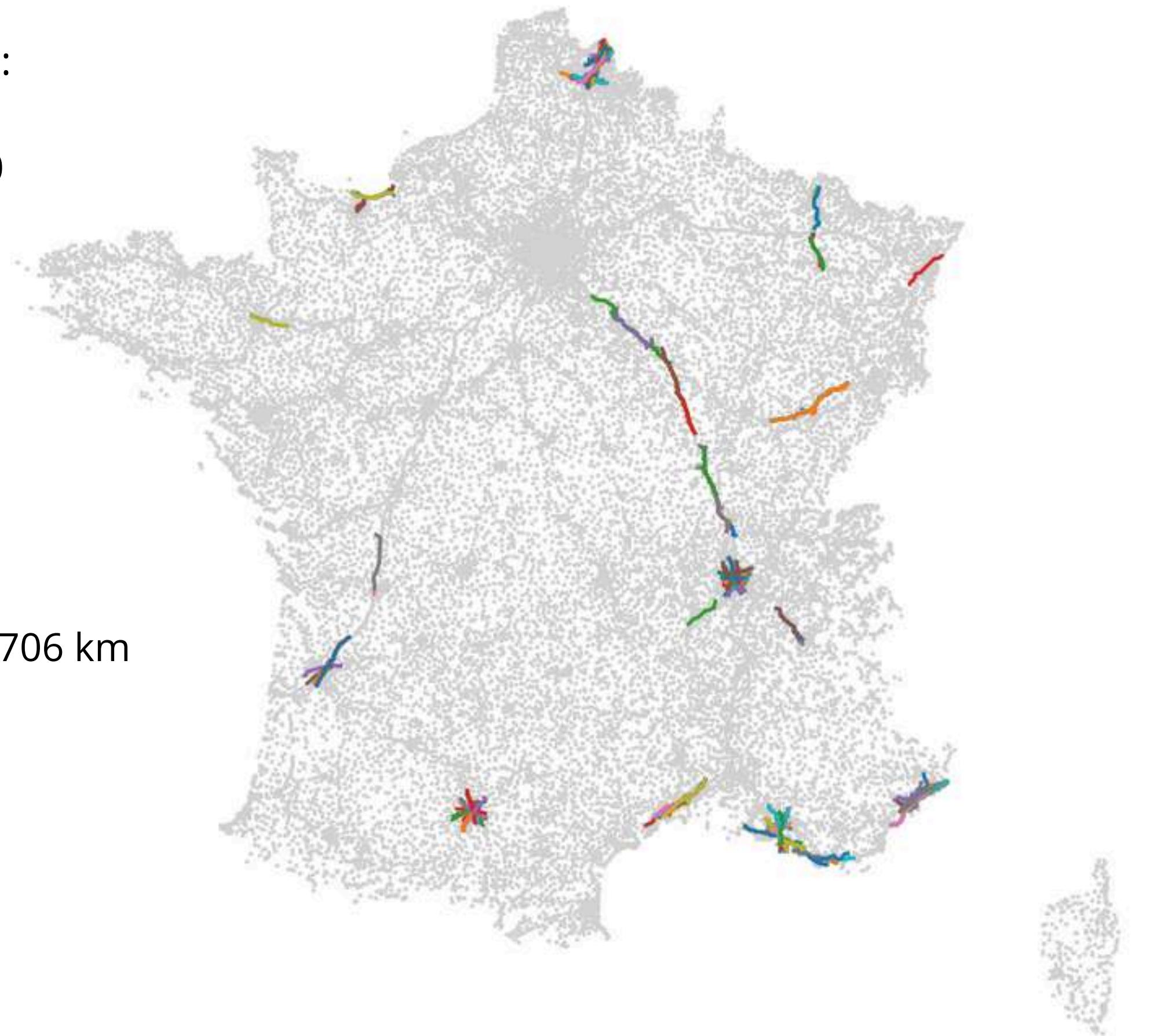
# **Fonction d'erreur**

# Fonction d'erreur

- Avant :
  - Compare la longueur totale trouvée à la longueur totale de référence (erreur relative)
  - Problème : pas de considération géographique, le bruit compense le manque d'information ailleurs.
- Après :
  - Considère l'erreur relative **et** l'écart type des erreurs relatives par département
  - La minimiser = forcer à détecter la bonne longueur globale **et** à ce que tous les départements soient autant bien détectés

# Paramètres optimaux

- Stations Orange
- Intervalles des paramètres :
  - $rayon = 20\ 000 - 60\ 000$
  - $n\_neighbors = 10 - 30$
  - $dist\_min = 1 - 70$
  - $dist\_max = 4\ 000 - 10\ 000$
  - $angle\_min = 0 - 80$
  - $angle\_max = 60 - 90$
  - $n\_directions = 2 - 5$
  - $angle\_tol = 1 - 20$
  - $min\_len = 10 - 55$
- Paramètres trouvés :
  - $rayon = 58\ 000$
  - $n\_neighbors = 20$
  - $dist\_min = 12$
  - $dist\_max = 4000$
  - $angle\_min = 35$
  - $angle\_max = 70$
  - $n\_directions = 3$
  - $angle\_tol = 5$
  - $min\_len = 20$
- Erreur globale : 408 706 km  
(94%)
- Score : 1,26



# Fonction d'erreur

- Score =  $0.5 \times \text{erreur globale} + 0.5 \times \text{écart type}$   
→ Modifier les poids pour donner plus d'importance à l'erreur globale.

## Remarque :

- Résultats toujours soit bruités mais avec toutes les routes, soit très lisses mais manque des routes/tronçons de routes  
→ Sans doute impossible de faire mieux, compromis entre les deux.

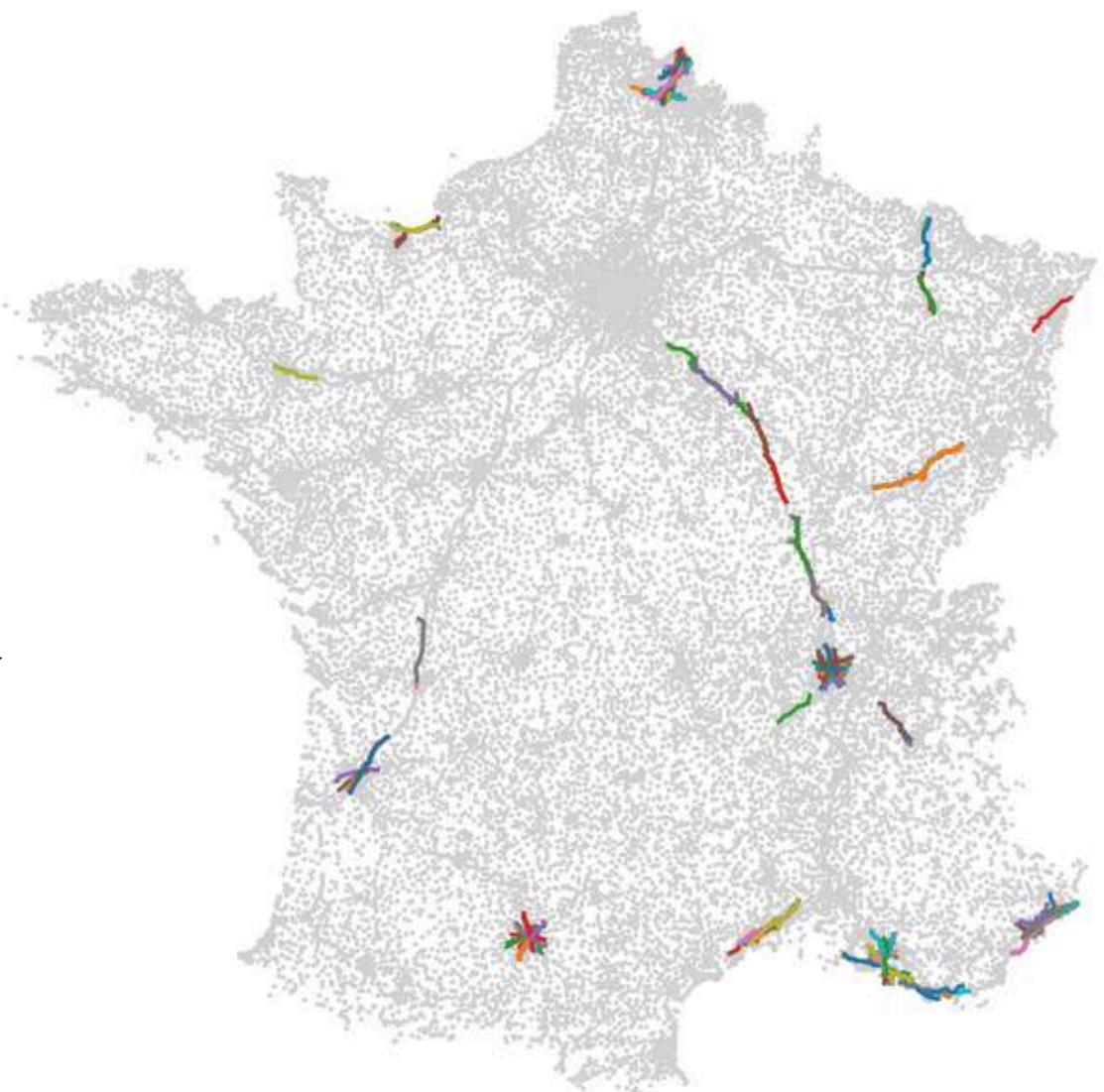
**22 - 28 Juillet**

# **Score**

Rappel...

# Score

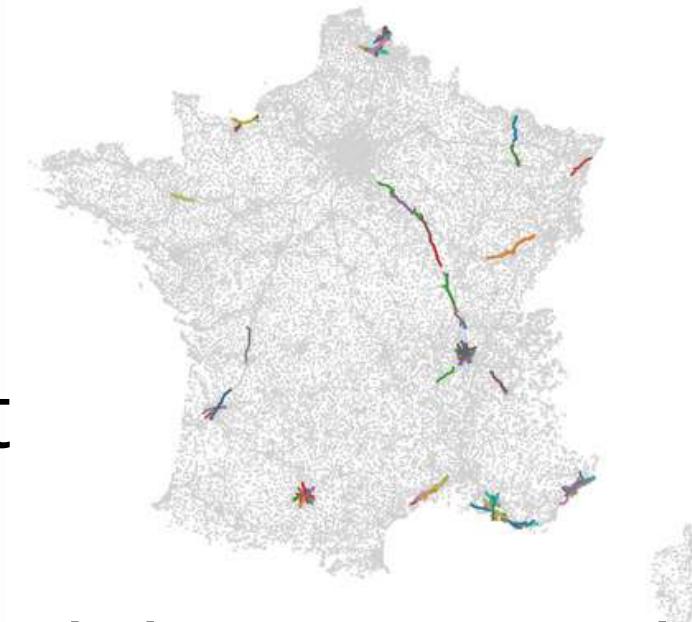
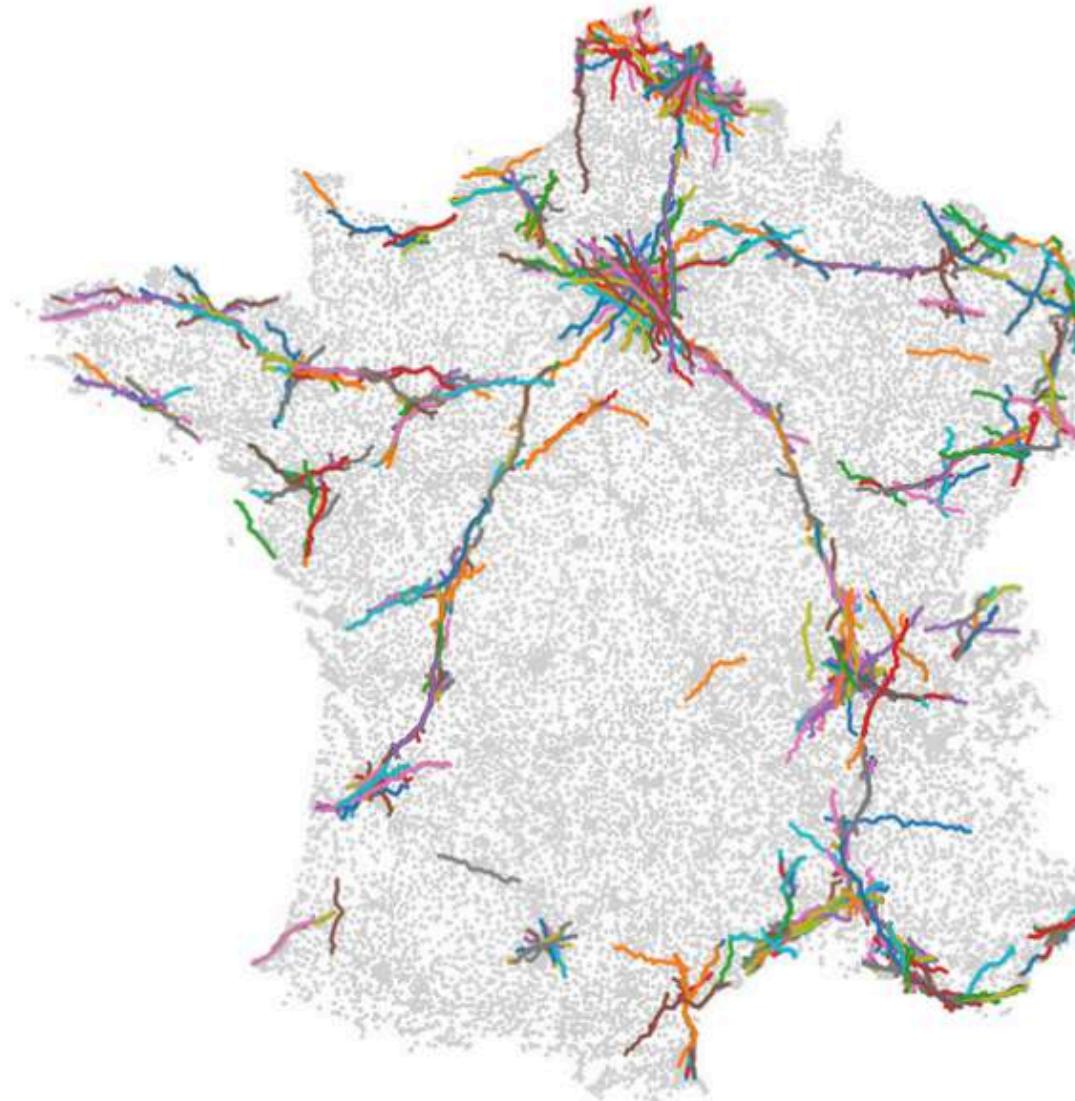
- Erreur globale = écart entre longueur totale réelle et détectée
- Ecart-type des erreurs relatives par département : permet une considération géographique
- Score =  $\alpha \times$  erreur globale +  $\beta \times$  écart type
- La semaine dernière :  $\alpha = 0.5$  et  $\beta = 0.5$  →



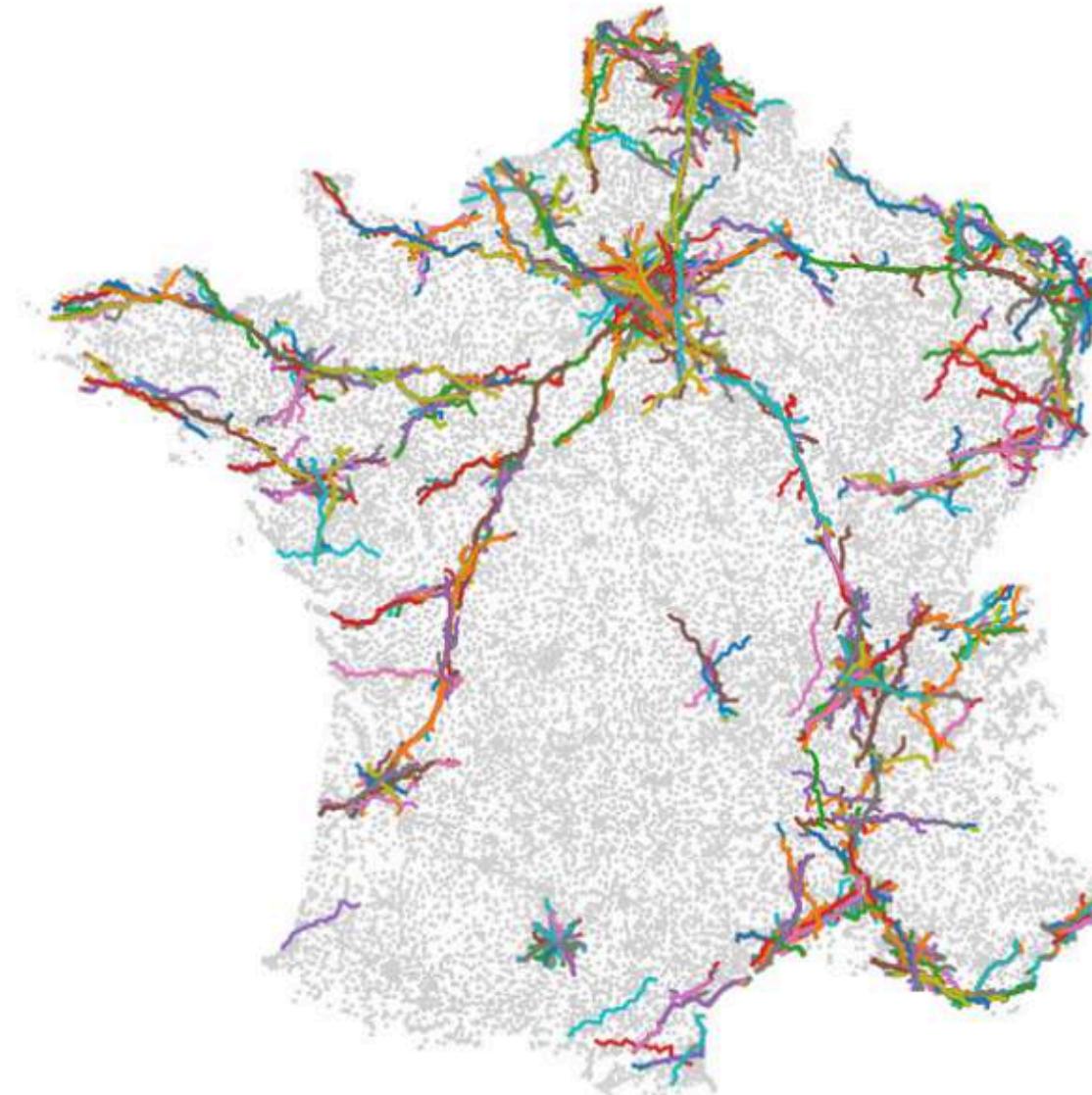
→ Modifier les poids pour donner plus d'importance à l'erreur globale.

# Score

- $\alpha = 0.75$  et  $\beta = 0.25$  : Même résultat
- $\alpha = 0.9$  et  $\beta = 0.1$  :
  - Erreur = 254101 km (58.7%)



- $\alpha = 0.97$  et  $\beta = 0.03$  :
  - Erreur = 2416 km (0.56%)



# **Sans les villes**

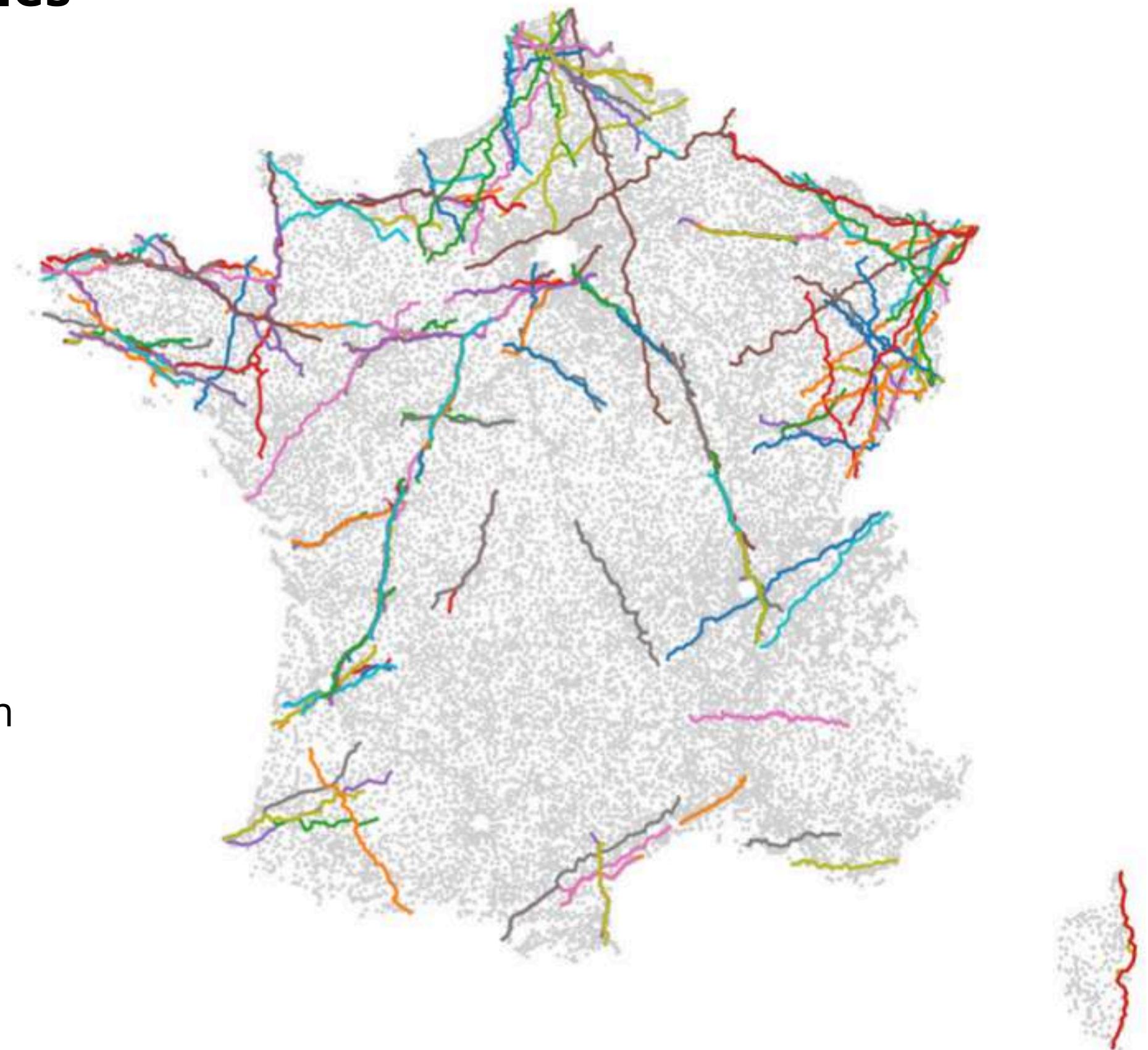
# Paramètres optimaux Sans les villes

- Stations Orange
- Intervalles des paramètres :
  - $rayon = 10\ 000 - 60\ 000$
  - $n\_neighbors = 10 - 30$
  - $dist\_min = 1 - 200$
  - $dist\_max = 4\ 000 - 10\ 000$
  - $angle\_min = 15 - 80$
  - $angle\_max = 70 - 90$
  - $n\_directions = 2 - 5$
  - $angle\_tol = 1 - 10$
  - $min\_len = 10 - 55$
- $\alpha = 0.9$  et  $\beta = 0.1$
- Longueur de référence : Férré + Autoroutes + Nationales + Départementales

- Paramètres trouvés :

- $rayon = 38\ 000$
- $n\_neighbors = 10$
- $dist\_min = 1$
- $dist\_max = 8500$
- $angle\_min = 65$
- $angle\_max = 85$
- $n\_directions = 3$
- $angle\_tol = 5$
- $min\_len = 30$

- Erreur globale : 551km  
(1.10%)



# **Indicateurs**

# Indicateurs

- Erreur absolue et erreur relative globale : écart entre longueur réelle et détectée
- Erreurs relatives des départements
  - Écart-type
  - Département avec l'erreur relative maximale
  - Nombre de départements mal détectés ( $\text{err relative} > 50\%$ )
- Taux de redondance : proportion de chaînes presque identiques (à 90%)

→ Ajout d'autres indicateurs si besoin

# Indicateurs

## Suivi de chemin selon densité de stations, hors villes

Évaluation globale :

Longueur détectée : 50740.3 km

Longueur réelle (référence) : 50189.0 km

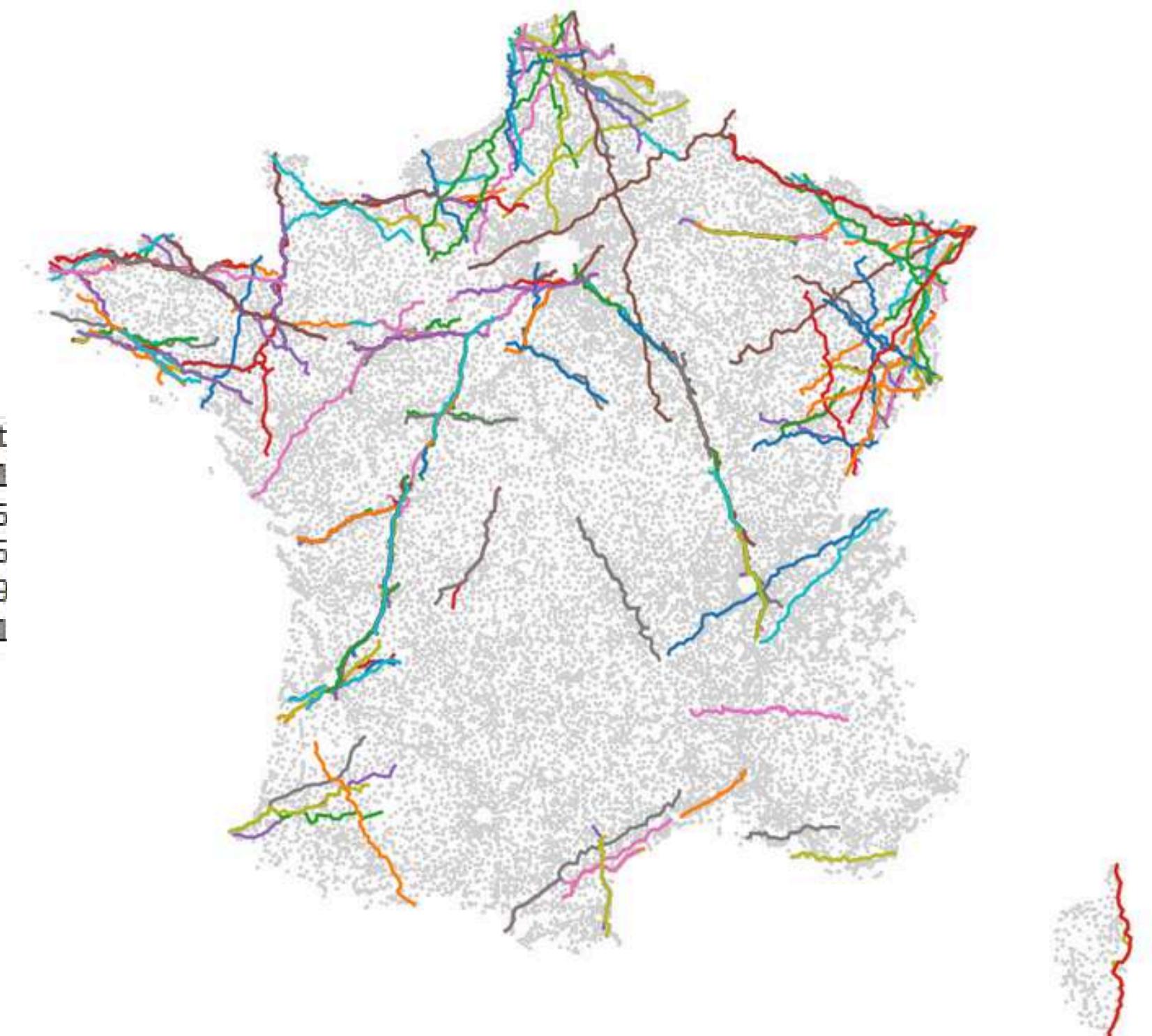
Erreur absolue : 551.3 km

Erreur relative : 1.10%

Évaluation par département :

	Département	Réel (km)	Détecté (km)	Erreur (km)	Erreur relat
71	Paris	73	1522.9	1449.9	19.861
80	Corse-du-Sud	51	234.9	183.9	3.605
74	Hauts-de-Seine	155	665.8	510.8	3.295
95	Alpes-Maritimes	252	975.2	723.2	2.869
36	Var	355	1097.4	742.4	2.091

- Écart-type des erreurs départementales : 2.059
- Max erreur relative d'un département : 1986.15%
- Nombre de départements à forte erreur (>50%) : 35
- Taux de redondance (chaines quasi-identiques) : 9.58%



**29 - 1 Août**

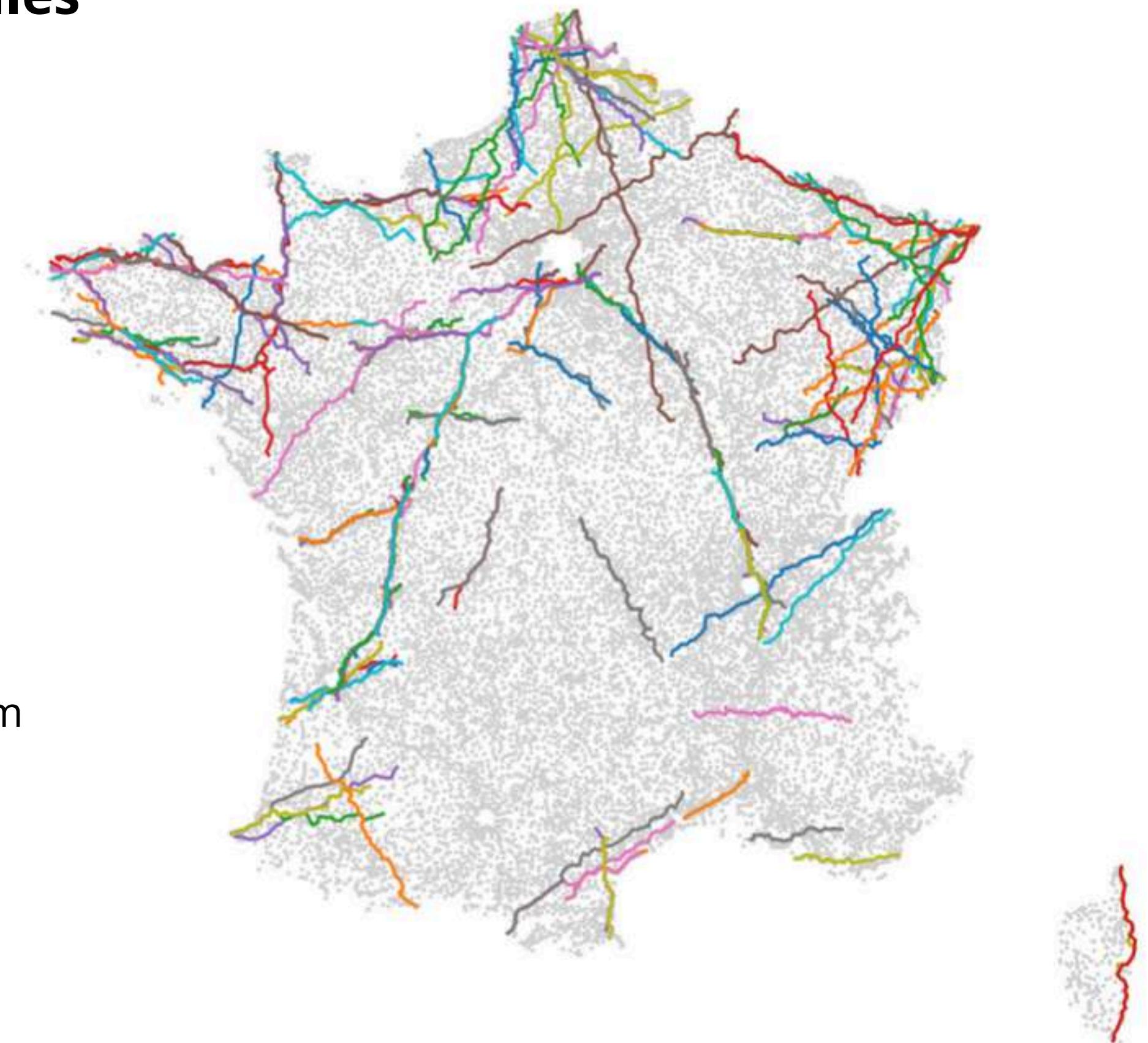
# **Sans les villes**

Orange

Rappel...

## Paramètres optimaux Sans les villes

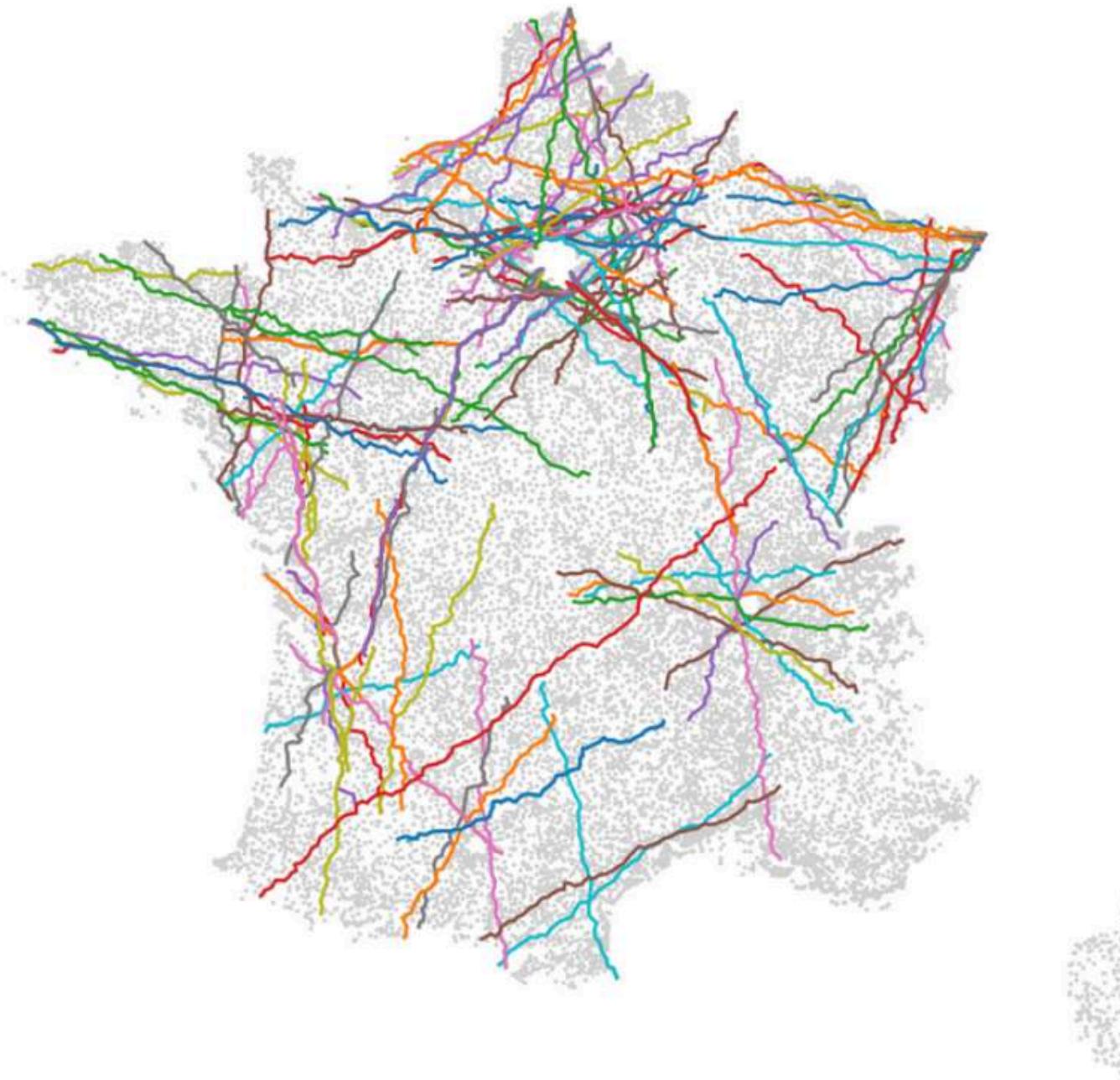
- Stations Orange
- Intervalles des paramètres :
  - $rayon = 10\ 000 - 60\ 000$
  - $n\_neighbors = 10 - 30$
  - $dist\_min = 1 - 200$
  - $dist\_max = 4\ 000 - 10\ 000$
  - $angle\_min = 15 - 80$
  - $angle\_max = 70 - 90$
  - $n\_directions = 2 - 5$
  - $angle\_tol = 1 - 10$
  - $min\_len = 10 - 55$
- $\alpha = 0.9$  et  $\beta = 0.1$
- Longueur de référence : Férré + Autoroutes + Nationales + Départementales
- Paramètres trouvés :
  - $rayon = 38\ 000$
  - $n\_neighbors = 10$
  - $dist\_min = 1$
  - $dist\_max = 8500$
  - $angle\_min = 65$
  - $angle\_max = 85$
  - $n\_directions = 3$
  - $angle\_tol = 5$
  - $min\_len = 30$
- Erreur globale : 551km  
(1.10%)



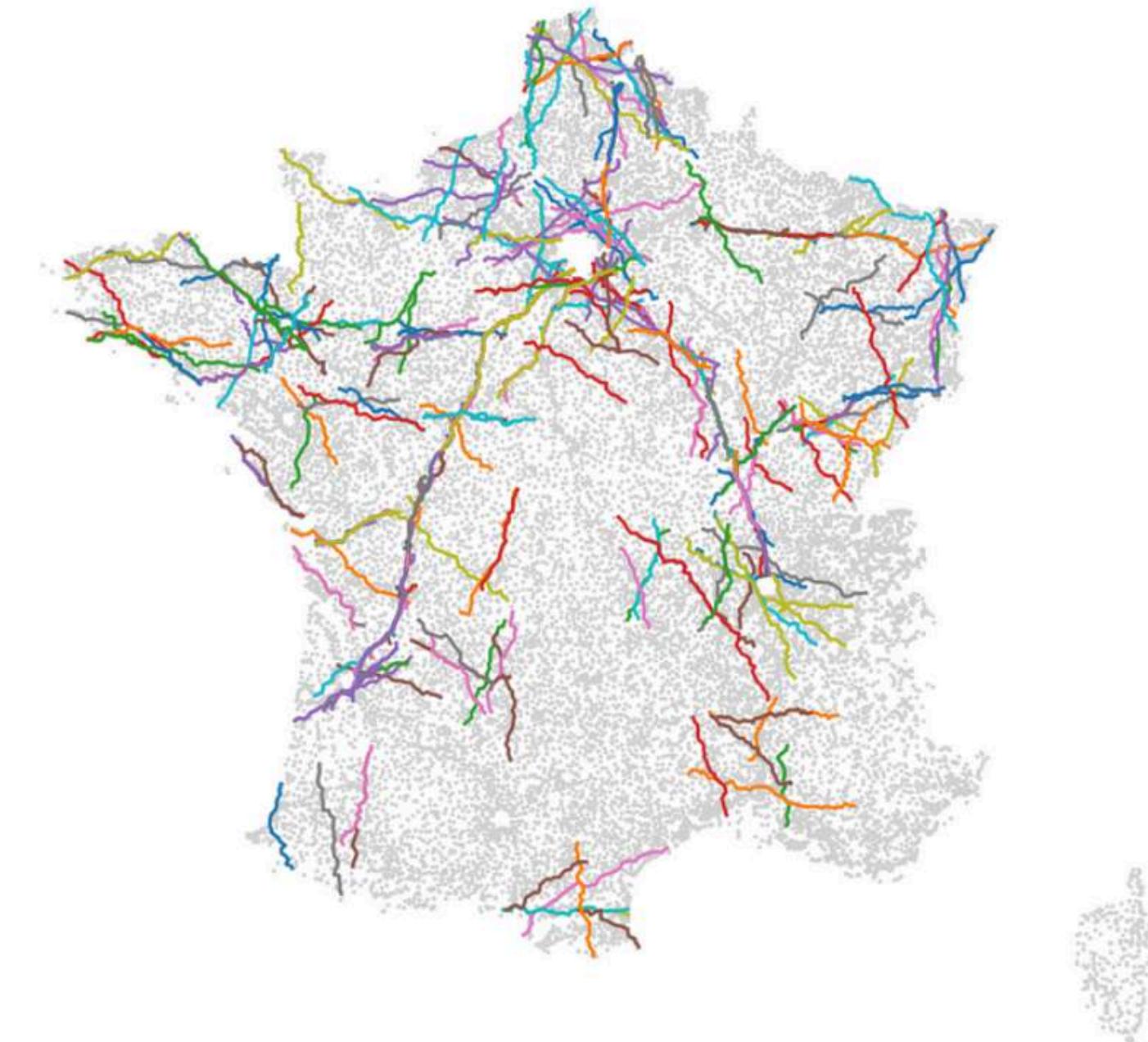
# Paramètres optimaux Sans les villes

$\alpha = 0.75$  et  $\beta = 0.25$

$dist_{max} = 4000 - 8500$  (~~4000 - 10 000~~)



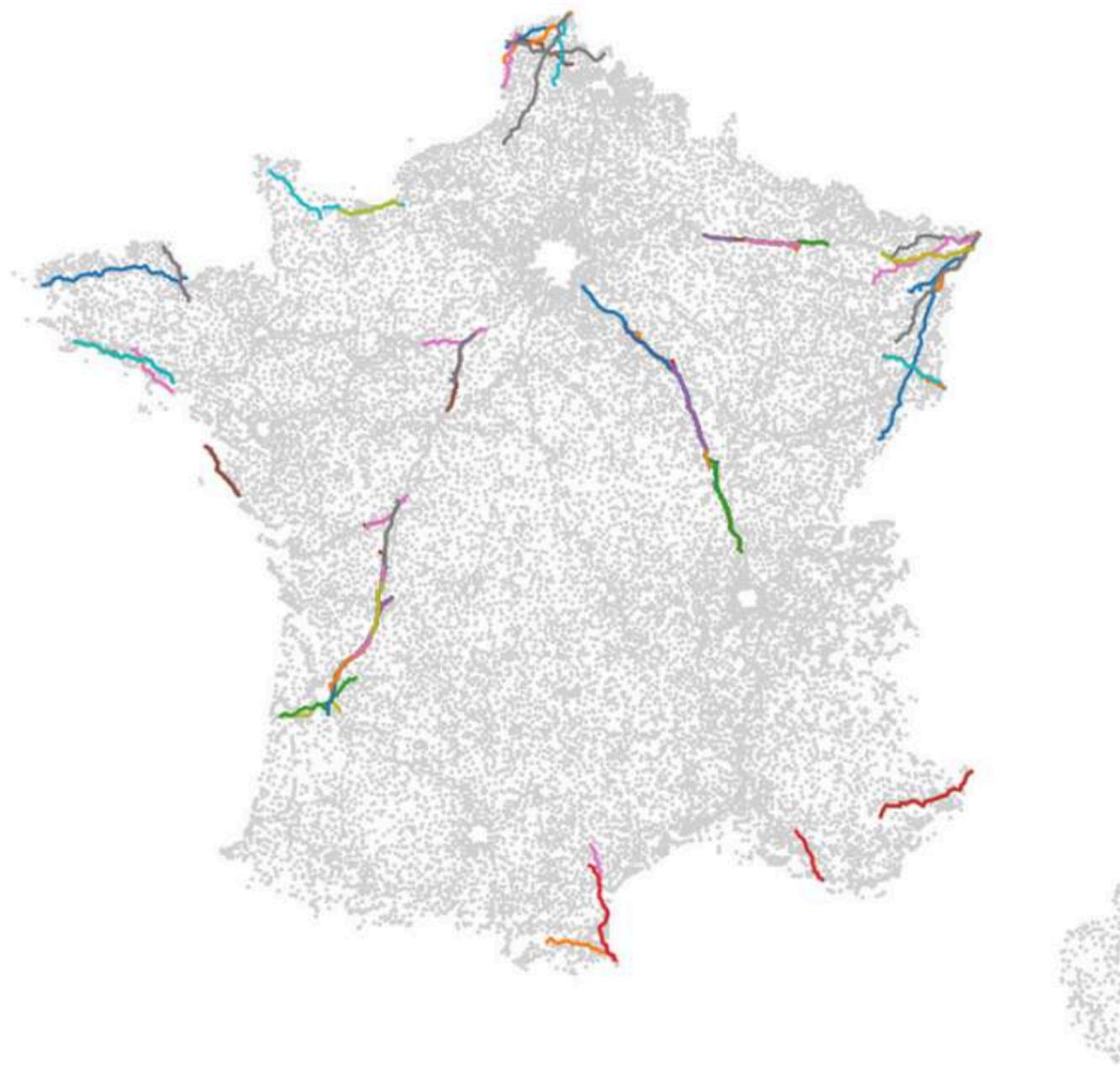
Erreur globale : 3041 km (6.06%)



Erreur globale : 93 km (0.19%)

# Paramètres optimaux Sans les villes

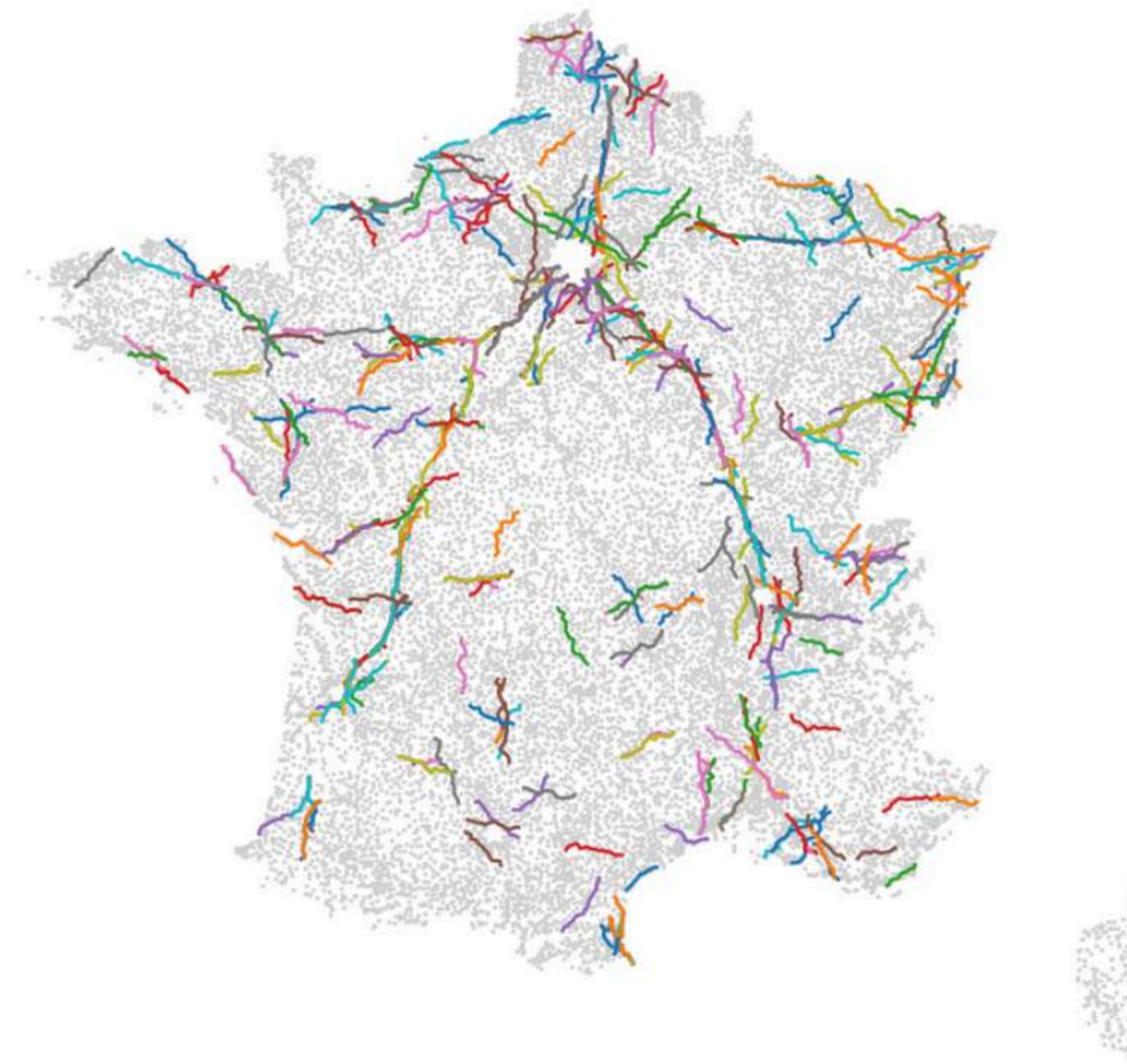
$dist\_max = 4000 - 7500$  (~~4000 - 10 000~~)



Erreur globale : 35334 km (59%)

$dist\_max = 4000 - 7500$  (~~4000 - 10 000~~)

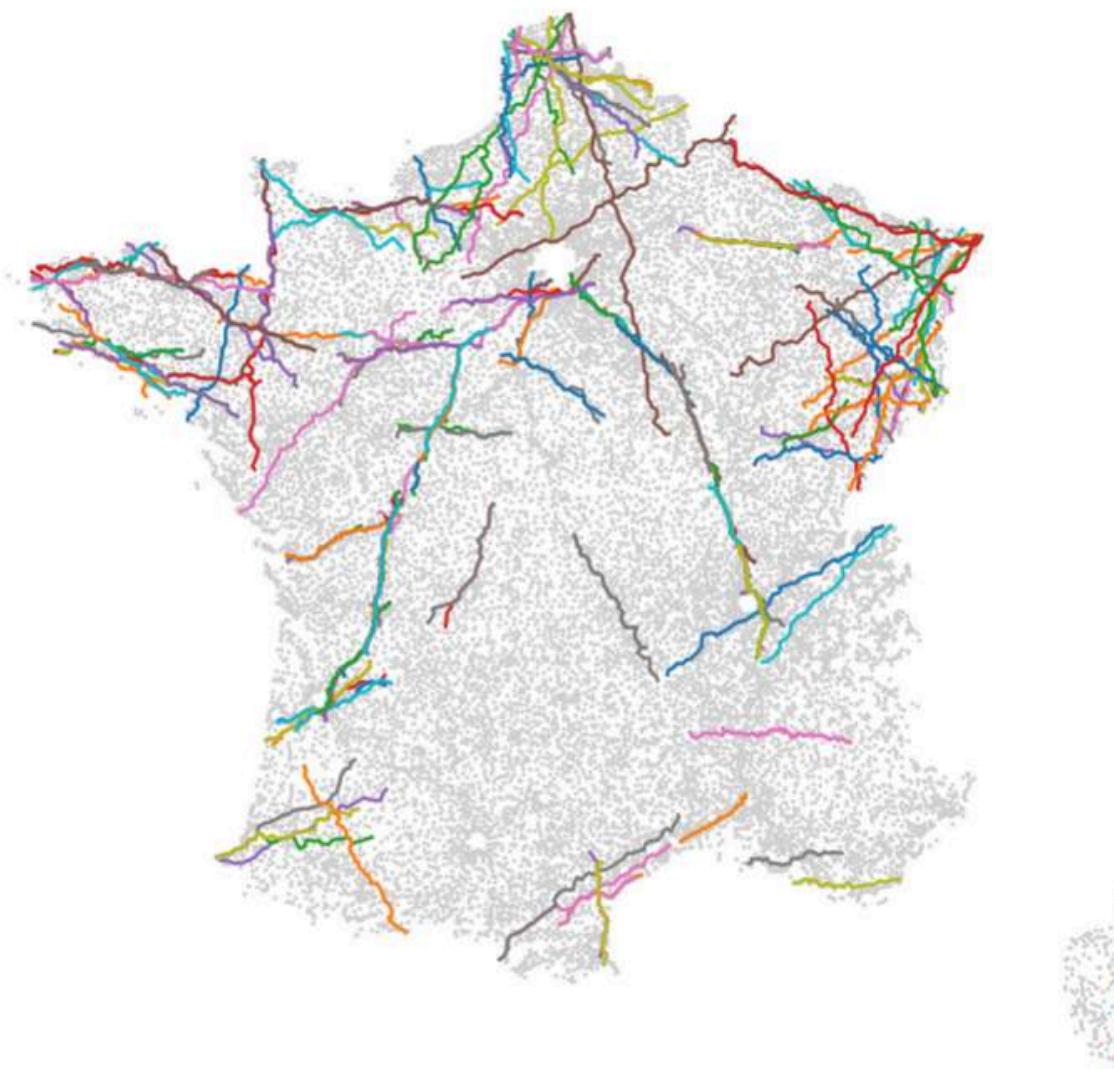
graine aléatoire = 42 (~~43~~)



Erreur globale : 2417 km (4.82%)

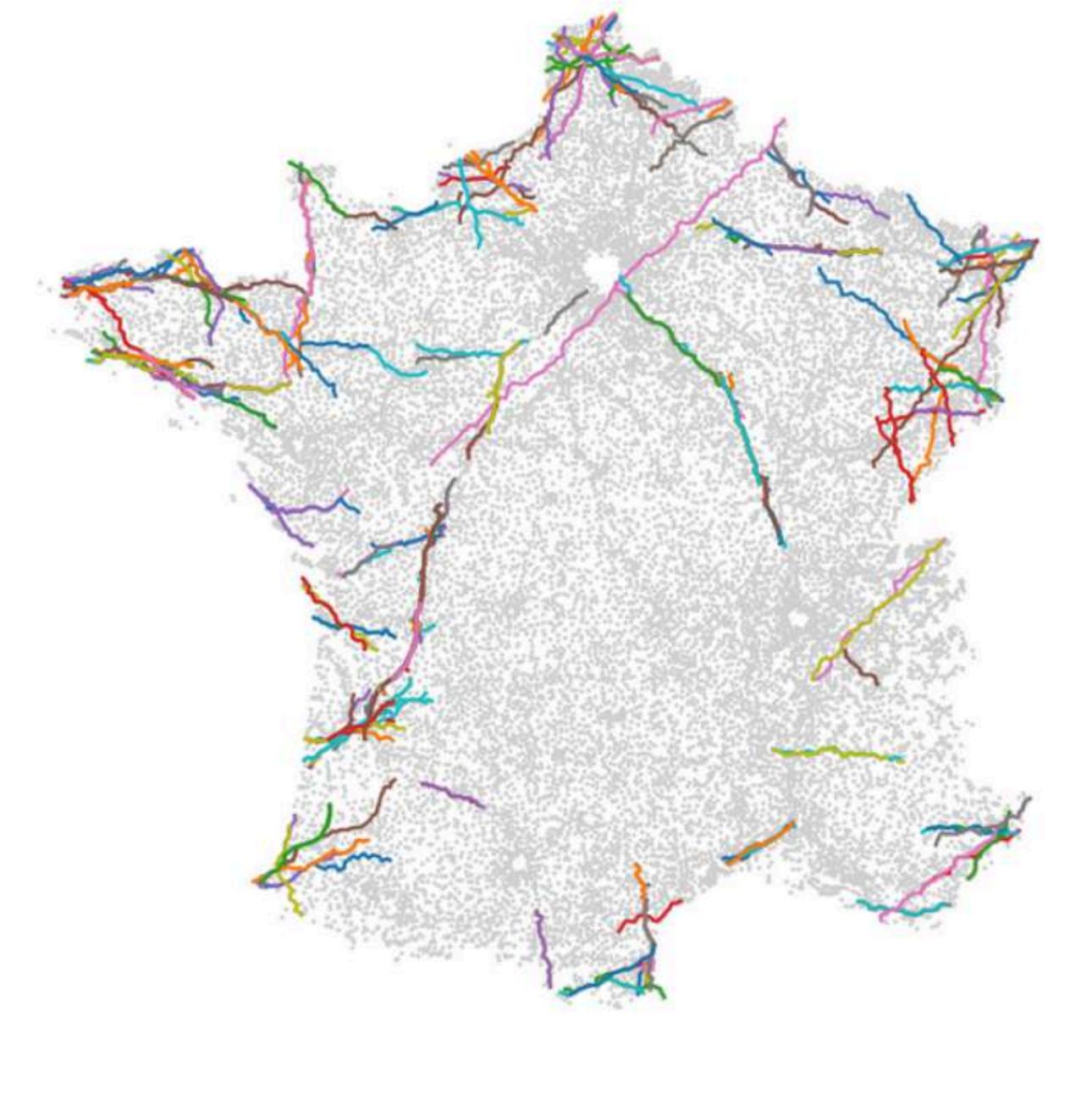
# Paramètres optimaux Sans les villes

Rappel... Au départ :



Erreur globale : 551 km (1.10%)

Enlever moins de points en ville



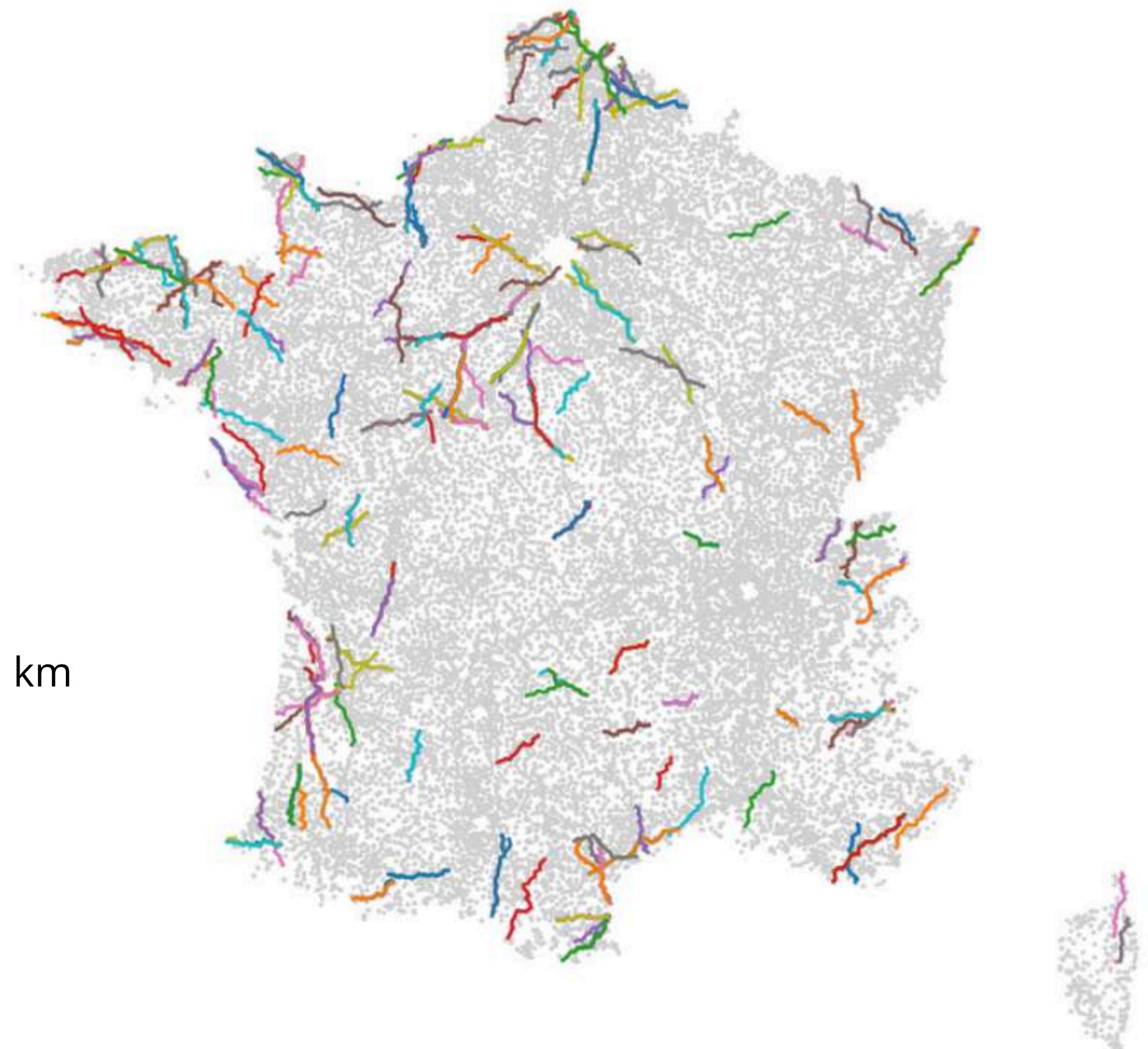
Erreur globale : 2046 km (4.08%)

# **Sans les villes**

Tous les opérateurs

# Paramètres optimaux Sans les villes, tout opérateur

- Tout opérateur
- Intervalles des paramètres :
  - $rayon = 10\ 000 - 60\ 000$
  - $n\_neighbors = 10 - 30$
  - $dist\_min = 1 - 400$
  - $dist\_max = 4\ 000 - 9\ 000$
  - $angle\_min = 15 - 80$
  - $angle\_max = 70 - 90$
  - $n\_directions = 2 - 5$
  - $angle\_tol = 1 - 10$
  - $min\_len = 10 - 55$
- $\alpha = 0.97$  et  $\beta = 0.03$
- Longueur de référence : Férré + Autoroutes + Nationales + Départementales
- Paramètres trouvés :
  - $rayon = 16\ 000$
  - $n\_neighbors = 10$
  - $dist\_min = 290$
  - $dist\_max = 8500$
  - $angle\_min = 55$
  - $angle\_max = 75$
  - $n\_directions = 5$
  - $angle\_tol = 5$
  - $min\_len = 25$
- Erreur globale : 9 961 km (19.85%)



# **Comparaison**

Rappel...

# Indicateurs

Suivi de chemin selon densité de stations, hors villes, Orange

Évaluation globale :

Longueur détectée : 50740.3 km

Longueur réelle (référence) : 50189.0 km

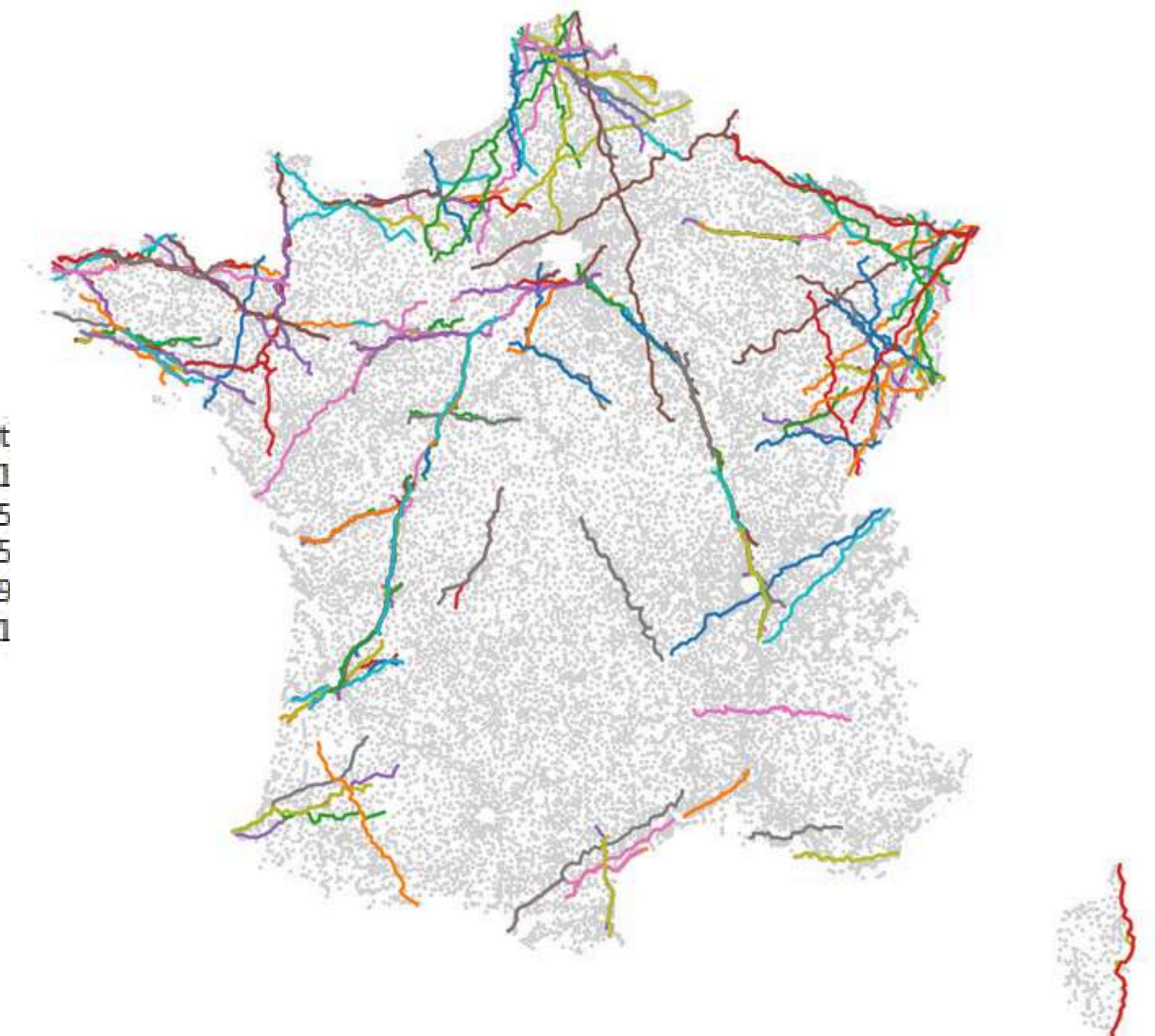
Erreur absolue : 551.3 km

Erreur relative : 1.10%

Évaluation par département :

	Département	Réel (km)	Détecté (km)	Erreur (km)	Erreur relat
71	Paris	73	1522.9	1449.9	19.861
80	Corse-du-Sud	51	234.9	183.9	3.605
74	Hauts-de-Seine	155	665.8	510.8	3.295
95	Alpes-Maritimes	252	975.2	723.2	2.869
36	Var	355	1097.4	742.4	2.091

- Écart-type des erreurs départementales : 2.059
- Max erreur relative d'un département : 1986.15%
- Nombre de départements à forte erreur (>50%) : 35
- Taux de redondance (chaines quasi-identiques) : 9.58%



# Indicateurs Suivi de chemin selon densité de stations, hors villes, Tout opérateur

Évaluation globale :

Longueur détectée : 40227.9 km

Longueur réelle (référence) : 50189.0 km

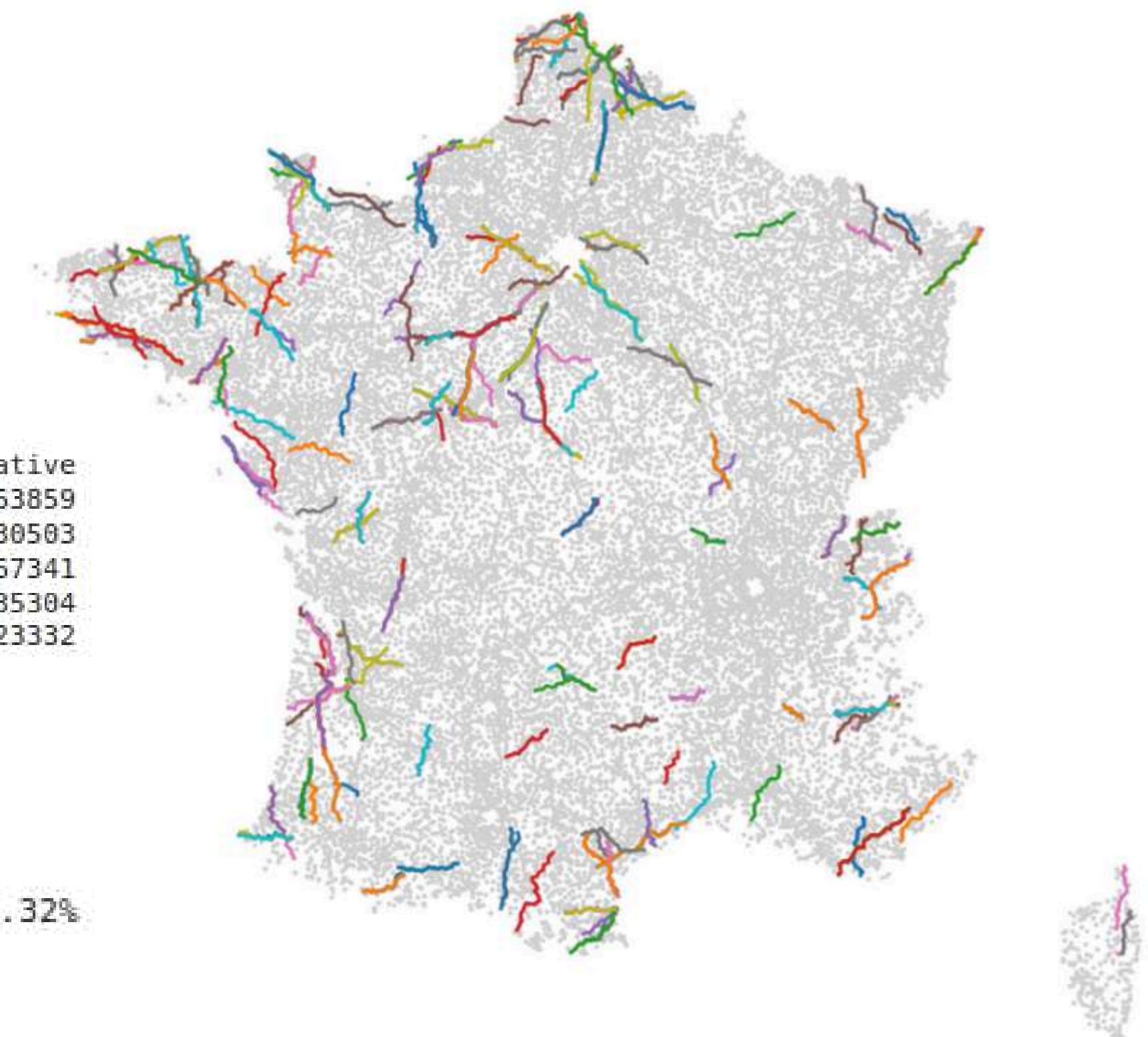
Erreur absolue : 9961.1 km

Erreur relative : 19.85%

Évaluation par département :

	Département	Réel (km)	Détecté (km)	Erreur (km)	Erreur relative
17	Paris	73	939.1	866.1	11.863859
24	Seine-Saint-Denis	178	681.8	503.8	2.830503
44	Corse-du-Sud	51	181.9	130.9	2.567341
10	Hauts-de-Seine	155	455.0	300.0	1.935304
80	Var	355	966.8	611.8	1.723332

- Écart-type des erreurs départementales : 1.243
- Max erreur relative d'un département : 1186.39%
- Nombre de départements à forte erreur (>50%) : 43
- Taux de redondance (chaines quasi-identiques, >90%) : 51.32%



**2 - 7 Août**

# **Amélioration**

Retrait des doublons

# Retrait des doublons

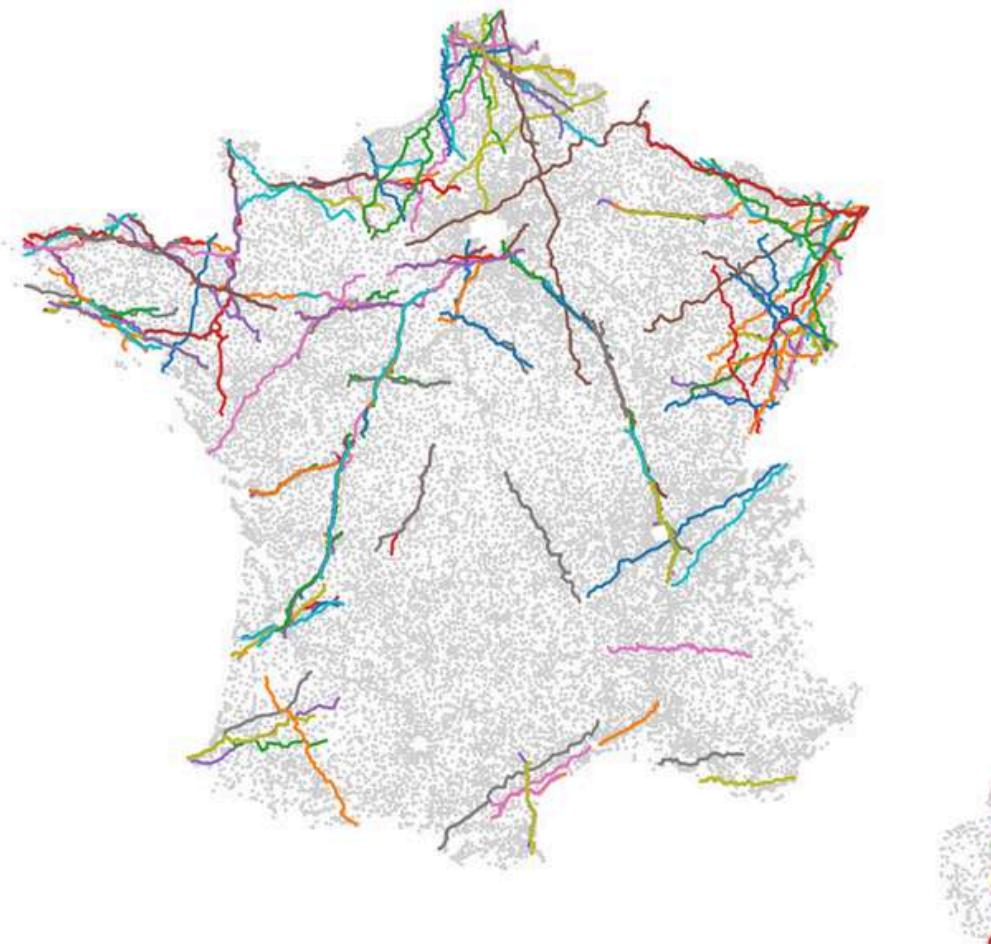
## Hors villes, Orange

### Avant :

Évaluation globale :  
Longueur détectée : 50740.3 km  
Longueur réelle (référence) : 50189.0 km  
Erreurs absolue : 551.3 km  
Erreurs relatives : 1.10%

Évaluation par département :

Département	Réel (km)	Détecté (km)	Erreurs (km)	Erreurs relatives
71 Paris	73	1522.9	1449.9	19.861482
80 Corse-du-Sud	51	234.9	183.9	3.605215
74 Hauts-de-Seine	155	665.8	510.8	3.295539
95 Alpes-Maritimes	252	975.2	723.2	2.869709
36 Var	355	1097.4	742.4	2.091188



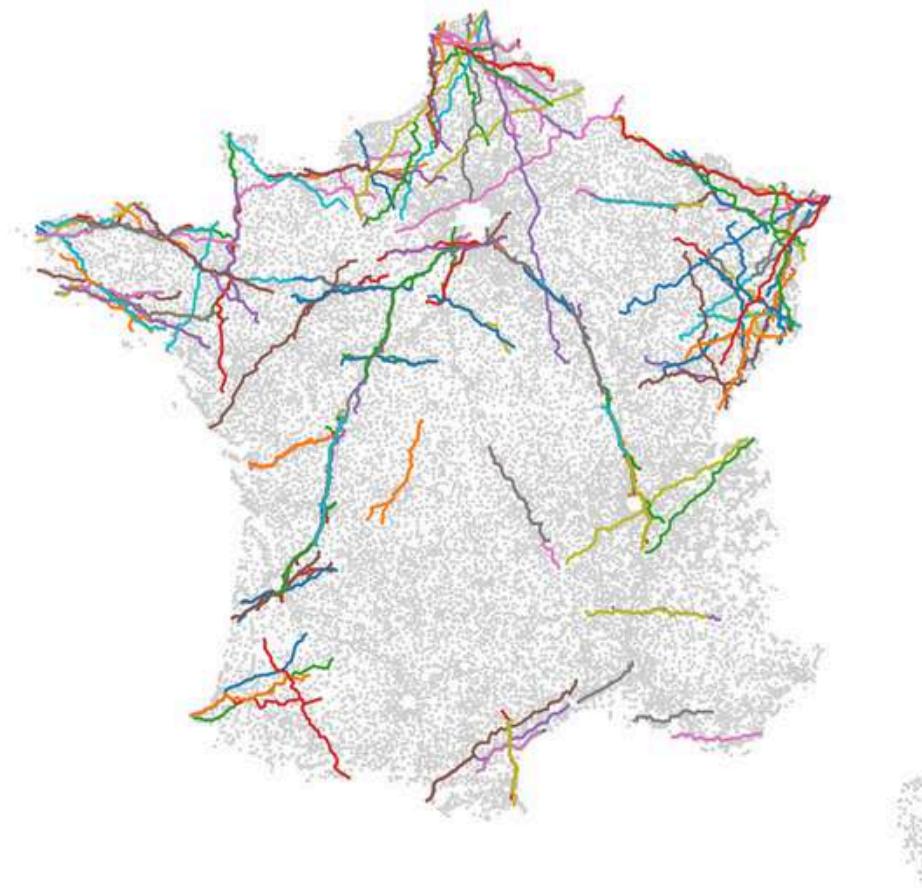
- Écart-type des erreurs départementales : 2.059
- Max erreur relative d'un département : 1986.15%
- Nombre de départements à forte erreur (>50%) : 35
- Taux de redondance (chaînes quasi-identiques) : 9.58%

### Après :

Évaluation globale :  
Longueur détectée : 46123.9 km  
Longueur réelle (référence) : 50189.0 km  
Erreurs absolues : 4065.1 km  
Erreurs relatives : 8.10%

Évaluation par département :

Département	Réel (km)	Détecté (km)	Erreurs (km)	Erreurs relatives
57 Paris	73	1349.4	1276.4	17.484925
58 Corse-du-Sud	51	211.7	160.7	3.151365
4 Hauts-de-Seine	155	583.0	428.0	2.761201
89 Alpes-Maritimes	252	887.0	635.0	2.519971
67 Var	355	974.5	619.5	1.745155



- Écart-type des erreurs départementales : 1.801
- Max erreur relative d'un département : 1748.49%
- Nombre de départements à forte erreur (>50%) : 28
- Taux de redondance (chaînes quasi-identiques, >90%) : 2.72%

# Retrait des doublons

Hors villes, Tout opérateur

## Avant :

Évaluation globale :

Longueur détectée : 40227.9 km

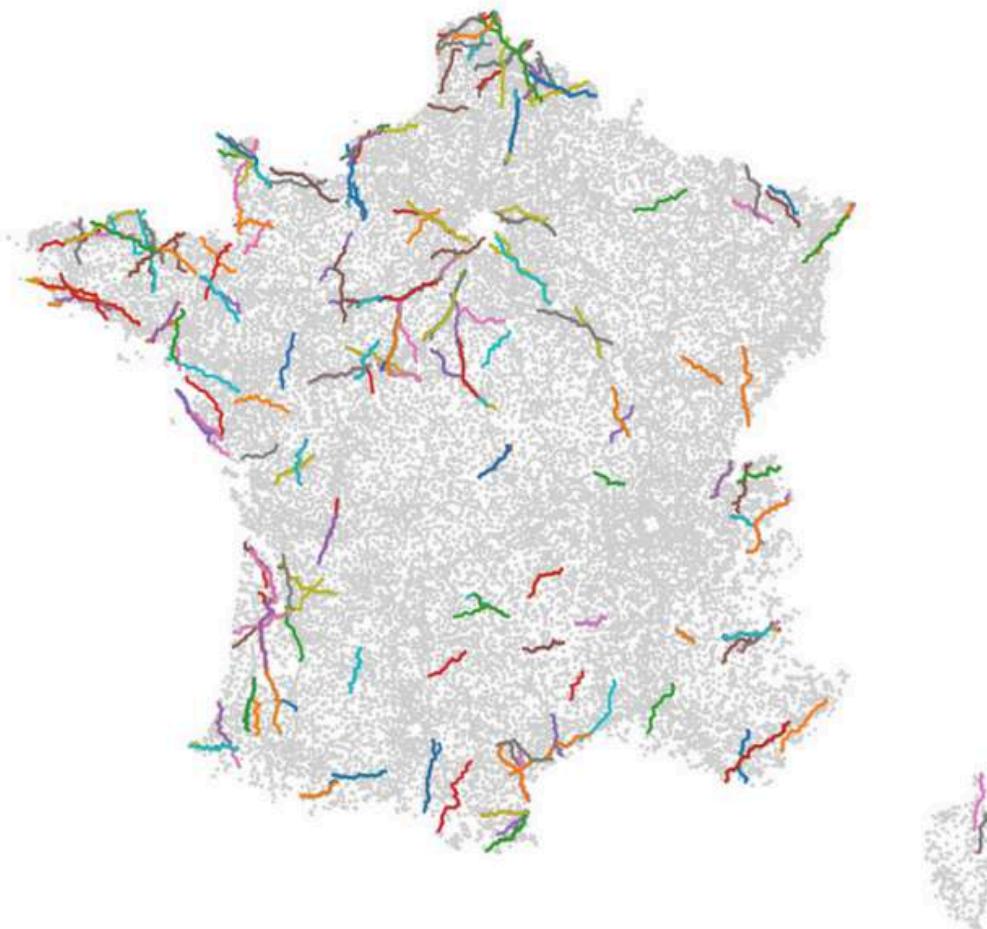
Longueur réelle (référence) : 50189.0 km

Erreur absolue : 9961.1 km

Erreur relative : 19.85%

Évaluation par département :

	Département	Réel (km)	Détecté (km)	Erreur (km)	Erreur relative
17	Paris	73	939.1	866.1	11.863859
24	Seine-Saint-Denis	178	681.8	503.8	2.830503
44	Corse-du-Sud	51	181.9	130.9	2.567341
10	Hauts-de-Seine	155	455.0	300.0	1.935304
80	Var	355	966.8	611.8	1.723332



- Ecart-type des erreurs départementales : 1.243
- Max erreur relative d'un département : 1186.39%
- Nombre de départements à forte erreur (>50%) : 43
- Taux de redondance (chaines quasi-identiques, >90%) : 51.32%

## Après :

Évaluation globale :

Longueur détectée : 31428.8 km

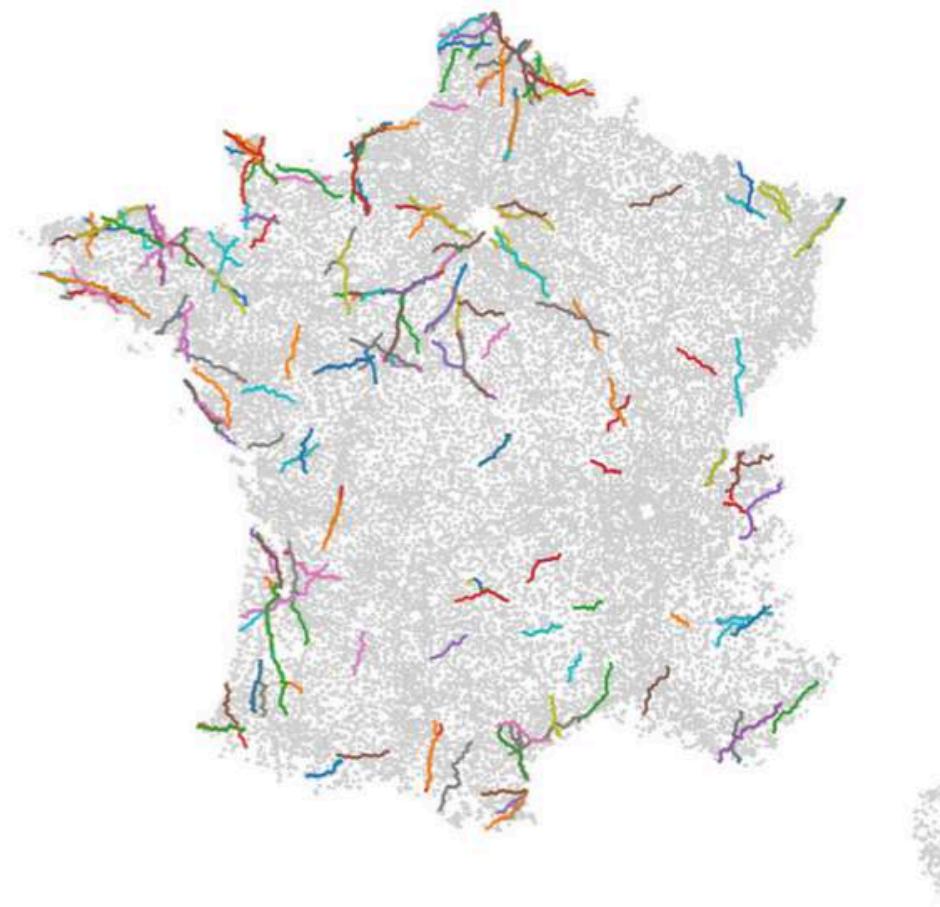
Longueur réelle (référence) : 50189.0 km

Erreur absolue : 18760.2 km

Erreur relative : 37.38%

Évaluation par département :

	Département	Réel (km)	Détecté (km)	Erreur (km)	Erreur relative
57	Paris	73	715.1	642.1	8.795843
5	Seine-Saint-Denis	178	603.1	425.1	2.387951
88	Corse-du-Sud	51	155.9	104.9	2.056059
93	Hauts-de-Seine	155	398.0	243.0	1.567686
34	Var	355	822.9	467.9	1.318118



- Ecart-type des erreurs départementales : 0.912
- Max erreur relative d'un département : 879.58%
- Nombre de départements à forte erreur (>50%) : 55
- Taux de redondance (chaines quasi-identiques, >90%) : 18.54%

# **Zones mal détectées**

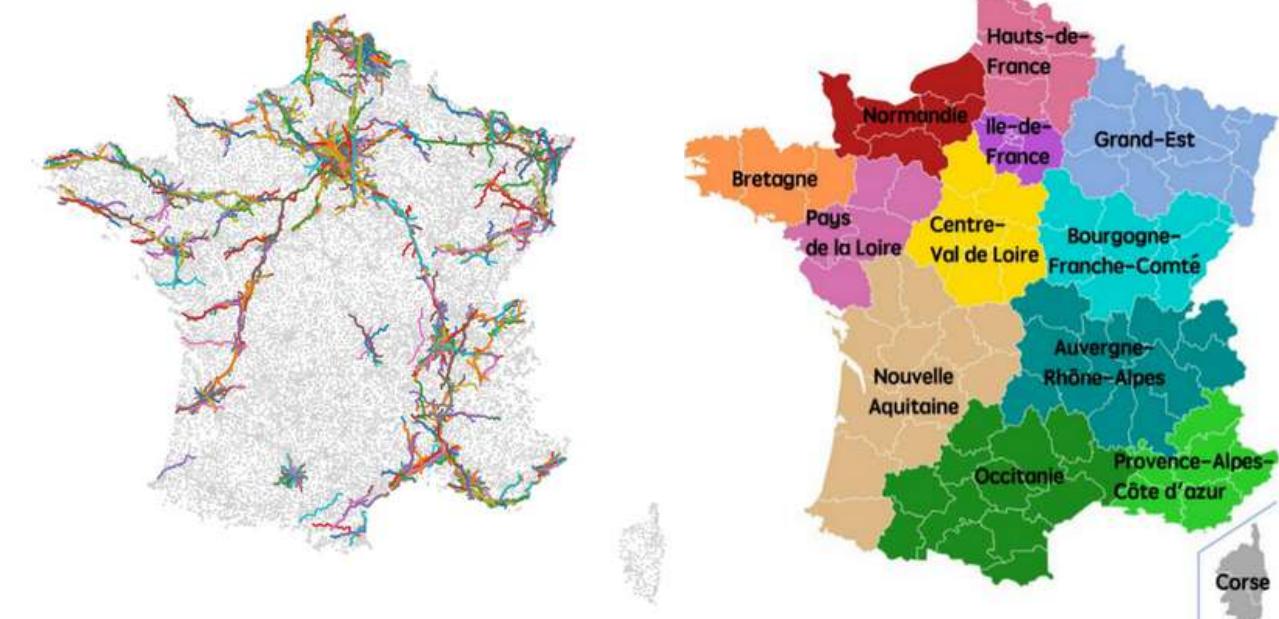
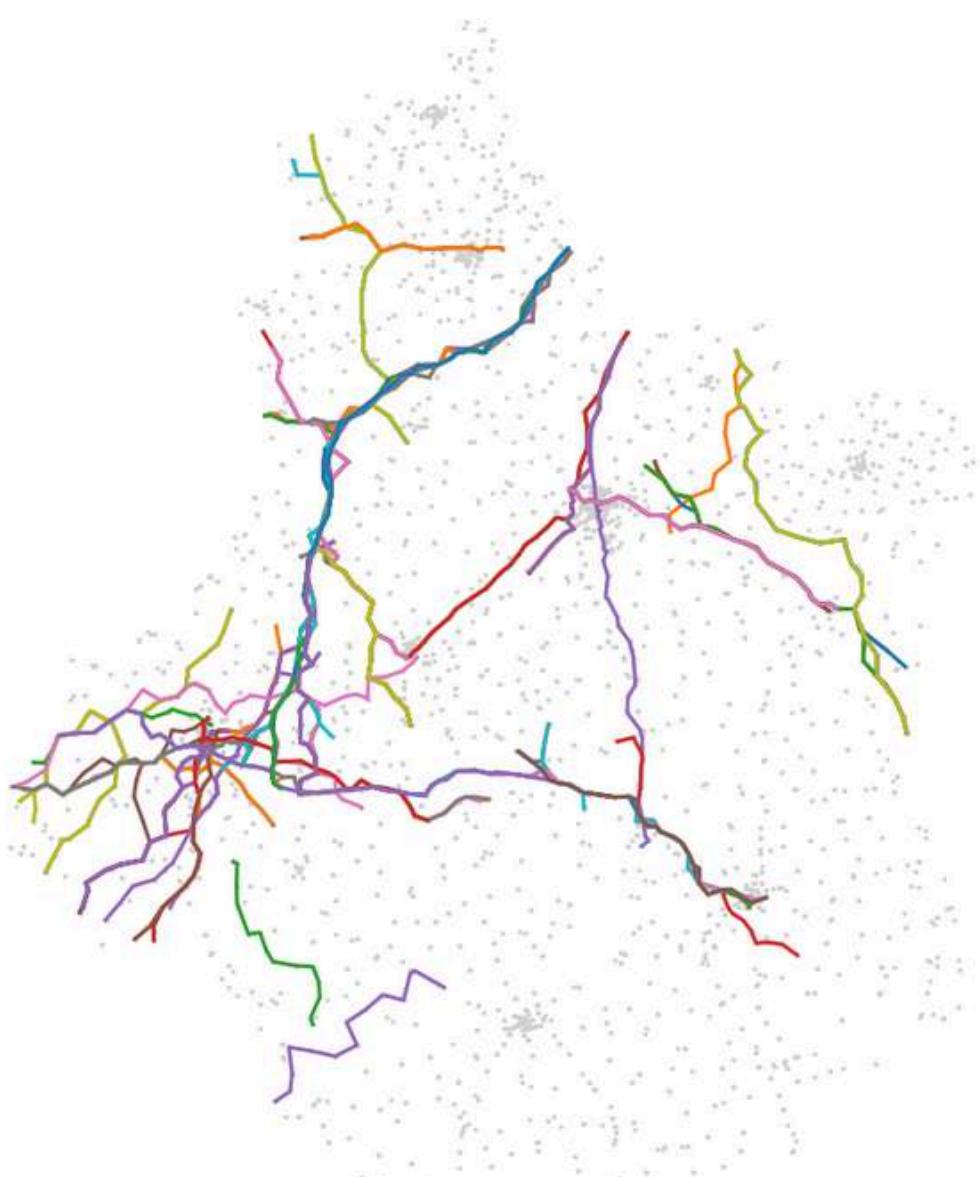
Zoom

# Zones mal détectées Centre-Val de Loire

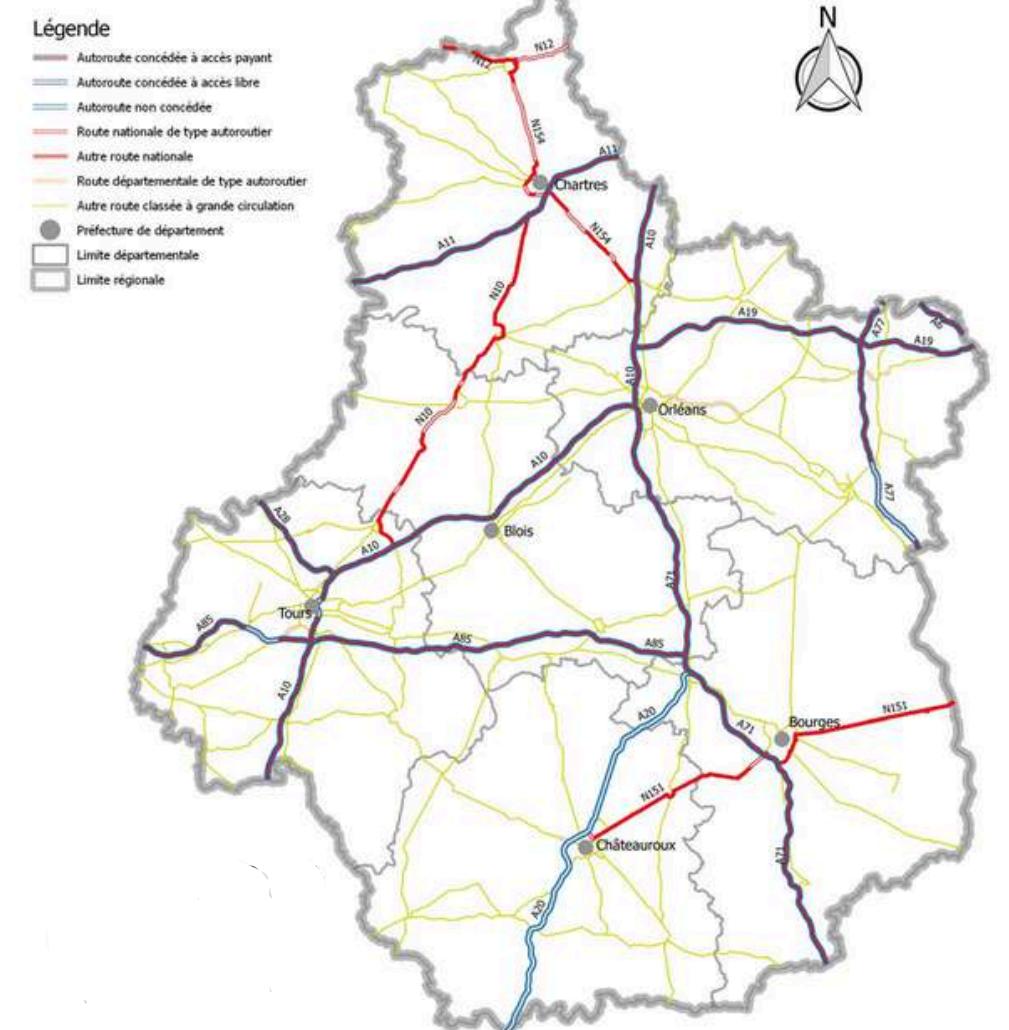
Tout opérateur :

- rayon = 35 000
- n\_neighbors = 15
- dist\_min = 1
- dist\_max = 7000
- angle\_min = 55
- angle\_max = 85
- n\_directions = 5
- angle\_tol = 0
- min\_len = 30

Orange seul :



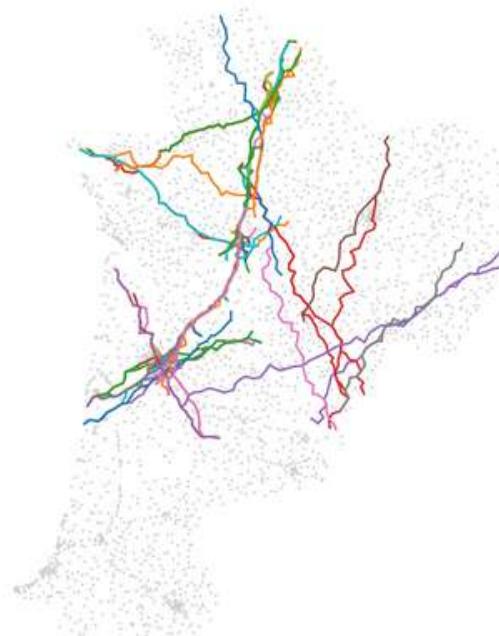
**Routes à grande circulation en région Centre-Val de Loire**



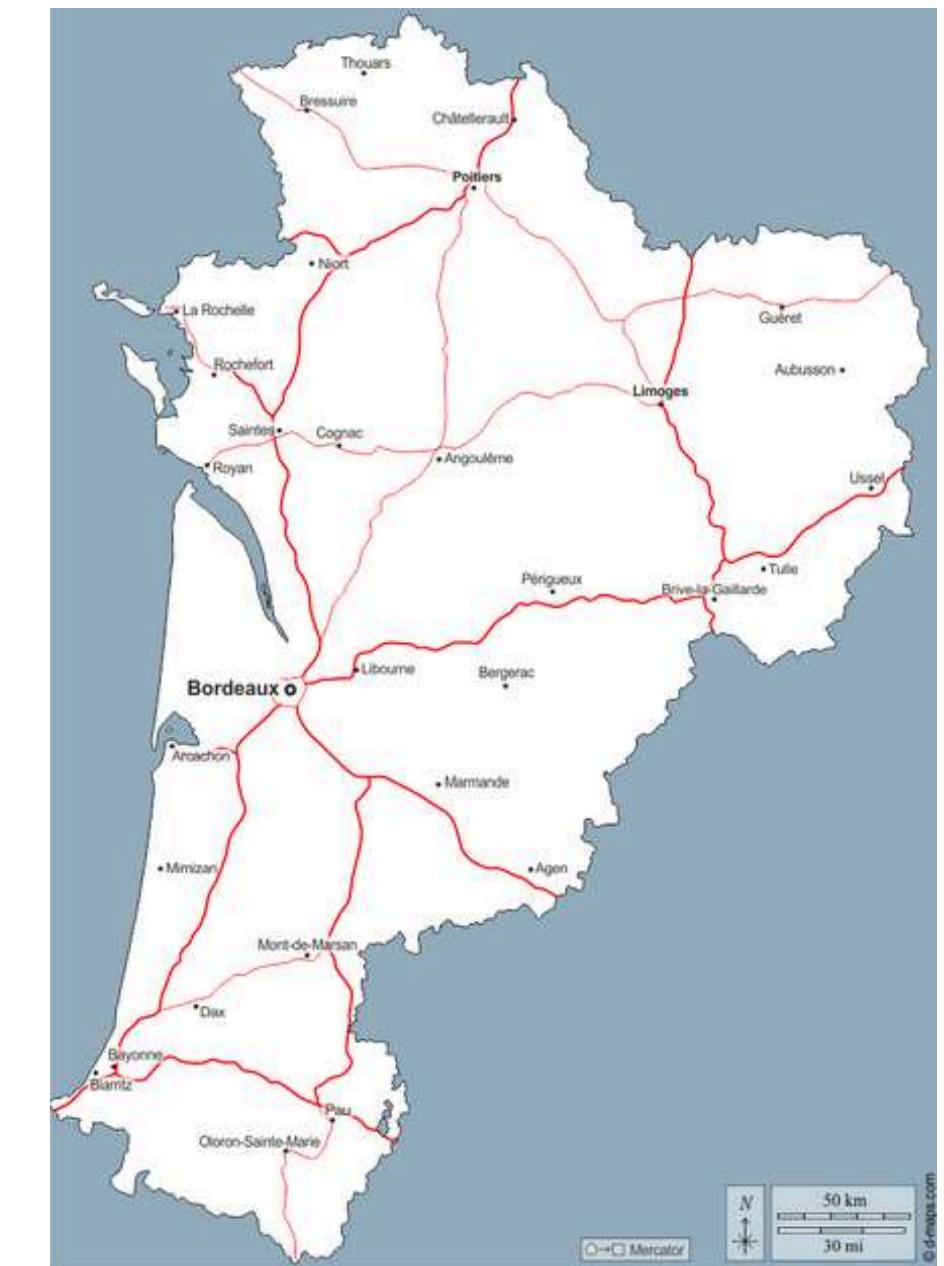
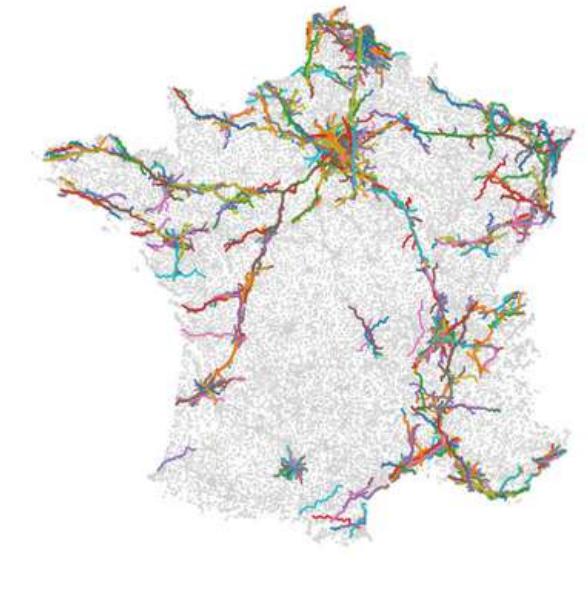
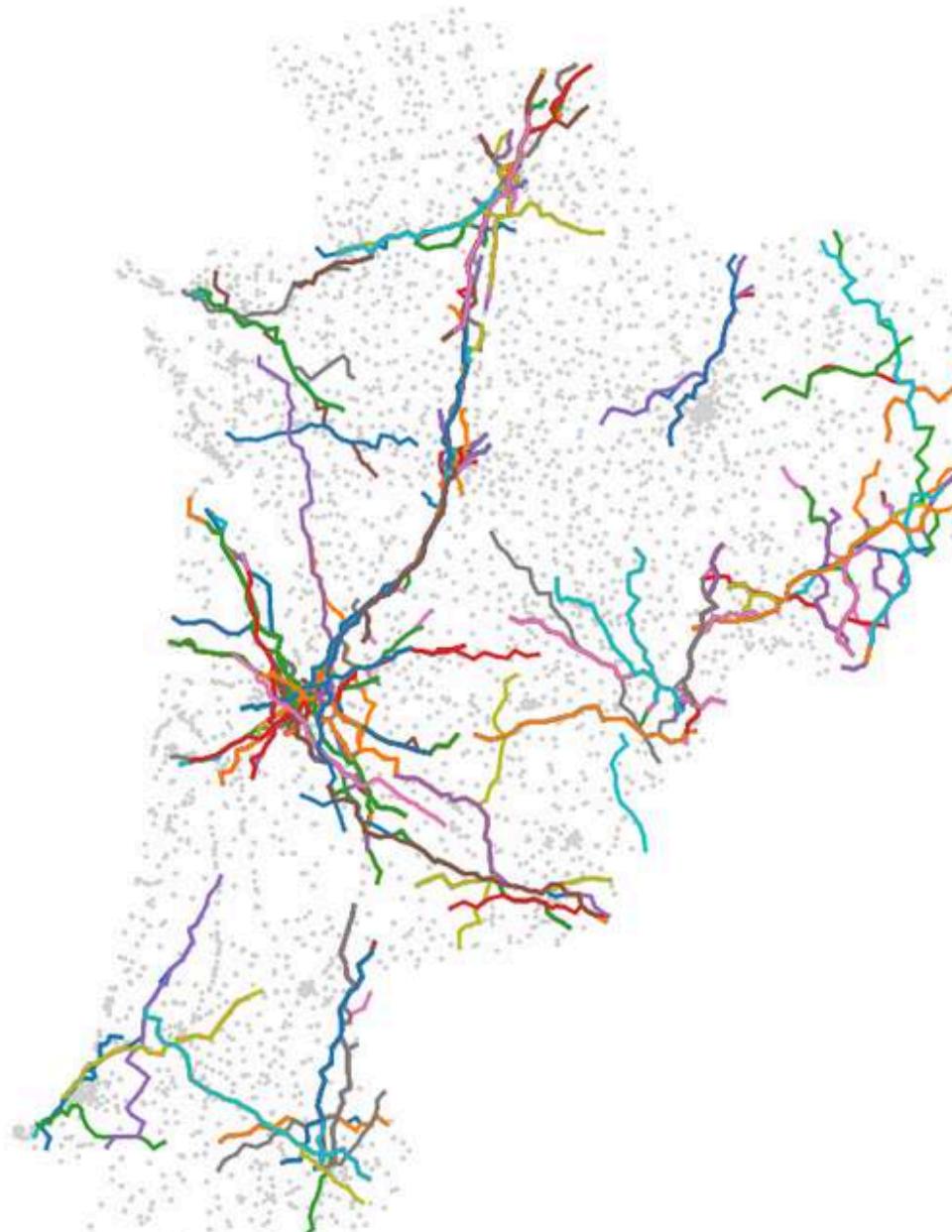
# Zones mal détectées Nouvelle Aquitaine

Tout opérateur :

- rayon = 40 000
- n\_neighbors = 15
- dist\_min = 1
- dist\_max = 7000
- angle\_min = 55
- angle\_max = 85
- n\_directions = 5
- angle\_tol = 0
- min\_len = 35



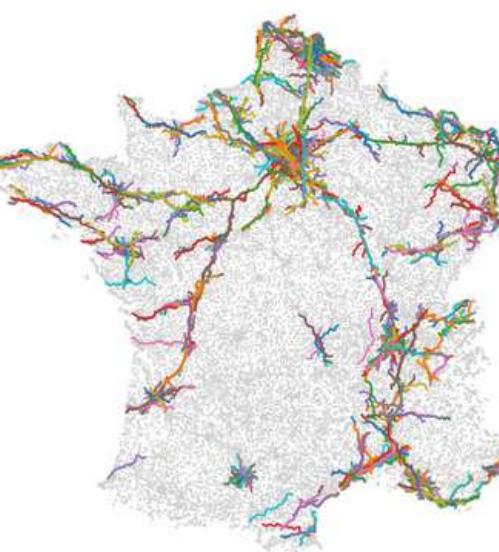
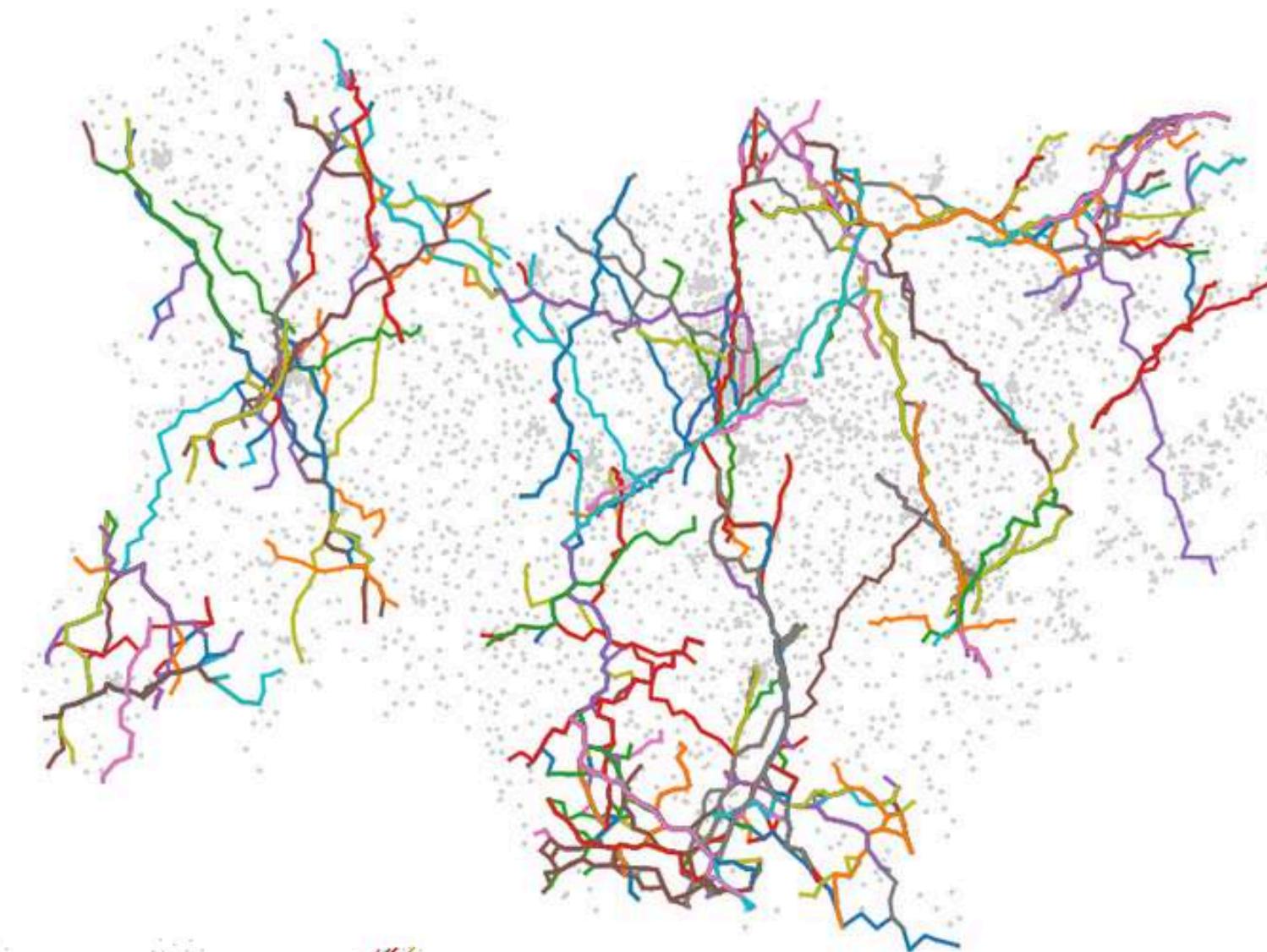
Orange seul :



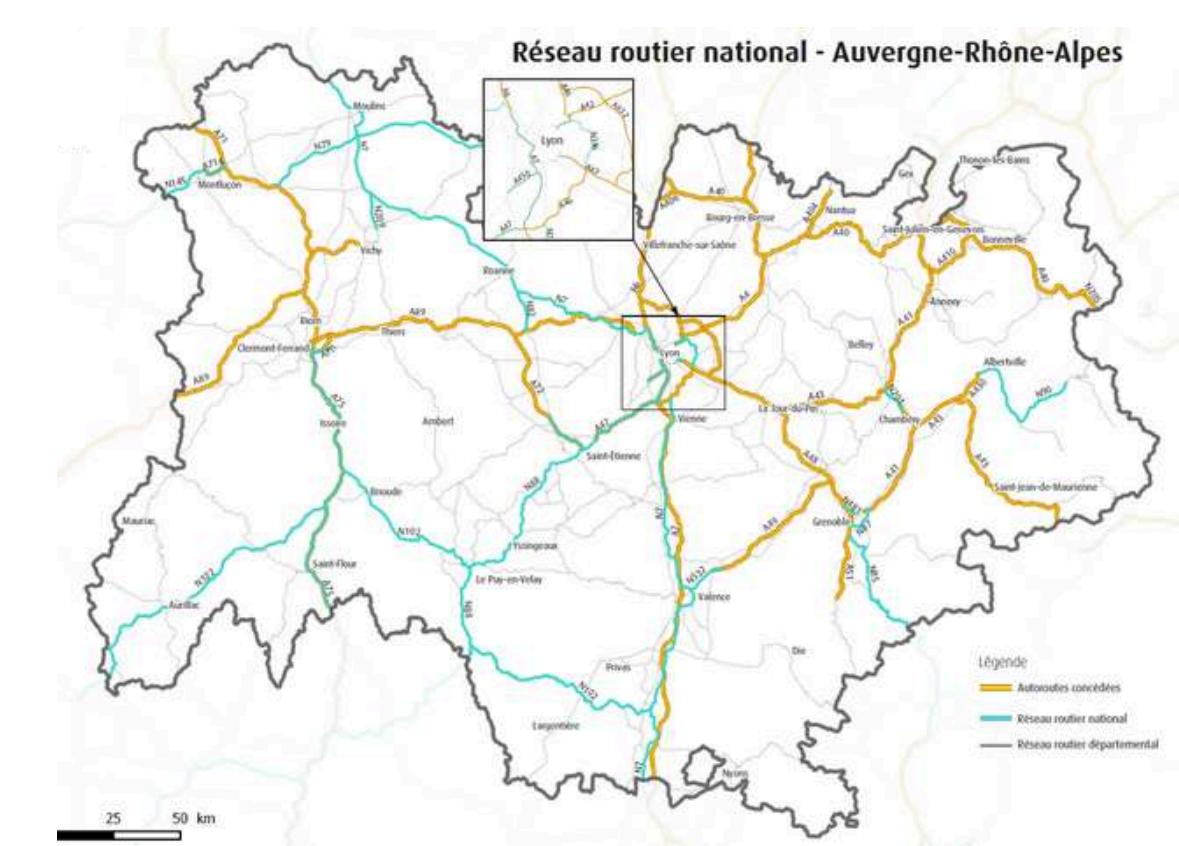
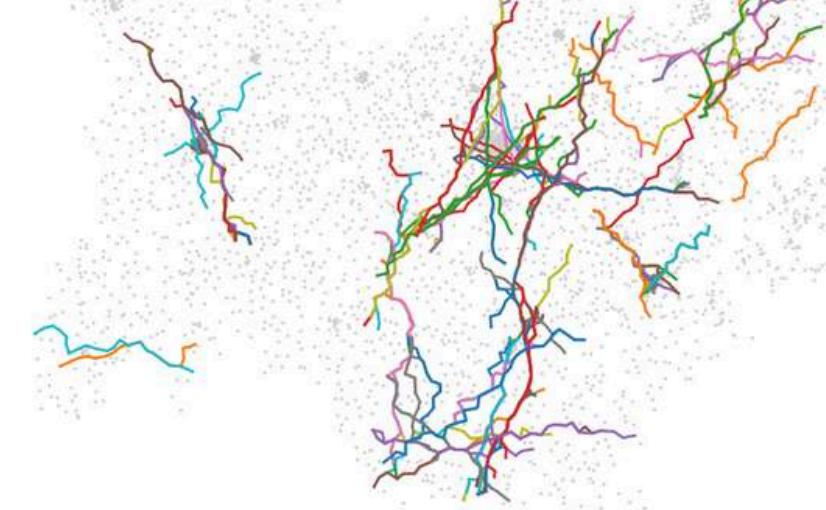
# Zones mal détectées Auvergne-Rhône-Alpes

Tout opérateur :

- rayon = 50 000
- n\_neighbors = 15
- dist\_min = 1
- dist\_max = 6000
- angle\_min = 55
- angle\_max = 85
- n\_directions = 5
- angle\_tol = 0
- min\_len = 40



Orange seul :



**8 - 14 Août**

# **Meilleur calcul de densité**

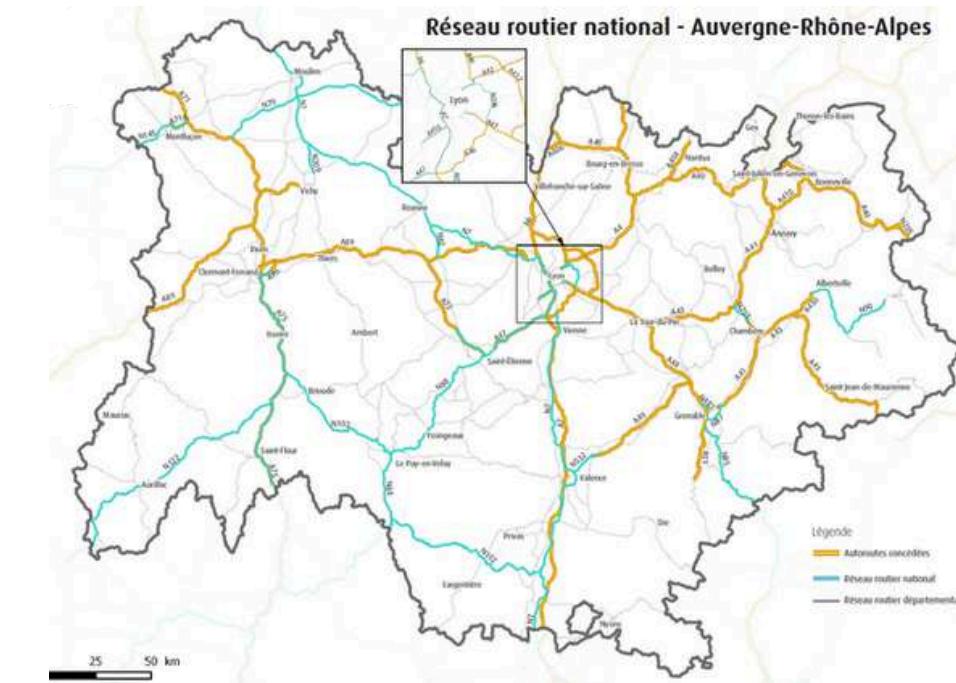
# Calcul de densité

- Avant : Nombre de voisins dans le rayon
- Après : Nombre de voisins dans le rayon sur surface occupée (enveloppe convexe de ces points)

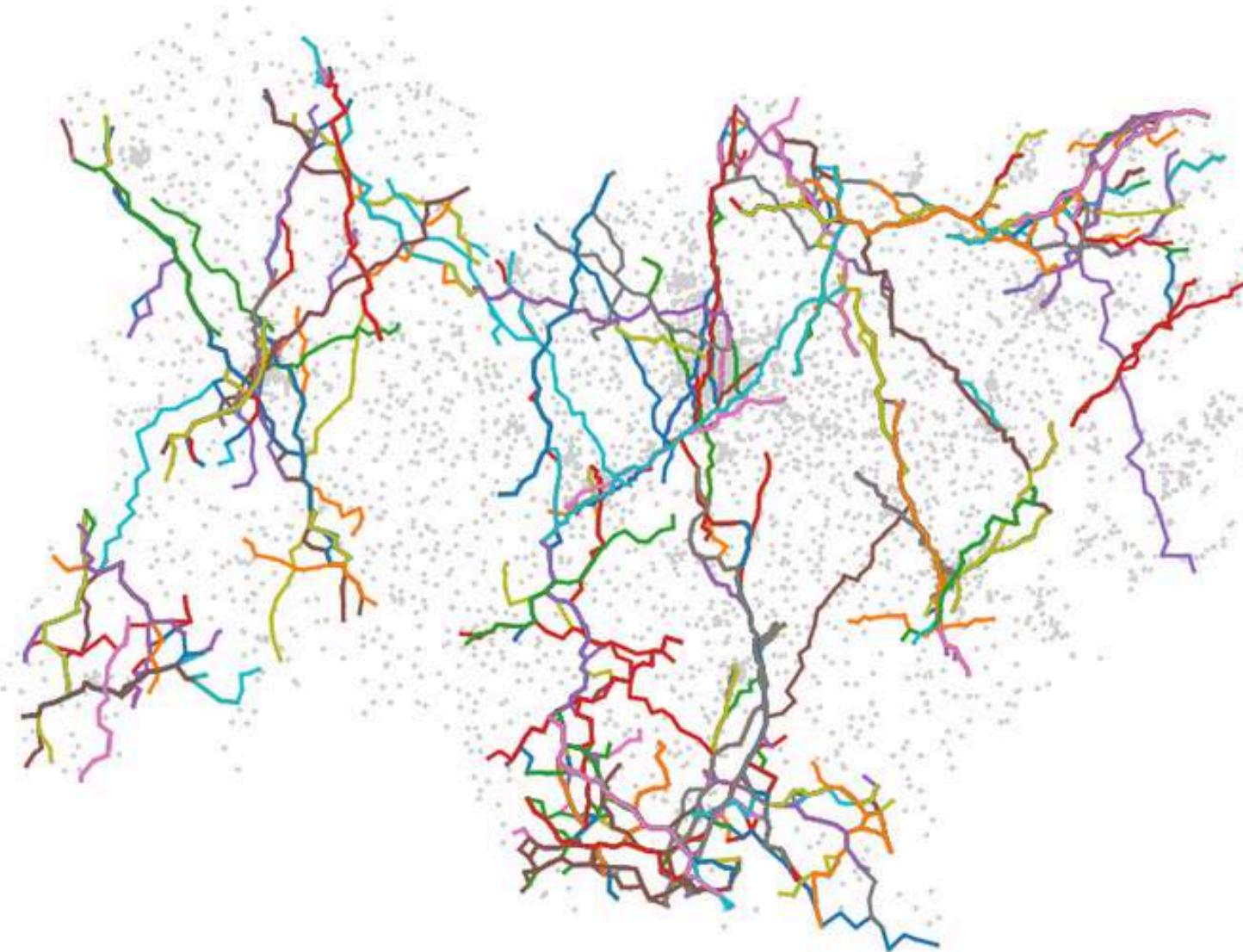
# Régions mal détectées

Zoom

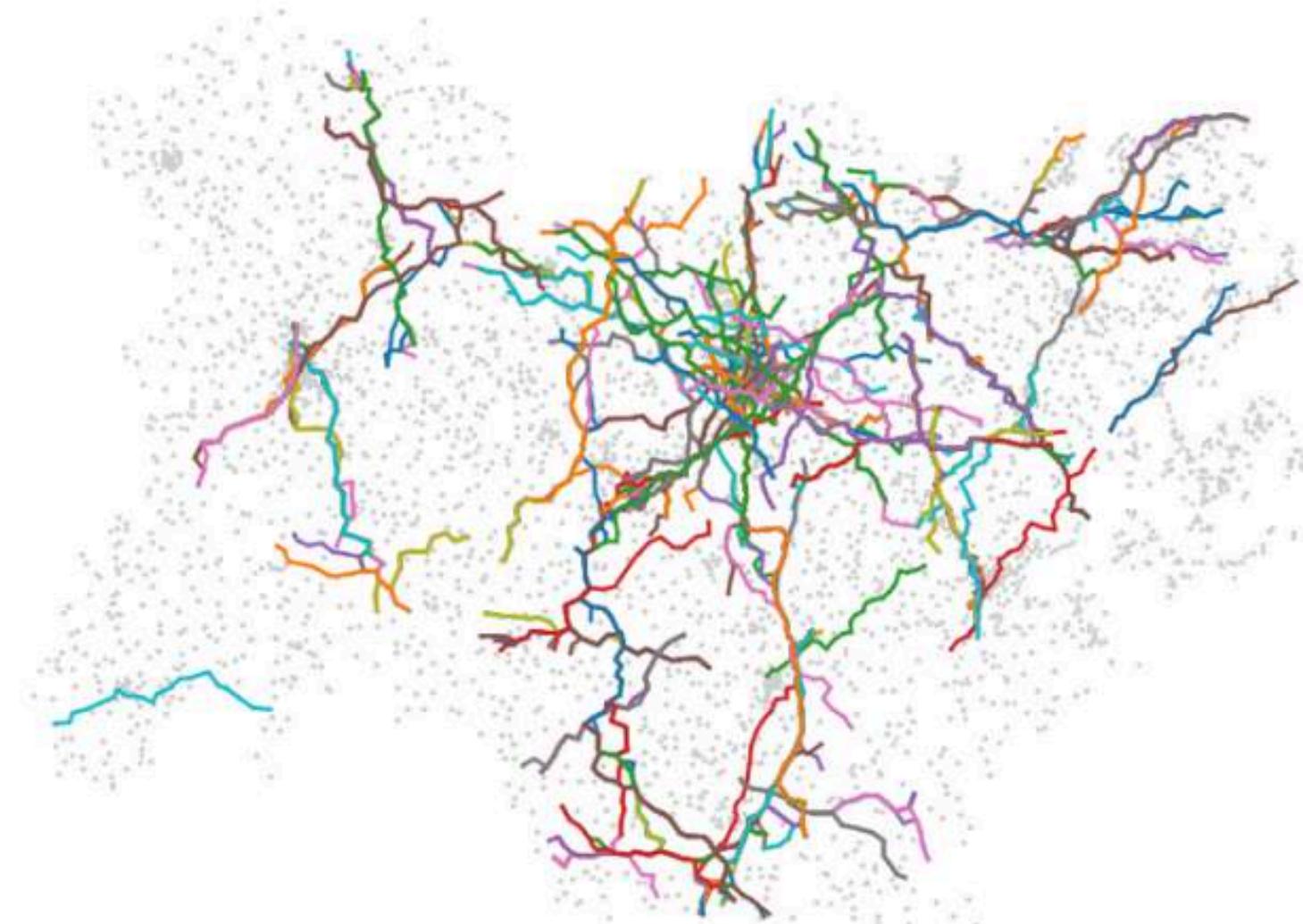
# Région Auvergne-Rhône-Alpes



Avant :

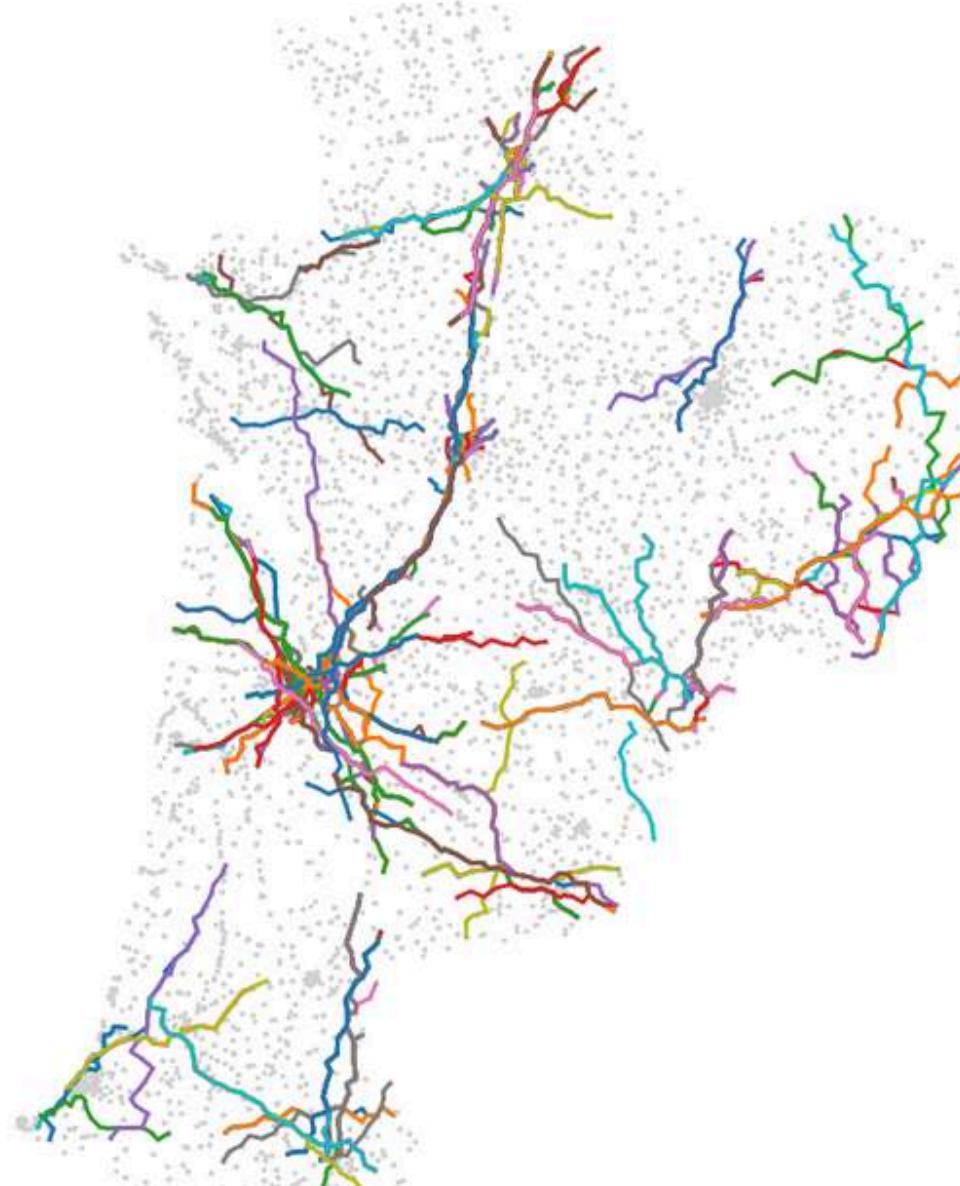


Après :

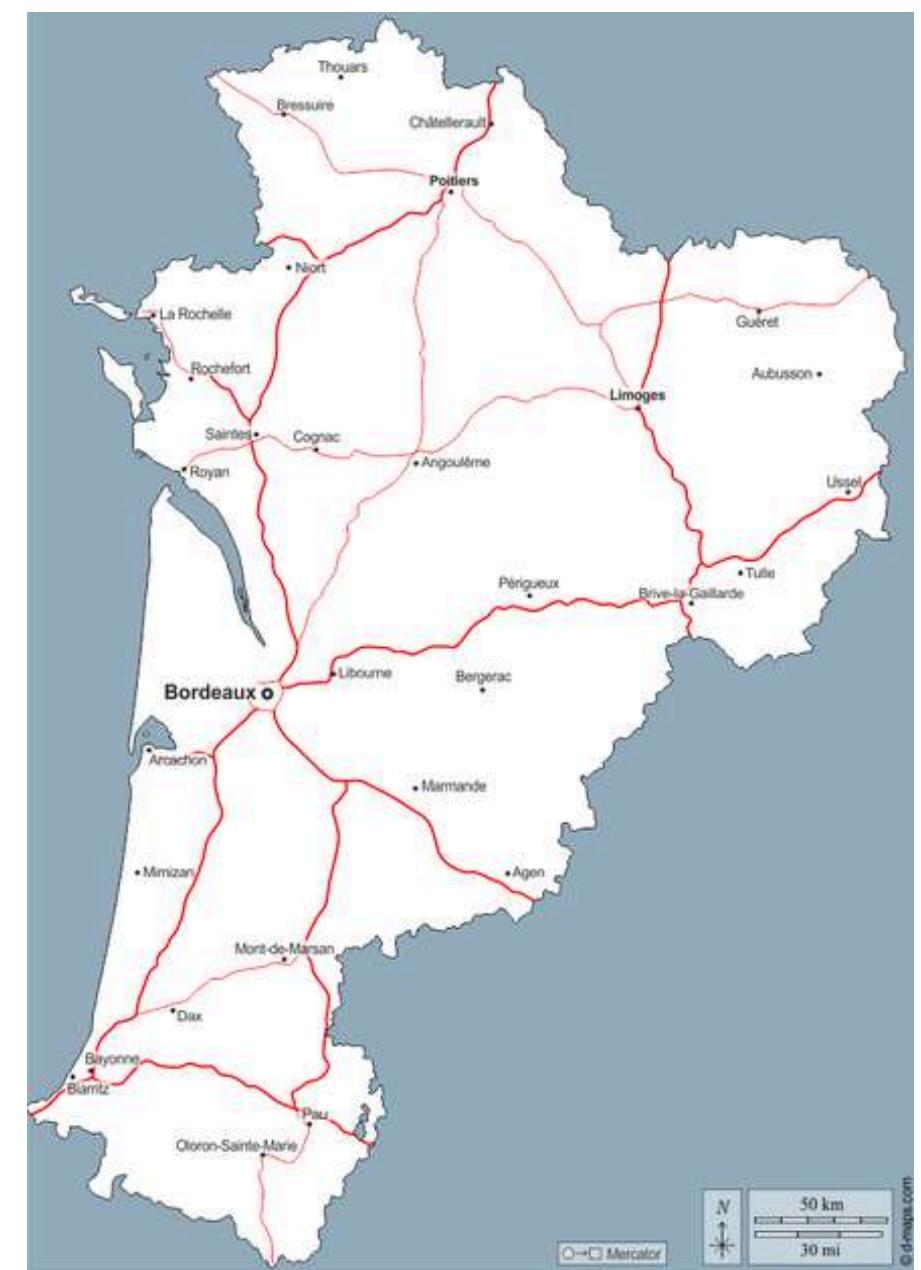


# Région Nouvelle Aquitaine

Avant :

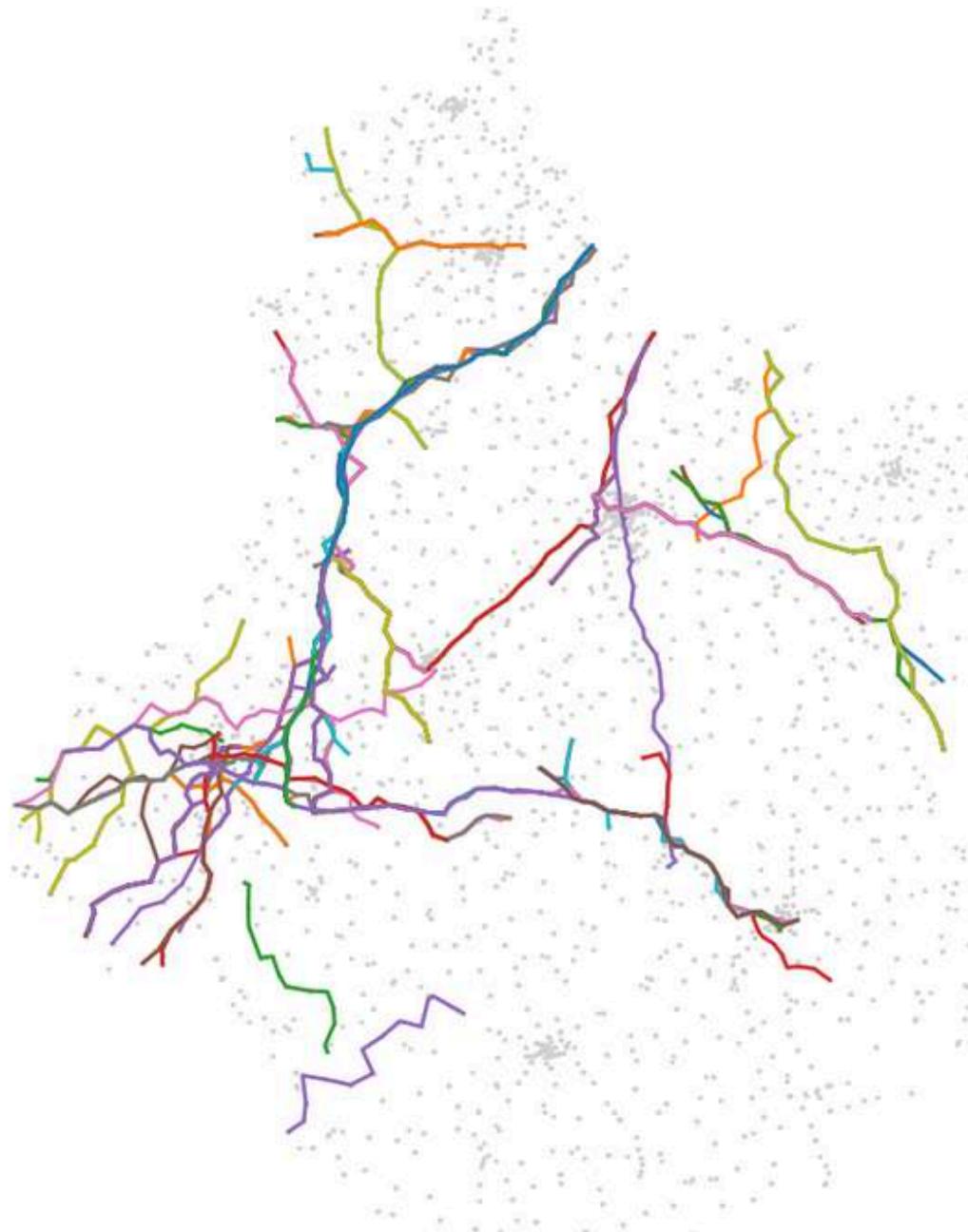


Après :

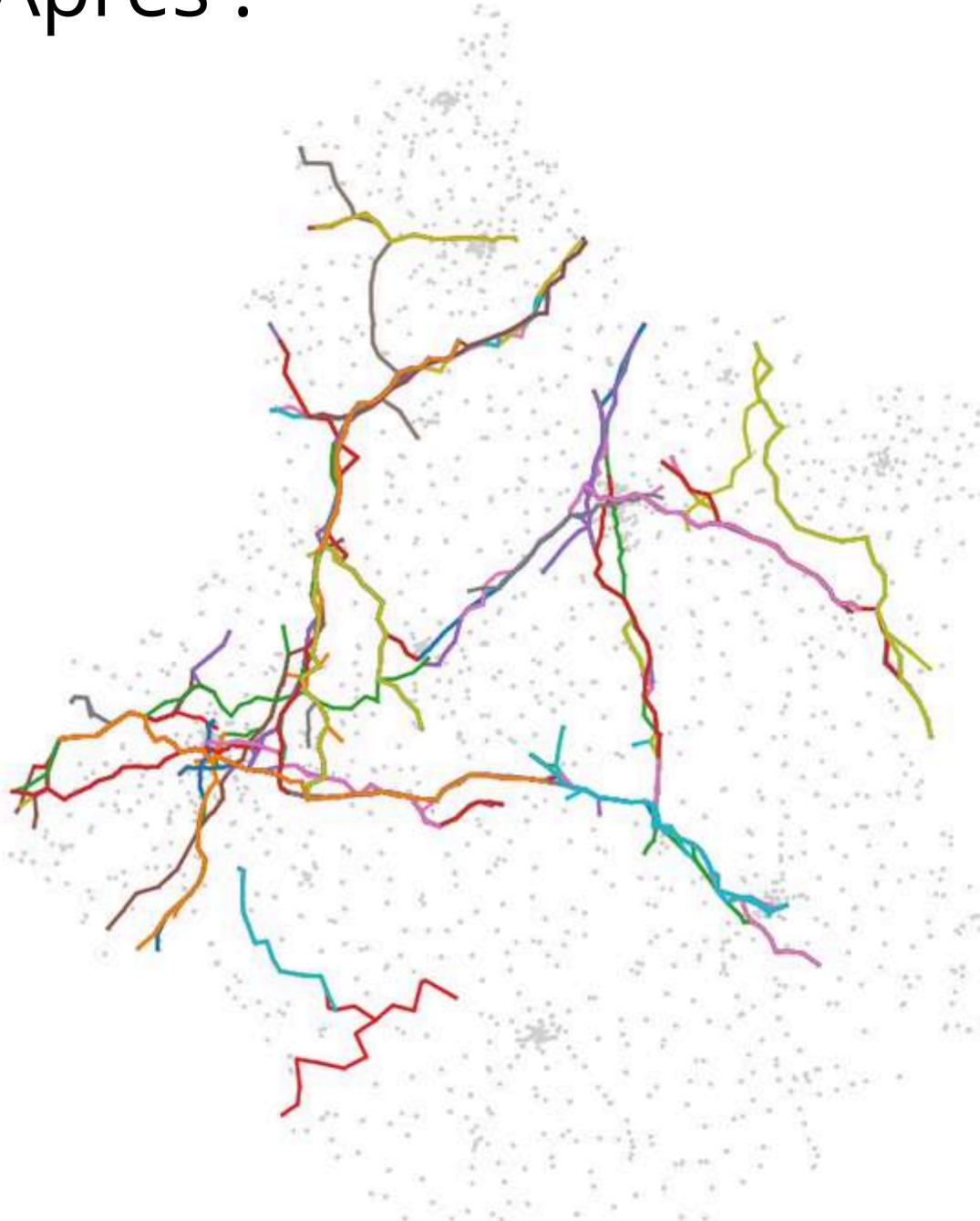


# Région Centre-Val de Loire

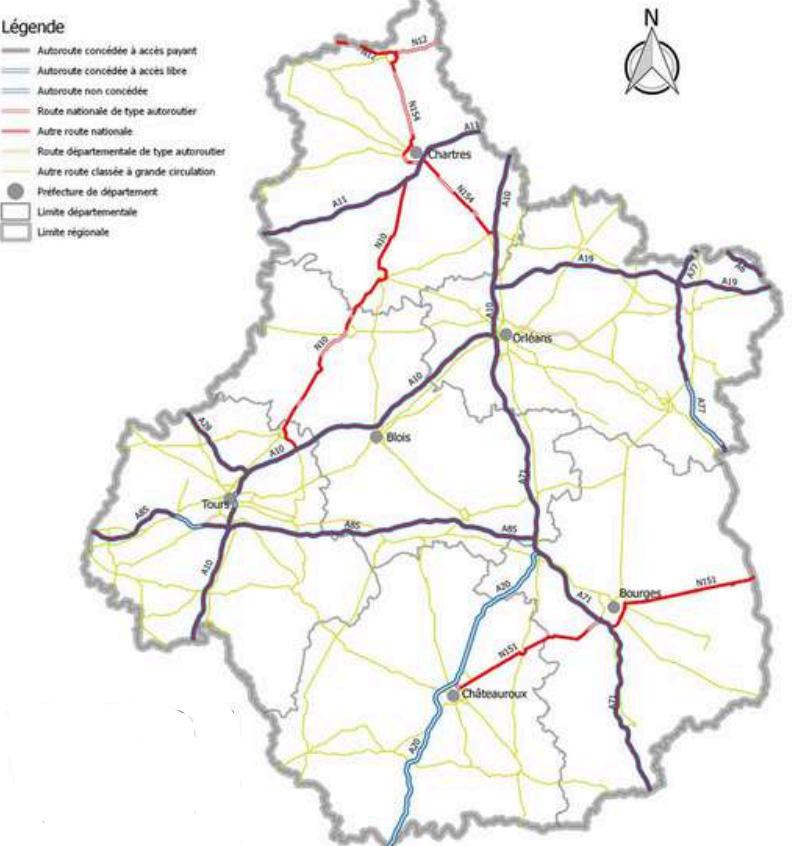
Avant :



Après :



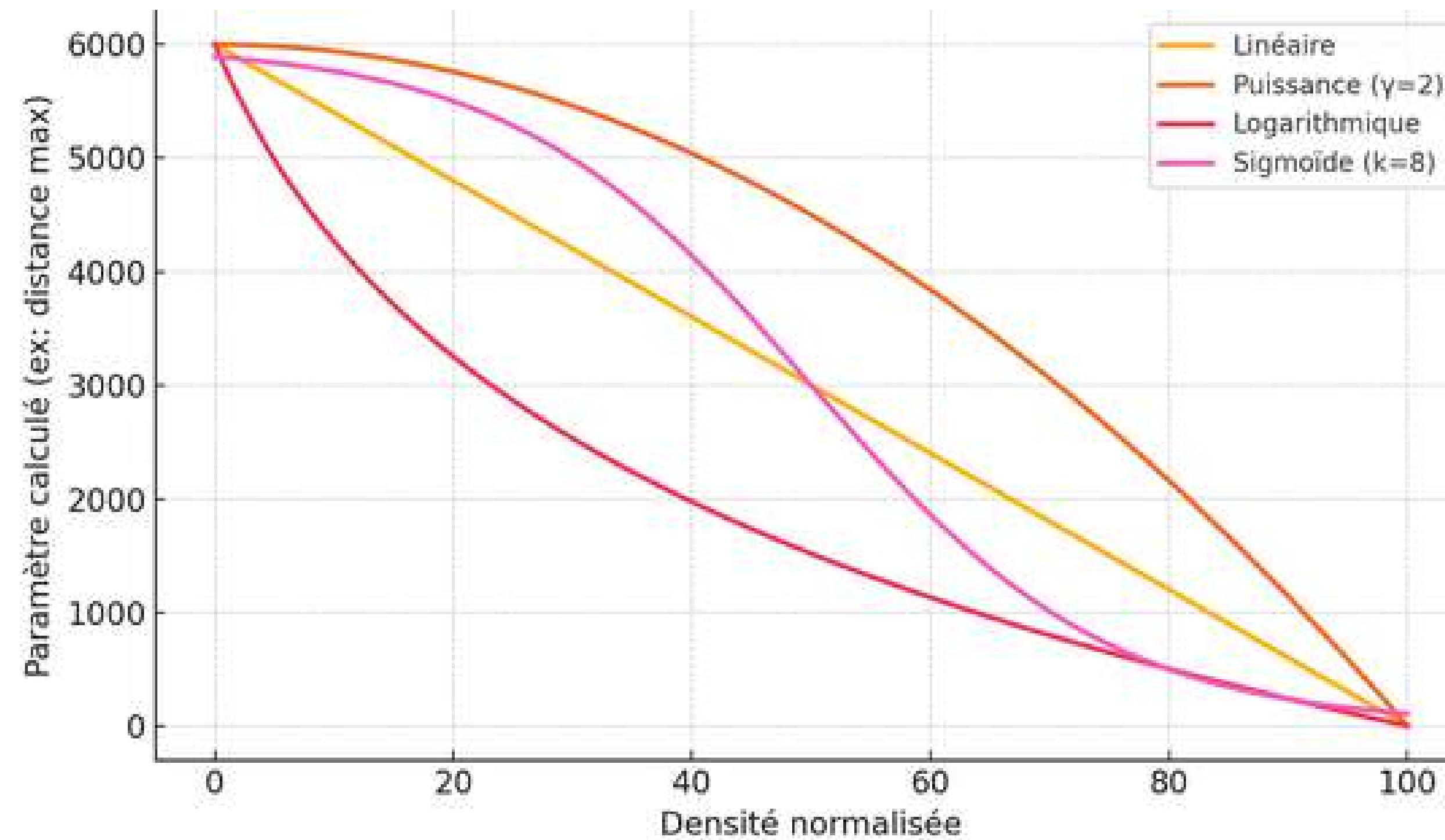
Routes à grande circulation en région Centre-Val de Loire



# **Paramètres selon la densité**

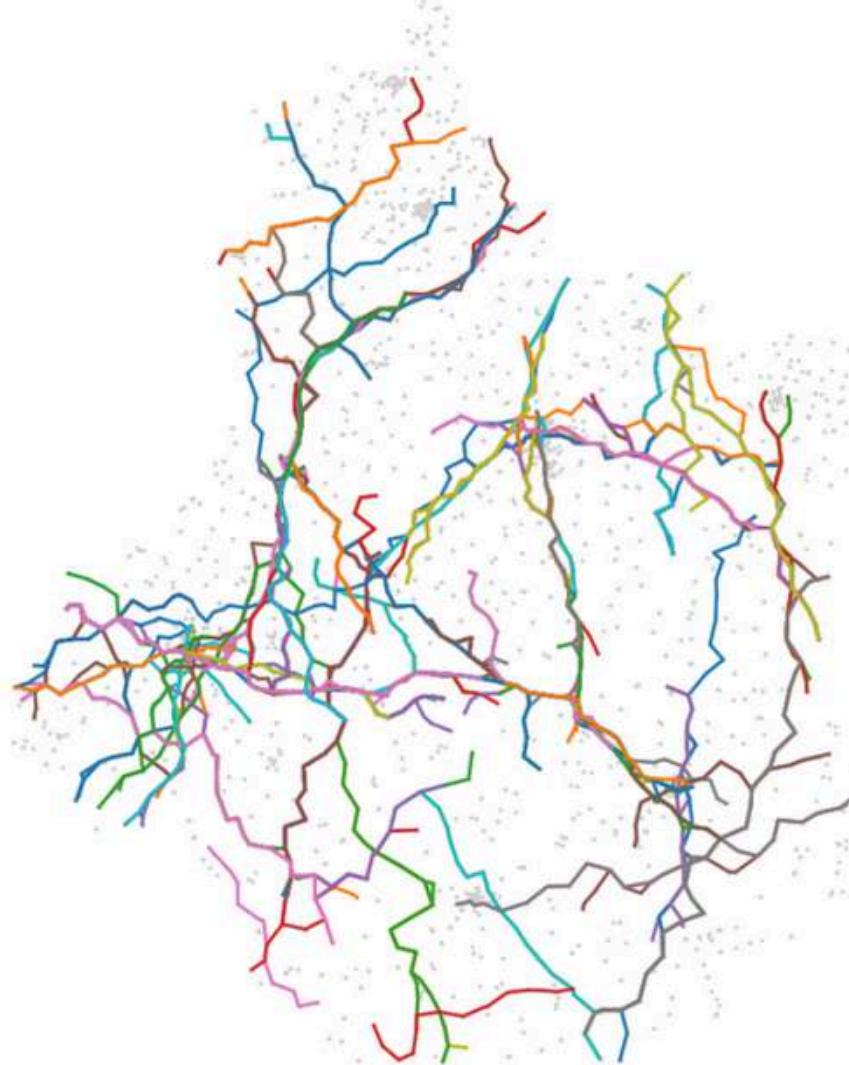
Non linéaire

# Adaptation des paramètres selon la densité

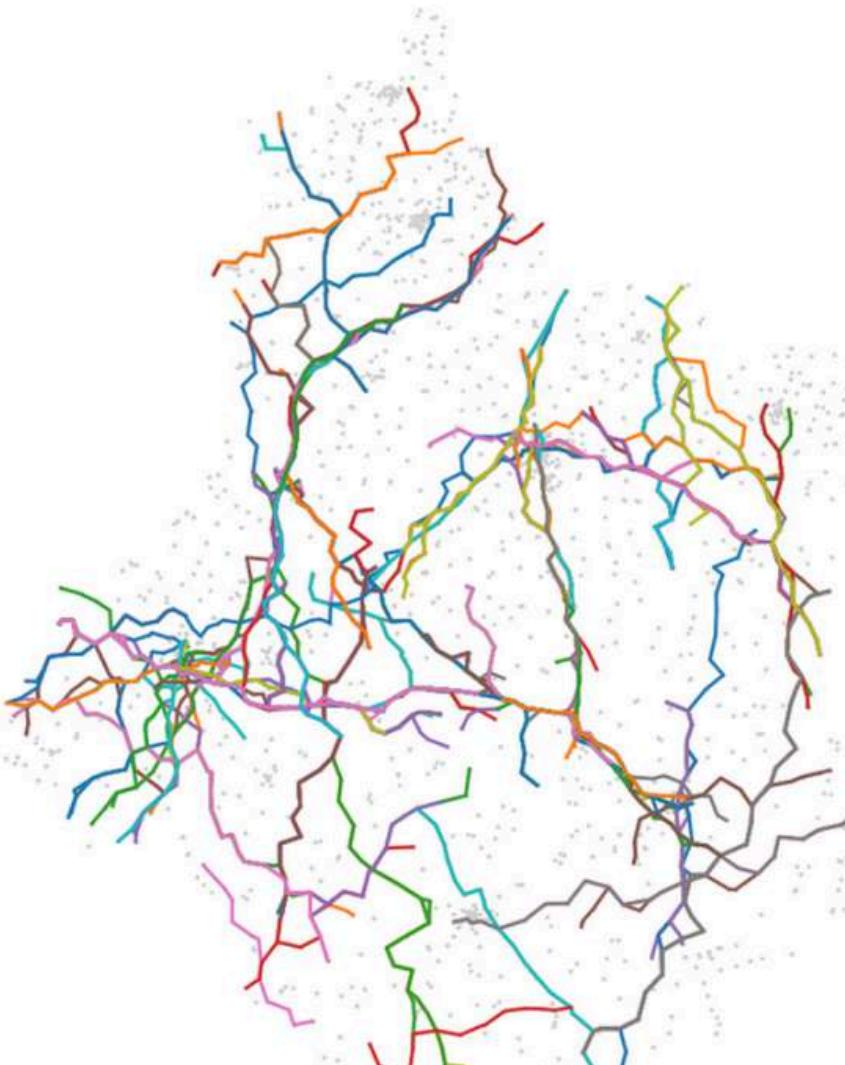


# Adaptation des paramètres selon la densité Centre-Val de Loire

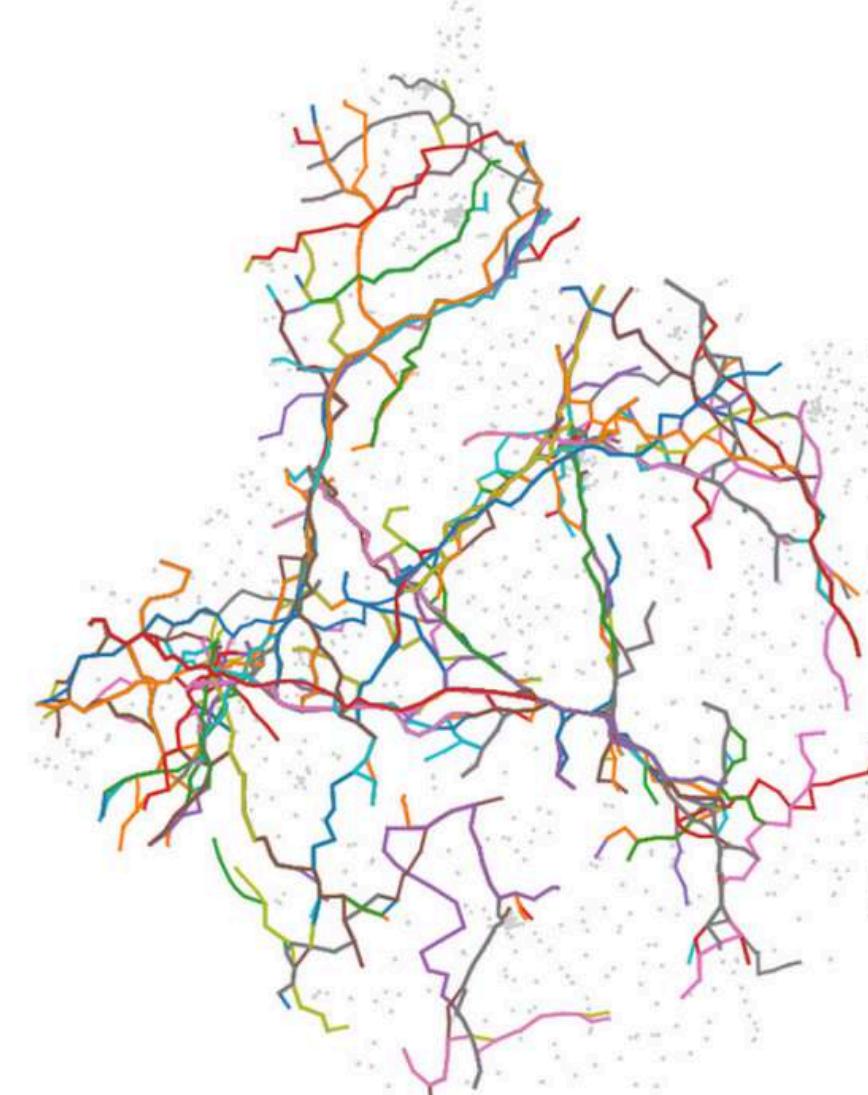
Linéaire



Puissance



Logarithme

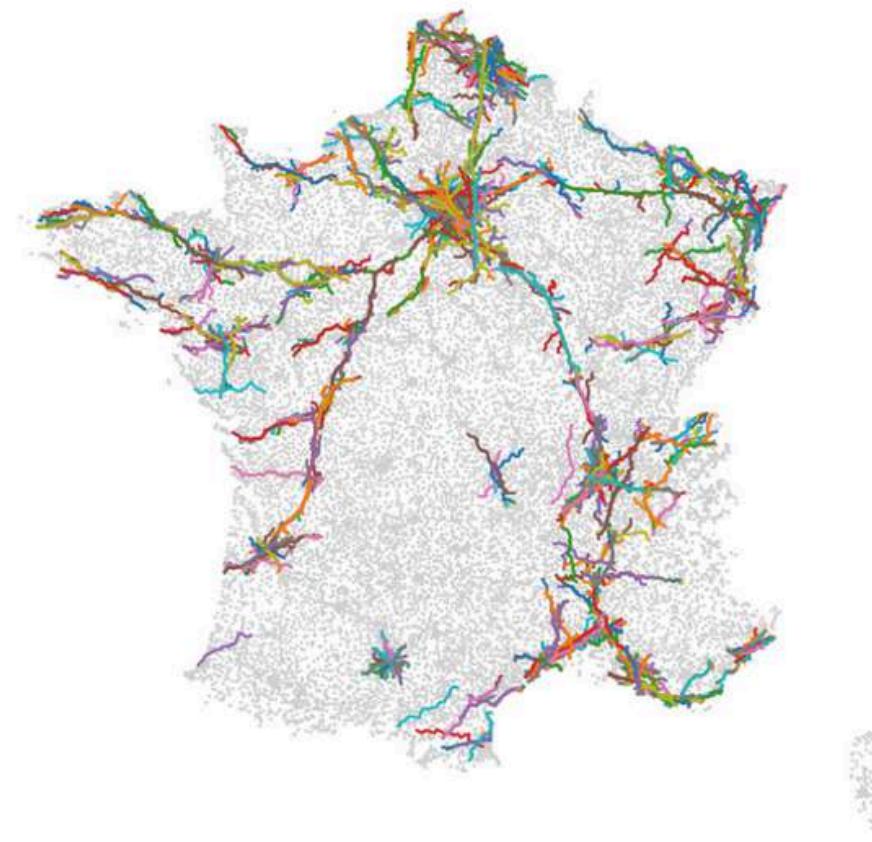


Sigmoïde

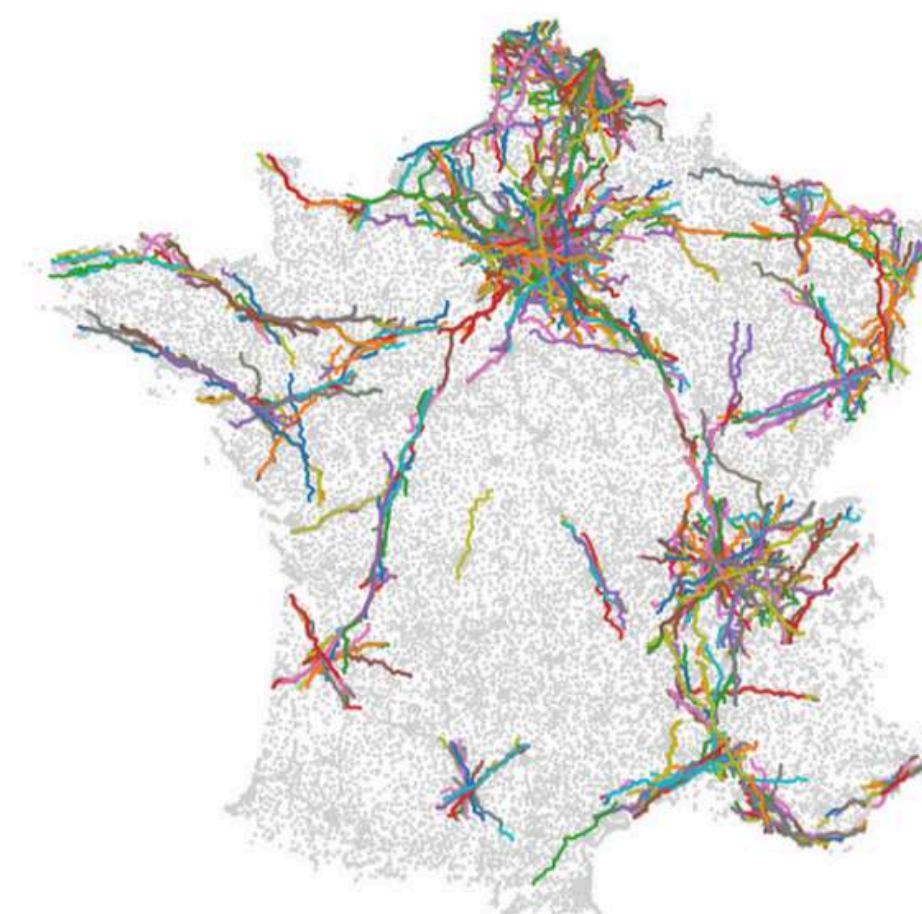


# Adaptation des paramètres selon la densité France

Linéaire :

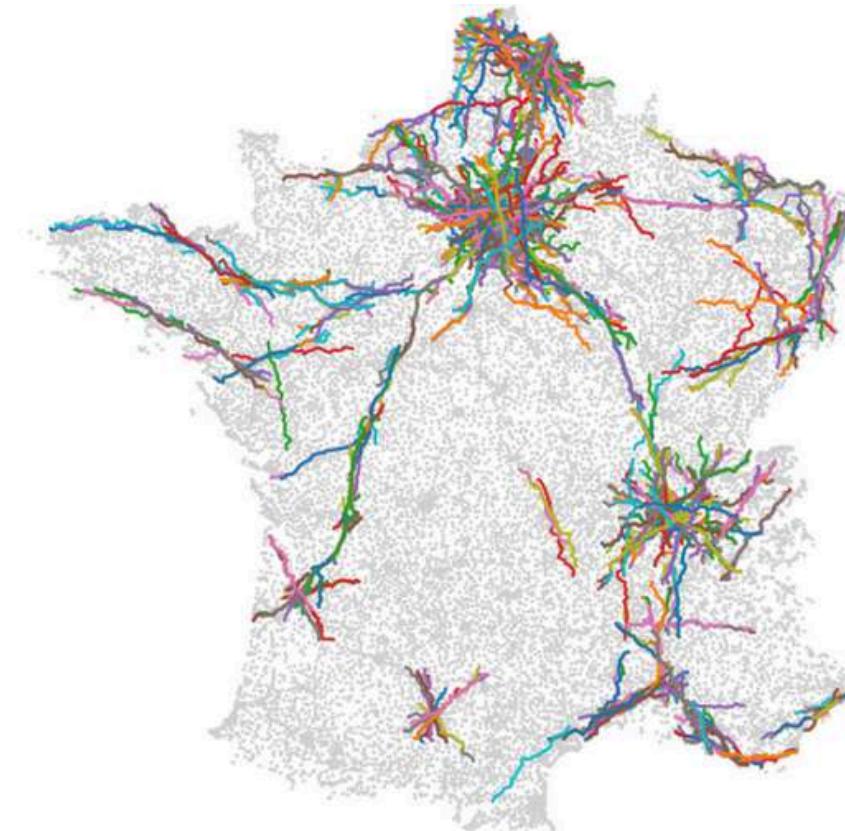


Puissance :



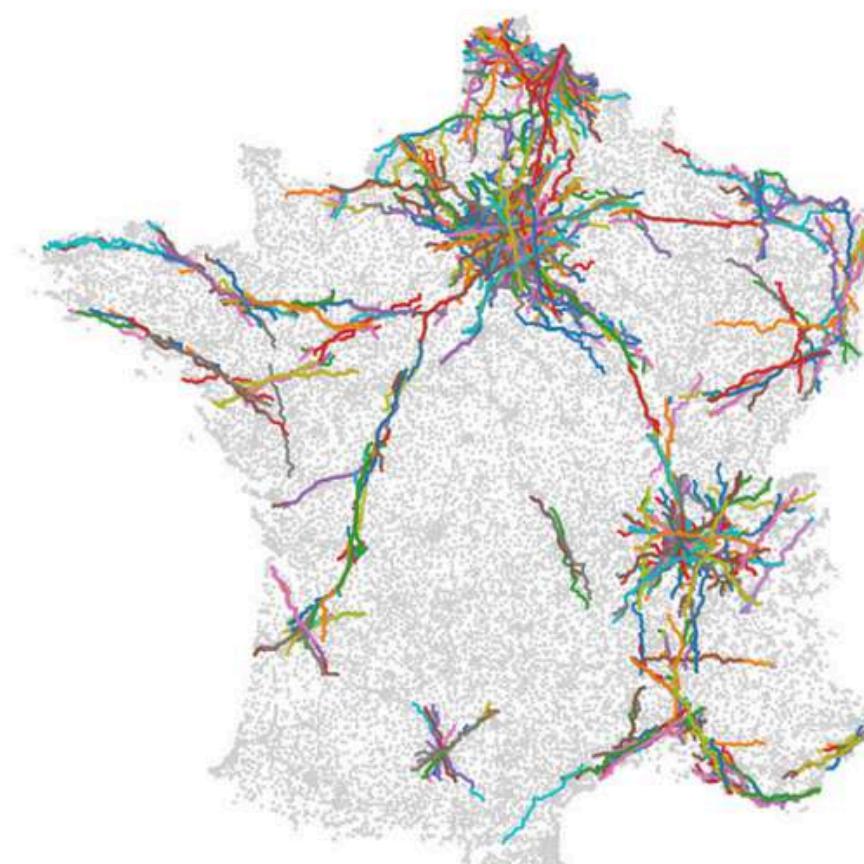
3495 routes

Logarithme :



3165 routes

Sigmoïde :



3205 routes

**15 - 20 Août**

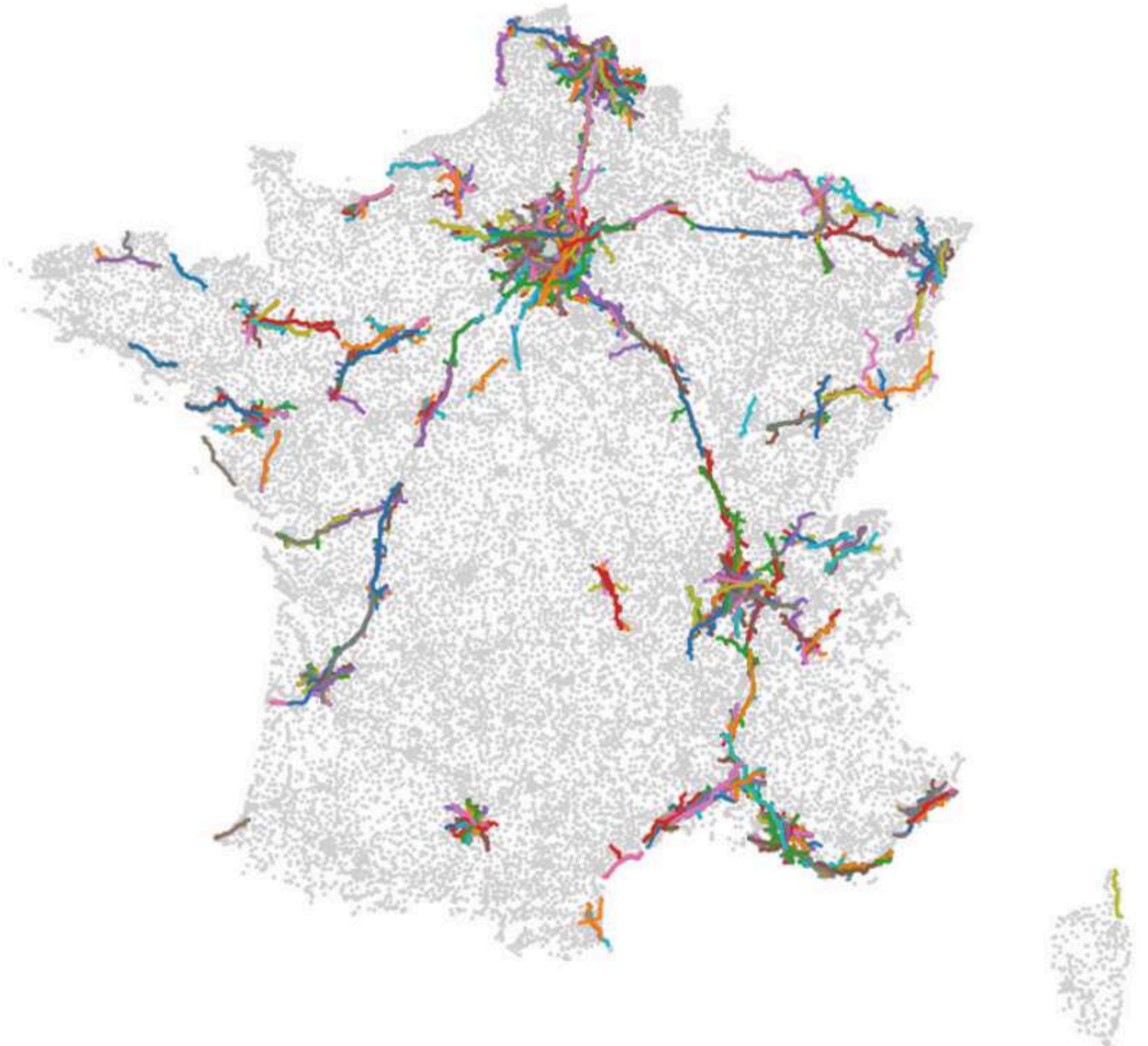
# **Correction d'une erreur**

# Correction d'erreur

- But du calcul des paramètres de distance et d'angle selon la densité : chaque station a ses paramètres adaptés.
- Problème jusqu'ici : Utilisation des paramètres de la 1ère station de la chaîne pour toute la chaîne...
- Correction pour qu'au long de la construction de la chaîne on utilise les paramètres adaptés à la station en cours

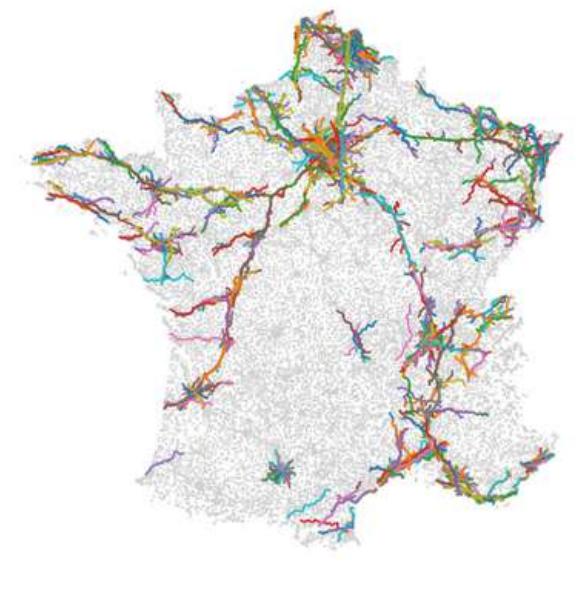
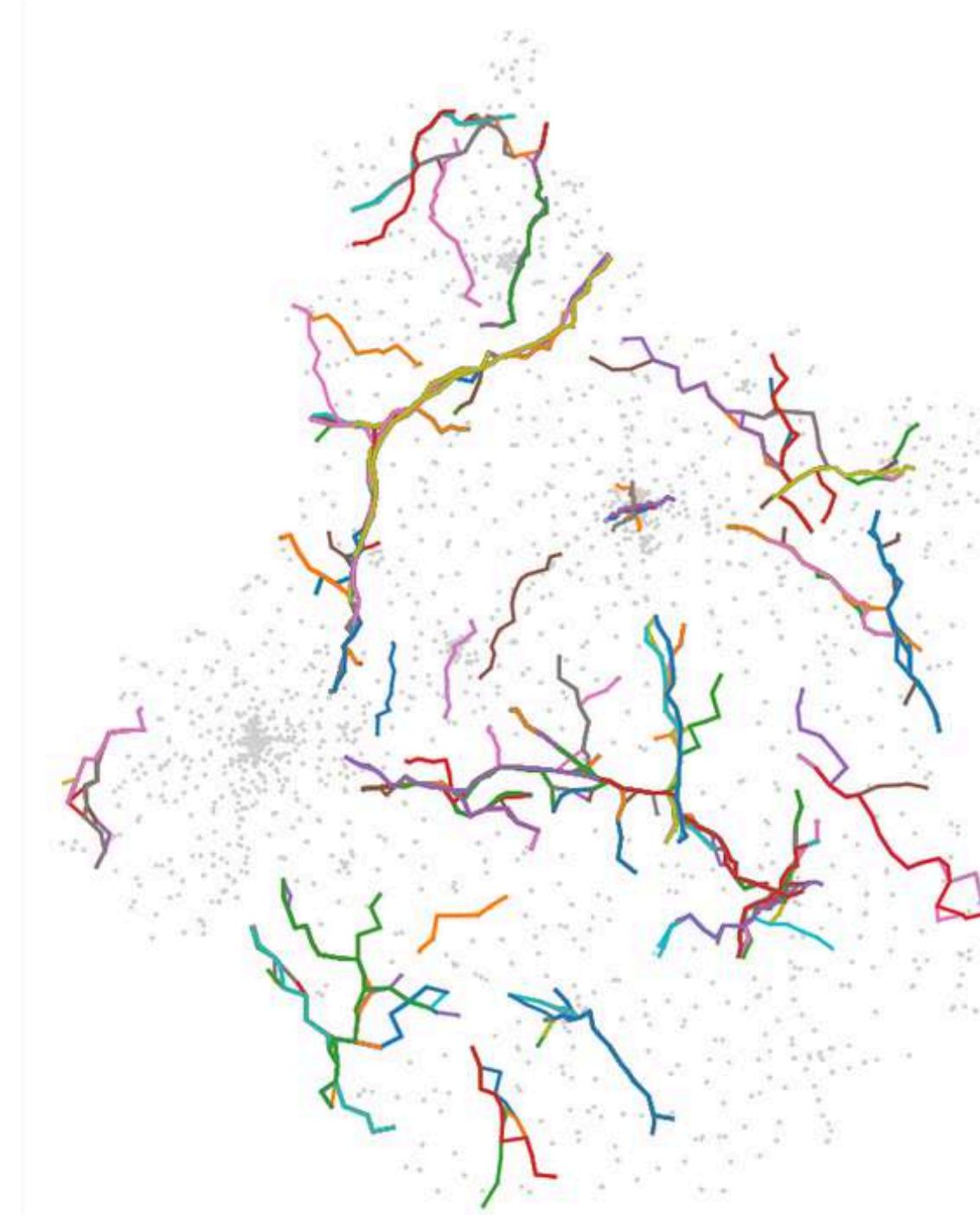
# Paramètres optimaux tout opérateur

- Paramètres trouvés :
  - $rayon = 16\ 000$
  - $n\_neighbors = 10$
  - $dist\_min = 50$
  - $dist\_max = 4500$
  - $angle\_min = 50$
  - $angle\_max = 85$
  - $n\_directions = 5$
  - $angle\_tol = 5$
  - $min\_len = 20$
- Erreur globale : 33 238 km  
(7%)



# Centre-Val de Loire

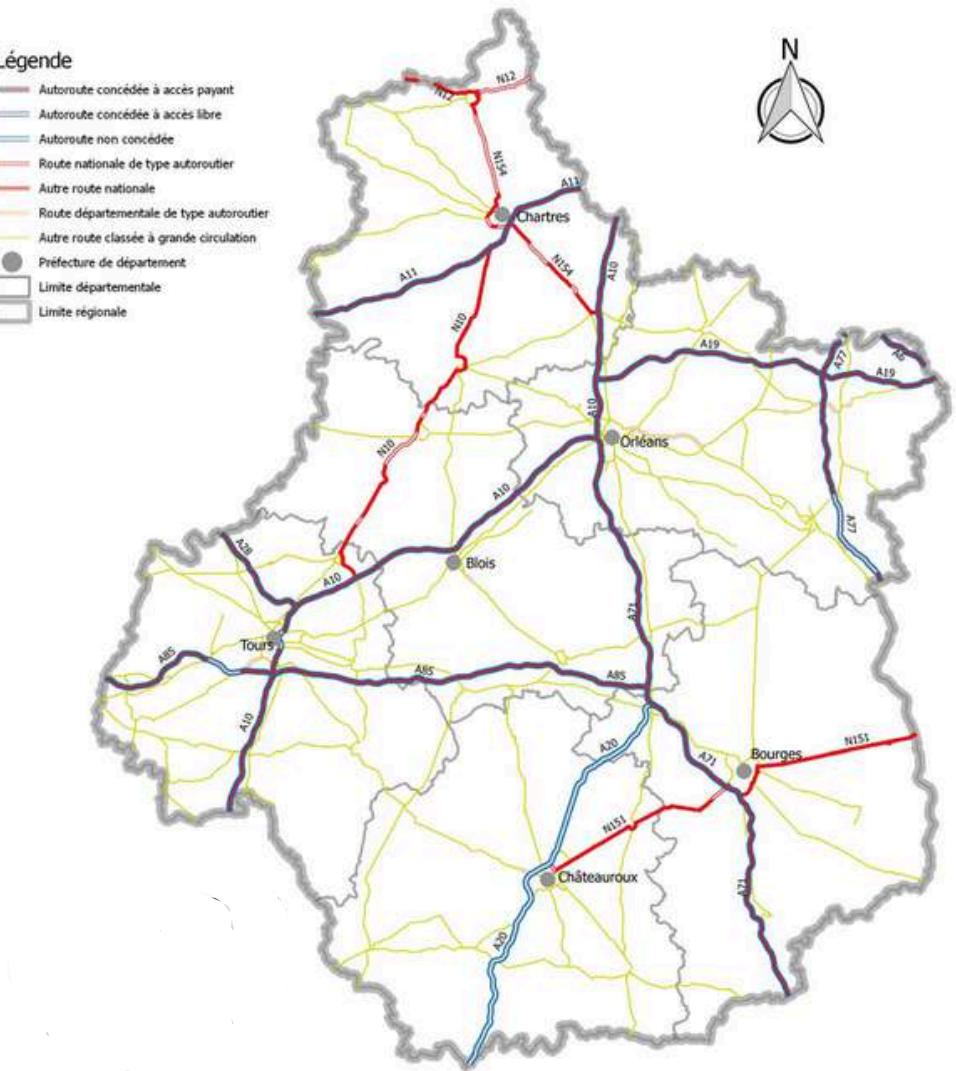
- rayon = 25 000
- n\_neighbors = 15
- dist\_min = 100
- dist\_max = 7000
- angle\_min = 75
- angle\_max = 85
- n\_directions = 4
- angle\_tol = 0
- min\_len = 20



**Routes à grande circulation en région Centre-Val de Loire**

Légende

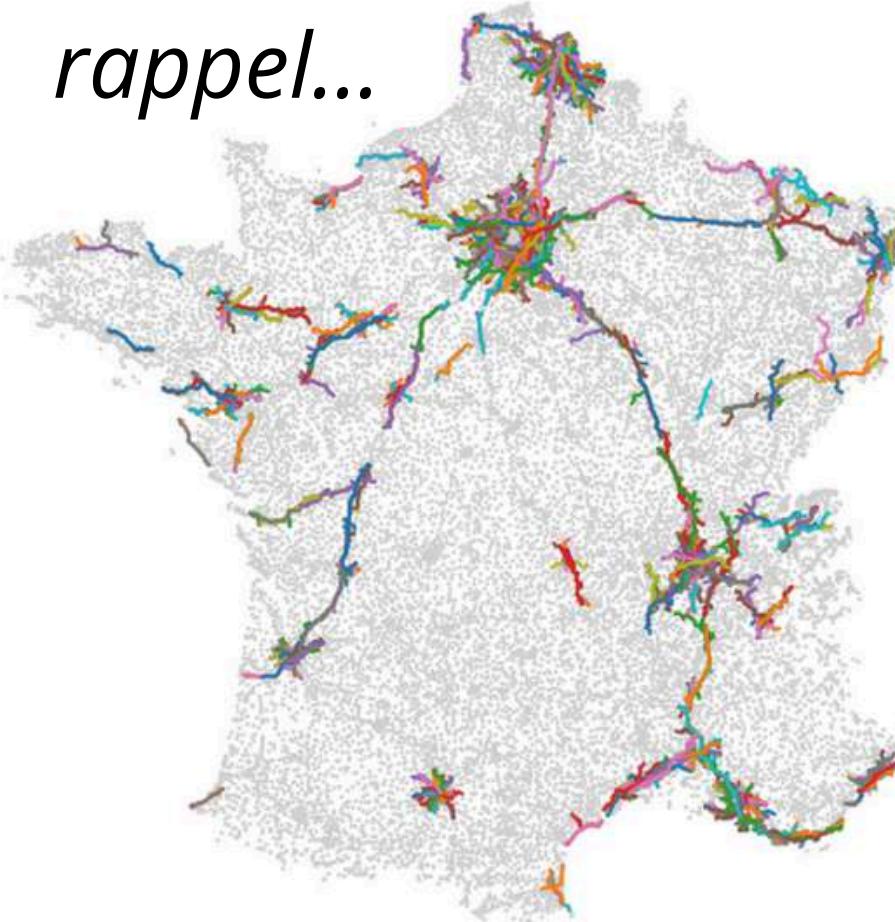
- Autoroute concédée à accès payant
- Autoroute concédée à accès libre
- Autoroute non concédée
- Route nationale de type autoroutier
- Autre route nationale
- Route départementale de type autoroutier
- Autre route classée à grande circulation
- Prefecture de département
- Limite départementale
- Limite régionale



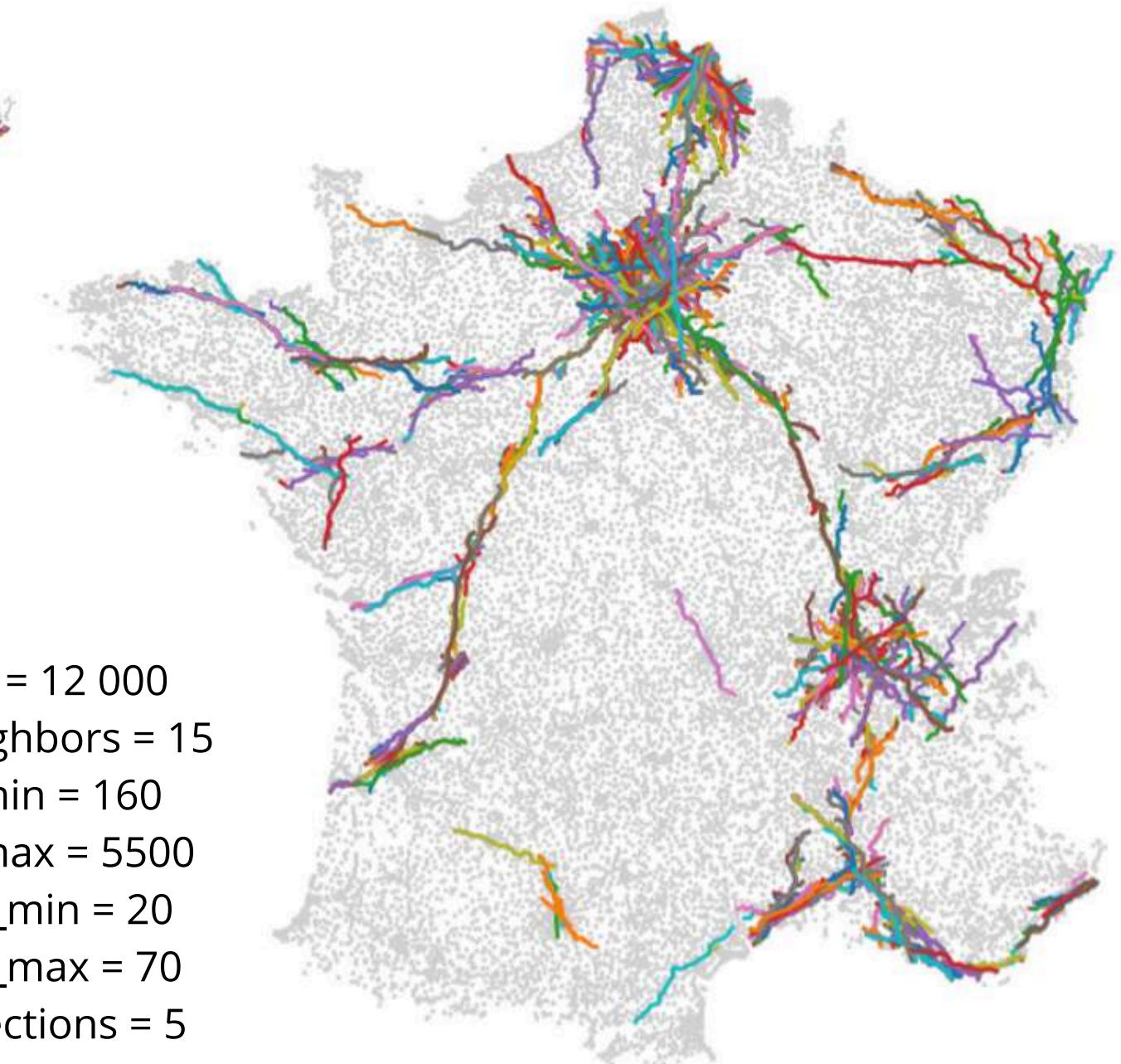
**21 - 26 Août**

# **Optimisation après correction d'erreur**

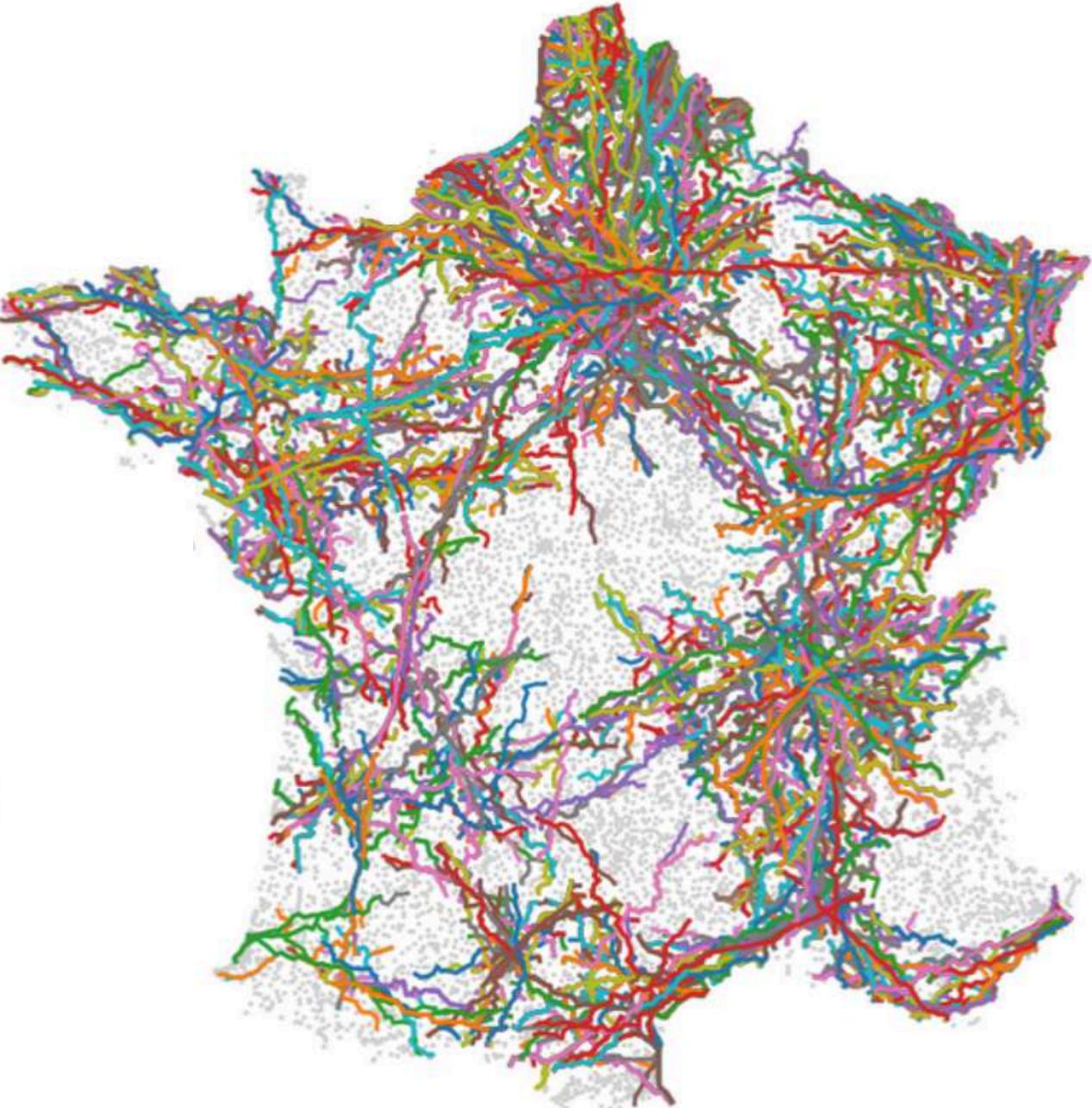
*rappel...*



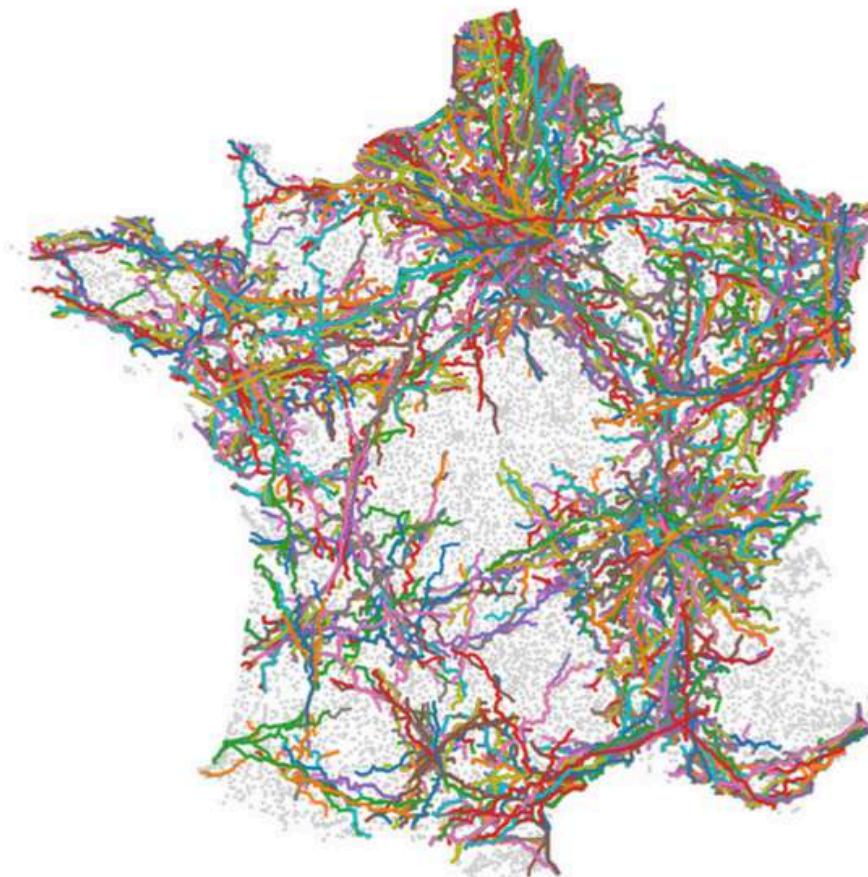
- rayon = 10 000
- n\_neighbors = 15
- dist\_min = 1
- dist\_max = 7000
- angle\_min = 25
- angle\_max = 85
- n\_directions = 5
- angle\_tol = 0
- min\_len = 35



- rayon = 12 000
- n\_neighbors = 15
- dist\_min = 160
- dist\_max = 5500
- angle\_min = 20
- angle\_max = 70
- n\_directions = 5
- angle\_tol = 0
- min\_len = 30



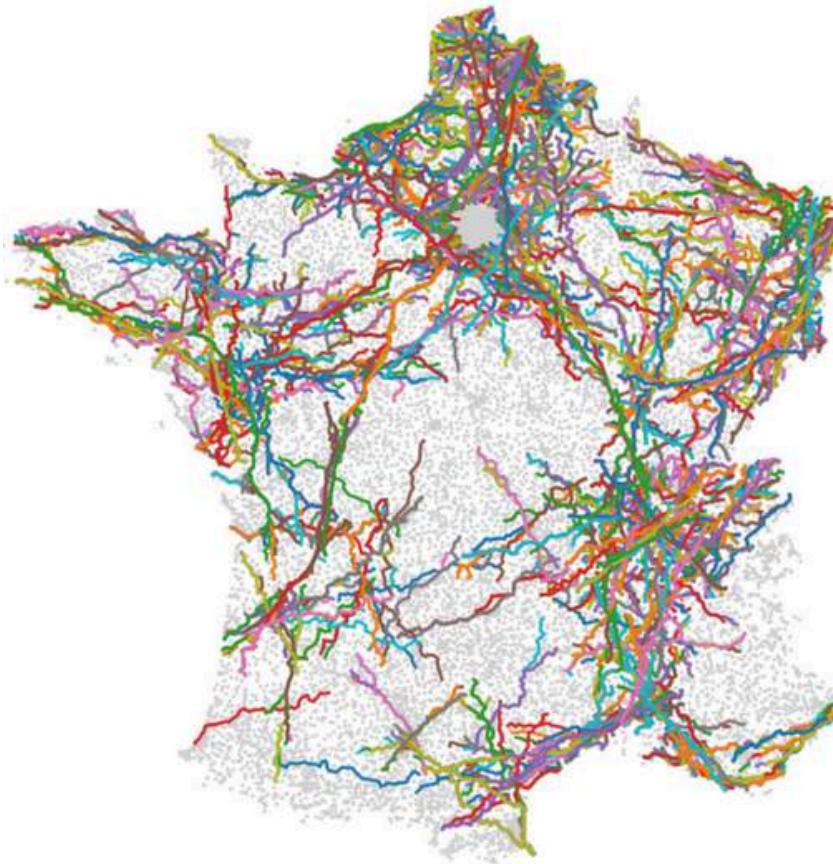
- n\_neighbors = 15
- dist\_min = 1
- dist\_max = 7000
- angle\_min = 25
- angle\_max = 85
- n\_directions = 4
- angle\_tol = 0
- min\_len = 40



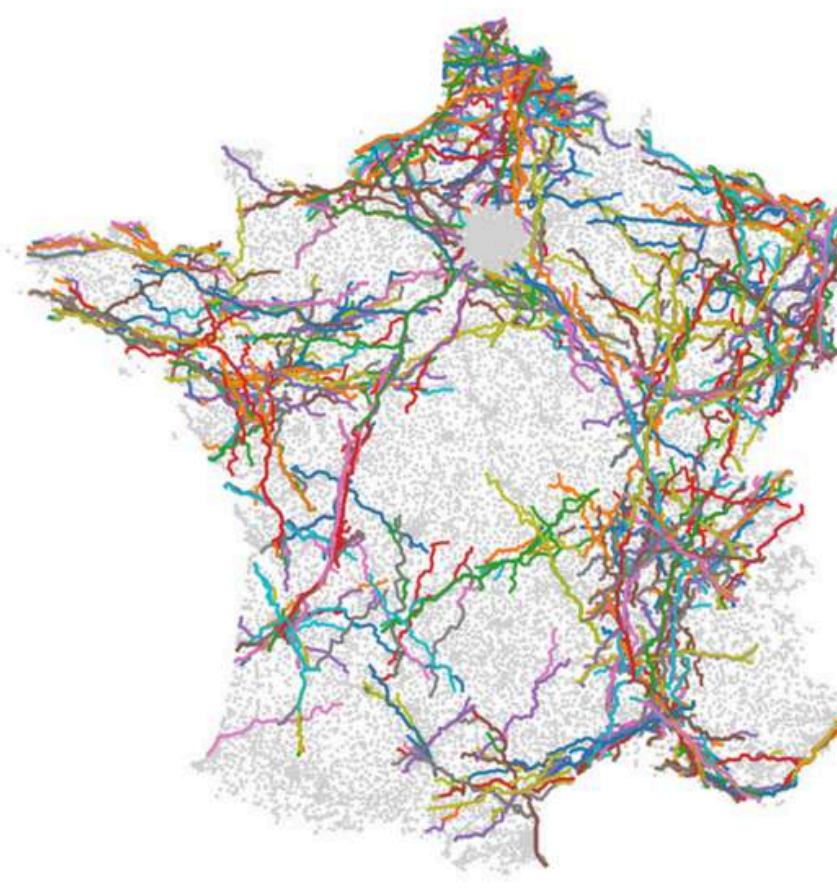
rayon = 10 000



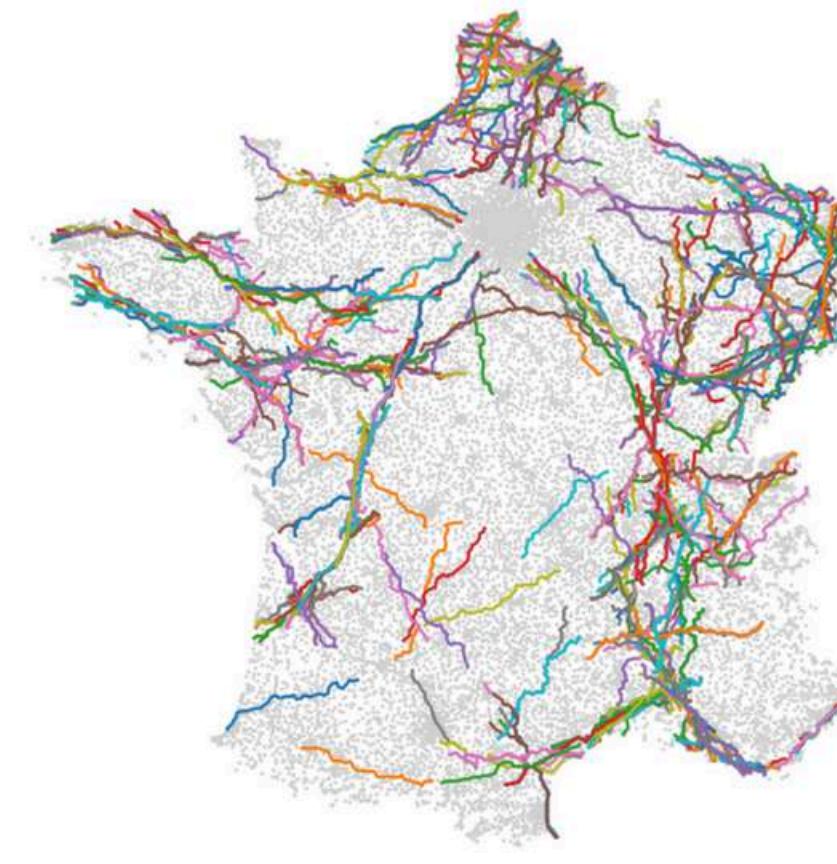
rayon = 20 000



rayon = 30 000



rayon = 40 000

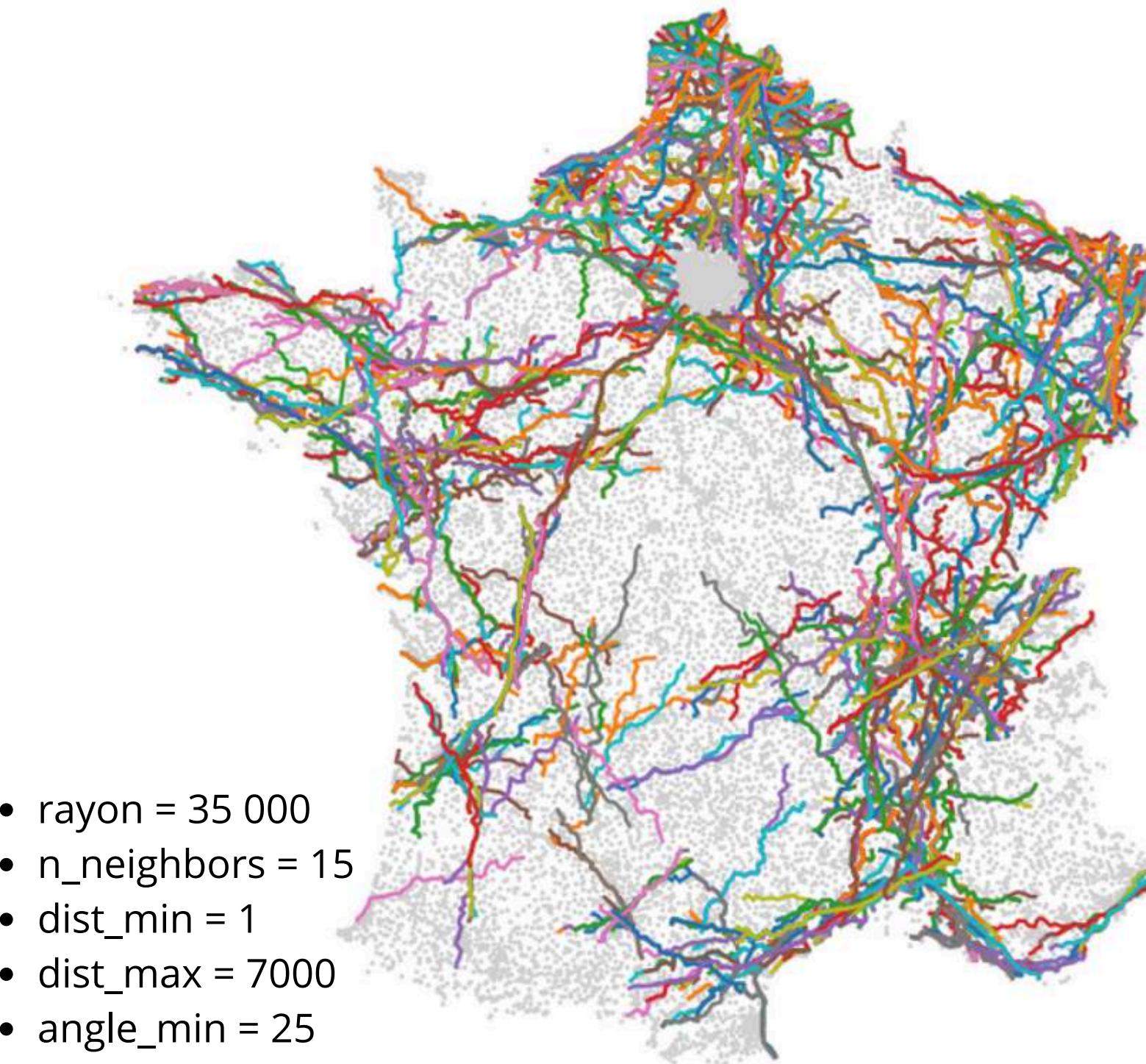


rayon = 50 000

# **Indicateurs**

# Indicateurs

## Suivi de chemin selon densité de stations, Orange



- rayon = 35 000
- n\_neighbors = 15
- dist\_min = 1
- dist\_max = 7000
- angle\_min = 25
- angle\_max = 85
- n\_directions = 5
- angle\_tol = 0
- min\_len = 35

- Évaluation globale :  
Longueur détectée : 735945.5 km  
Longueur réelle (référence) : 432680.0 km  
Erreur absolue : 303265.5 km  
Erreur relative : 70.09%

- Évaluation par département :

Département	Réel (km)	Détecté (km)	Erreur (km)	Erreur relative (%)
65 Ardèche	238	17357.2	17119.2	71.929254
92 Hérault	464	19571.1	19107.1	41.179143
9 Haut-Rhin	261	10762.9	10501.9	40.237241
87 Somme	719	27396.4	26677.4	37.103541
28 Vaucluse	362	13230.7	12868.7	35.548998

	Erreur relative (%)
65	7192.9%
92	4117.9%
9	4023.7%
87	3710.4%
28	3554.9%

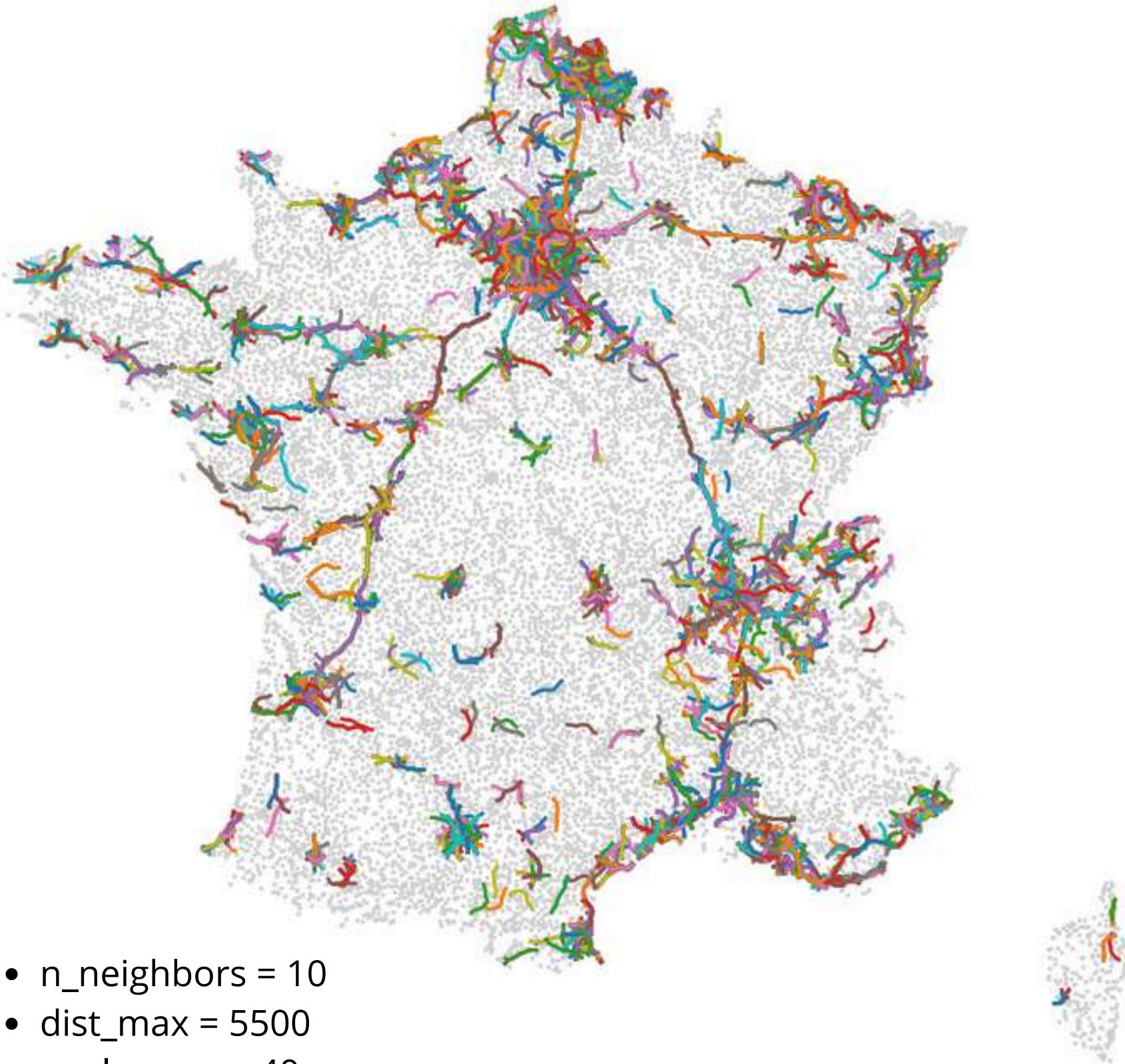
Écart-type des erreurs départementales : 12.785  
Max erreur relative d'un département : 7192.93%  
Nombre de départements à forte erreur (>50%) : 86  
Départements non détectés : Cher, Corse-du-Sud, Haute-Corse, Hautes-Pyrénées, Hauts-de-Seine, Paris, Seine-Saint-Denis, Val-de-Marne

- Taux de redondance (chaines quasi-identiques, >90%) : 8.91%
- Taux de doublons (Proportion de segments dans plusieurs chaines) : 67.94%
- Taux de doublons (Proportion de points dans plusieurs chaines) : 82.90%
- Proportion de points non utilisés : 45.47%



# Indicateurs

Triplets alignés amélioré, Orange



- n\_neighbors = 10
- dist\_max = 5500
- angle\_max = 40
- min\_len = 9