



Prévision du rendement

éolien chez



Komlan Katakou

&

Jana Taleb

2024 - 2025

Pourquoi ce projet ?



Contexte

La transition énergétique



Pourquoi l'éolien ?

Ressource abondante, renouvelable et bien répartie sur le territoire français.



Objectif

Prédire l'indice de rendement éolien à partir des données météorologiques et géographiques sur 10 zones climatiques

Régime de fonctionnement des éoliennes

1

Démarrage

En dessous de la vitesse de démarrage, pas de production ; au-delà, la puissance augmente rapidement avec la vitesse du vent.

2

Plein régime

Puissance nominale atteinte (à $V \approx 12$ m/s) : production maximale constante au-delà de cette vitesse jusqu'à la vitesse d'arrêt.

3

Arrêt

Au-delà de la vitesse d'arrêt, l'éolienne se met en arrêt de sécurité pour éviter les dommages.

Deux indicateurs techniques

Puissance installée (p_i)

- **Définition** : puissance maximale que produirait l'installation dans des conditions de vent idéales
- **Rôle** : sert de référence technique pour évaluer la capacité du parc, indépendamment des variations météo.
- **Remarque** : il s'agit d'une valeur théorique, différente de la production réelle observée.

Facteur de charge (FC)

- **Définition** : ratio entre la production réelle et la production maximale théorique

$$FC = e_{\text{olien}} / p_i$$

- **Utilité** : indicateur clé pour évaluer le rendement des éoliennes, utilisé par RTE et d'autres organismes énergétiques.
- **Interprétation** : reflète le rendement réel de l'éolienne et l'adéquation entre l'équipement et le vent local.

Le plan

1

Les données

2

Etude descriptive

3

La démarche de RTE

4

Notre apport

5

Conclusion et Perspectives

Les données

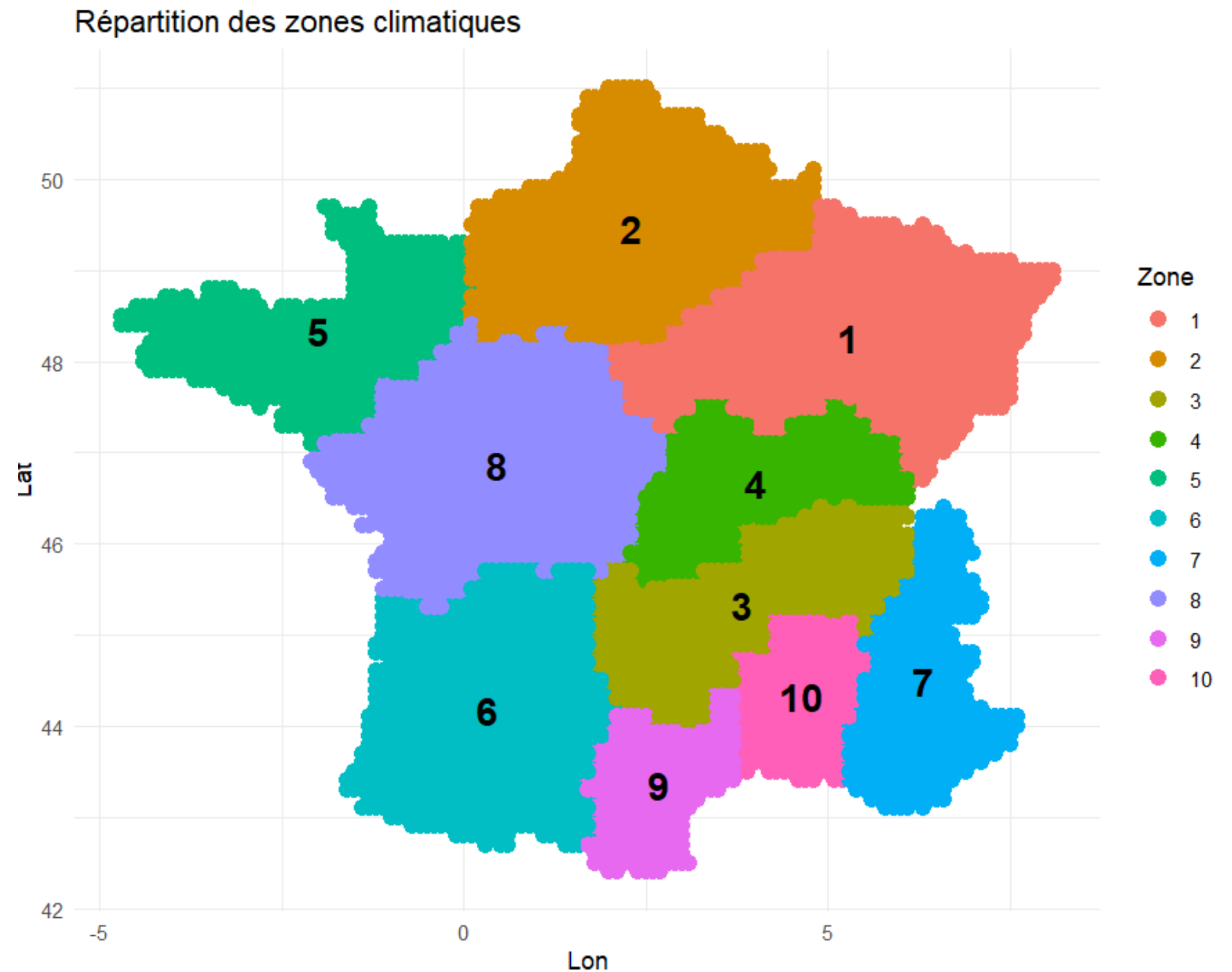
1

Données cartographiques

Grille de coordonnées géographiques (Longitude, Latitude) de résolution 6351 x 6351 points et
Classification en dix zones climatiques



ZC_grille_01_elargie.csv



Les données

2

Données de production

Quantité d'énergie
électrique produite par
l'éolien

eCO2mix_RTE_Annuel-
Definitif_2017.xls, ...



C	D	E	F	G	H	I	J	K	L
Date	Heures	Consommation	Prévision J-1	Prévision J	Fioul	Charbon	Gaz	Nucléaire	Eolien
2017-01-01	00:00	76259	76100	76500	164	1579	7617	53528	267
2017-01-01	00:15		75500	75800					
2017-01-01	00:30	74981	74900	75100	601	1506	6626	53456	311
2017-01-01	00:45		74150	74250					
2017-01-01	01:00	72912	73400	73400	596	1291	6593	53431	351
2017-01-01	01:15		73650	73650					
2017-01-01	01:30	72682	73900	73900	587	1028	6595	53429	367
2017-01-01	01:45		73600	73600					
2017-01-01	02:00	72166	73300	73300	599	1024	6548	53381	360
2017-01-01	02:15		72350	72550					
2017-01-01	02:30	71790	71400	71800	634	1056	6404	53322	354
2017-01-01	02:45		70450	70750					
2017-01-01	03:00	69709	69500	69700	629	993	6240	53010	340
2017-01-01	03:15		68750	68900					
2017-01-01	03:30	68048	68000	68100	480	1046	5841	53212	341
2017-01-01	03:45		67250	67300					
2017-01-01	04:00	66476	66500	66500	400	937	5625	52940	376
2017-01-01	04:15		65900	66000					
2017-01-01	04:30	65338	65300	65500	339	1026	5684	53239	392
2017-01-01	04:45		65100	65100					
2017-01-01	05:00	64321	64900	64700	179	1014	5676	53200	410
2017-01-01	05:15		65000	64750					
2017-01-01	05:30	64364	65100	64800	194	1046	5719	53264	428
2017-01-01	05:45		65000	64550					

Annuel-Definitif_2017.xls ▾

Les données

3

Données météorologiques

Composantes U et V de la vitesse du vent enregistrées toutes les heures en chaque point d'une grille de coordonnées de résolution 71 x 51 points



wind_2017.nc, ...,
wind_2022.nc

```
2 variables (excluding dimension variables):
  short u100[longitude,latitude,time]
    scale_factor: 0.000829374955909171
    add_offset: 4.05863110414802
    _FillValue: -32767
    missing_value: -32767
    units: m s**-1
    long_name: 100 metre U wind component
  short v100[longitude,latitude,time]
    scale_factor: 0.00080198685220234
    add_offset: -0.96512553413655
    _FillValue: -32767
    missing_value: -32767
    units: m s**-1
    long_name: 100 metre V wind component
```

```
3 dimensions:
  longitude  Size:71
    units: degrees_east
    long_name: longitude
  latitude  Size:51
    units: degrees_north
    long_name: latitude
  time      Size:8760
    units: hours since 1900-01-01 00:00:00.0
    long_name: time
    calendar: gregorian
```

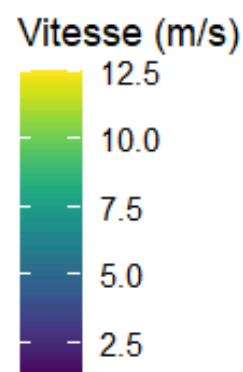
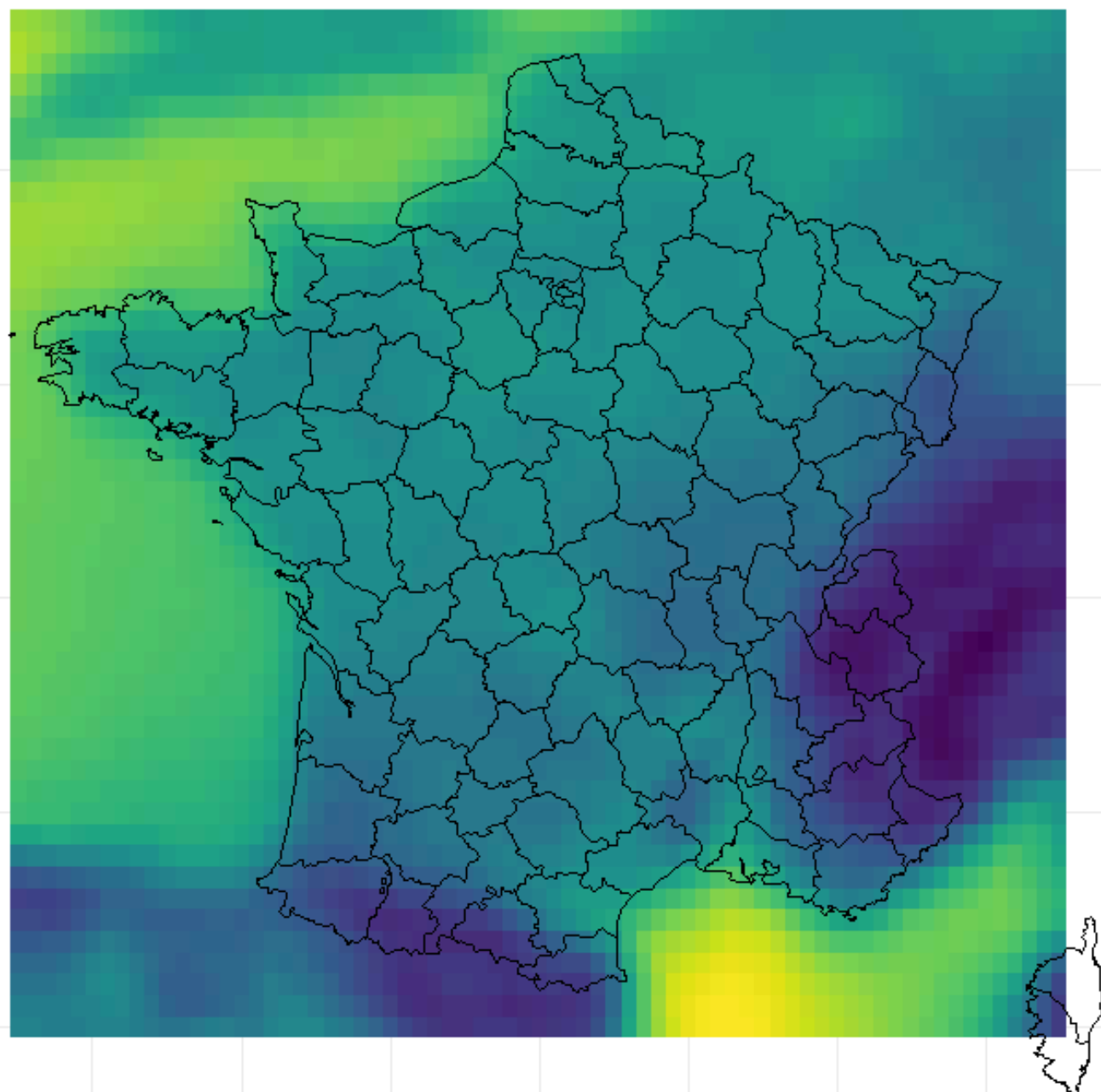
Structure du fichier wind_2017.nc

Format NetCDF

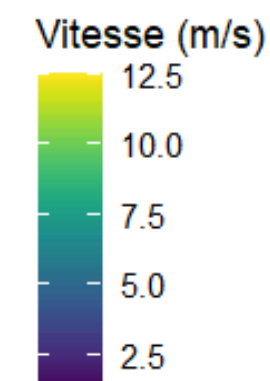
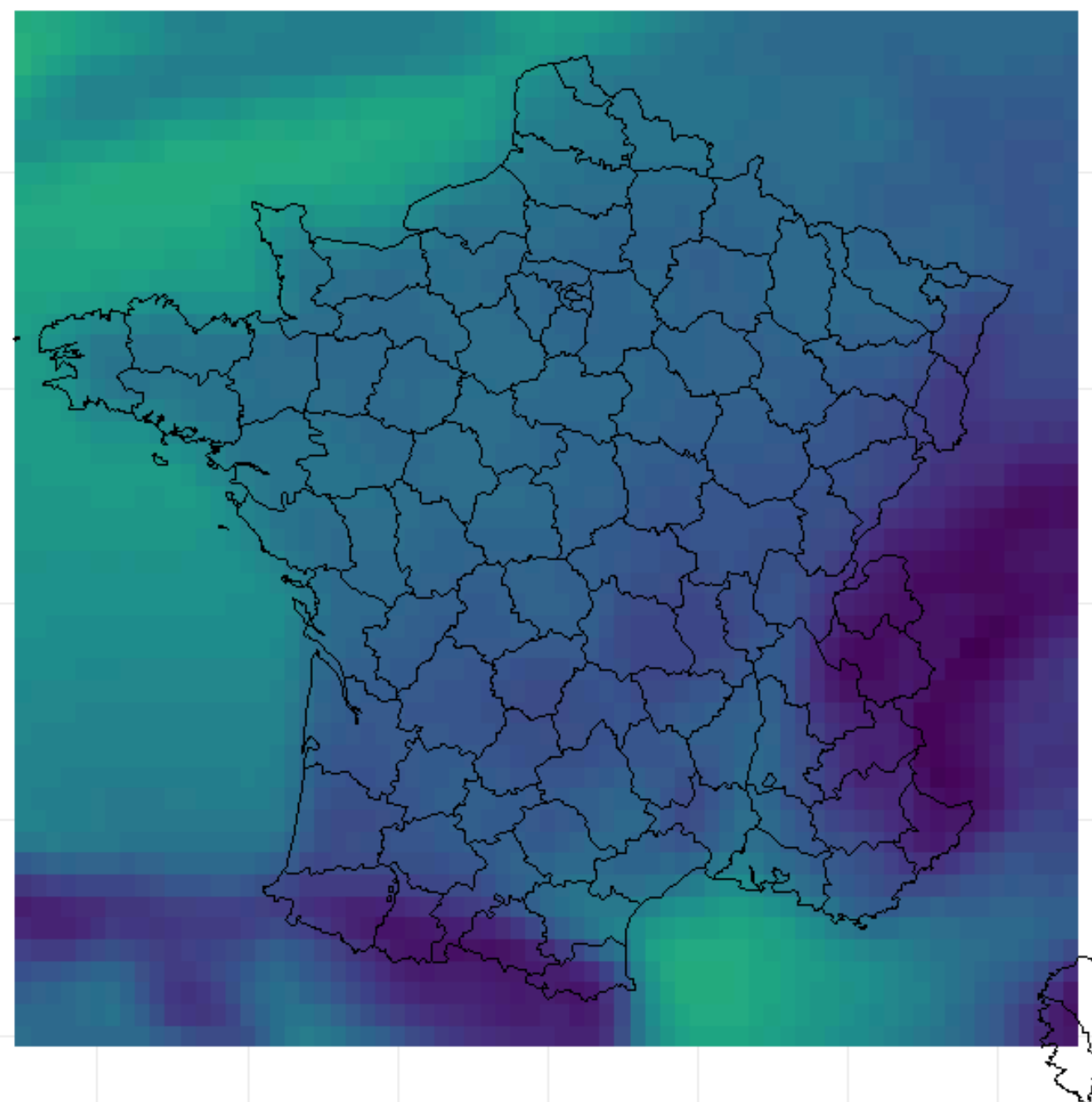
- Utilisé pour stocker des données multidimensionnelles sous l'extension **.nc**
- Package ncdf4 : permet de manipuler des fichiers de la version 4 du format NetCDF, très adaptée pour des ensembles de données très grands.

Etude descriptive : Approche saisonnière

Vitesse moyenne du vent (Hiver 2017)

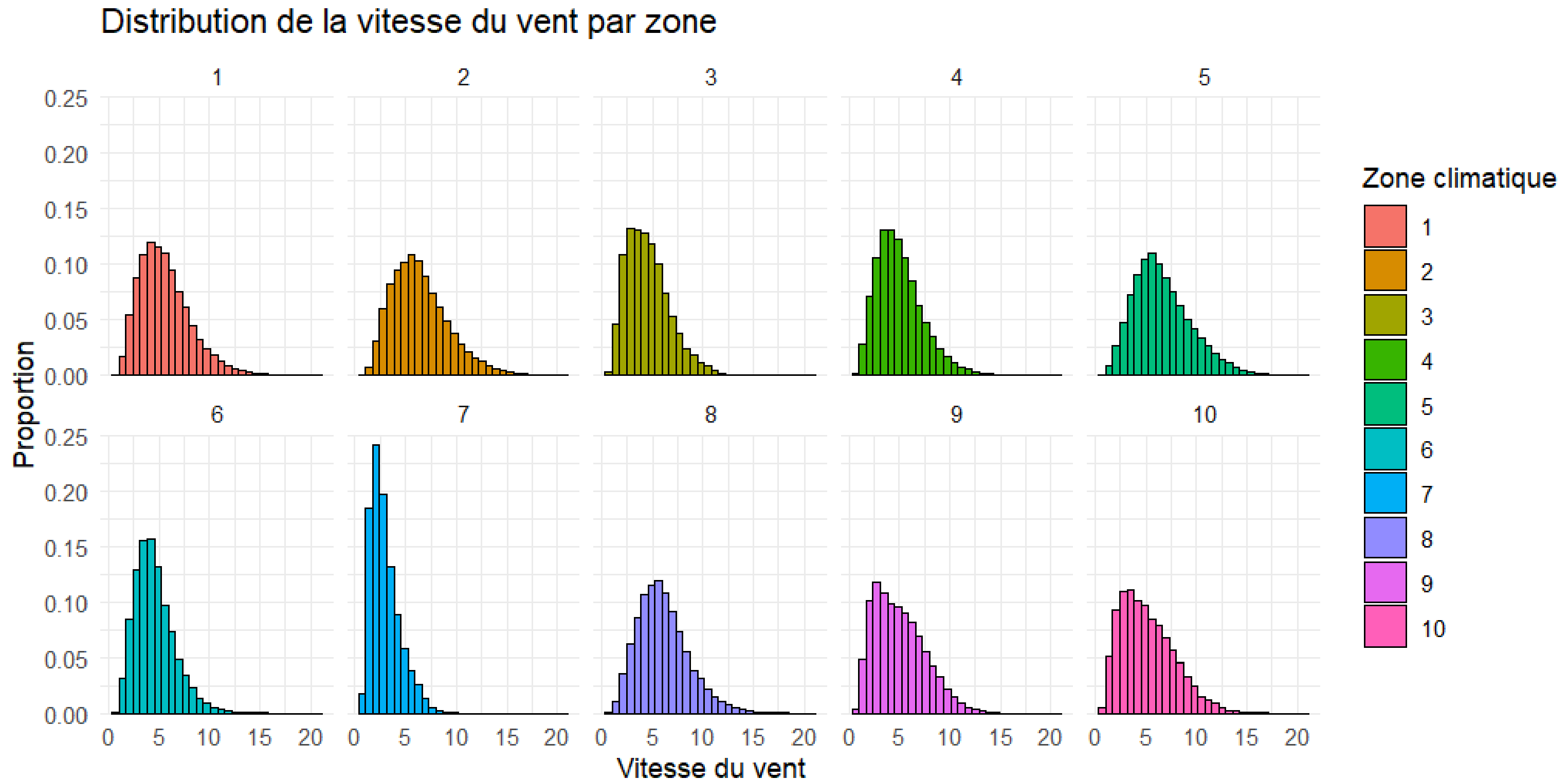


Vitesse moyenne du vent (Été 2017)



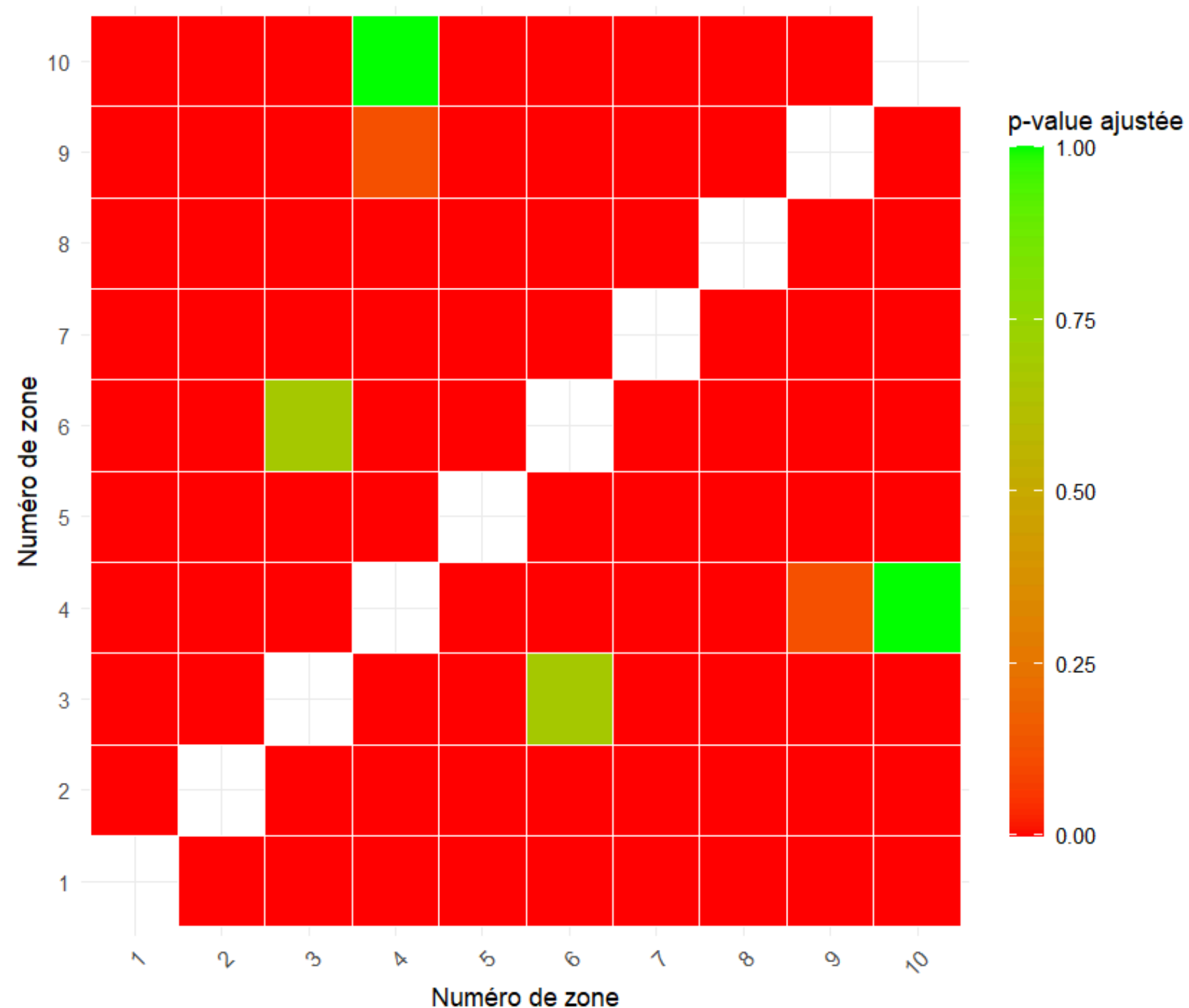
- **Hiver** : vitesses élevées sauf sur la côte sud-ouest et la frontière italo-suisse.
- **Été** : vitesses plus faibles surtout dans la moitié sud.
- **Toute l'année** : de très grandes vitesses de vent sur le pourtour méditerranéen.

Etude descriptive : Approche zonale



Constat : différence significative d'une zone à l'autre.

Approche zonale : Test statistique



Test de Kruskal Wallis et de Dunn

Toutes les zones ont-elles la même distribution de la vitesse du vent (**Hypothèse H0**) ou non (**Hypothèse H1**) ?

Modélisation statistique : GAMs

- Les GAMs autorisent des relations **non linéaires** entre les prédicteurs et la variable réponse.

$$g(\mathbb{E}(Y_i)) = X_i^* \theta + f_1(x_{i1}) + f_2(x_{i2}) + \cdots + f_p(x_{ip})$$

- Estimation des f_j : Splines de régression
- Ajustement du modèle : Maximisation de vraisemblance pénalisée (optimisation de la pénalisation par validation croisée)

Partie 1

Le code

1

Agrégation horaire des données de production

2

de

Traitement des données du vent :
Extraction des composantes U et V
Agrégation zonale

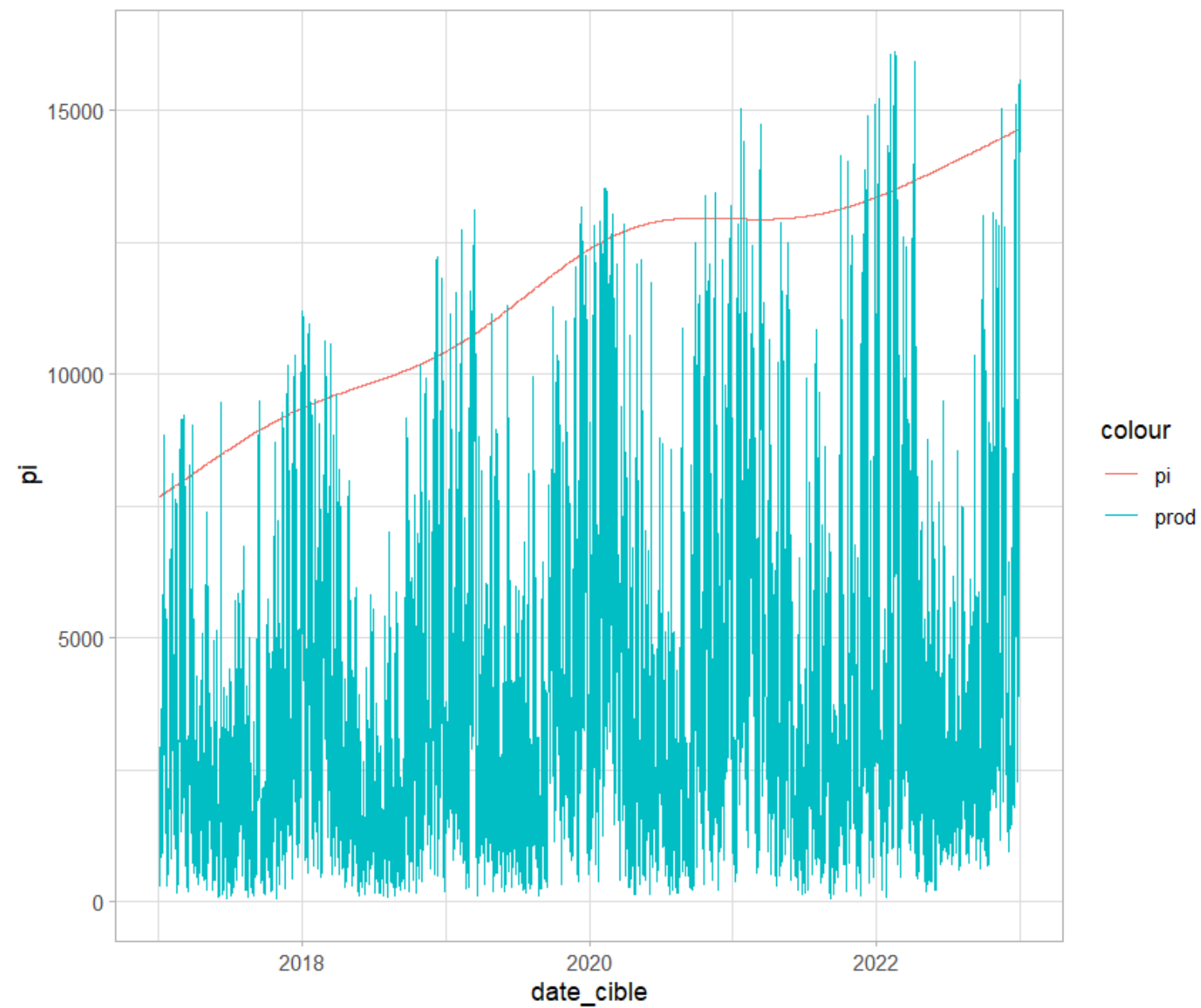
3

Modélisation GAM (eolien ~ ff100)
Estimation de la puissance
installée

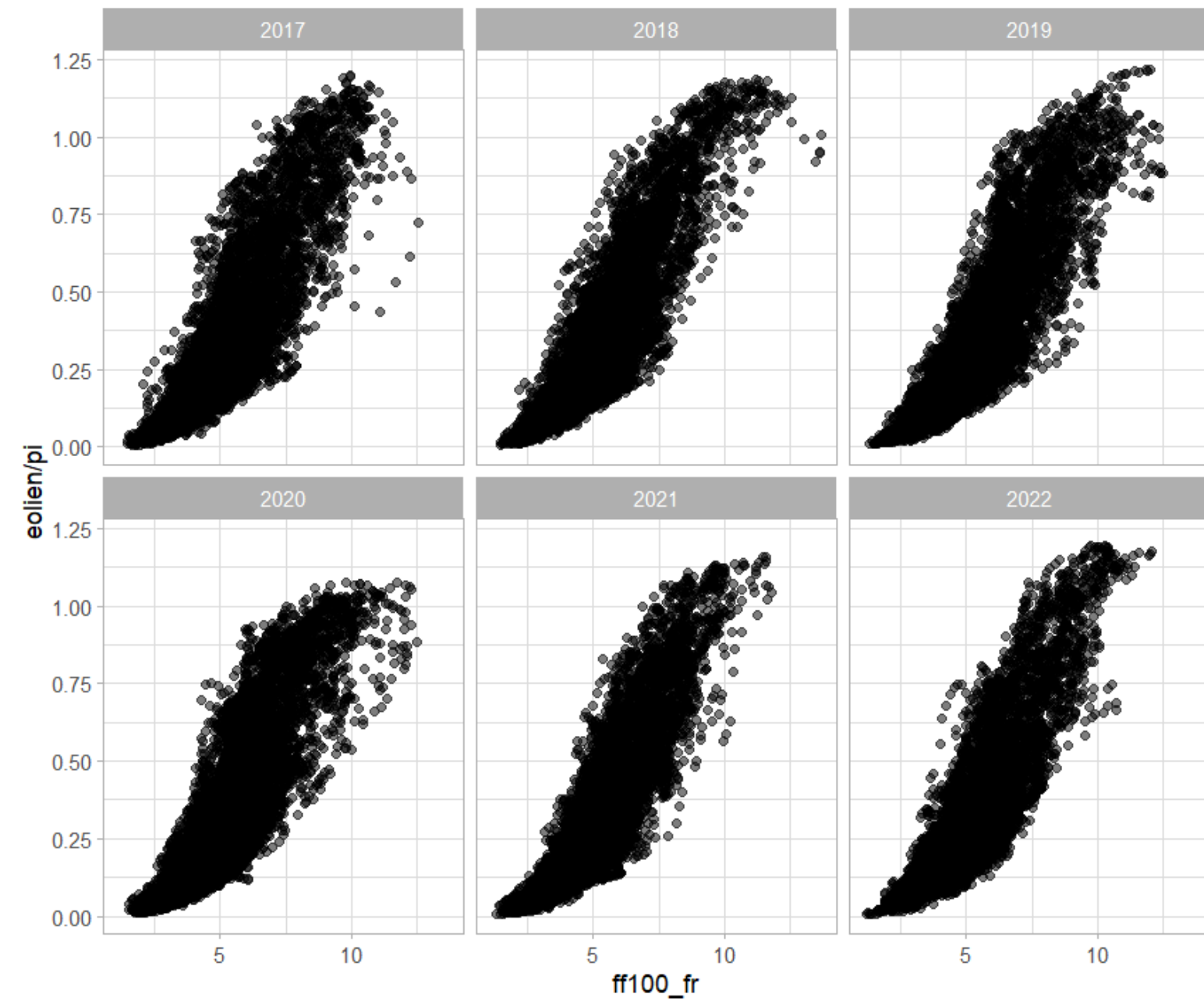
RTE

R

Résultats de l'estimation



Production réelle et Puissance installée



Ratio eolien / pi (FC)

Partie 2

Le code

1

Définition de la métrique d'évaluation (RMSE)

2

de

3

Modélisation GAM par validation croisée et analyse des résidus

Listing 6.1 – Calcul du RMSE entre les colonnes FC et prev

```
df_prev[, .(rmse = rmse(FC, prev))]  
  rmse  
  <num>  
1: 0.04707577
```

RTE

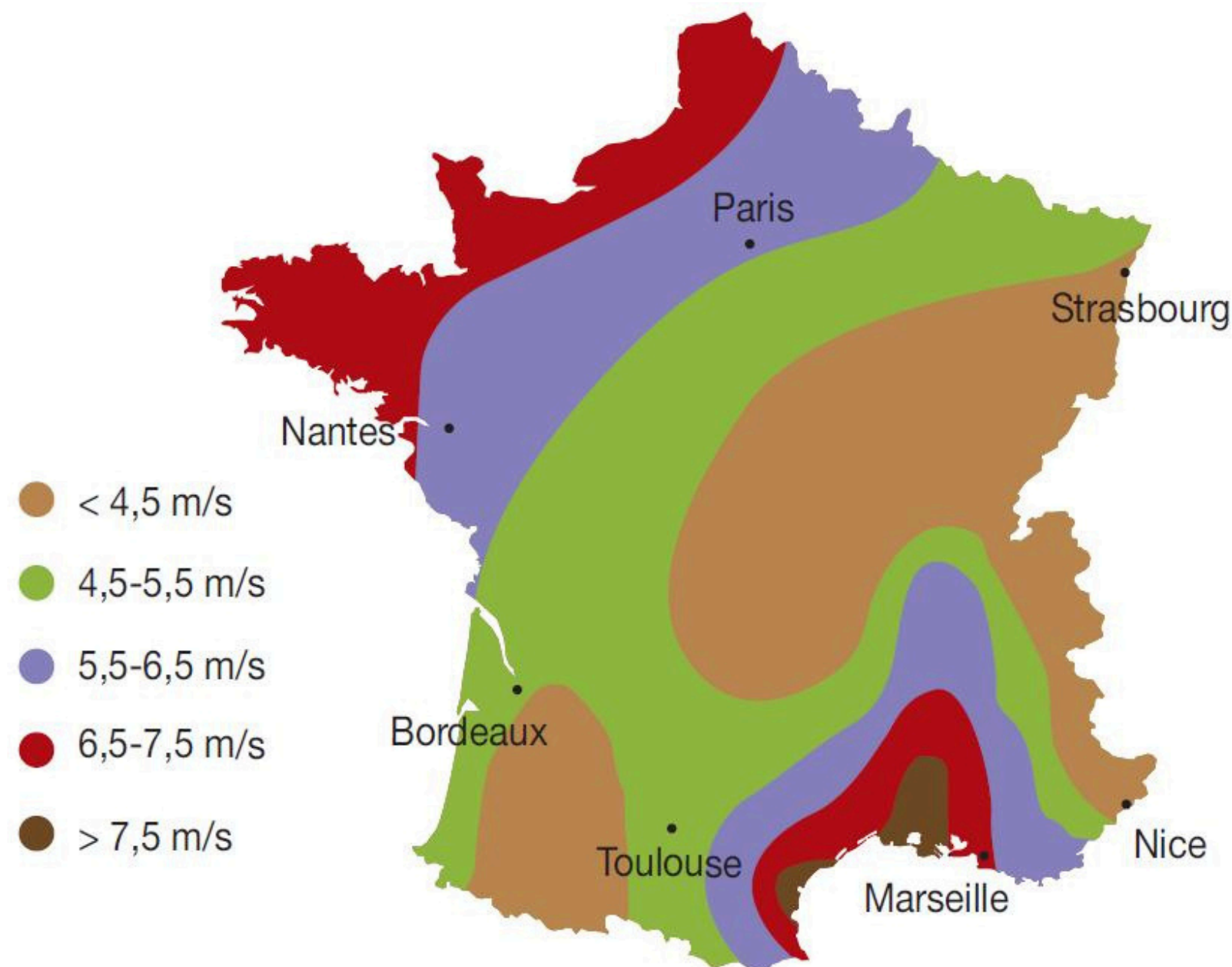
The R logo, consisting of a large white letter 'R' inside a blue circle.

Amélioration des performances

Le modèle à 5 zones

Idée : voir l'impact du nombre de zones sur la qualité des prédictions.

Le découpage ci-contre de Météo France est une classification zonale selon la vitesse du vent à 50 m d'altitude. Il faudra donc appliquer une transformation sur nos vitesses car elles sont mesurées à 100 m.



$$u_x = u_r \left(\frac{z_x}{z_r} \right)^\alpha$$

Résultat obtenu :

Listing 7.4 – RMSE du modèle GAM à 5 zones

```
df_prev[,.(rmse = rmse(FC, prev))]  
      rmse  
    <num>  
1: 0.1532901
```


Amélioration des performances

Transformation des prédicteurs

Puisque la production éolienne est proportionnelle au cube de la vitesse du vent, une approche convenable serait d'expliquer le FC à partir du cube de la vitesse du vent sur chacune des zones.

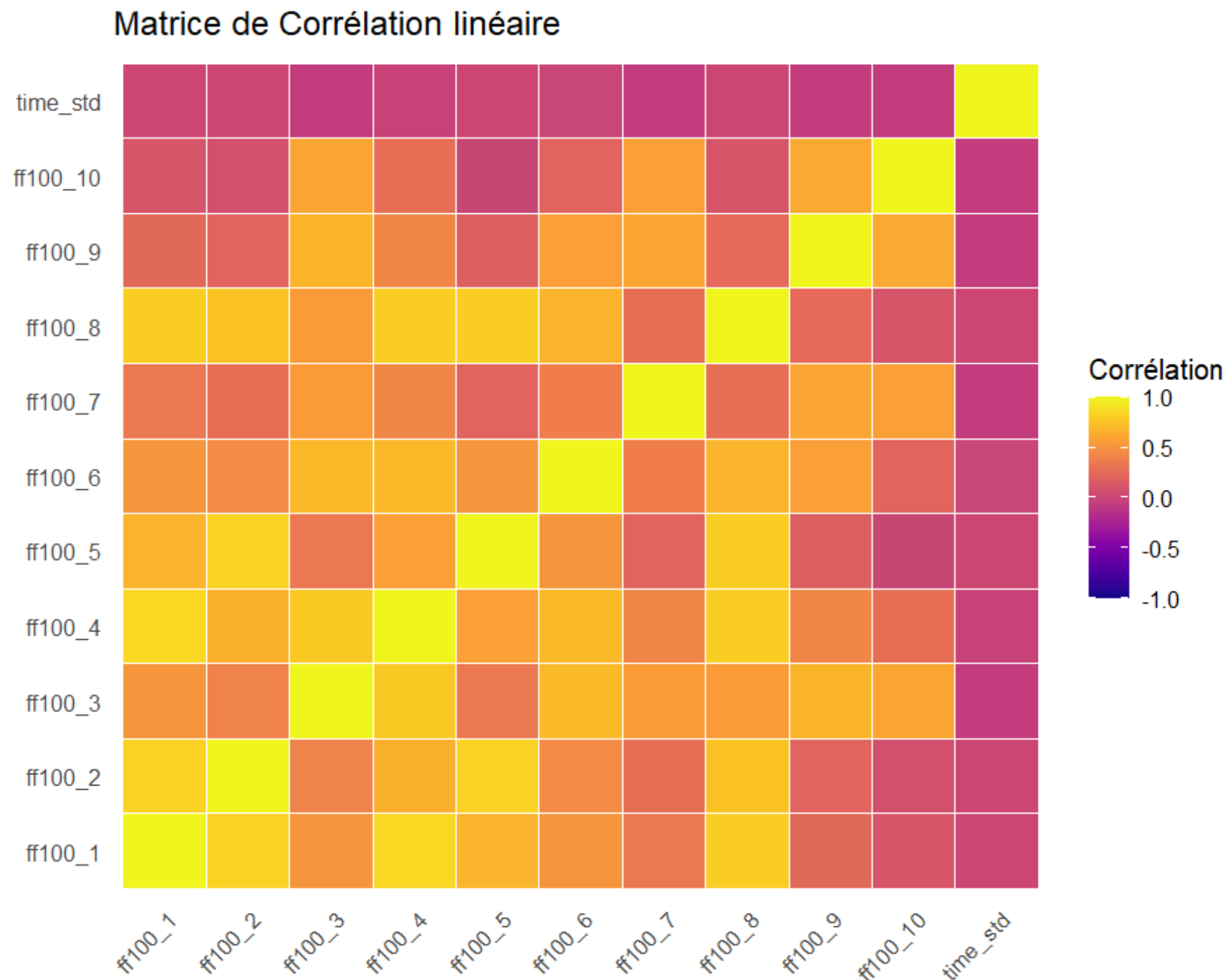
En transformant les prédicteurs et en reprenant la modélisation GAM comme précédemment, on obtient la RMSE donnée ci-après.

Listing 7.6 – RMSE du modèle GAM à 10 covariables transformées

```
df_prev[,.(rmse = rmse(FC, prev))]  
      rmse  
    <num>  
1: 0.04707577
```

Amélioration des performances

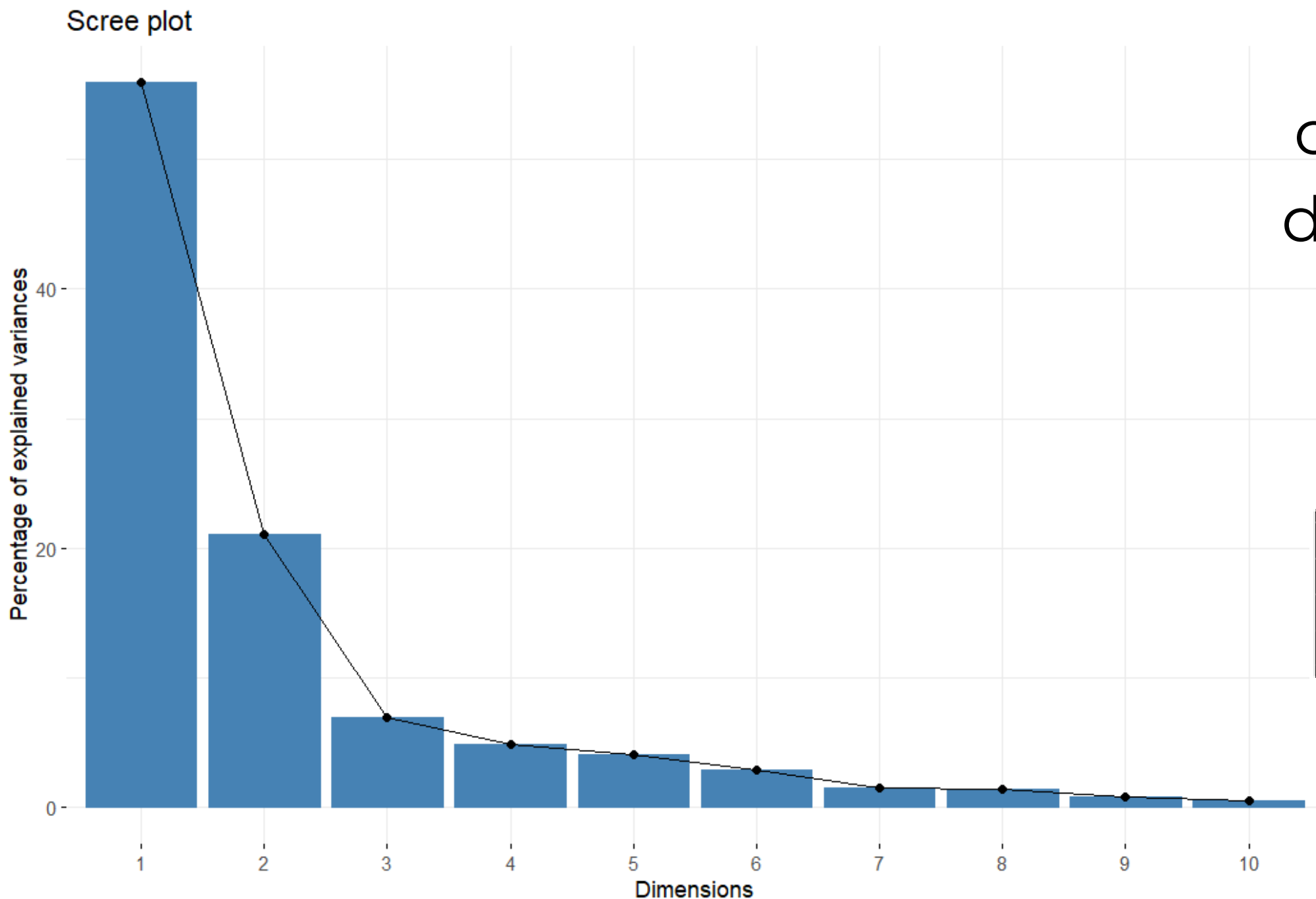
Le constat



Pas raisonnable d'ignorer
la dépendance spatiale !

Amélioration des performances

Décorrrelation par ACP



On considère les 6 premières composantes de l'ACP (plus de 90% de variance expliquée) et on refait la modélisation GAM.

Listing 7.9 – RMSE - Modèle ACP

```
df_prev_pca[, .(rmse = rmse(FC, prev))]  
      rmse  
    <num>  
1: 0.07552623
```

Amélioration des performances

Modèles de régularisation

- Ridge (L2)
- Lasso (L1)
- Elastic-Net (L1 + L2)

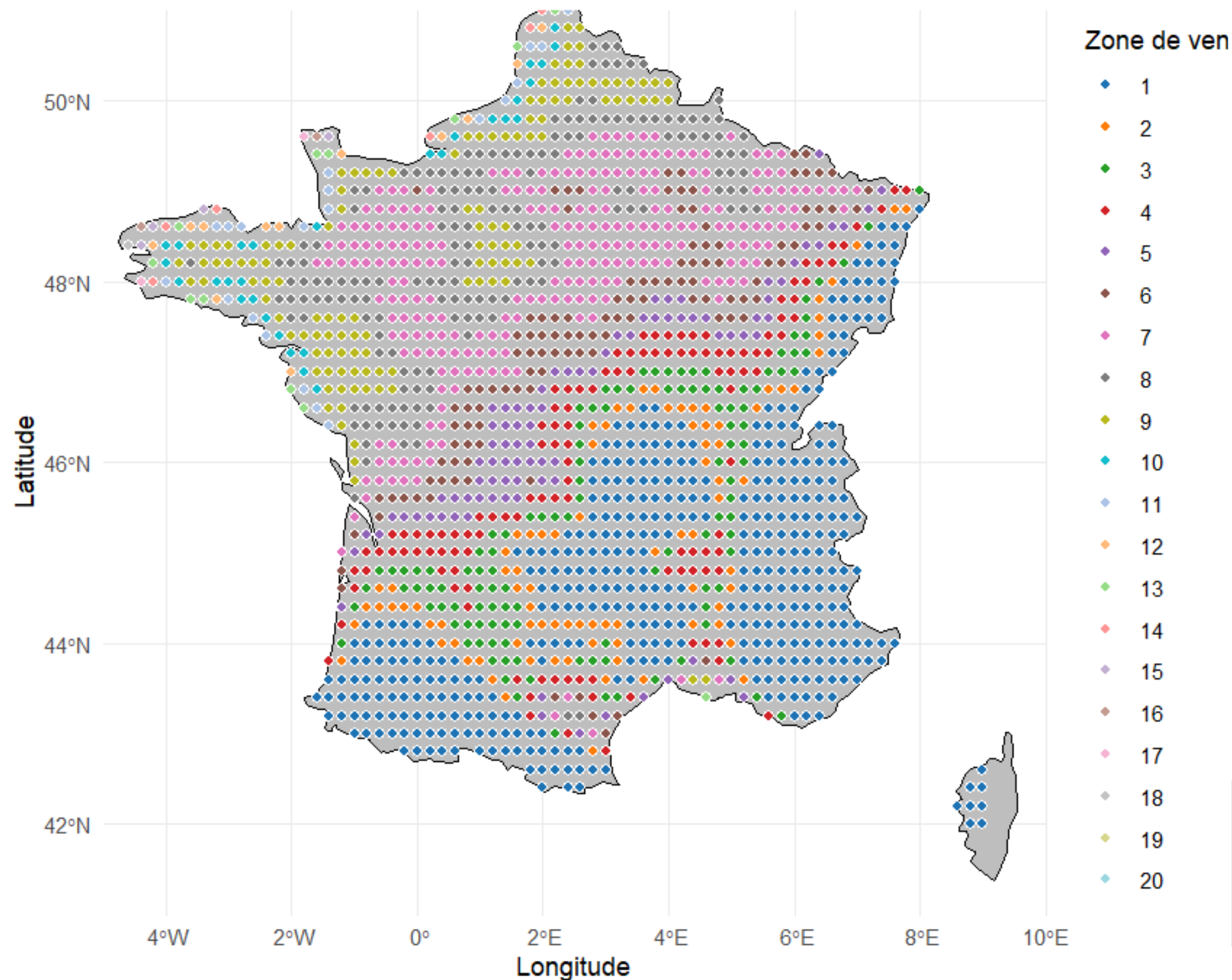
Listing 7.10 – RMSE des modèles de régularisation

```
print(c(mean(erreurs$rmse_elastic), mean(erreurs$rmse_lasso), mean(erreurs$rmse_
_ridge)))
[1] 0.04066285 0.04087566 0.04532456
```


Amélioration des performances

Redécoupage en 20 zones

Découpage en 20 zones selon la vitesse à 50 m



On reprend le découpage de Météo France et on refait le zonage à des intervalles plus fins selon la vitesse du vent.. Par exemple, 20 zones comme sur la carte ci-contre.

En reprenant la modélisation comme précédemment, on obtient finalement une **RMSE de 0.0004, soit 100 fois mieux que le modèle de RTE.**

Listing 7.11 – RMSE du modèle à 20 zones

```
df_prev[,.(rmse = rmse(FC, prev))]  
      rmse  
    <num>  
1: 0.000425589
```

Conclusion et Perspectives

- ✅ **Ce qu'on a appris** : gérer des données massives dans un contexte réel, manipuler des packages variés et des fichiers de diverses extensions
- 🔧 **Ce qu'on a apporté** : des améliorations pour des prédictions plus précises, allant jusqu'à 100 fois mieux que le modèle de RTE
- 🔍 **Autres approches à explorer** : Modèles de Machine Learning, Deep Learning (RNN) et autres stratégies de découpage connexe



**Merci pour
votre attention**

