



The background is a Halloween-themed illustration. It features a large, bright yellow full moon in the center. Two dark, leafless trees stand on either side of the moon. A black bat is flying in the sky above the moon. In the foreground, there is a dark blue ground with a white fence. Three hands are reaching up from the ground, and a tombstone with a white cross is on the right.

BOUKHDIR Saïf Eddine
HAFID Fayçal
SANT'ANNA Louis

ELU 501 - Data Science
Défi 3

- 
1. Connaissance du métier
 2. Connaissance des données
 3. Préparation des données
 4. Modélisation des données
 5. Evaluation
 6. Déploiement et conclusion
- 

1. CONNAISSANCE DU MÉTIER

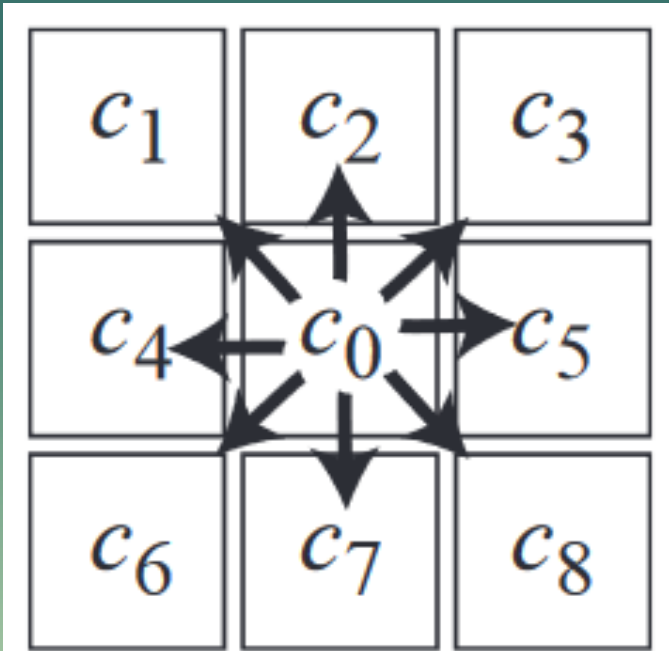


Apocalypse de zombies à partir de Rize (Turquie) !
Arriveront-ils à Brest ? Si oui, quand ?
Pourrons-nous les arrêter ?
Comment !?

Jour 0 : Environ 256 000 zombies



1. CONNAISSANCE DU MÉTIER



Un zombie peut aller dans 8 directions différentes de sa position actuelle

La dispersion de n zombies présents sur C_0 vers les voisins C_i :

- Le nombre d'humains présents sur C_i
- La différence d'élévation entre C_0 et C_i

$$Z_{j+1}(c_i) = Z_j(c_i) + \frac{H_j(c_i)}{\sum_{k=1}^8 H_j(c_k)} Z_j(c_0) \lambda_d(c_i),$$

where λ_d is a factor of geographical slope;

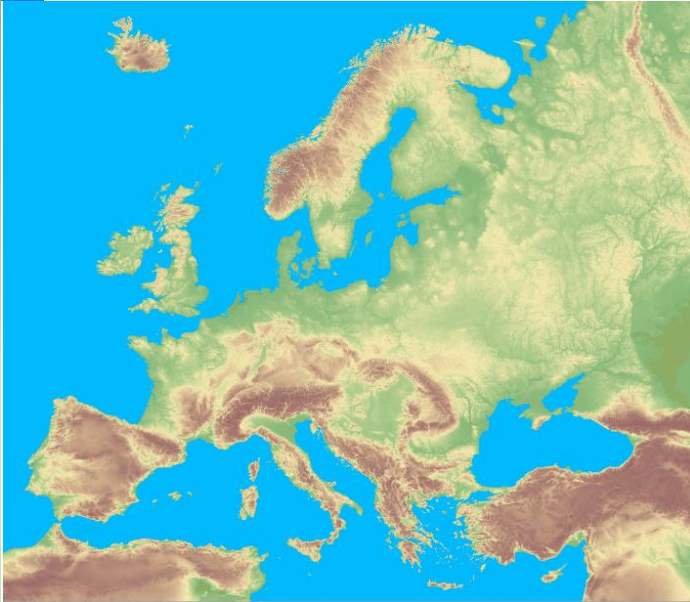


2. CONNAISSANCE DES DONNÉES



Une carte pour la densité de population par km²

- 4830 x 3510 pixels (RGB)
- 30 couleurs différentes -> 30 densités de population différentes
- Projection utilisée : World Mollweide



Une carte pour l'élévation (altitude) par km²

- 4901 x 4251 pixels (RGB)
- 189 couleurs -> 189 niveaux d'élévation différents
- Projection utilisée : Lambert Azimuthal Equal Area



3. PRÉPARATION DES DONNÉES

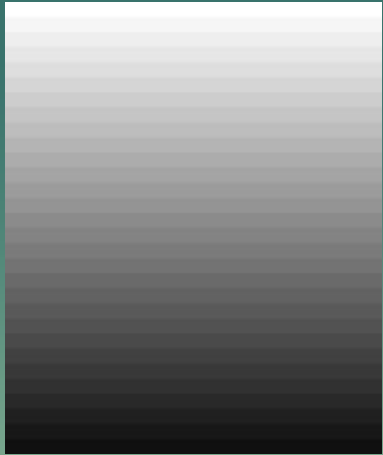


Image 1 : densité de population

- Ordonner les couleurs (ordre croissant des valeurs RGB)
- Relation linéaire entre la couleur et la densité de population (de 0 à 3000)
- L'eau (niveau 0) est représentée par des pixels (16,16,16) et (24,24,24), La densité pour ces DEUX valeurs vaut donc 0

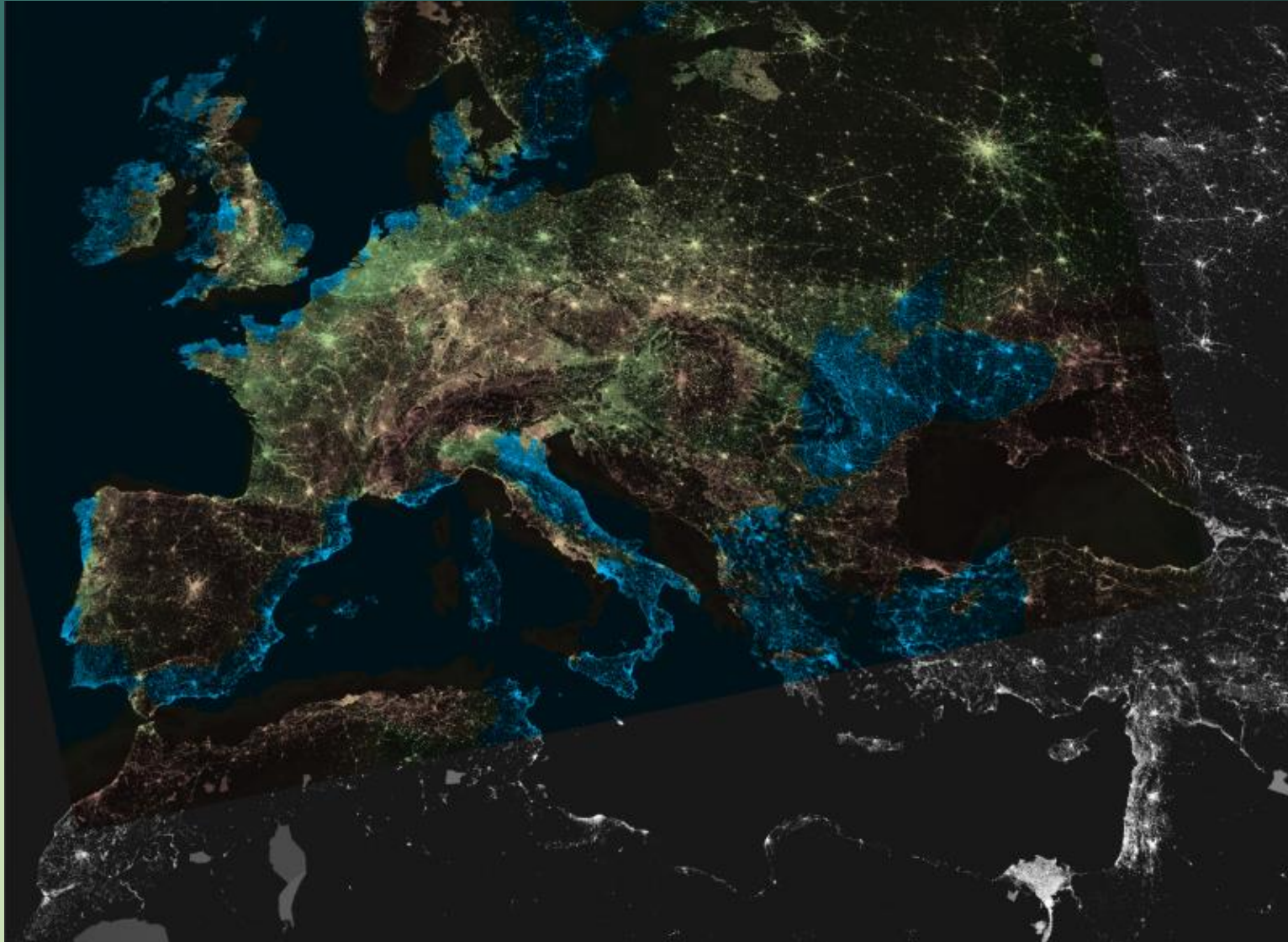
Image 2 : élévation



- Ordonner les couleurs : plus compliqué. Usage d'une heuristique (Nearest Neighbours + Shortest Path) en considérant les valeurs R, G et B du pixel comme étant des coordonnées spatiales (-> calcul de distances)
- Relation linéaire entre la couleur et l'élévation de 0 à 4,180 km (Mont Blanc)



3. PRÉPARATION DES DONNÉES



Les images n'étant pas dans la même projection, le pixel (x_0, y_0) de l'image 1 ne correspond pas au pixel (x_0, y_0) de l'image 2.

Pour pouvoir connaître, **pour un même pixel à la fois sa densité de population et son altitude**, il est nécessaire de **transformer la 2nd image** pour la faire correspondre à la première.

Les documentations en ligne et les outils opensource de projection géographique ont révélé qu'il n'y a **pas de transformation affine** possible entre les deux projections.

Il est donc **impossible de superposer PARFAITEMENT** l'image 2 à l'image 1 en la transformant



3. PRÉPARATION DES DONNÉES



SOLUTION :

Approximation de la fonction (non déterministe) permettant de passer de l'image 2 à l'image 1 par du **Machine Learning** (régression non linéaire, computer vision) :

- Etape 1 :

Dresser une liste de points P1 dans l'image 1, avec une liste de points P2 de l'image 2 qui y correspondent géographiquement (entre 50 et 60 points ont été pris)

- Etape 2 :

Utiliser l'**homographie** (fonction donnée par la librairie cv2 de Python) pour avoir la transformation



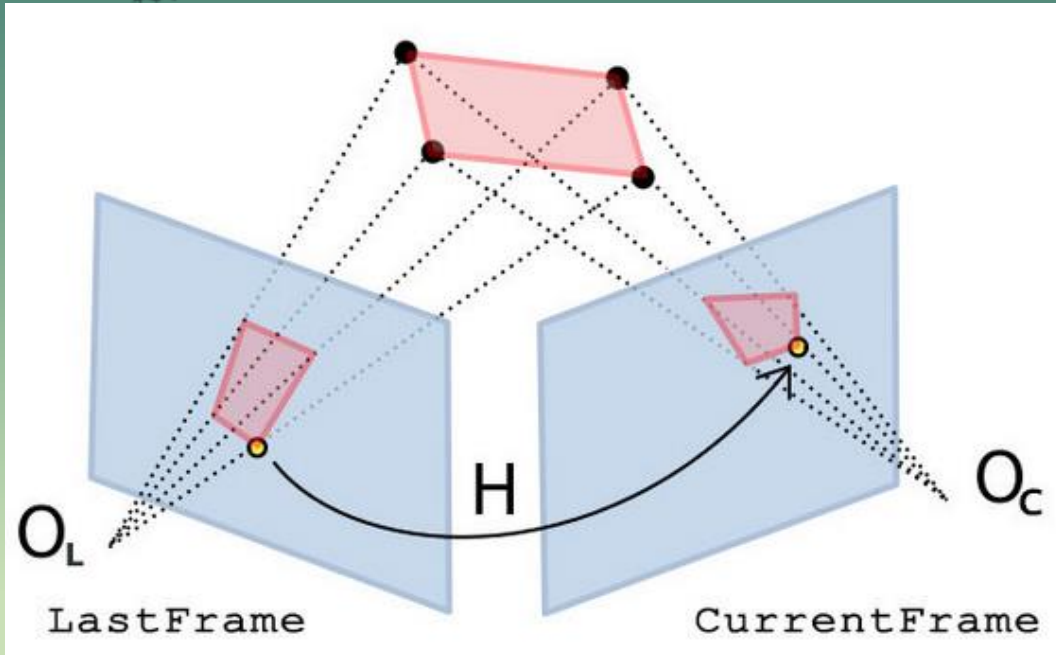
3. PRÉPARATION DES DONNÉES

HOMOGRAPHIE :

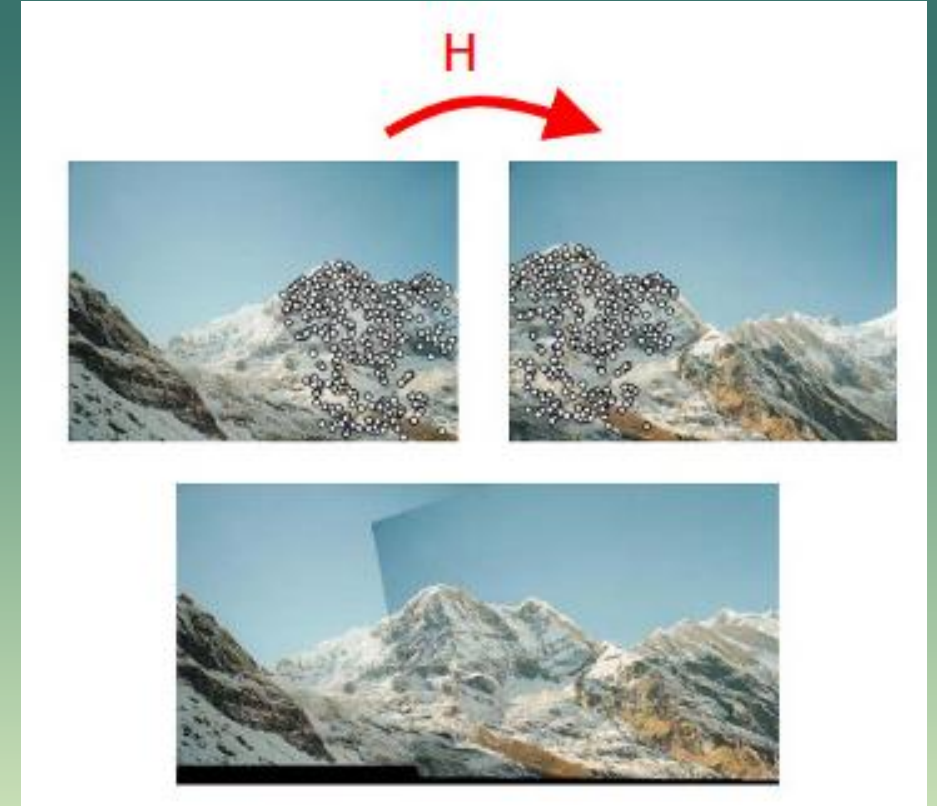
Définition : homographie 2D

Il s'agit d'une transformation linéaire entre deux plans projectifs

Source : Vision par ordinateur : Homographie et calibration, Jean-Philippe Tardif et Sébastien Roy



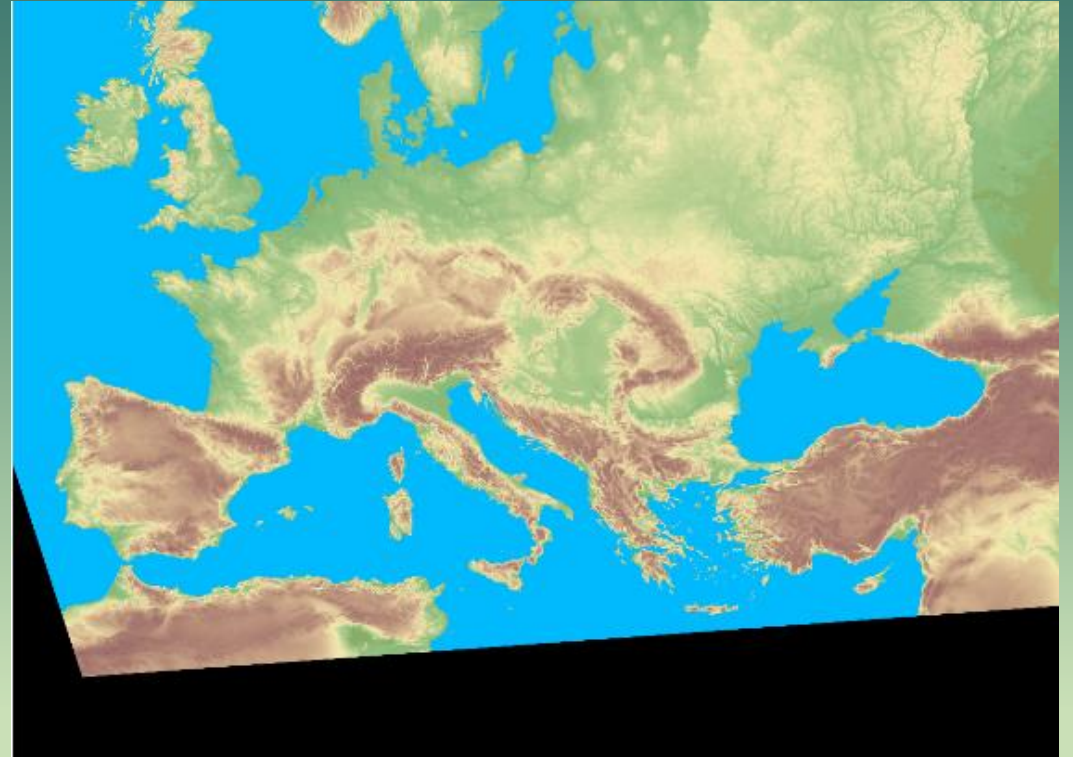
Source : [Removing Dynamic 3D Objects from Point Clouds of a Moving RGB-D Camera](#)



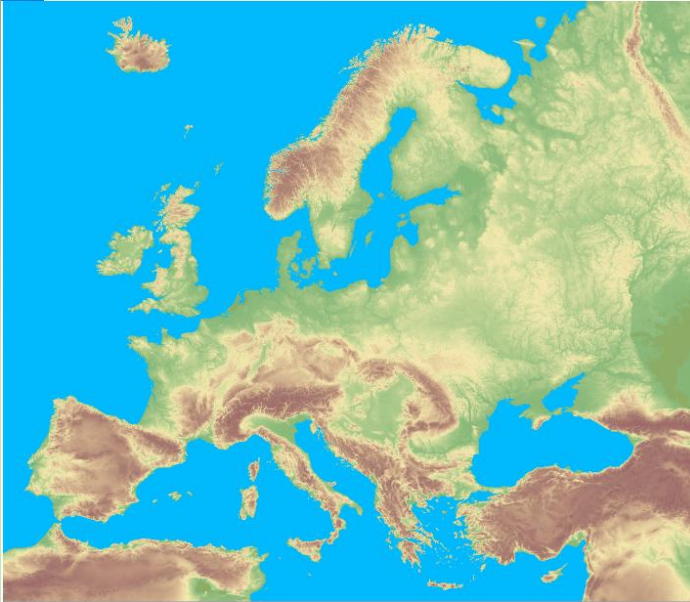
Source : documentation opencv

3. PRÉPARATION DES DONNÉES

Résultat :



3. PRÉPARATION DES DONNÉES



Compression de 15x15 pixels en un seul (une cellule)

----->

Obtention de 2 images de dimension 234 x 322 cellules

Problème : le pixel résultant par la fusion de 15×15 (225) pixels aura comme couleur la valeur RGB qui sera la moyenne des valeurs RGB des 225 pixels

→ nous obtenons de nouvelles couleurs qui n'étaient pas présentes dans les images originales

Solution:

Implémentation d'un algorithme qui, de la même manière que Nearest Neighbour, affectera à chaque nouvelle couleur, la couleur la plus « proche »/similaire qu'on possédait dans les images originales. Ca sera donc cette couleur approximée qui donnera la valeur de la densité de population et de l'altitude du nouveau pixel



4. MODÉLISATION DES DONNÉES

Modélisation par un graphe

On crée un nœud (i,j) pour chaque pixel de l'image compressée.

Chaque nœud possède comme attributs :

- pop_dens : densité de population
 - elev : élévation
- zombies : zombies présents sur la cellule
- color : couleur du graphe (« green » initialement, « red » pour les cellules infectées par les zombies, « yellow » pour les cellules bombardées et « blue » pour les cellules sécurisées)

Arrêtes

On crée une arrête entre deux nœuds si :

- Les pixels qu'ils représentent sont adjacents dans l'image
- Aucun des deux n'a une densité de population nulle (réduction de la complexité, car les zombies n'iront pas vers une cellule sans humains)

4. MODÉLISATION DES DONNÉES

Modélisation des zombies sur une cellule

Chaque cellule est un nœud. Chacune possède un attribut appelé « zombies », qui est une liste initialement vide. Au départ, pour la cellule de Rize, cette liste contiendra l'élément « 1 » (représente l'âge en jours du zombie) n fois (n étant le nombre d'humains qui y étaient initialement).

→ Chaque jour, pour chaque cellule, on incrémente chaque élément de sa liste « zombies ». Si la valeur d'un élément dépasse 15, on le supprime (le zombie meurt naturellement)

→ Le choix des zombies qui se déplaceront vers une cellule voisine se fait aléatoirement

→ Lorsque M humains sont transformés en zombies dans une cellule, la liste « zombies » de cette cellule se verra s'y rajouter M fois l'élément 1.

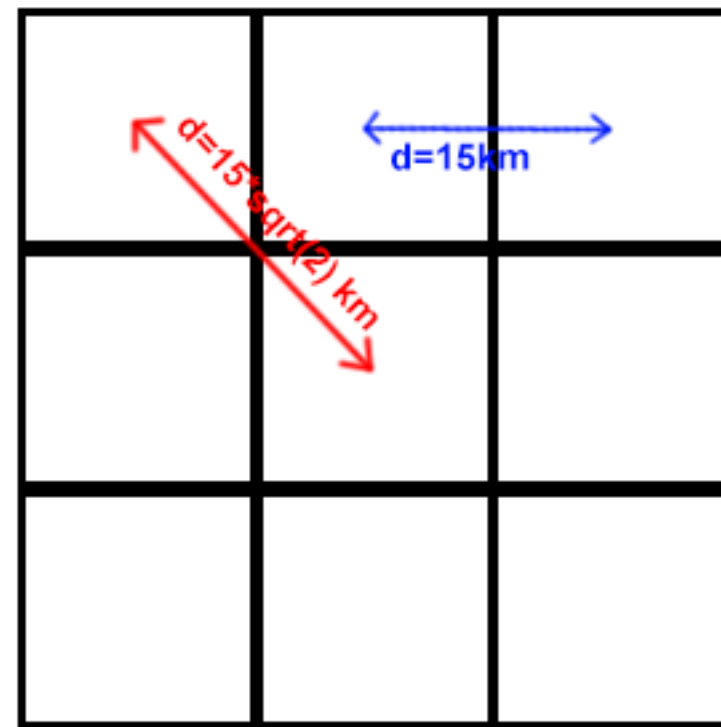
4. MODÉLISATION DES DONNÉES

Déroulement de l'apocalypse

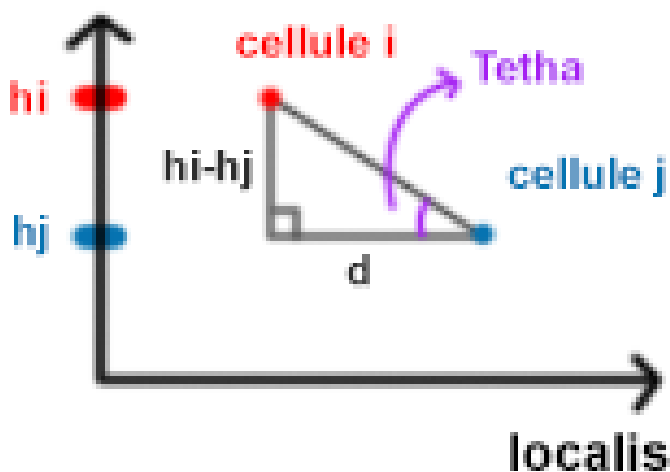
$$Z_{j+1}(c_i) = Z_j(c_i) + \frac{H_j(c_i)}{\sum_{k=1}^8 H_j(c_k)} Z_j(c_0) \lambda_d(c_i),$$

$$\theta = \arctan\left(\frac{|Elev[cell1] - Elev[cell2]|}{d}\right),$$

d=15km si les cellules sont sur la même ligne ou colonne,
d=15*√2 km sinon



Elevation



$$\lambda_d = \begin{cases} 1 & \text{si } \theta = 0^\circ \\ 0 & \text{si } \theta > 20^\circ \\ 1 - \frac{\theta}{20} & \text{sinon} \end{cases}$$

4. MODÉLISATION DES DONNÉES

Déroulement de l'apocalypse

**1**

On commence par transformer tous les humains présents à Rize en zombies

2

Les zombies de la cellule se divisent pour aller aux cellules voisines selon l'équation donnée (Application sur toutes les cellules avant de passer en 3)

3

Sur la nouvelle cellule se trouvent alors n zombies et m humains.
 $n \leftarrow \text{Min}(m, 10 * n) + n$
 $m \leftarrow m - \text{min}(m, 10 * n)$
(Application sur toutes les cellules avant de passer en 4)

4

Les humains tuent les zombies :
 $n \leftarrow n - \text{Min}(n, 10 * m')$
 $m \leftarrow m$

Conditions d'arrêt :

- Les zombies sont arrivés à Brest
OU
- Les zombies sont tous morts

Boucle (2;3;4)
d'itérations



5. ÉVALUATION

Question 0:

Oui, les zombies arrivent à Brest... En 251 jours !



Day 15



Day 60



Day 120



Day 180



Day 225



Day 251 : arrivés à Brest !

Question 0:



251, c'est aussi la longueur du chemin le plus court reliant Rize à Brest. Le résultat est donc acceptable vu que les zombies se propagent à une cellule voisine par jour.

Deux chemins (bleu et vert) calculés par l'algorithme de Dijkstra tel que deux cellules sont voisines si :

- Elles sont adjacentes
- Les deux possèdent une densité de population non nulle
- L'arrête est pondérée proportionnellement à la densité de population

Le chemin en bleu prend en compte l'élévation (pondération du graph selon λ) → Ce n'est pas le même chemin, mais les deux chemins obtenus ont la même longueur

Question 1:

Jour 60 : sélection de 20 cellules à sécuriser.
Comment ?

Méthode :

1

Pondérer le graphe tel
que le poids entre deux
nœuds n_i et n_j vaut :

$\text{Weight}(n_i, n_j) =$
 $\text{pop_dens}(n_j) \times \lambda(n_j, n_i)$

- Car on veut que :
- Plus la cellule voisine possède d'humains, plus la cellule courante est importante
 - Plus $\lambda(\text{cellule courante}, \text{cellule voisine})$ est grand, plus la cellule courante est importante

Question 1:

Jour 60 : sélection de 20 cellules à sécuriser.
Comment ?

Méthode :

1

Pondérer le graphe tel
que le poids entre deux
nœuds n_i et n_j vaut :
 $\text{Weight}(n_i, n_j) =$
 $\text{pop_dens}(n_j) \times \lambda(n_j, n_i)$

- Car on veut que :
- Plus la cellule voisine possède d'humains, plus la cellule courante est importante
 - Plus $\lambda(\text{cellule courante}, \text{cellule voisine})$ est grand, plus la cellule courante est importante

2

Application de
l'algorithme
PageRank sur le
graphe pondéré →
Obtention d'un
classement des
cellules de la plus
importante à la
moins importante
par les critères
d'importances
définis en 1

Question 1:

Jour 60 : sélection de 20 cellules à sécuriser.
Comment ?

Méthode :

1

Pondérer le graphe tel
que le poids entre deux
nœuds n_i et n_j vaut :
 $\text{Weight}(n_i, n_j) =$
 $\text{pop_dens}(n_j) \times \lambda(n_j, n_i)$

- Car on veut que :
- Plus la cellule voisine possède d'humains, plus la cellule courante est importante
 - Plus $\lambda(\text{cellule courante}, \text{cellule voisine})$ est grand, plus la cellule courante est importante

2

Application de
l'algorithme
PageRank sur le
graphe pondéré →
Obtention d'un
classement des
cellules de la plus
importante à la
moins importante
par les critères
d'importances
définis en 1

3

Filtrer la liste en ne gardant que
les cellules :

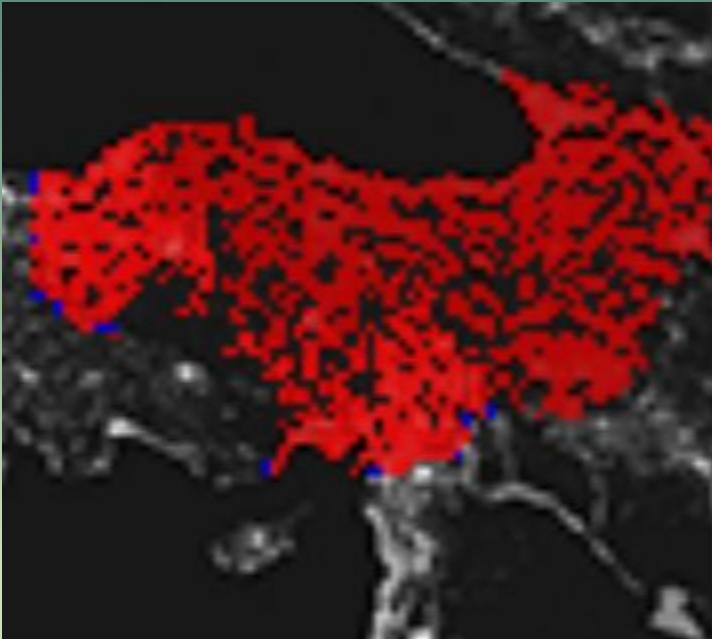
- Qui contiennent des zombies
- Qui possèdent au moins une cellule voisine non infectée et qui possède des humains

→ Prendre les 20 mieux
classées

Question 1:

Jour 60 : sélection de 20 cellules à sécuriser.
Comment ?

Résultat :



5. ÉVALUATION

Question 1:

Jour 60 : sélection de 20 cellules à sécuriser.
Comment ?

Résultat : Pas très efficace, les zombies continuent à se propager malgré les cellules sécurisées



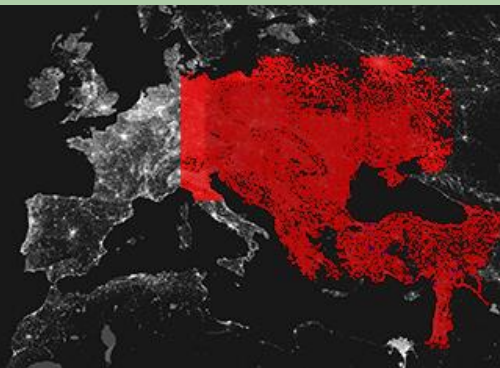
Day 60



Day 70



Day 120



Day 180

→ Ils finiront quand même par arriver à Brest !



5. ÉVALUATION

Question 1:

Jour 60 : sélection de 20 cellules à sécuriser.
Comment ?

Résultat : Le nombre de zombies diminue au jour 60 puis reprend progressivement et compense rapidement les pertes



```
Day 58 Number of infected cells : 806
Total number of zombies : 37788841
Day 59 Number of infected cells : 769
Total number of zombies : 38165176
Day 60 Number of infected cells : 737
Total number of zombies : 37682513
Day 61 Number of infected cells : 727
Total number of zombies : 36176469
Day 62 Number of infected cells : 704
Total number of zombies : 36220811
Day 63 Number of infected cells : 686
Total number of zombies : 35907540
Day 64 Number of infected cells : 670
Total number of zombies : 36098370
Day 65 Number of infected cells : 674
Total number of zombies : 37996483
Day 66 Number of infected cells : 696
Total number of zombies : 40699298
Day 67 Number of infected cells : 700
Total number of zombies : 44153710
```

Avant jour 60 :
Augmentation

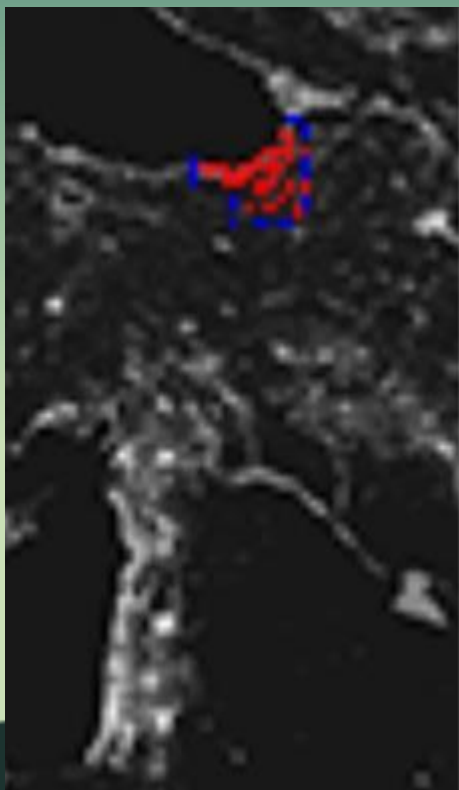
De 60-63 :
Diminution

Après le jour 63:
Augmentation et
compensation

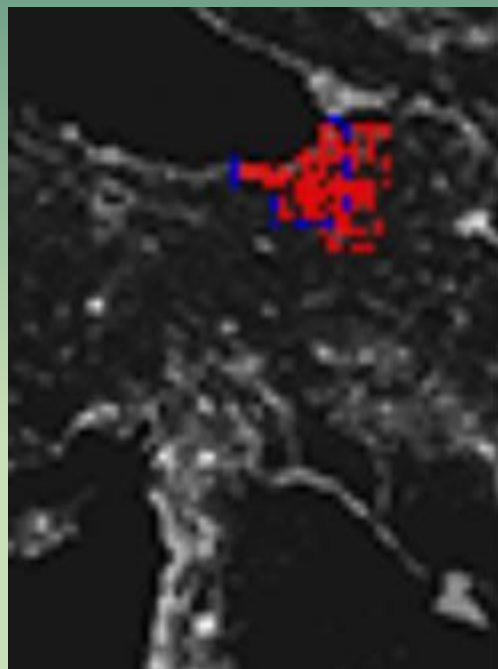
Question 1:

En retestant cette solution au jour 8 au lieu du jour 60, elle est au faite efficace, les zombies finissent par mourir au jour 127 même si ils ont continué à se propager dans d'autres directions (cela montre que la solution du PageRank est bonne). Le problème est donc au fait que cette solution est utilisée tardivement et que le nombre de cellules sécurisées n'a pas été suffisant

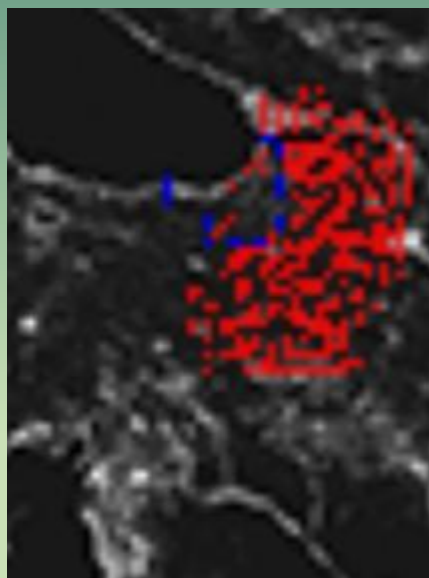
Day 8



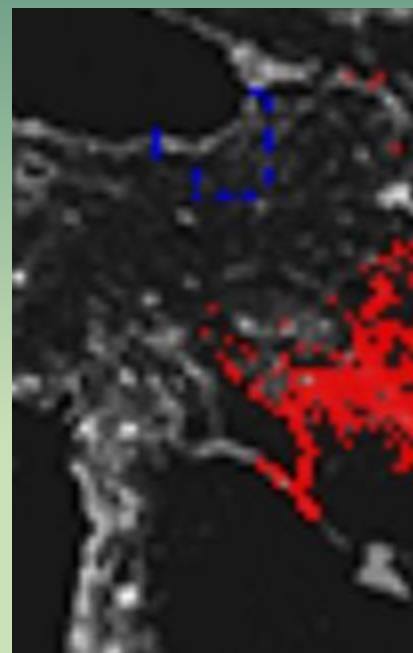
Day 15



Day 30



Day 60



Day 75



5. ÉVALUATION

Question 2:

Jour 120 : sélection de cellules à bombarder.
Lesquelles et combien ?

Méthode :

Idem que question 1, sauf qu'on prend toutes les cellules après le filtrage au lieu de s'en limiter à 20 seulement.

Résultat de l'algorithme → Toutes les cellules infectées et qui possède au moins une cellule voisine qui contient des humains et qui soit accessible (élévation) est bombardée.



5. ÉVALUATION

Question 2:

Jour 120 : sélection de cellules à bombarder.
Lesquelles et combien ?

Résultat : Les zombies finissent par mourir, le reste de l'humanité est sauvé. Aucun humain n'a été touché par les bombes



Day 120



Day 125



Day 130



Day 135

Il ne reste aucun survivant parmi les
zombies au bout du jour 145

6. DÉPLOIEMENT ET CONCLUSION

1- Solution idéale :

Intervention le plus tôt possible (avant le jour 30) en sécurisant assez de cellules
(une 30aine)

6. DÉPLOIEMENT ET CONCLUSION

2- Back-up solution :

La solution des bombes nucléaires est **d'une efficacité certaine**, car elle stop tous les zombies, qui finiront par dépérir en 15 jours au maximum, et elle peut être utilisée à n'importe quel moment : il n'est **jamais trop tard**.

Cependant elle nécessite l'usage **de beaucoup de bombes** dont le nombre croît avec le numéro du jour actuel.

(nombre de bombes = nombres de cellules infectées adjacentes à au moins une cellule non infectée, habitée et accessible)

Gros inconvénient : ces cellules ne pourront être traversées par les humains pendant 50 ans. Mais elle les aura sauvés, ils n'ont pas à se plaindre ;)

6. DÉPLOIEMENT ET CONCLUSION

Pour conclure, si jamais une apocalypse de zombie se lance :

- 1- L'UE doit intervenir le plus tôt possible en sécurisant le maximum de cellules (sécuriser 15 cellules au 8^{ème} jour sont suffisantes)
- 2- En cas d'échec de 1, l'UE doit user de bombes nucléaires, de préférence le plus tôt possible.
- 3- En cas d'échec de 1, prenez abri dans un point élevé de la carte.
- 4- OUT OF THE BOX SOLUTION, peu réaliste car difficile de diriger autant de gens d'une manière bien organisée et efficace :
Les humains qui se trouvant à moins de 15km d'une zone infectée de zombies doivent avancer d'au moins 15km (une cellule) dans la direction opposée. Les zombies seront bloqués et finiront par mourir.