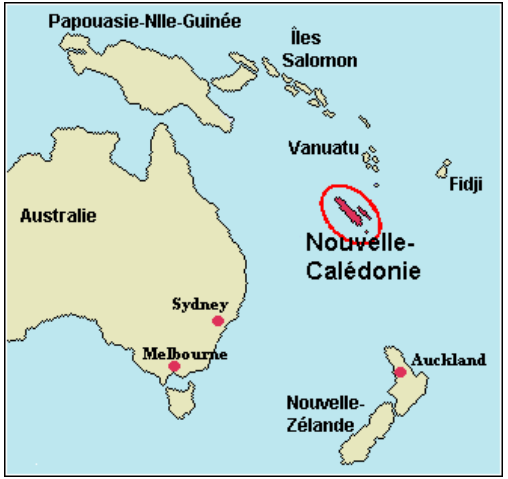
Traitement et analyse de données de positions en mer à différentes échelles : cas d'application au trafic maritime

Loïc Salmon
Docteur en informatique
loic.salmon16@unc.nc

Présentation et contexte



- Eléments de contexte sur l'Université de la Nouvelle-Calédonie (UNC)

  -3700 étudiants et 3 départements de formation (Droit, Economie, Gestion ; Lettres, Langues, Sciences Humaines ; Sciences et Techniques)

  -7 équipes de recherche

- Membre de L'ISEA (Institut des Sciences Exactes et Appliquées)

  -Equipe pluridisciplinaire (mathématiques, informatique, géologie, géographie, biologie, physique et chimie)

  -Science des données appliquée aux données environnementales

Synthèse des activités de recherche

Domaines de recherche et d'intérêt

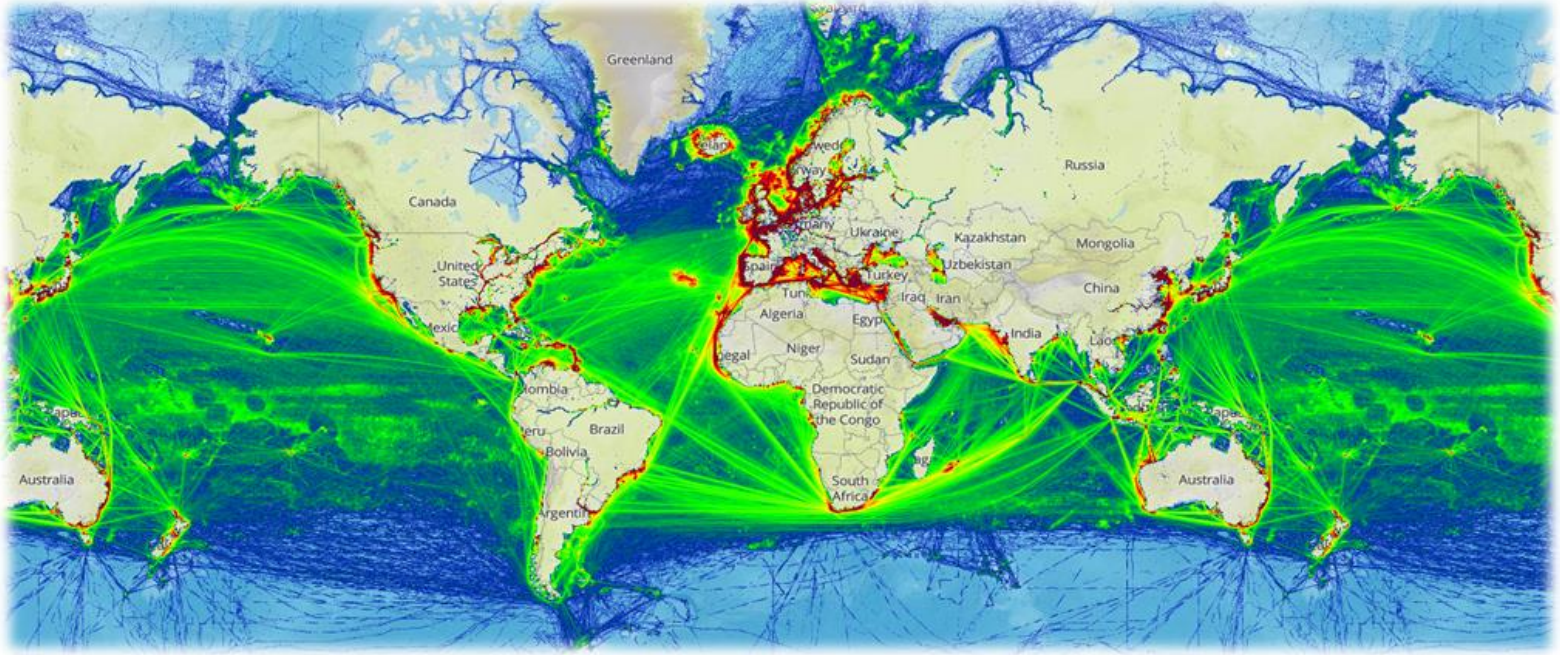
- Science des données et Systèmes d'Informations Géographiques (SIG), Big Data pour des données spatio-temporelles
- Observation de l'environnement

Travaux de recherche et réalisations

- Mise en place d'une approche hybride pour le suivi des mobilités maritimes
  - **DéAIS** projet ANR concernant la détection de messages AIS falsifiés (2014-2017)
  - **DataCRON** projet H2020 pour le suivi et la prédiction du trafic maritime (2014-2018)
  - **ISOLA** Projet H2020 concernant la sécurisation de navires de croisières (2020-2021)
- Conception d'un simulateur de trajectoires pour des drones sous-marins
  - **Navidro** projet co-financé par le SHOM et la DGA (2019-2020)
- Travaux pour le suivi de déchets plastiques en mer
  - Dépôt d'un projet en cours

Mise en place d’une approche hybride pour le suivi des mobilités maritimes

Un nombre grandissant de données maritimes à traiter



Maritime traffic (marinetraffic.com)

19 millions de messages quotidiens(18 700 000 AIS, 47 000 VMS, 39 000 LRIT), de 80 600 bateaux [EMSA, 2016]. Toutes les données reçues représentent 22 Go par jour

Objectif et proposition

- **Objectif :** Élaborer un système de gestion et de traitement des données permettant l'analyse de trajectoires en continu dans une démarche exploitant conjointement des données archivées et des flux continus de données

Ceci nécessite de :

- **Modéliser** un système dit hybride (i.e. mêlant connaissances extraites de données archivées et des flux temps-réel) pour l'analyse d'objets mobiles
- **Développer** une **architecture** et des **algorithmes** "**hybrides**" pour l'analyse de données de positions de navires

Approche Offline

- Données stockées (avec des volumes conséquents) , ce qui nécessite des techniques appropriées pour faciliter l'accès aux données et limiter les temps de traitement (e.g. Index, partitionnement, vues…)

- Permet de répondre à des requêtes "compliquées"
- Réponse de bonne qualité
- Données **entièrement** conservées

Avantages

- Mises à jour périodiques
- Temps de traitement long
- Analyse à posteriori

Inconvénients

Approche Online

- Traitement en mémoire (limitée) qui nécessite des techniques appropriées pour réduire le nombre de données (e.g. filtrage, agrégat, échantillonnage…) et traiter rapidement les requêtes (eg. one pass, approximation). Un des paradigmes de traitement régulièrement utilisé est celui d'un **graphe direct acyclique** (ou workflow), acheminant les données représentées par les arêtes et les nœuds représentant les opérateurs intervenant dans le processus.

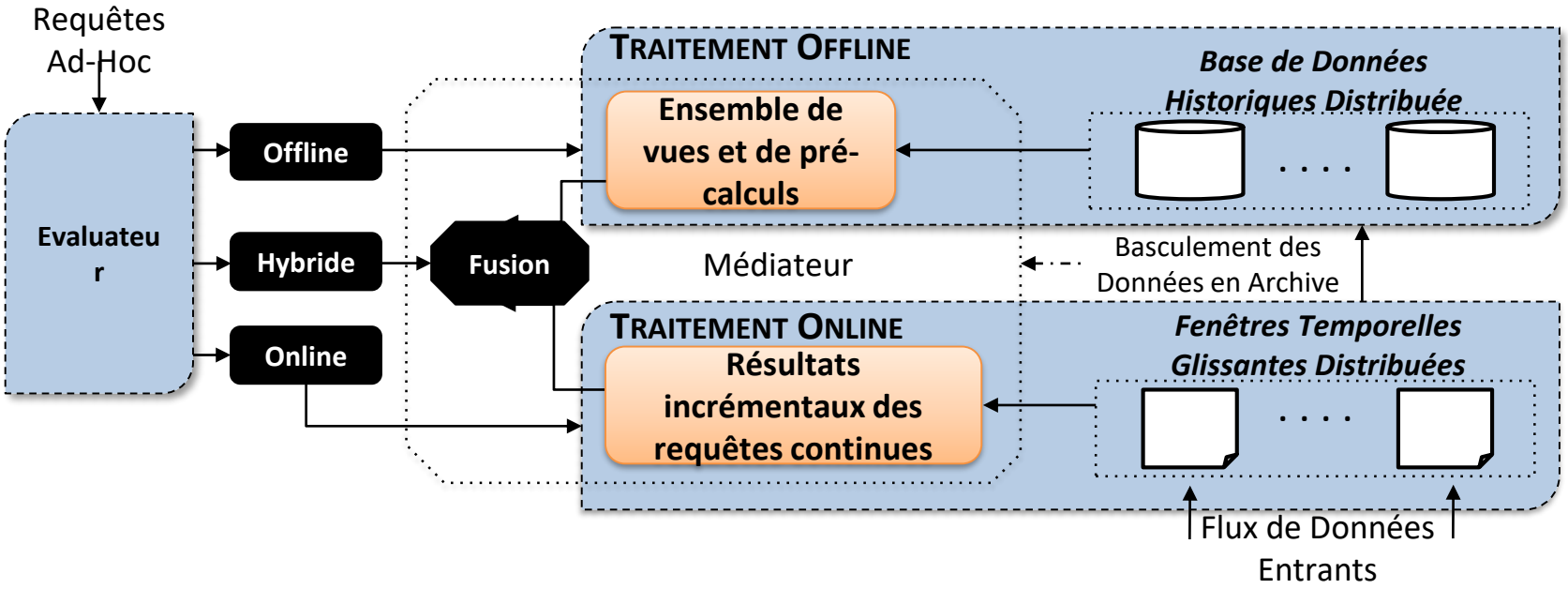
- **Traitement des données « à la volée »**
- **Temps de réponse rapide**
- **Système réactif (requêtes continues)**

Avantages

- **Absence d'historique** pour répondre aux questions
- **Réponse** parfois **approximative**
- **Perte des données** après traitement

Inconvénients

Principes d'une approche hybride pour le traitement d'objets mobiles



Architecture proposée dans [Salmon *et al*, 2015] et dérivant de l'architecture lambda [Marz, 2013]

Application d'une approche hybride pour identifier les zones de non couverture de l'AIS

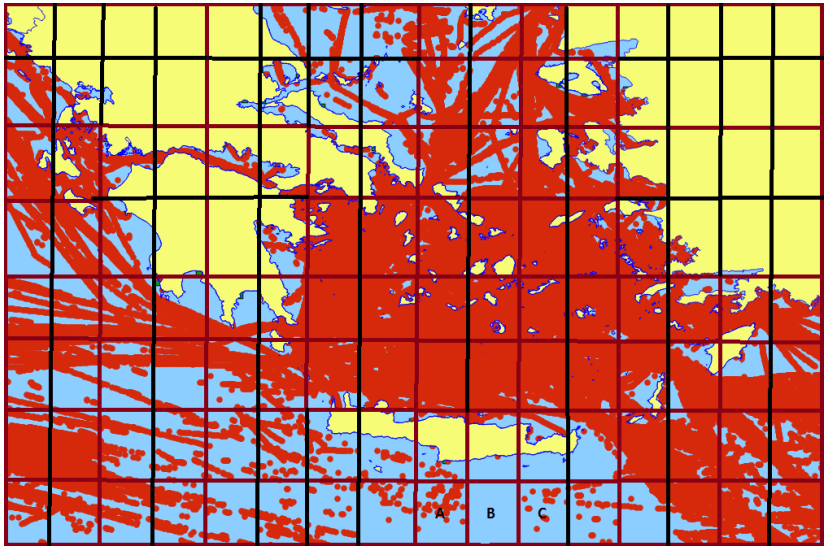
Les navires se déplacent "librement" en **milieu semi-ouvert**. Contrairement à une mobilité sur des réseaux contraints cela engendre une **incertitude** plus grande quant à la position des bateaux

Limites des systèmes de localisation de navires (notamment de l'AIS utilisé dans notre étude) :

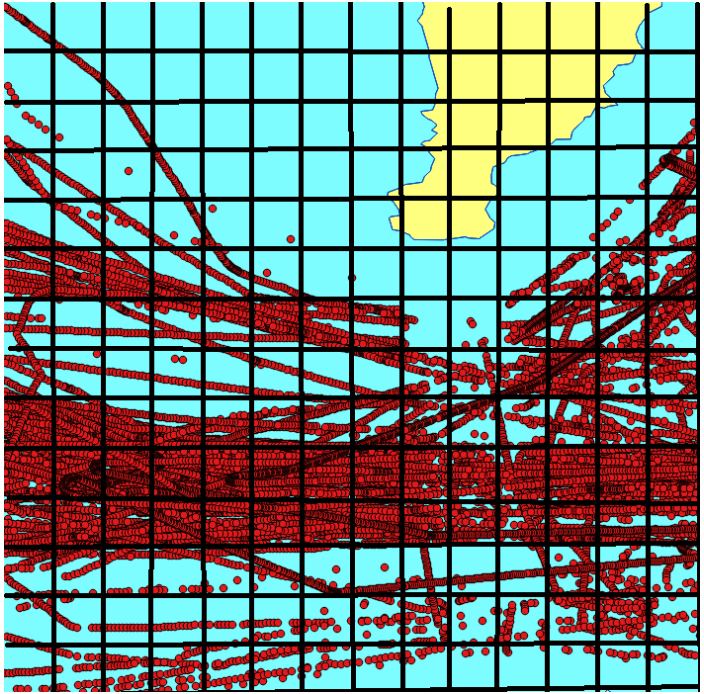
- Messages **non délivrés ou perdus** dû aux limites du système ou à des erreurs techniques ou humaines
- Les limites de **couverture** des récepteurs et qui **varie en fonction des conditions météorologiques**
- Des **vulnérabilités** du système AIS [Balduzzi et al., 2014], [Ray et al., 2014], [Windward, 2014].
- Des malversations sont possibles de la part des équipages à bord qui souhaitent cacher des activités illégales (pêche dans des zones interdites, contrebande etc.)

- L'objectif est donc d'identifier les **zones de couverture du signal** AIS qui varient en fonction du temps et de la météo

Application d'une approche hybride pour identifier les zones de non couverture de l'AIS



Exemple de zone non couverte observée en mer Egée



Exemple de zone non couverte observée au niveau du "cap de la chèvre"

Définitions

- Une **grille uniforme** est utilisée pour diviser l'espace en cellules C(i,j) pour gérer de manière plus pratique le traitement hybride.

  Pour chaque cellule C(i,j) et sur une période de temps donnée T, les mesures suivantes sont calculées :

  - Npos(i,j)(T) **le nombre de positions** qui ont été enregistrées dans C(i,j) pendant la période T.
  - Nobj(i,j)(T) **le nombre d'objets mobiles qui ont enregistré au moins une position** dans la cellule pendant la période T.
  - Ncross(i,j)(T) **le nombre d'objets mobiles qui ont traversé la cellule**. Traversé signifie que l'objet mobile a enregistré une position dans C(i,j) ou que la polyligne correspondant à la trajectoire de l'objet mobile a une intersection avec la cellule.

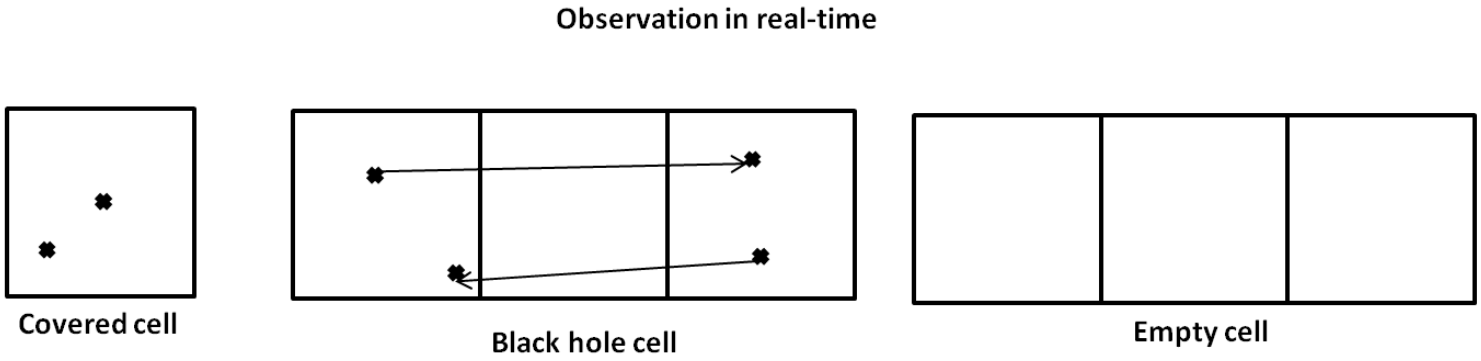
Définitions

- Les **cellules traversées** correspondent aux cellules qui ont été traversées par un objet mobile (i.e. Ncross(i,j) > 0) tandis que les **cellules vides** correspondent à celles qui n'ont pas été traversées (i.e. Ncross(i,j) = 0)

- Les **cellules couvertes** correspondent aux cellules depuis lesquelles les positions sont reçues tandis que les **cellules** dites **"Black Holes"** sont non couvertes.

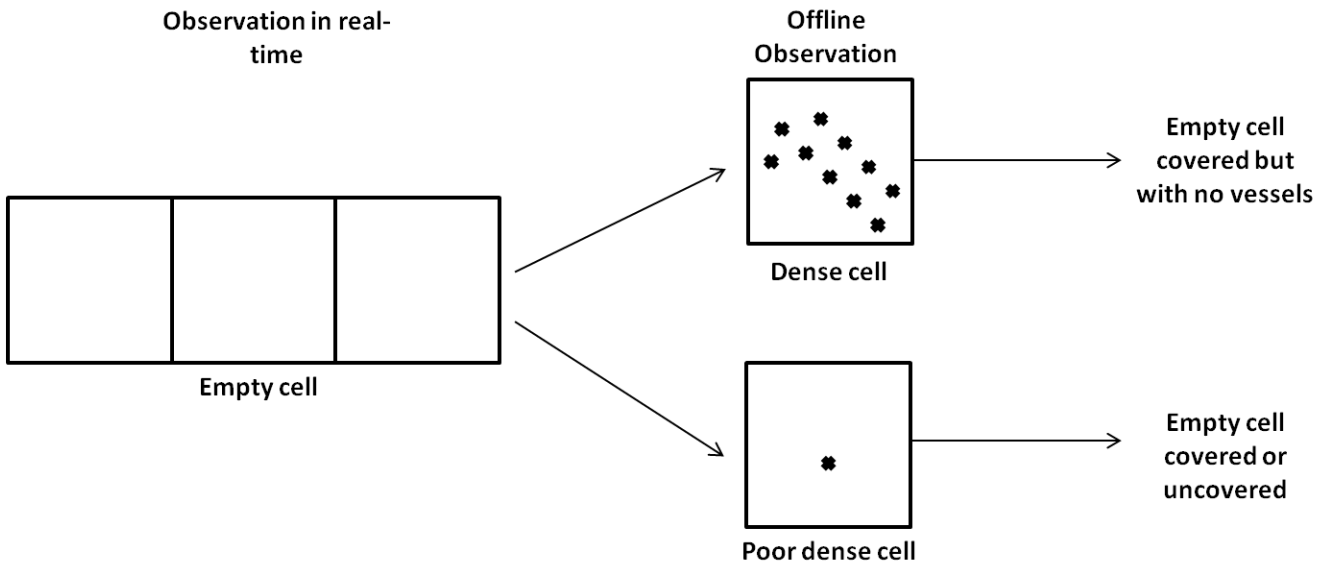
- Pour chacun de ces ensembles, des cellules peuvent être identifiées **formellement** respectivement comme couvertes ou non couvertes **utilisant uniquement les flux de données** reçus "au fil de l'eau".

Catégorisation des cellules utilisant la partie online seulement



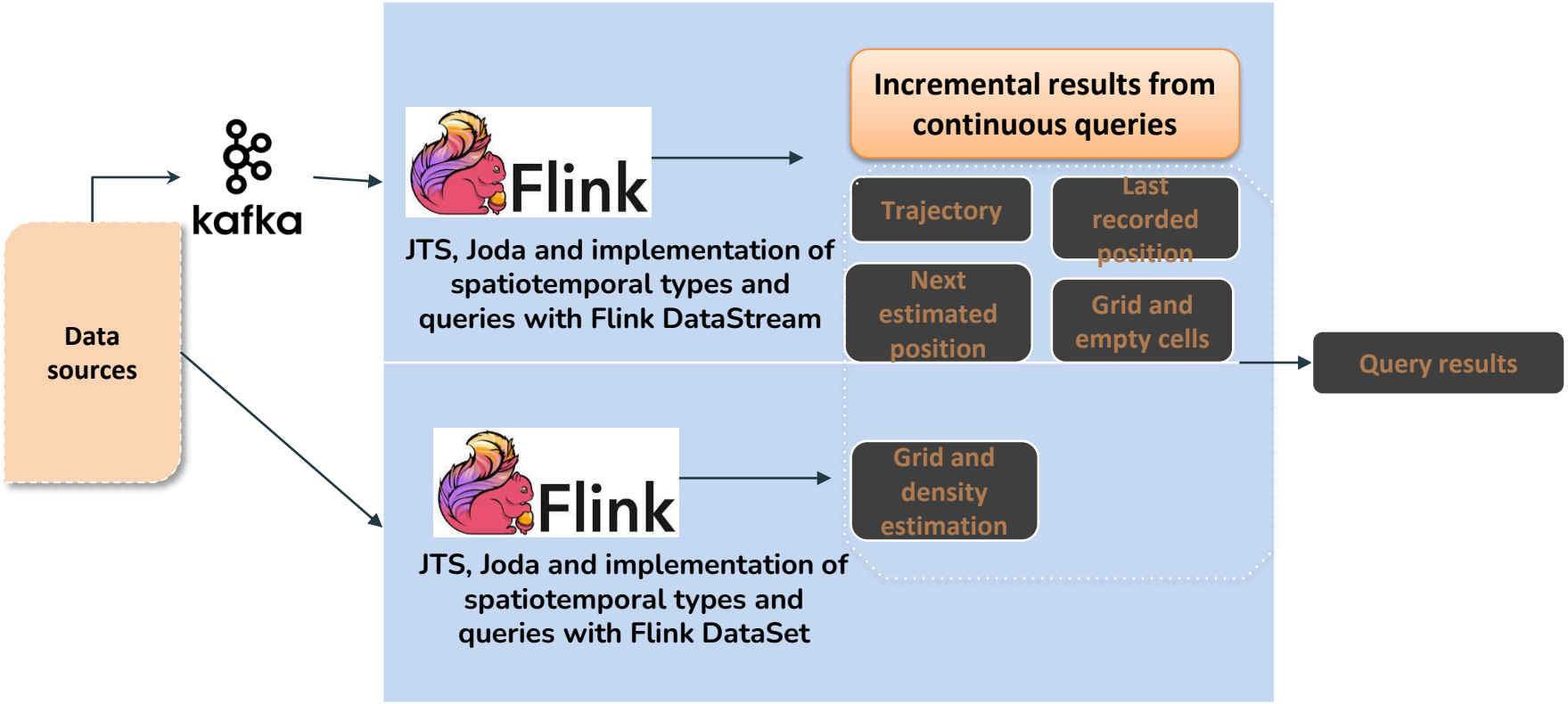
Différents cas peuvent être identifiés considérant seulement les flux temps-réel à savoir cellule vide, cellule couverte ou non couverte (“Black Hole”)

Identification par l'utilisation d'un traitement hybride

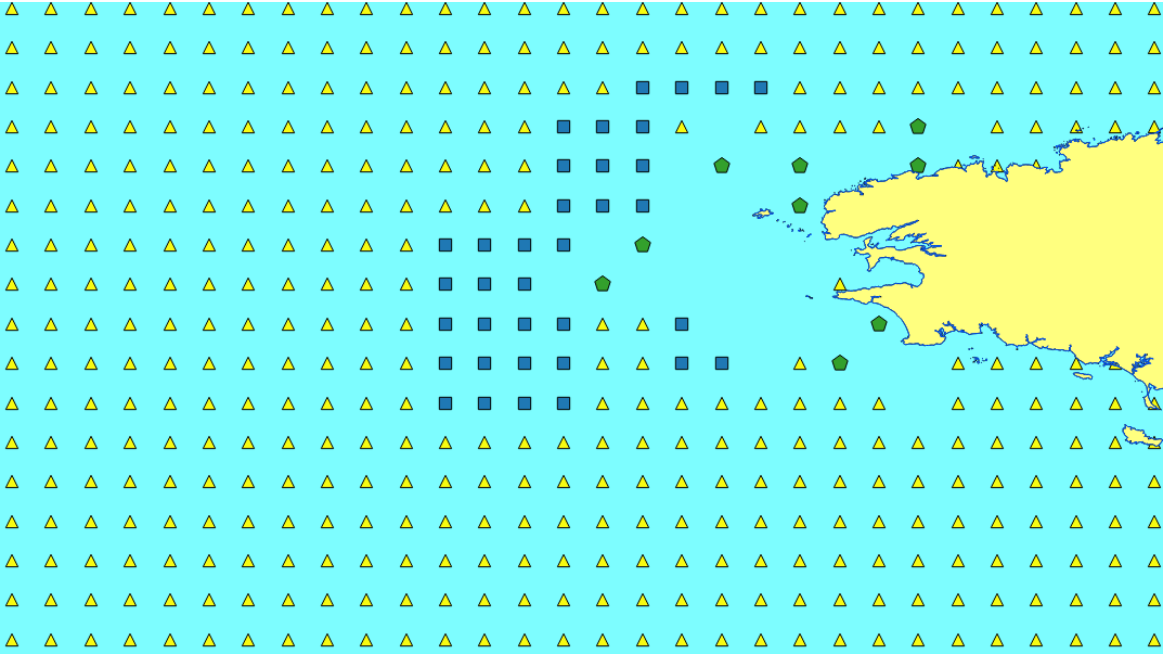


La connaissance extraite de la partie offline permet d'identifier les cellules qui sont couvertes mais qui n'ont pas été traversées pendant la dernière période de temps T considérée.

Architecture logicielle de “Geoflink”



Analyse des résultats expérimentaux



Carte de résultats pour une fenêtre temporelle de 6h. Les triangles correspondent à des cellules vides(sans certitude sur la couverture), les carrés représentent les cellules supposées couvertes et les pentagones les cellules identifiées non couvertes

Principes d’une approche hybride pour le traitement d’objets mobiles

Pour résumer l’approche hybride pour le traitement d’objets mobiles proposée repose sur plusieurs éléments :

- La notion de vues ou **synopses** qui constituent un **agrégat de données** qui sont mis à jour au fur et à mesure que les données entrent dans le système

- L’utilisation d’**une grille pour mailler l’espace** et **croiser plus facilement** des informations issues de la **partie offline avec la partie online**

- Une **plateforme distribuée** permettant de **traiter si bien des données en online qu’en offline** (flink)

Généralisation pour la recherche de motifs fréquents sur des données de trajectoires

Modélisation à base d’événements et recherche de motifs

On souhaite généraliser la notion de grille utilisée précédemment afin de :

- **Rechercher des motifs fréquents** afin d’étudier le trafic maritime à une échelle macroscopique
- Permettre d’adopter l’**approche hybride** présentée précédemment en comparant facilement les données en temps-réel et les résultats extraits de la partie offline
- **Modéliser les trajectoires à base d’événements** pour caractériser les mobilités

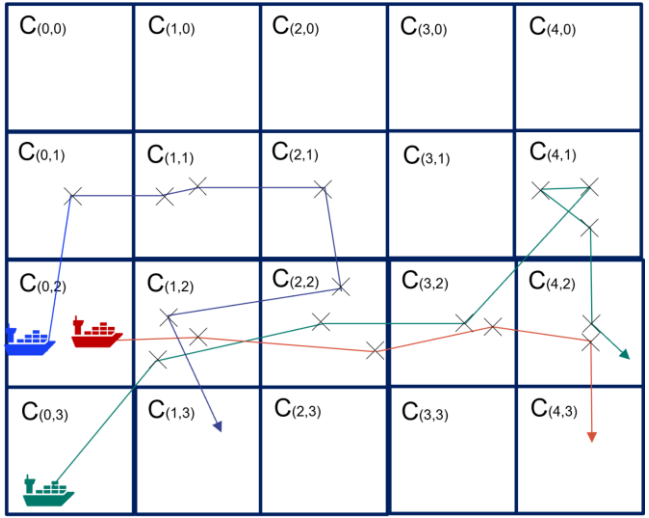
Une analyse des mobilités maritimes à différentes échelles

Analyse des mobilités maritimes

Un des enjeux scientifiques est le **suivi des mobilités**, celles-ci peuvent être analysées à **différents niveaux de représentation**. Une **approche générique** à l'aide d'une **grille** et de notions d'**événements spatio-temporels** est privilégiée :

- A un niveau **macroscopique** pour étudier les **flux de navires** à grande échelle, identifier les **chemins fréquents** et **couloirs spatio-temporels**

- A un niveau **microscopique** pour identifier des **comportements** qui peuvent être liés aux types de navires et aussi aux zones traversées pour de la **détection d'anomalies** (pêche illégale, dégazage en mer, activités illicites)



Maillage de l'espace et notion d'événements spatio-temporels pour caractériser le déplacement des navires

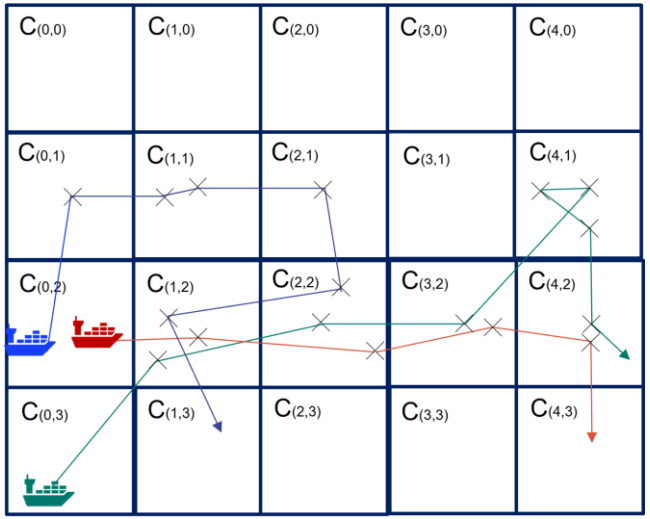
Une analyse des mobilités maritimes à différentes échelles

Analyse des mobilités maritimes à un niveau macroscopique

A un niveau **macroscopique** on peut :

Résumer le déplacement du bateau bleu à : C(0,2), C(0,1), C(1,1), C(2,1), C(2,2), C(1,2), C(1,3) à un niveau de détail macro.

On peut considérer également la succession des ports dans lesquels un navire s'est arrêté.

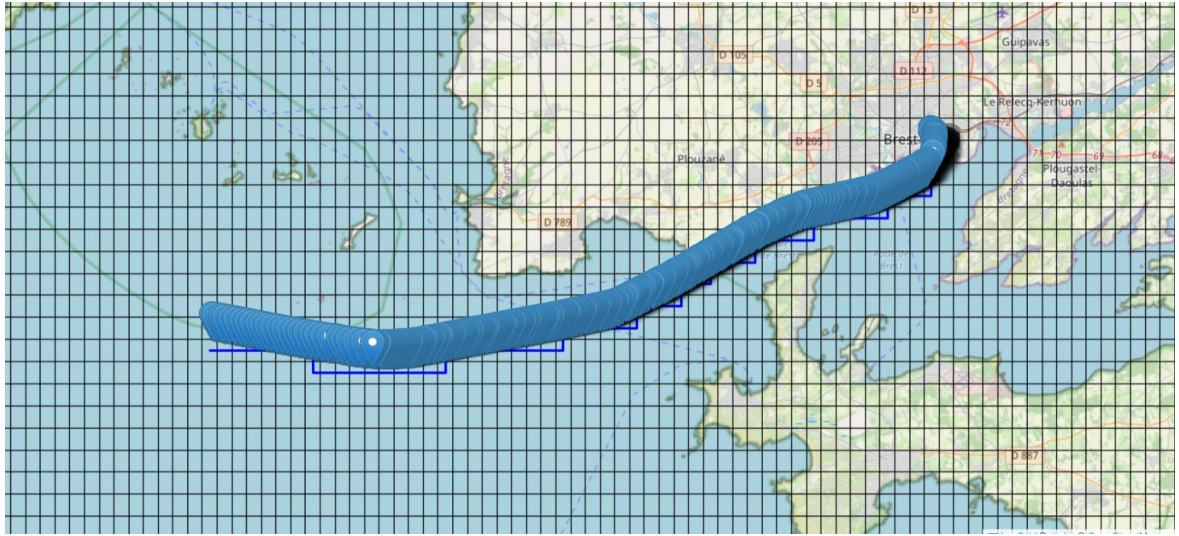


Maillage de l'espace pour caractériser le déplacement des navires

Une analyse des mobilités maritimes à différentes échelles

Analyse des mobilités maritimes à un niveau macroscopique

Exemples sur des données réelles



Des milliers de positions successives d'un navire peuvent être résumées à un nombre limité de cellules

Une analyse des mobilités maritimes à différentes échelles

Analyse des mobilités maritimes à un niveau macroscopique

Intérêt de l'approche :
-Gain significatif en termes de stockage de données et comparaison de trajectoires via des algorithmes de **comparaison de motifs** (i.e. distance de Jaccard)
-Déterminer des **chemins fréquents** à l'aide d'algorithmes de recherche de motifs fréquents (fpgrowth, prefixspan)
-Comparer les motifs par jour, par semaine afin d'**identifier des périodicités** et l'**évolution du trafic maritime**

Inconvénients :
-Fortement dépendant de la **taille des cellules** considérées
-Nécessite une forte densité du trafic pour avoir une analyse pertinente

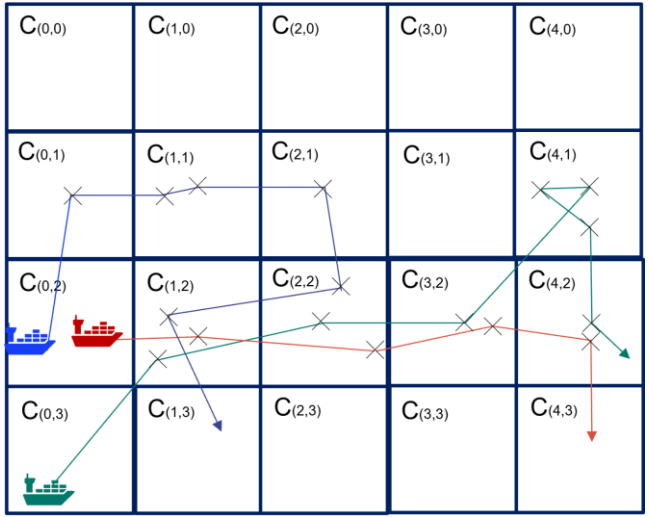
Une analyse des mobilités maritimes à différentes échelles

Analyse des mobilités maritimes à un niveau microscopique

A un niveau **microscopique** on peut : identifier des **comportements** par types de navires et par zones traversées

A un niveau de détail plus fin, on peut modéliser la trajectoire du navire bleu comme suit :
C(0,2) (vFixe, cFixe), C(0,1) (vBaisse,cChangeDroiteF, Vaugmente,cFixe), C(1,1) (vFixe, cChangeGaucheL, vAcc, cFixe), C(2,1) (vBaisse,cChangeDroiteF, Vaugmente,cFixe)

->Utilisation de la notion d'événements pour caractériser le déplacement des navires



Notion d'événements spatio-temporels pour caractériser le déplacement des navires

Une analyse des mobilités maritimes à différentes échelles

Analyse des mobilités maritimes à un niveau microscopique

Trois types d'événements pour des objets mobiles sont distingués :

Les **événements basiques pour un objet mobile**, correspondent à une action de ce dernier : entrer dans une zone, accélérer, décélérer, changer de direction.

Les **événements complexes pour un objet mobile**, sont une concaténation de plusieurs événements basiques, et permettent de décrire et caractériser le comportement d'un objet mobile dans une zone. Par exemple, le bateau fait demi-tour, le bateau est en situation de pêche ...

Les **événements complexes pour un groupe d'objets mobiles**, ceux-ci permettent d'observer des effets de groupe ou des interactions entre différents objets mobiles et se composent de différents événements complexes pour un objet mobile.

Une analyse des mobilités maritimes à différentes échelles

Analyse des mobilités maritimes à un niveau microscopique

Intérêt de l’approche :

- Caractériser le déplacement des navires à une échelle plus fine afin de détecter des anomalies
- Permet d’identifier des comportements types de navires suivant la zone traversée
- Utilisation de la même approche par recherche de motifs fréquents

Conclusion et perspectives

Une analyse des mobilités maritimes à différentes échelles

Conclusion

Une approche à l'aide d'une **grille** et d'**événements** a été proposée afin de permettre à la fois un **traitement hybride des données d'objets mobiles** et d'extraire de l'information à l'aide d'algorithmes de **recherche de motifs fréquents**

Perspectives

Utilisation d'index h3 utilisés par uber afin de mailler l'espace

Adaptation d'algorithmes de recherche de motifs fréquents pour les données de positions

Croisement avec des données de nature différente (courants, vents, marées, météo)

Algorithmes de prédiction