

ARTICLE**Valorisation des bases de données trafic pour la prévision de la congestion**

Thomas Courbon, Adjoint au chef de service, Cerema Méditerranée
Étienne Hans, Chargé d'étude et de recherches, Cerema Centre-Est

Communication : com099

Thématique : Infrastructures intelligentes pour une mobilité connectée

1/ Enjeux et objectifs

Les gestionnaires de réseaux ont fréquemment à leur disposition d'importantes bases de données de trafics historiques. Celles-ci sont peu utilisées hormis pour les études de trafic occasionnelles. Pourtant, ces données permettent d'élaborer de puissants indicateurs pour guider les politiques d'exploitation sur ces réseaux. Encore faut-il être en mesure de mettre au point des indicateurs pertinents.

Cet article présente quelques exemples d'indicateurs relatifs à la probabilité d'occurrence de la congestion. Nous étudions d'une part la faisabilité du calcul de tels indicateurs à partir des sources de données existantes, d'autre part leur apport par rapport aux indicateurs usuels.

Objectif : utiliser les données pour étudier la congestion historique et anticiper les congestions futures en temps réel sans utiliser de modèle d'écoulement du trafic.

2/ Etat de l'art, descriptif de la situation, données existantes

Les données possédées par les gestionnaires de réseaux sont principalement celles issues de stations de comptage du trafic, le plus souvent des boucles électromagnétiques. Ces données dites QTV renseignent sur le débit, le taux d'occupation et la vitesse moyenne du flot de véhicule. Elles apportent parfois des informations plus riches en distinguant les trafics par voie ou encore en distinguant les catégories de véhicules. Depuis peu, les gestionnaires ont parfois des données issues de systèmes FCD.

3/ Méthodologies, idées, techniques et méthodes innovantes

L'utilisation des données historiques permet d'estimer la probabilité d'occurrence de congestion en différents points d'observation. Cet indicateur, la fréquence de congestion historique, est très simple à calculer. Il peut être visualisé sur un planning de congestion, outil que les exploitants n'ont pas toujours à leur disposition.

Les plannings de congestion ponctuels peuvent être généralisés sur l'ensemble du réseau. Des cartes du réseau avec la probabilité d'apparition de la congestion peuvent alors être générées. Elles pourraient être particulièrement utiles aux exploitants, notamment pour la répartition *a priori* des activités des agents sur le réseau.

Les données sont tout d'abord traitées indépendamment pour chaque boucle. Ce traitement comporte une phase de filtrage des données aberrante, une imputation des données manquantes, et d'un lissage des chroniques.

Plusieurs diagrammes peuvent alors être générés pour mettre en évidence les trafics mesurés

par une boucle donnée :

- Les chroniques moyennes des débits mesurés par jour de semaine ;
- Un planning des fréquences de ralentissements et de congestions ;
- Un diagramme fondamental dans le plan (concentration, débit) ;
- Un diagramme de répartition des trafics par voie, le cas échéant.

Ces diagrammes sont regroupés sur un "tableau de bord" d'une boucle, comme montré sur la Figure 1.

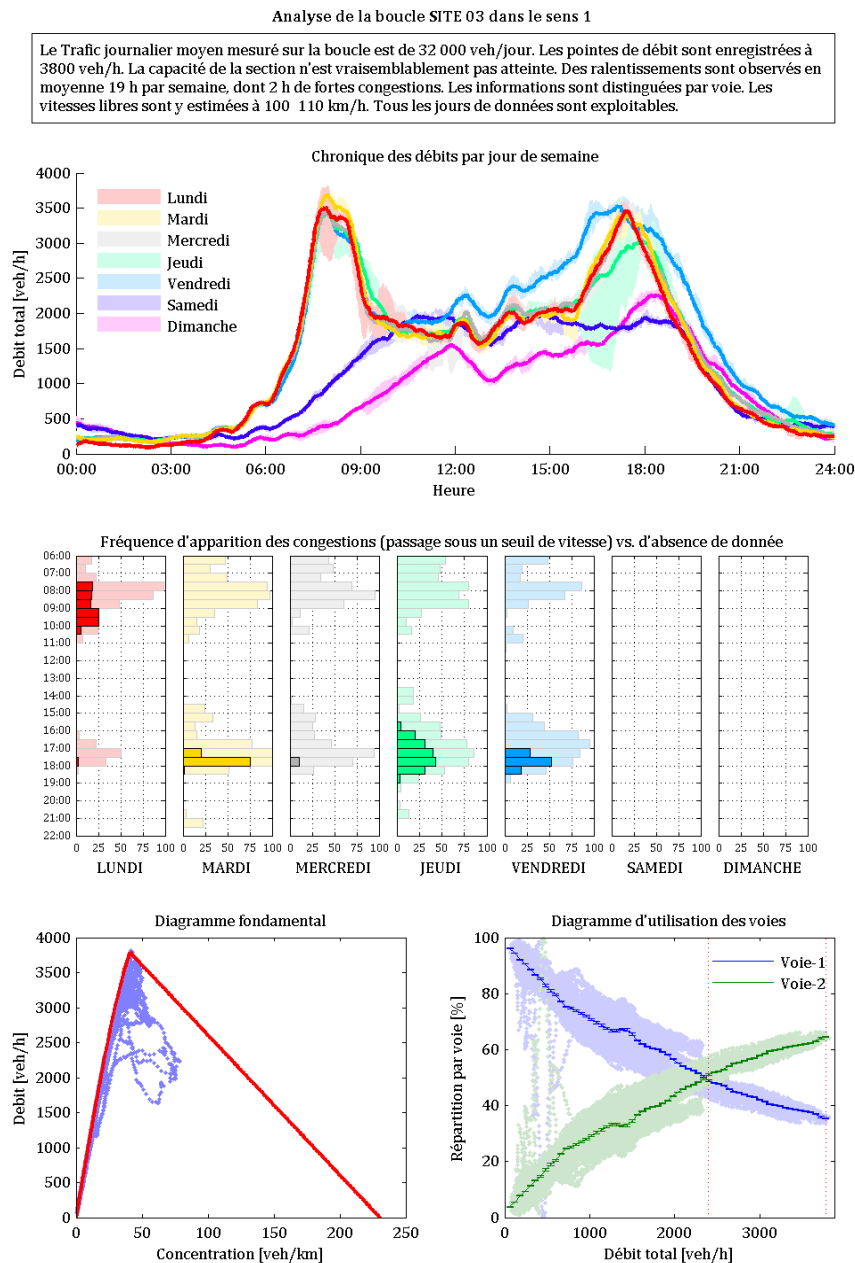


Figure 1. Exemple de Tableau de bord pour l'analyse d'une boucle isolée.

Pour faire apparaître la dynamique du trafic sur les différents jours, les données doivent être traitées dans leur globalité. Le choix du diagramme espace-temps réalisé à partir des vitesses mesurées a été fait pour sa lisibilité.

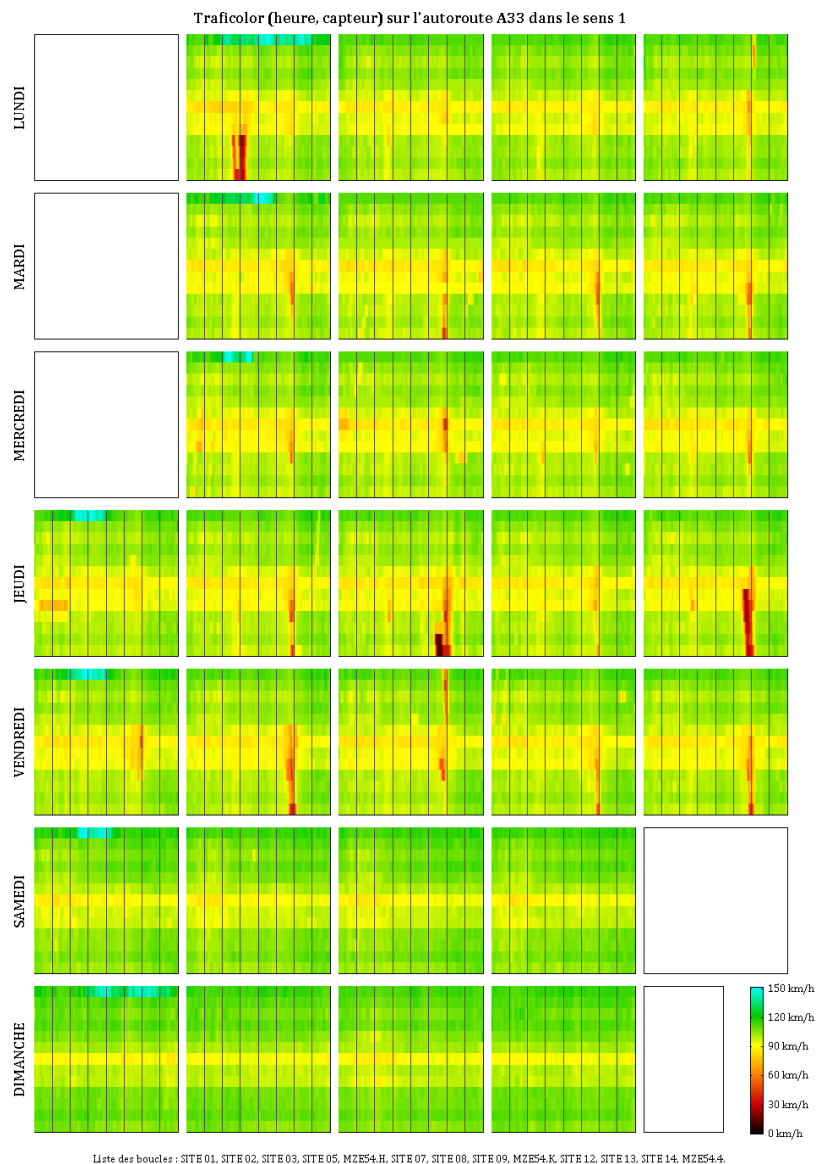


Figure 2. Diagrammes espace-temps journaliers pour un sens de circulation sur l'A33 sur l'ensemble du mois de septembre.

4/ Résultats théoriques ou expérimentaux, interprétations, déploiement et voies de progrès

1/ Tableau de bord des boucles

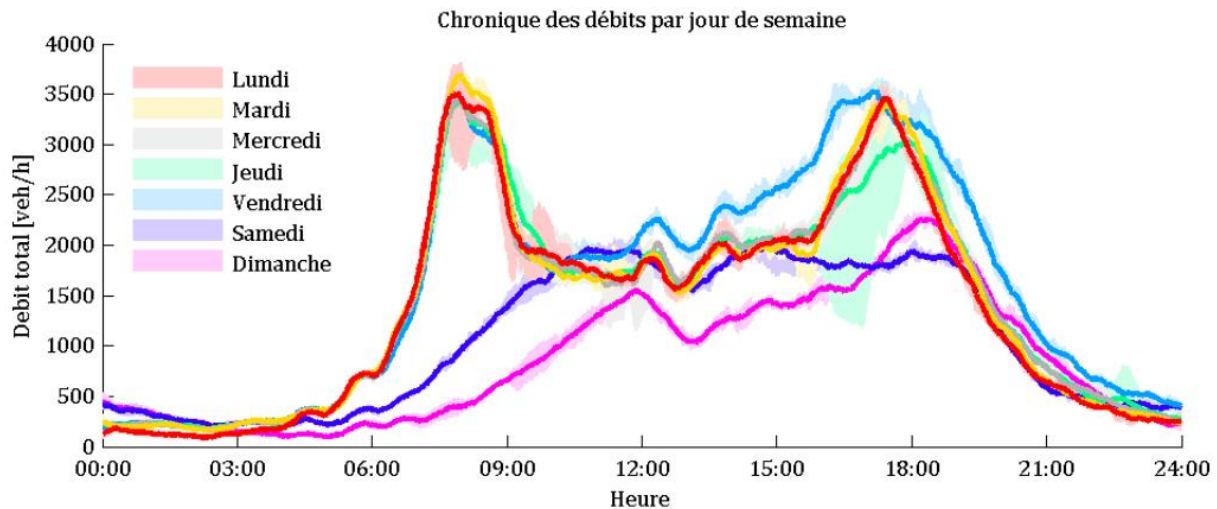


Figure 3. Chronique des débits (moyenne, enveloppe min - max) par jour de semaine.

L'indicateur de fréquence de congestion historique est une information nouvelle que les exploitants n'avaient jusqu'à présent que grâce à leur expertise de leur réseau. Il est facilement améliorable par l'utilisation des techniques décrites précédemment. Enfin, il peut être utilisé sur les données détenues par les exploitants, sans nécessiter le recours à des recueils de données complémentaires, souvent coûteux.

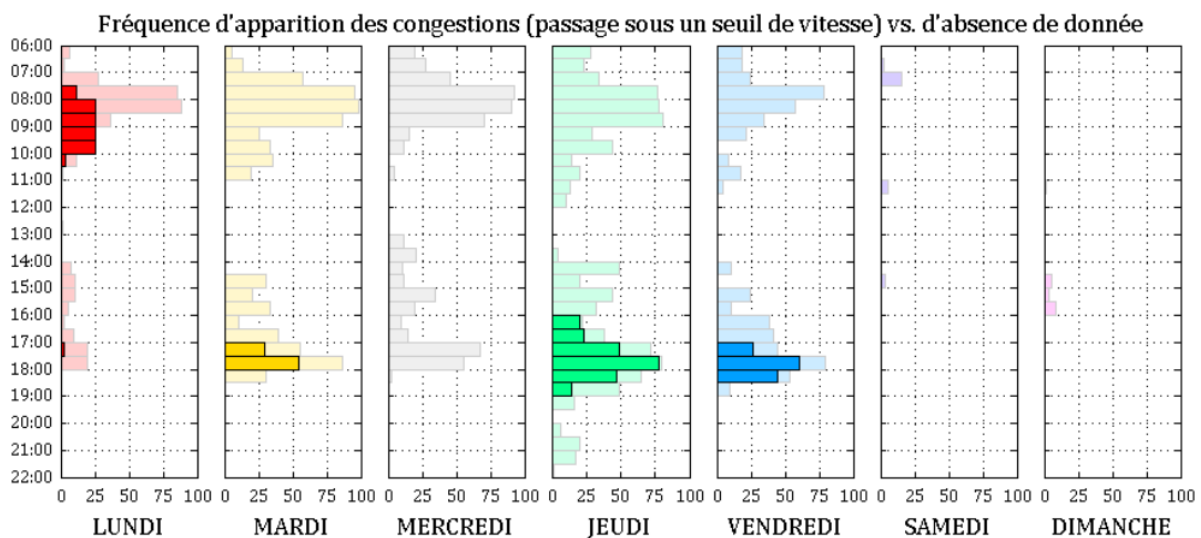
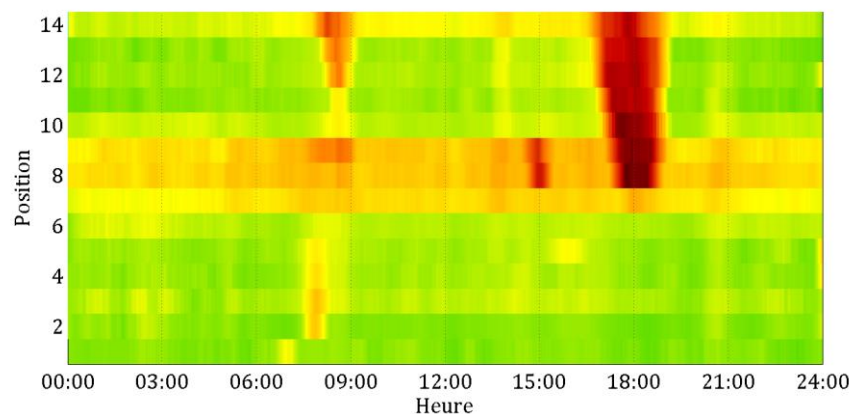


Figure 4. Planning des fréquences de congestion.

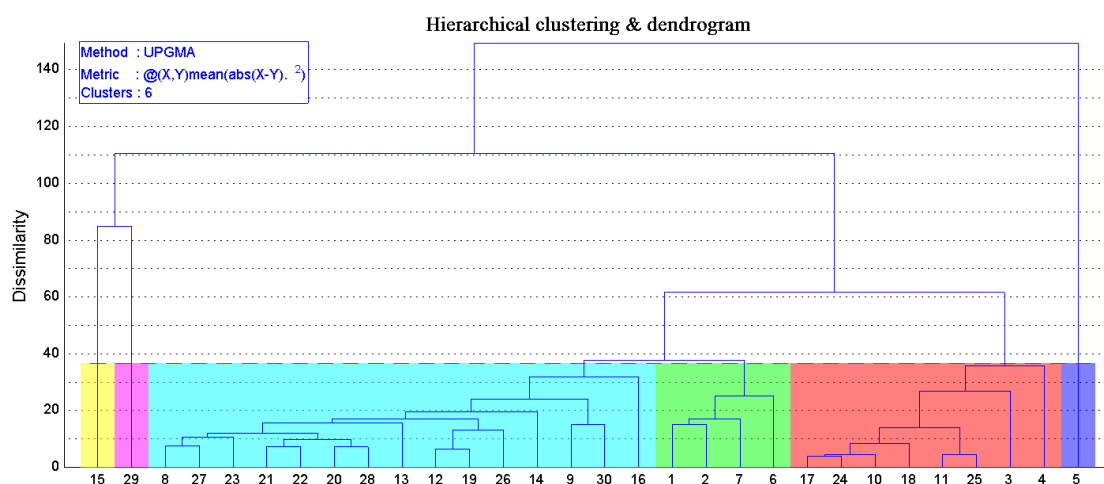
Les deux derniers graphiques du tableau de bords sont des indicateurs classiques et bien connu, le diagramme fondamental ainsi que la répartition par voie en fonction du débit observé.

2/ *Interprétation des diagrammes espace-temps***Figure 5.** Diagramme pour un jour donné : détection des congestions.

Les diagrammes espace-temps permettent d'observer la formation d'une congestion ainsi que son évolution spatiale et temporelle. Sur l'exemple présenté en Figure 5, une congestion se forme à l'aval de la section observée et remonte sur l'A33 jusqu'au niveau de la boucle n°8. La congestion atteint cette position aux alentours de 17h30 et perdure jusqu'à environ 18h30 avant de se dissiper en remontant progressivement jusqu'à la fin de la section.

3/ *Réalisation d'une Classification Ascendante Hiérarchique (CAH)*

Il est intéressant de regrouper des jours entre eux. Intuitivement, on aimerait que les jours où les conditions de trafic sont similaires soient regroupés. De nombreuses méthodes de partitionnement (*clustering*) existent. Dans cet article, nous utilisons une classification ascendante hiérarchique, CAH. Cet algorithme consiste à regrouper les points qui sont les plus proches selon une certaine métrique, c'est-à-dire selon une certaine méthode de calcul de la proximité entre deux points. La distance euclidienne est un exemple de métrique possible. Le partitionnement initial place chaque jour dans une classe indépendante. Les classes sont alors itérativement fusionnées (les deux plus proches à chaque itération) jusqu'à n'obtenir plus qu'une seule classe contenant tous les jours. La distance entre les deux classes fusionnées est appelée similarité. On peut alors tracer un dendrogramme, comme celui de la Figure 6, synthétisant tous les résultats de l'algorithme.

**Figure 6.** Dendrogramme obtenu à partir d'une CAH pour partitionner les jours de semaine.

Les résultats de la CAH obtenue sur le jeu de données utilisé également pour la Figure 2 peuvent ainsi être analysés aisément. Trois jours isolés (5, 15 et 29), sont des jours avec des congestions exceptionnelles. Les jours 15 et 29 peuvent être regroupés sachant qu'ils ont les deux une congestion le soir, alors que le 5 a une congestion le matin. Les jours en bleu sont uniquement des jours de semaine. Les jours en vert sont également des jours de semaine, mais sans aucun ralentissement. Les jours en rouge sont les jours de weekend.

Le procédé de CAH est donc très efficace pour classer automatiquement les jours de semaine. En effet, la discrimination entre jours ouvrés et week-end est obtenu dès la première segmentation après isolement des jours avec congestion exceptionnelle. Notons que seules les vitesses sont utilisées pour ce procédé. De plus, l'algorithme n'a pas connaissance de la structure temporelle de ces vitesses. Sa mise en oeuvre est donc extrêmement aisée. Enfin, il serait possible de le rendre plus efficace en raffinant la métrique permettant de définir la proximité entre deux jours donnés.

5/ Conclusions

Les données actuelles ne sont pas encore exploitées à leur potentiel maximal. Hors, dans un contexte de raréfaction des moyens, l'optimisation de l'usage des infrastructures routières passe par l'utilisation optimale des sources et bases de données existantes. Les progrès techniques permettent des exploitations des données et la production d'indicateurs auparavant hors d'atteinte. Néanmoins, il est nécessaire de s'assurer que l'utilisation d'indicateurs complexes est utile et que l'amélioration apportée justifie les moyens engagés. Les indicateurs présentés dans cette contribution peuvent tous être déterminés automatiquement et à moindre coût.

L'utilisation de données de différents types (données FCD) et l'enrichissement des données de trafic par d'autres types de données contextuelles (météo, événements susceptible de générer un trafic particulier) sont des pistes d'améliorations envisageables. Cette démarche n'a pu être effectuée ici, la zone étudiée ici ne présente pas d'événement météorologique ayant un impact mesurable sur le trafic.

CURRICULUM VITAE DU ou DES ORATEUR(S)



Thomas Courbon est adjoint à la chef du service Connaissance des Trafics et Accidentologie de la Direction Territoriale Méditerranée du Cerema. Diplômé de l'ENTPE, spécialité transport et infrastructures, il a effectué toute sa carrière dans le domaine du transport et du trafic routier. Depuis 4 ans, Il travaille sur la thématique de la donnée trafic et des équipements dynamiques de la route. L'équipe qu'il pilote a une activité centrée sur les équipements de recueil de données trafic. En particulier, il a développé l'utilisation des technologies Bluetooth et LAPI pour la production de données de temps de parcours pour répondre aux besoins des maîtres d'ouvrage locaux. Il est également impliqué dans les champs de l'*open-data* et des ITS, notamment les véhicules à délégation de conduite, et assure des formations à destination des professionnels de l'ingénierie du trafic.



Étienne Hans est chargé d'études et de recherche en régulation dynamique des réseaux de transport au Département Mobilité du Cerema Centre-Est. Il a commencé son parcours professionnel dans le milieu du transport par une thèse préparée et soutenue au LICIT (UMR ENTPE/IFSTTAR) sur l'exploitation dynamique conjointe des transports en commun et du trafic routier en milieu urbain. Aujourd'hui au Cerema, son domaine d'activité est l'étude du trafic routier en milieu périurbain. Ses travaux visent à évaluer, a priori ou a posteriori, des stratégies de régulation dynamique du trafic. Pour cela, ses activités s'orientent autour de l'analyse de données et de la modélisation à moyenne ou grande échelle. En plus de ces activités, il donne des cours à l'ENTPE où il est responsable du cours de Probabilités et de Statistiques. Cela lui permet d'apporter une expertise plus poussée sur l'analyse de données de trafic.