



# Desjardins

## Groupe d'assurances générales

---

### Problèmes types dans le domaine des assurances

Atelier de maillage industriel du CRM  
27 janvier 2017

---

{ C E M R } *Étienne Plante-Dubé, FICA, FCAS  
DP Modélisation et Recherche*

**COOPÉRER**  
POUR CRÉER L'AVENIR



## Centre d'expertise en modélisation et recherche

- Qui sommes nous?
- Qu'est-ce qu'on fait?
- Quelles applications?

# Qui sommes nous?

3

Le centre d'expertise en modélisation et recherche c'est;

- 2 directions, 16 professionnels
  - actuaires
  - statisticiens
  - spécialistes de données
  - spécialistes en Systèmes d'Information Géographique
- 12 MSc, 1 PhD, 3 FCAS
  - Recrutement via de nombreux stages et collaborations universitaires

Une équipe au service de nos partenaires internes

# Qu'est-ce-qu'on fait?

4

Mandat : innovation en analytique

Ce qu'on fait:

- Modèles prédictifs ou explicatifs
- Systèmes d'information géographique (SIG)
- Vigie et innovation en analytique avancée
- rôle-conseil/choix d'outils et formation

# Quelles applications?

5

Tous les secteurs d'activités qui peuvent bénéficier de notre expertise pour exploiter leurs données

- Marketing
- Tarification
- Actuariat corporatif
- Gestion du risque
- Distribution
- Indemnisation
- Souscription
- Expérience-client

~ 50 initiatives/an (50 j/p par initiative)

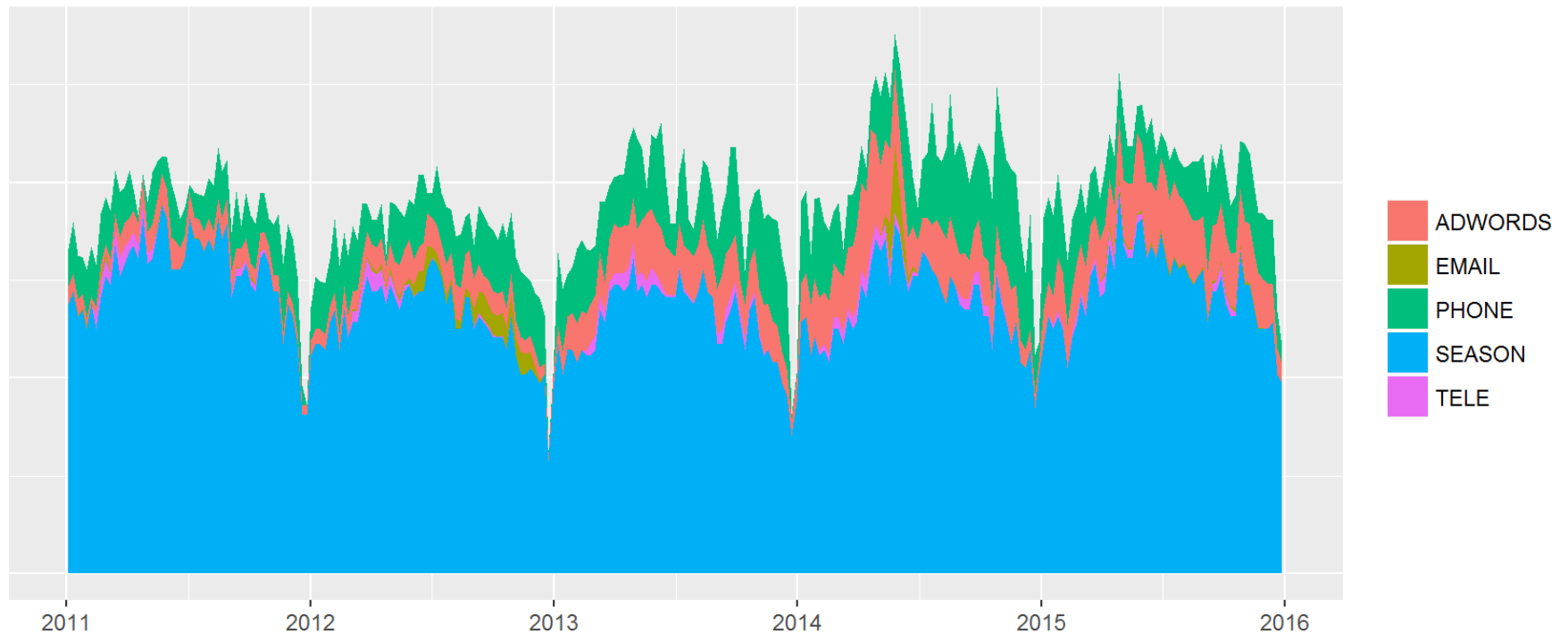
# Problème Type #1

## Marketing Return on Investment (MROI)

*"Half the money I spend on advertising is wasted;  
the trouble is I don't know which half"*  
- John Wanamaker, US department store merchant  
(1838 - 1922)

- Nos investissements en marketing représentent plusieurs millions de dollars annuellement.
  - Médias de masses
  - Marketing relationnel
  - Digital
- Ce qu'on peut « tracker » est facilement mesurable/optimisable via des modèles prédictifs (réponse, uplift, CLV...)
- Certains investissements sont difficiles à « tracker »

Mesurer la contribution des différents médias à nos soumissions





Soumissions ou ventes  $\sim f(\text{saison, notoriété, effort marketing, compétition})$

Tellis (2006) *Modeling Marketing Mix:*

- Effet courant et rémanence (impact qui dure dans le temps)
- Forme de la réponse (retours décroissants)
- Effets de la compétition (bruit réduit efficacité de la campagne)
- Effets dynamiques (wearin, wearout)
- Effet de contenu (émotions – pas de données)
- Effet de média (TVA vs RDS – pas de données)

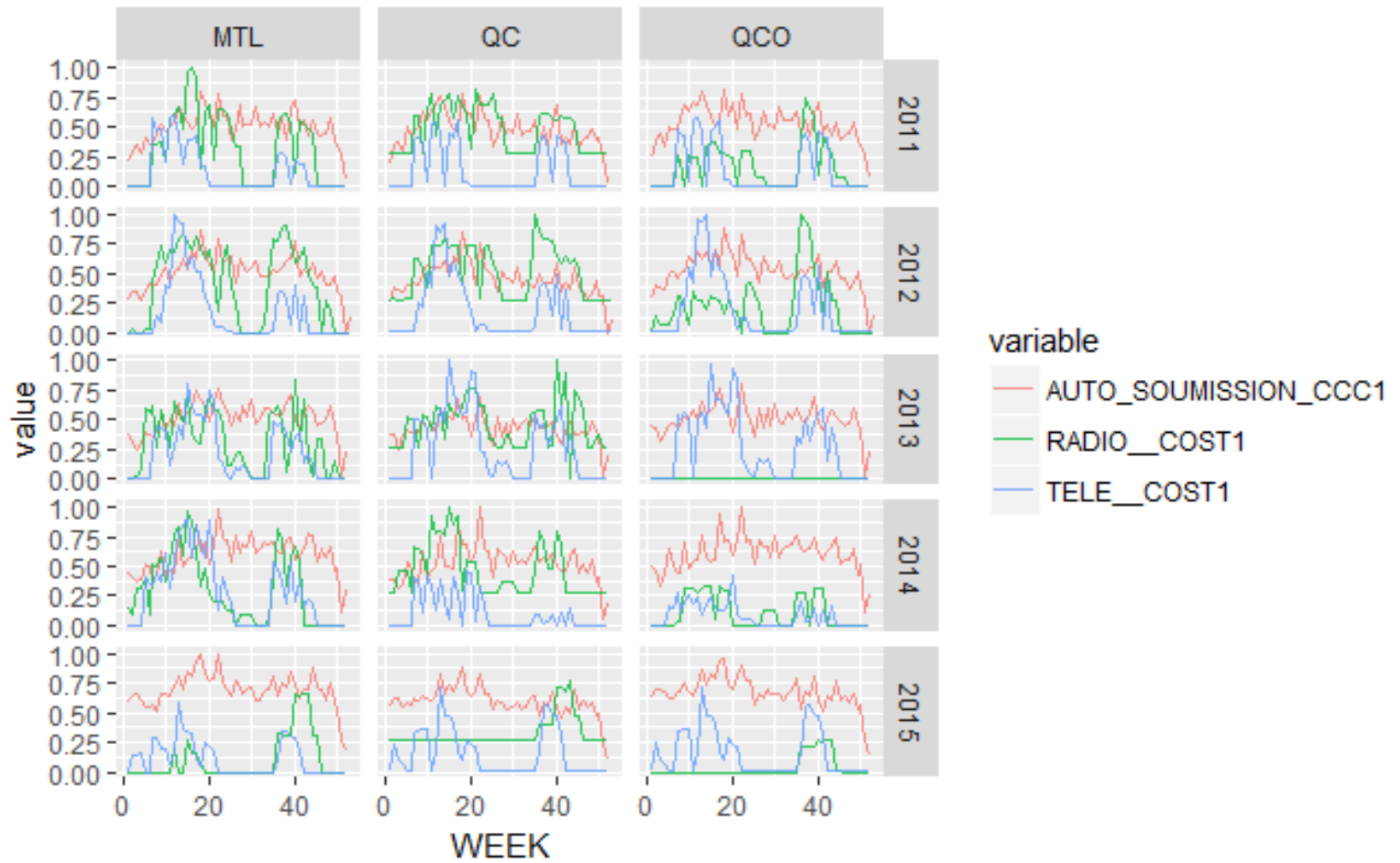
En plus :

- Synergies (interaction entre 2 investissements)
  - Modèle d'attribution (pub télé -> recherche google -> clic adwords -> soum)
-

- Info détaillées de nos soumissions
- Investissements marketing
  - \$, occurrences, pebs, audience, clics, impressions
  - géographie
    - Télé & radios par marché
    - Magazines/journaux sont nationaux
    - Emplacement géographique des panneaux
    - Codes postaux des envois postaux
    - Web via l'adresse Ip ou le pixel
  - Compétition

# Aperçu de quelques séries chronologiques

11



- Plusieurs influences externes à considérer
  - Notoriété ( Qc vs. Ontario)
  - Effet de Halo (State Farm & Desjardins)
  - Cycle d'assurance et compétition
- Saisonnalité et colinéarité : modèle instable
- Processus d'acquisition: first-touch vs. last-touch  
(TV -> Recherche Web -> PPC -> Soumission)
- Effet court-terme vs. long-terme

- Régressions (GLM, GAM...)
- Séries chronologiques
- Équation structurelles?
- Modèles mixtes?
- Approches bayésiennes?

## Problème Type #2

*ajusto*

# Contexte

Ajusto est notre programme d'assurance télématique  
L'information est recueillie via une app sur le téléphone

## Données

- Heure de conduite
- Distance parcourue
- Statistiques sur les accélérations, les freinages, les dépassement de vitesse, ...

## Objectifs

- Récompenser les bons conducteurs par des rabais sur leur prime
- Identifier les conducteurs les plus susceptibles d'avoir des accidents
- Améliorer notre segmentation du risque
- Identifier les comportements à risque
- Réduire les accidents

# Données disponibles

## Données d'assurance

- Pour chaque véhicule, on a
  - l'unité d'exposition sur la période d'assurance
  - Les variables traditionnelles de tarification en assurance (âge conducteur, groupe de véhicule, dossier de conduite, code postal, etc)
  - Le nombre de sinistres et le coût moyen des sinistres observés.

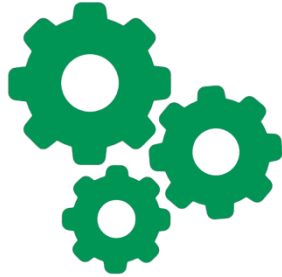
## Données télématique

Les données collectées suite à l'adhésion au programme

- Par conducteur
- Un record par trajet (du data plus granulaire serait aussi disponible)
- Sur chaque trajet, on relève des variables reliées à la conduite (accélération, vitesse, freinage, distance parcourue, ...)
- Pour certains trajets, aucune donnée télématique n'est collectée



# Estimation et scoring



## Estimation du modèle

- Construction d'une base de données par période de police.
- Agrégation des données de télématiques pour construire un profil de conducteur
- Estimation d'un modèle de fréquence



## Scoring du modèle

Contrainte de donner un score par trajet. Donc on agrège d'abord les variables de télématiques pour construire un profil « instantané » du trajet



**Problème : Entre l'estimation et le scoring, l'ordre des agrégations (variables et trajets) est inversé.**

# Problématique

## Défis :

Comment utiliser le mécanisme d'agrégation des trajets scorés dans le processus d'estimation du modèle ?

Comment tout traiter en une fois pour ne plus faire d'approximation ?

## Contraintes :

- Indépendance entre le score donné à chaque trajet
- Enjeu sur la taille de la base de données