

# LES MÉTHODES DE L'INDICE DES RISQUES DE CORRUPTION ET DE LA DÉTECTION D'ANOMALIES DANS LES MARCHÉS PUBLICS À PARTIR DES DONNÉES DU SEAO

**César Garzon**

*École nationale d'administration publique (ENAP)*

**Louis-Alexis Pelletier-Dubé**

*Université Laval*

**Pierre-André Hudon**

*Université Laval*

**Michael Morin**

*Université Laval*

# LE CONTEXTE

- Trois raisons pour vouloir appliquer des méthodes de détection de la corruption aux marchés publics basées sur l'analyse des données massives :
  - 30 % des investissements gouvernementaux (13 mille milliards de US dollars en 2020; 15 à 20 % du PIB-pays de l'OCDE) <sup>1</sup>
  - Les risques de corruption sont élevés dans l'approvisionnement public (20 à 25 % des montants engagés) <sup>2</sup>
  - Le secteur est riche en données et en *big data*

1. <https://www.open-contracting.org/why-open-contracting/>

2. <https://www.oecd.org/gov/ethics/Corruption-Public-Procurement>

# LE PROBLÈME

- Traditionnellement, la détection de la corruption et sa quantification s'est avérée très difficile :
  - Les mesures de corruption étaient basées sur les perceptions des individus (IPC de TI; *Worldwide Governance Indicators* de la Banque mondiale) <sup>3,4</sup>
  - C'est seulement au début des années 2010 qu'apparaissent les premiers indicateurs objectifs, notamment avec les travaux de Fazekas *et al.*, (2013a et 2013b) du *Corruption research center* de Budapest <sup>5</sup>
- Cette difficulté d'obtenir des mesures objectives de la corruption a été à la source des faiblesses lorsqu'il s'agit d'évaluer les risques de corruption et de déployer des mesures d'atténuation dans le cadre de la gestion des contrats publics (Johnsøn, 2015)

3. [Transparency International's 2020 Corruption Perceptions Index to be published January 28, 2021 \(corruptionanonymous.org\)](#)

4. [Worldwide Governance Indicators \(WGI\) | Data Catalog \(worldbank.org\)](#)

5. [The Institute | CRCB](#)

# L'OPPORTUNITÉ

- Les deux méthodes proposées ici, l'indice des risques de la corruption (CRI) et de la détection d'anomalies (IF), s'inscrivent dans les approches des indicateurs objectifs.
- Les méthodes :
  - utilisent essentiellement les informations sur les appels d'offres ouverts en provenance des bases de données gouvernementales
  - développent des mesures quantifiables aidant à détecter la corruption dans les processus de passation des marchés publics
  - contribuent à l'évaluation et la quantification des risques de corruption, par la détection des contrats suspects
  - servent à combler un vide actuel de la gestion de risques de corruption dans les contrats publics

# LES DONNÉES OUVERTES DU SEAO <sup>6</sup>

- Le secrétariat du Conseil du Trésor rassemble dans le SEAO les avis et informations sur les contrats publics au Québec depuis 2009:
  - Plus de 200 000 contrats
  - Plus de 600 000 soumissions
- La base de données contient six types de fichiers:
  - Contrats conclus (avis et fournisseurs)
  - Les dépenses supplémentaires
  - L'information finale des contrats
  - Trois fichiers révisions

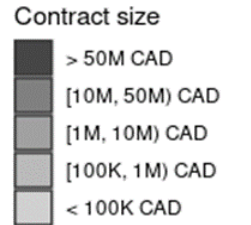
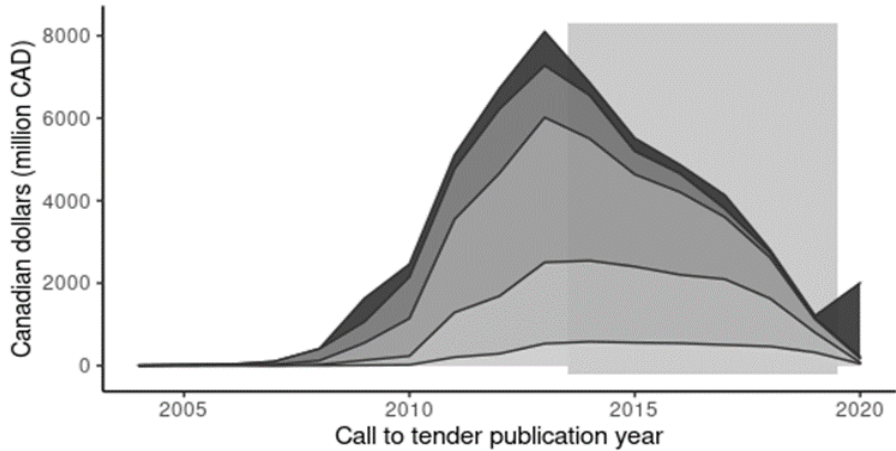
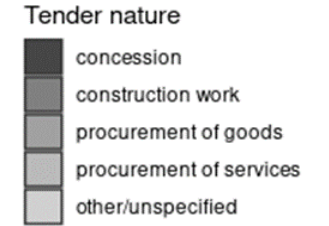
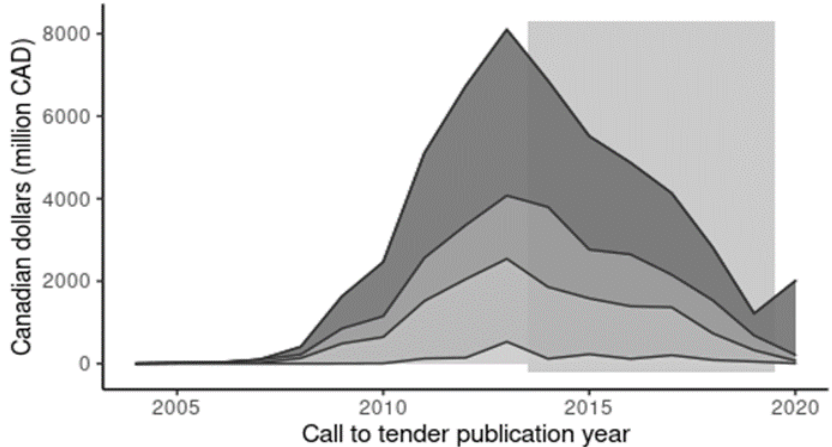
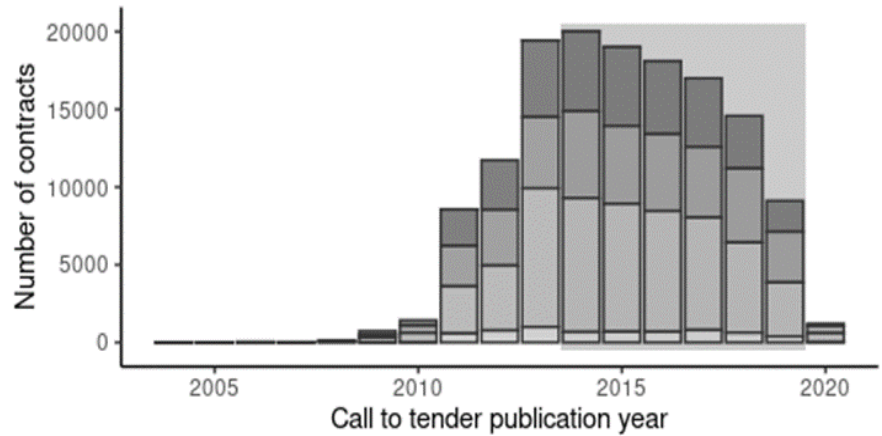
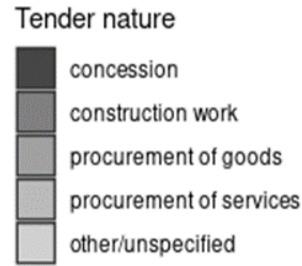
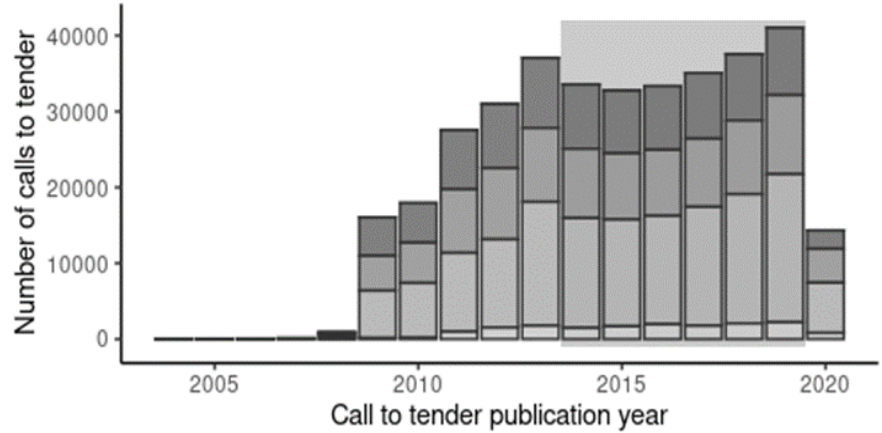
6. <https://www.donneesquebec.ca/recherche/dataset/d23b2e02-085d-43e5-9e6e-e1d558ebfdd5/resource/af41596c-b07f-4664-82c8-577e1ef9a6f3/download/seao-specificationsxml-donneesouvertes-20171010>



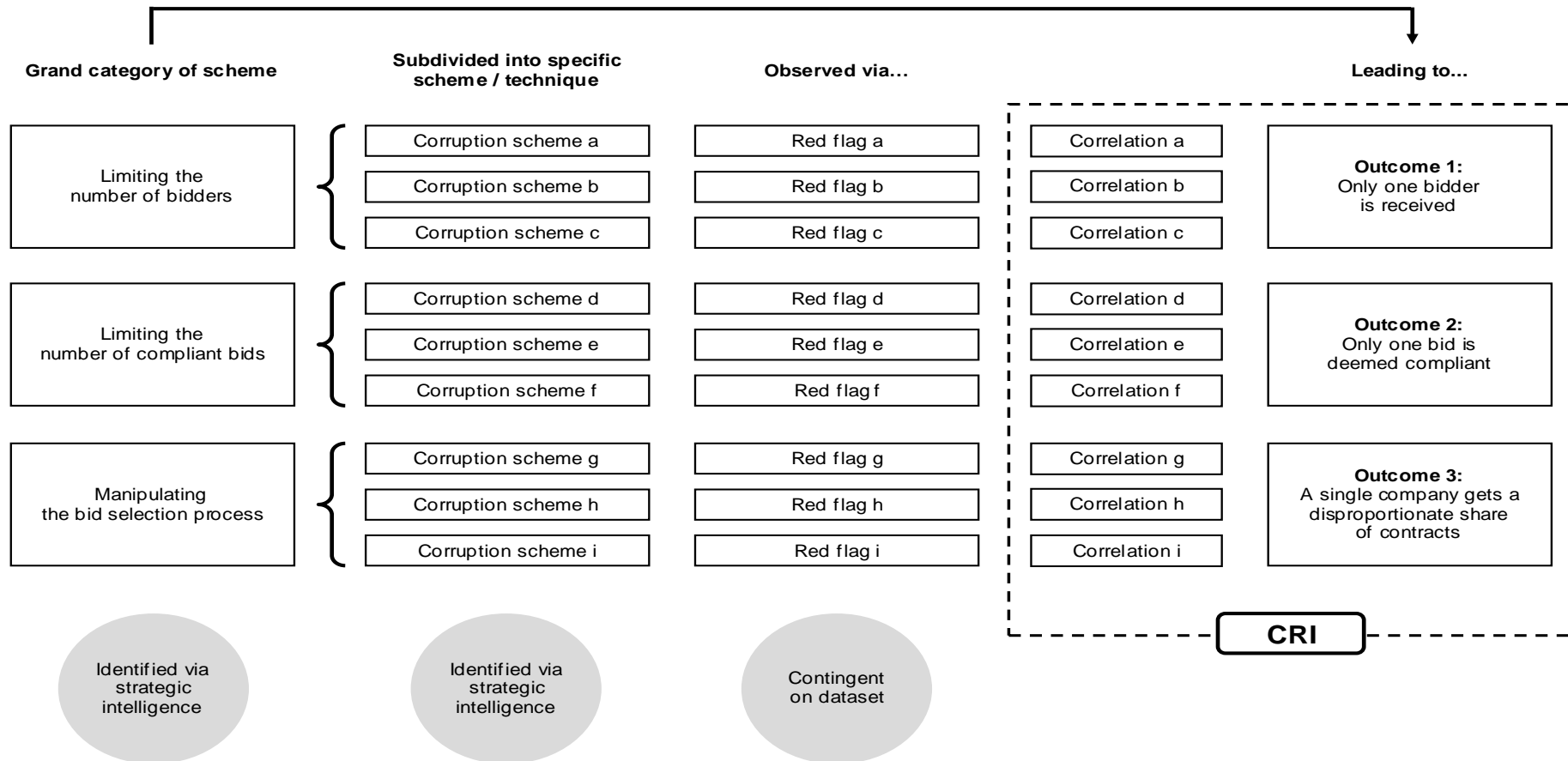
# PRÉSENTATION DES DONNÉES

- Pour qu'un contrat se retrouve dans cette étude, il doit être attaché à un appel d'offres publié sur la plateforme SEAO entre le 1 Janvier 2014 et le 31 décembre 2019.
- Après transformations des données, on se retrouve avec 4 tables:
  - Table des appels d'offres
    - SEAO ID, organisme qui émet l'appel d'offres, entité municipale, type, nature et la date de publication.
  - Table des contrats
    - SEAO ID, date de conclusion du contrat, montant du contrat et compagnie Id
    - Possible d'avoir plusieurs contrats pour un même appel d'offre. À l'inverse, possible qu'il n'y ait pas de contrats pour un appel d'offres
  - Table des dépenses supplémentaires
    - SEAO ID, date de la dépense, le montant de la dépense et compagnie Id
  - Table des fournisseurs
    - SEAO ID, compagnie ID, admissible ou non, conforme ou non, appel d'offres remporté ou non et montant de soumission.

# PRÉSENTATION DES DONNÉES

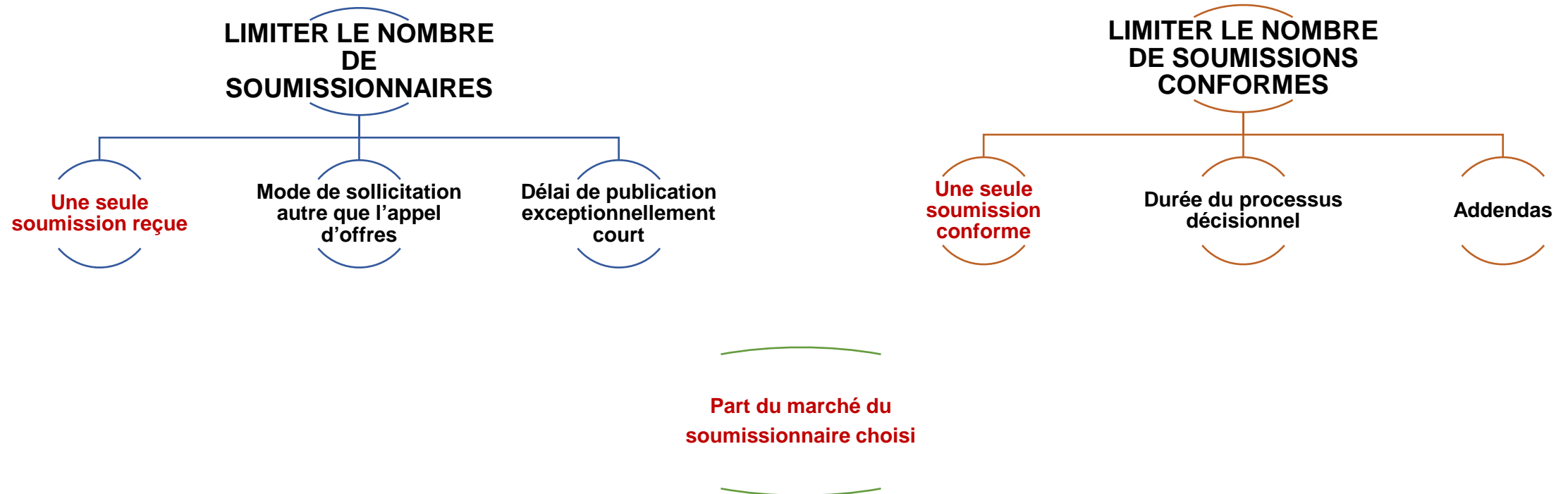


# LE CRI : MÉTHODE ET CALCUL

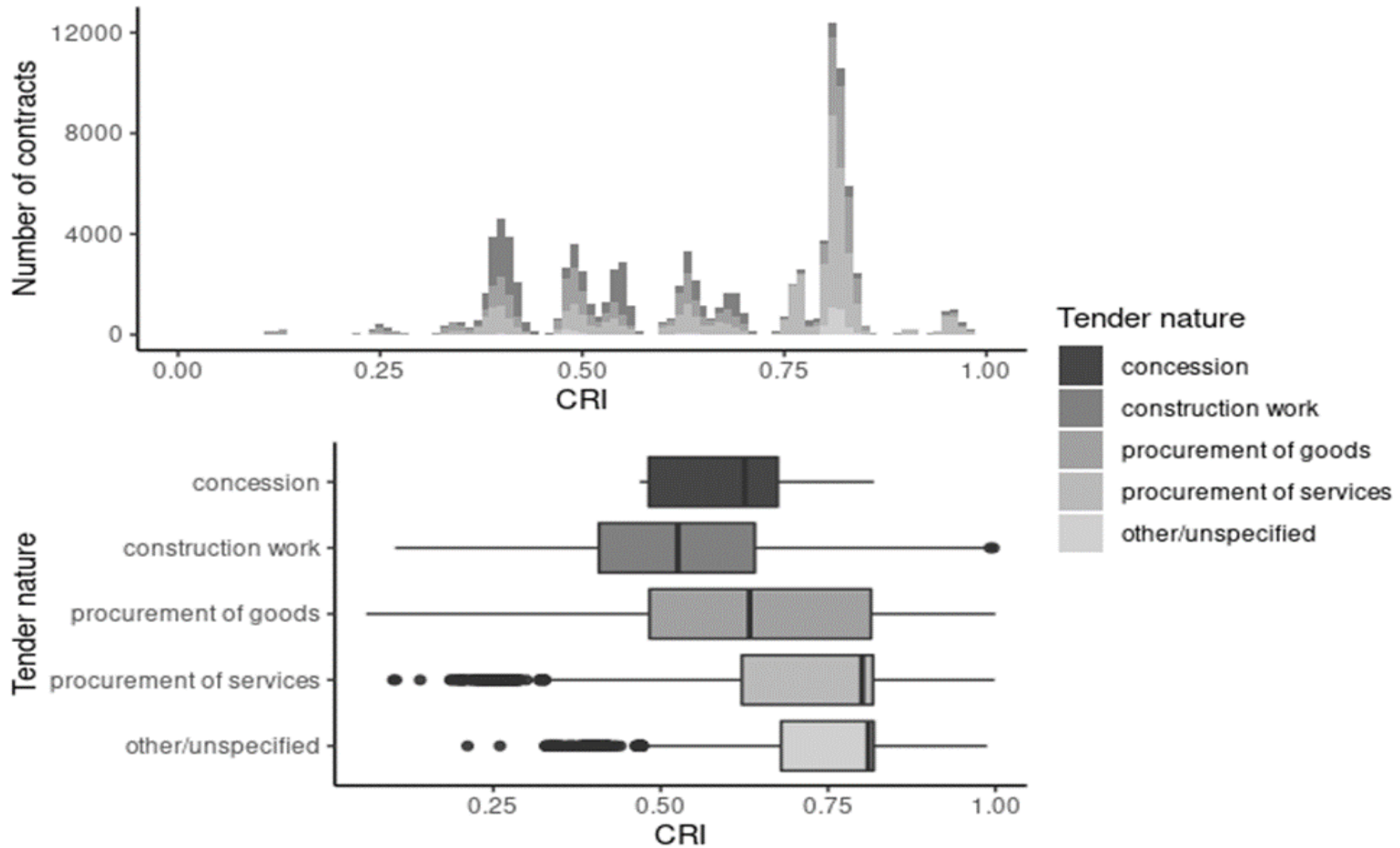




# LE CRI : INDICATEURS (INPUTS-OUTCOMES)

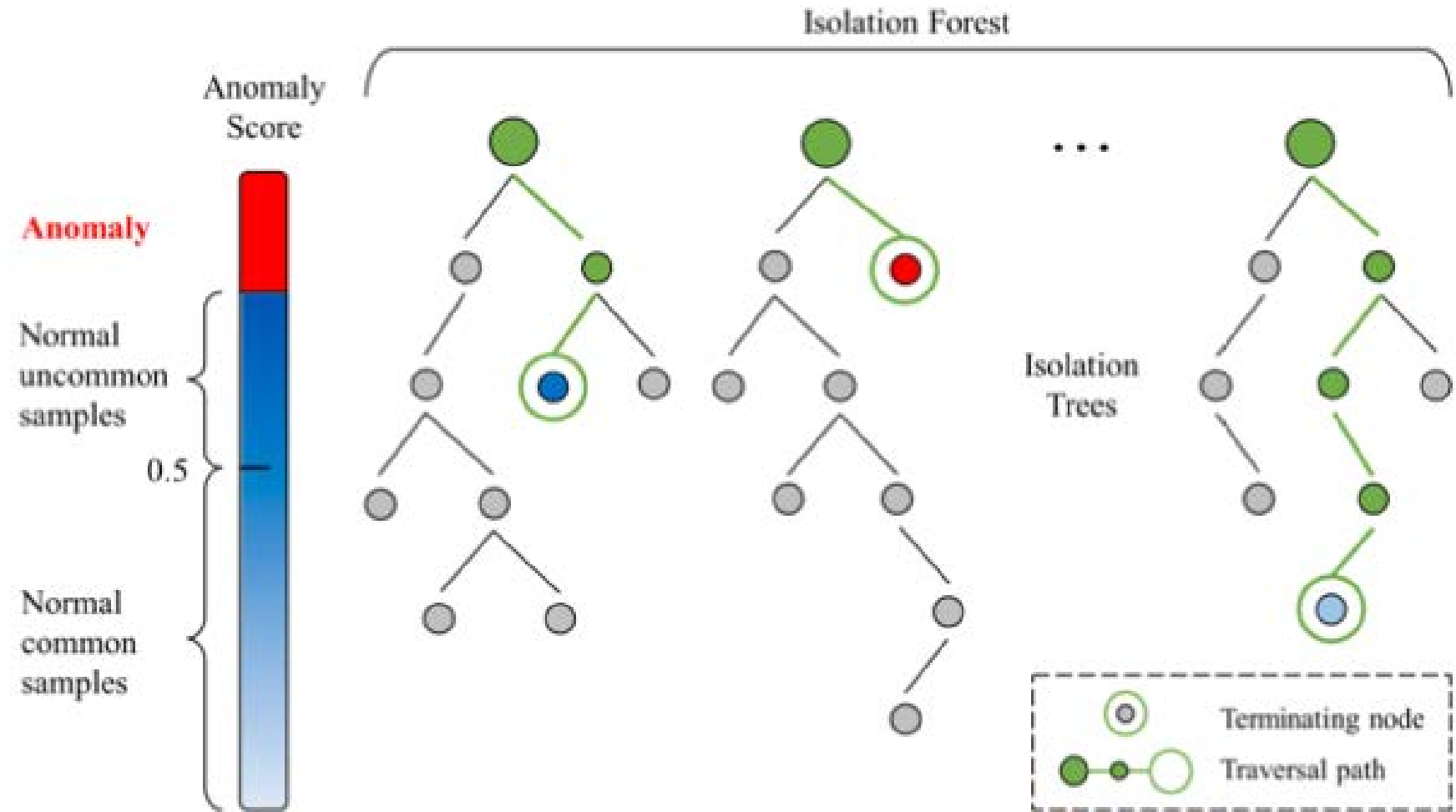


# Résultat du CRI

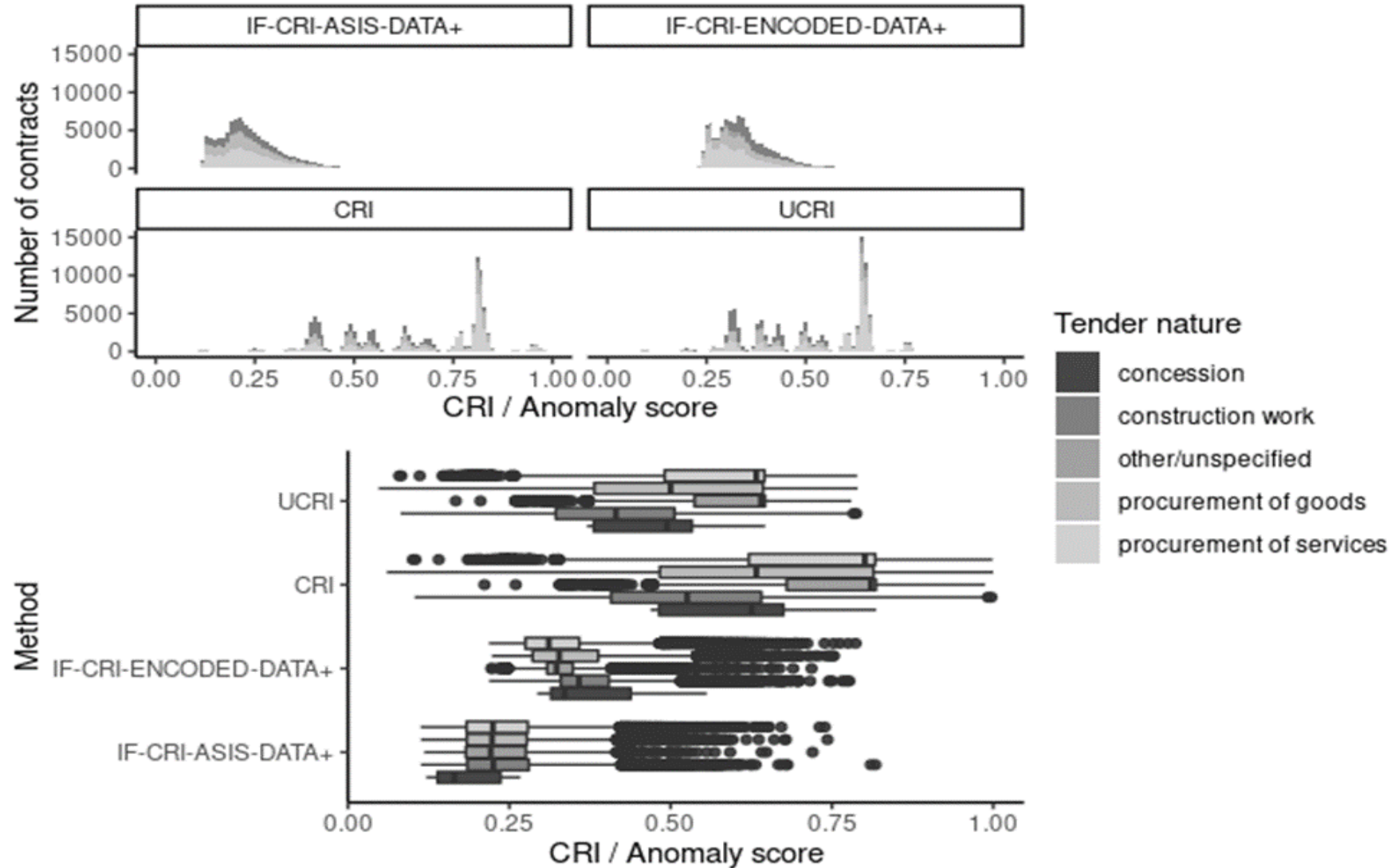


# L'ISOLATION FOREST (IF)

- Sous l'hypothèse que les contrats corrompus ne sont pas la norme, l'utilisation de cet algorithme prend beaucoup de sens.
- L'algorithme construit un ensemble d'arbres de décision en utilisant les données disponibles.

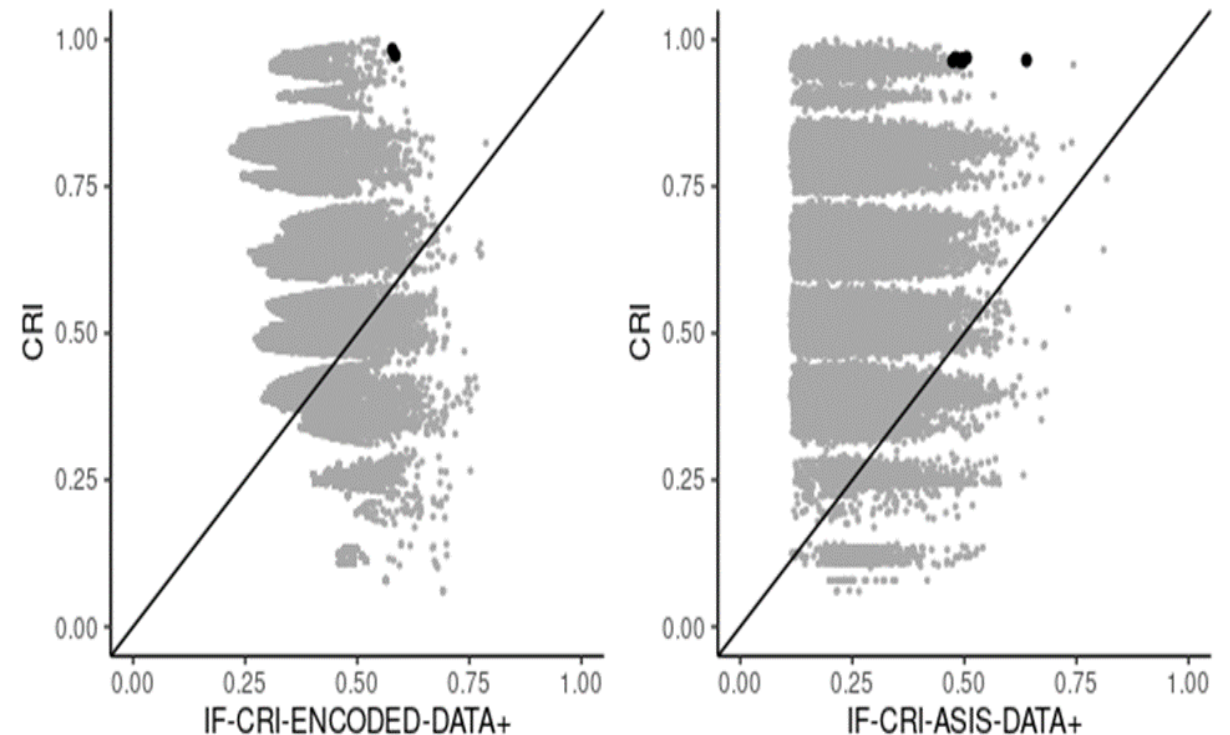


# COMPARAISON DES DEUX MÉTHODES



# COMPARAISON DES DEUX MÉTHODES

- Les contrats en noir sont des contrats qui ont un grand CRI et un grand score d'anomalie
- Le score d'anomalie de IF-CRI-ASIS-DATA+ apparaît non corrélé avec le CRI
- Le score d'anomalie de IF-CRI-ENCODED-DATA+ et le CRI ont une faible corrélation négative



# DISCUSSION

- Pour l'instant, nous ne pouvons pas produire d'évidence que l'un ou l'autre des modèles peut détecter avec certitude des contrats corrompus
  - En effet, sans un échantillon de contrats corrompus, il restera difficile d'identifier formellement de la corruption dans nos données
- Dans un premier temps, notre objectif était surtout de comparer ces deux modèles.
- En supposant que les hypothèses sur lesquels ces modèles sont construits tiennent:
  - Nous observons que les méthodes ne détectent pas les mêmes contrats
  - Cela nous laisse croire qu'il y a une valeur à utiliser les deux approches de concert



# APPLICATION DES MÉTHODES À LA GESTION DE RISQUES

- Dans le cadre de la gestion des risques, la détection des contrats soupçonnés de corruption permet:
  - Une intervention *ex-post* (corrective)
    - Réaliser des audits plus efficaces
    - Évaluer les effets d'interventions passées
    - Documenter les comportements pour enquêtes approfondies
  - Une intervention *ex-ante* (préventive)
    - Dévoiler des patterns de comportement déviants
    - Identifier les activités et fonctions vulnérables
    - Développer des mesures de mitigation des risques

# TRAVAUX À VENIR

- Tester la capacité prédictive des méthodes et établir leur complémentarité
- Intégrer des nouvelles données:
  - Des données SEO non structurées
  - Autres données issues des bases de données autres que le SEO (sur les entreprises, sur les liens entre le politique et les affaires, etc.)
- Procéder à des analyses comparatives dans le temps et entre les catégories (régions, organismes publics, nature de l'approvisionnement) et/ou entre différents pays.