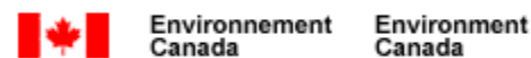


Perspectives d'analyse de données

Pascal Vincent et Marc-Élie Lapointe



Pascal Vincent

LISA



Laboratoire d'**I**nformatique des **S**ystèmes **A**daptatifs
Laboratoire d'**I**nformatique des **S**ystèmes d'**A**pprentissage

Notre expertise:

- ★ Apprentissage Statistique
(Machine Learning)
- ★ Data-mining (analyse de
grands ensembles de données)
- ★ Réseaux de neurones artificiels

L'apprentissage statistique

en quelques mots

- Apprendre une *fonction* effectuant des **prédictions**
- En se basant sur un **grand ensemble d'exemples** (base de données)
- Capable de **généraliser** à de nouveaux cas (pas vus dans les données)

Projet avec Environnement Canada: D'où nous sommes partis

- Données météo code
- Format XML

Données mises à notre
disposition sur le site
d'Environnement Canada.

<http://dd.weatheroffice.ec.gc.ca/meteocode/>

Données mises à notre disposition sur le site d'Environnement Canada.

<http://dd.weatheroffice.ec.gc.ca/meteocode/>

Index of /meteocode

http://dd.weatheroffice.ec.gc.ca/m

Google

Moodle EC programme EC EC remote

Index of /meteocode

<u>Name</u>	<u>Last modified</u>	<u>Size</u>
Parent Directory		-
doc/	25-Nov-2009 16:30	-
geodata/	25-Nov-2009 16:30	-
mtm/	25-Nov-2009 16:47	-
ont/	25-Nov-2009 16:30	-
pnr/	25-Nov-2009 16:30	-
pyr/	25-Nov-2009 16:30	-
que/	25-Nov-2009 16:30	-

Afficher un menu

Données mises à notre disposition sur le site d'Environnement Canada.

<http://dd.weatheroffice.ec.gc.ca/meteocode/>

Index of /meteocode

http://dd.weatheroffice.ec.gc.ca/meteocode/

Index of /meteocode

Name	Last modified	Size
Parent Directory	-	-
doc/	25-Nov-2009 16:30	-
geodata/	25-Nov-2009 16:30	-
mtm/	25-Nov-2009 16:47	-
ont/	25-Nov-2009 16:30	-
pnr/	25-Nov-2009 16:30	-
pyr/	25-Nov-2009 16:30	-
que/	25-Nov-2009 16:30	-

Afficher un menu

Index of /meteocode/que/cmml

http://dd.weatheroffice.ec.gc.ca/meteocode/que/cmml

Index of /meteocode/que/cmml

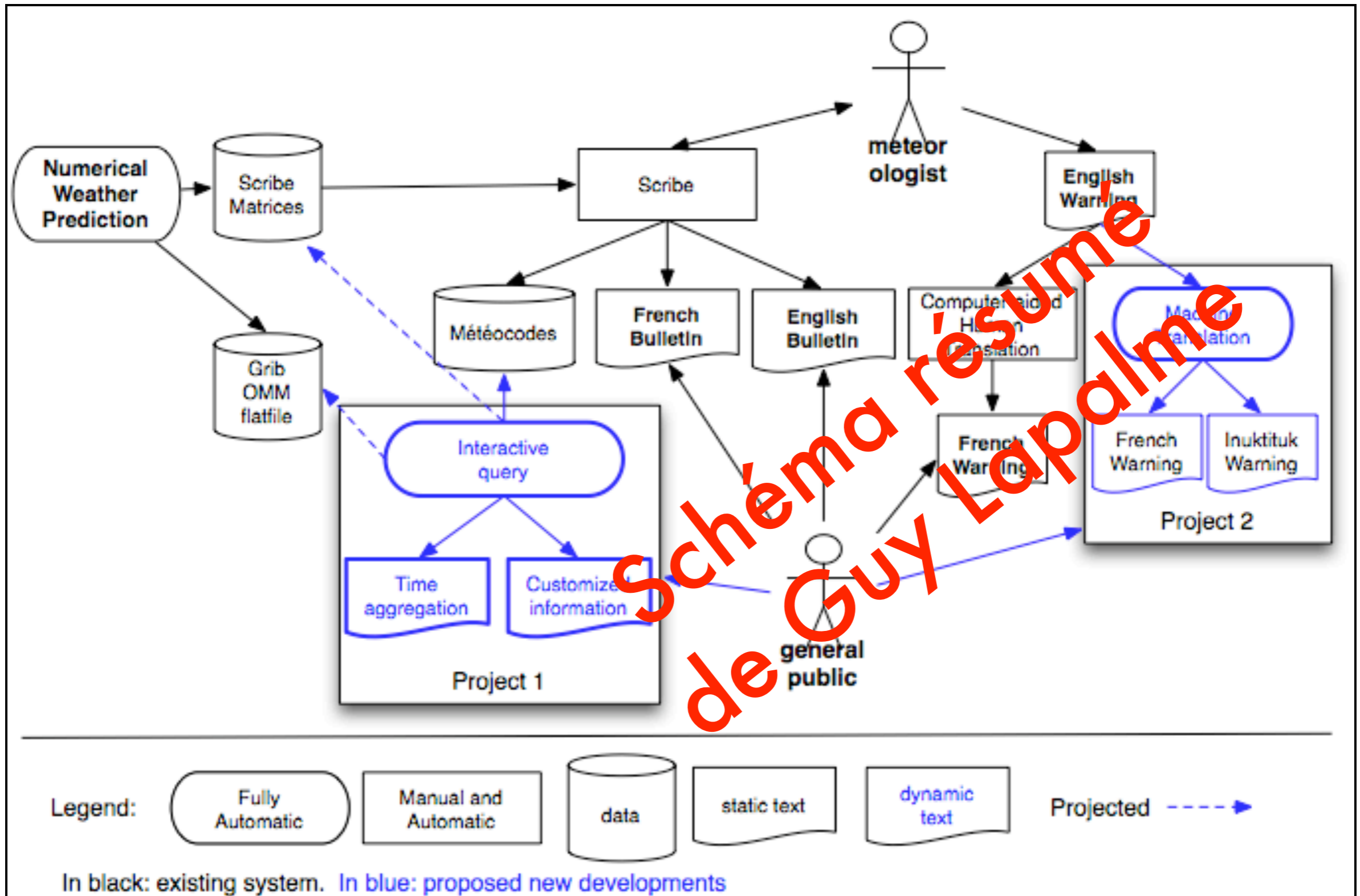
Name	Last modified	Size
Parent Directory	-	-
TRANSMIT.FPCN71.11.24.1226Z.AMD.xml	24-Nov-2009 12:33	463K
TRANSMIT.FPCN71.11.24.1630Z.xml	24-Nov-2009 16:43	413K
TRANSMIT.FPCN71.11.24.2045Z.xml	24-Nov-2009 20:51	478K
TRANSMIT.FPCN71.11.25.1000Z.xml	25-Nov-2009 09:51	542K
TRANSMIT.FPCN71.11.25.1630Z.xml	25-Nov-2009 17:02	468K
TRANSMIT.FPCN71.11.25.2045Z.xml	25-Nov-2009 21:08	475K
TRANSMIT.FPCN71.11.26.1000Z.xml	26-Nov-2009 09:54	508K
TRANSMIT.FPCN71.11.26.1630Z.xml	26-Nov-2009 16:29	441K
TRANSMIT.FPCN71.11.26.2045Z.xml	26-Nov-2009 20:49	244K
TRANSMIT.FPCN71.11.27.1000Z.xml	27-Nov-2009 10:40	278K
TRANSMIT.FPCN71.11.27.1423Z.AMD.xml	27-Nov-2009 14:37	264K
TRANSMIT.FPCN71.11.27.1630Z.xml	27-Nov-2009 16:27	389K
TRANSMIT.FPCN71.11.27.2045Z.xml	27-Nov-2009 21:03	427K
TRANSMIT.FPCN71.11.28.1000Z.xml	28-Nov-2009 09:56	502K
TRANSMIT.FPCN71.11.28.1630Z.xml	28-Nov-2009 16:37	426K
TRANSMIT.FPCN71.11.28.2045Z.xml	28-Nov-2009 21:05	446K
TRANSMIT.FPCN71.11.29.1000Z.xml	29-Nov-2009 10:36	524K
TRANSMIT.FPCN71.11.29.1630Z.xml	29-Nov-2009 16:45	460K
TRANSMIT.FPCN71.11.29.2045Z.xml	30-Nov-2009 01:54	492K
TRANSMIT.FPCN71.11.30.0111Z.AMD.xml	30-Nov-2009 02:01	436K

Afficher un menu

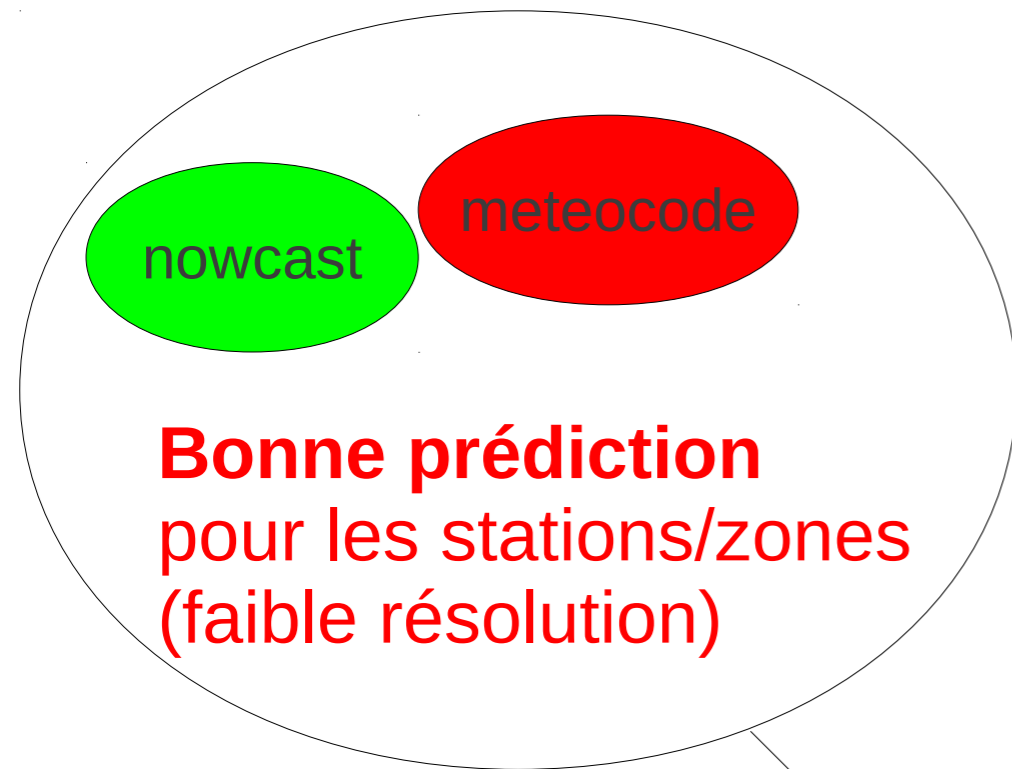

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<cmml xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="/data/wxodfeed/dd/public
  <head>
    <product operational-mode="developmental">
      <title>FPCN71</title>
      <field>meteorological</field>
      <category>forecast</category>
      <creation-date refresh-frequency="PT13H">2009-11-25T20:40:03Z</creation-date>
      <current-issue>2009-11-25T20:45:00Z</current-issue>
      <valid-begin-time>2009-11-25T21:00:00Z</valid-begin-time>
      <valid-end-time>2009-11-27T11:00:00Z</valid-end-time>
      <next-issue>2009-11-26T10:00:00Z</next-issue>
      <timezone>EST5EDT</timezone>
      <type>regular</type>
    </product>
    <source>
      <more-information>http://weatheroffice.ec.gc.ca</more-information>
      <production-center>Meteorological Service of Canada<sub-center>CWUL</sub-center></production-center>
      <feedback/>
      <credit/>
      <credit-logo/>
      <disclaimer>http://www.weatheroffice.ec.gc.ca/mainmenu/disclaimer_e.html</disclaimer>
    </source>
  </head>
  <data>
    <forecast>
      <meteo-code-forecast>
        <location>
          <msc-zone-code>r71.1</msc-zone-code>
          <msc-zone-name lang="fr">Montréal métropolitain - Laval </msc-zone-name>
          <msc-zone-name lang="en">Metro Montréal - Laval</msc-zone-name>
        </location>
        <parameters>
          <warning-list/>
          <cloud-list units="deci">
            <cloud-cover start="2009-11-25T21:00:00Z" end="2009-11-25T21:00:00Z" ceiling-code="2">9</cloud-cover>
            <cloud-cover start="2009-11-25T22:00:00Z" end="2009-11-25T22:00:00Z" ceiling-code="2">9</cloud-cover>
            <cloud-cover start="2009-11-25T23:00:00Z" end="2009-11-25T23:00:00Z" ceiling-code="2">9</cloud-cover>
            <cloud-cover start="2009-11-26T00:00:00Z" end="2009-11-26T00:00:00Z" ceiling-code="2">9</cloud-cover>
            <cloud-cover start="2009-11-26T01:00:00Z" end="2009-11-26T01:00:00Z" ceiling-code="2">9</cloud-cover>
            <cloud-cover start="2009-11-26T02:00:00Z" end="2009-11-26T02:00:00Z" ceiling-code="2">9</cloud-cover>
            <cloud-cover start="2009-11-26T03:00:00Z" end="2009-11-26T03:00:00Z" ceiling-code="2">9</cloud-cover>
            <cloud-cover start="2009-11-26T04:00:00Z" end="2009-11-26T04:00:00Z" ceiling-code="2">9</cloud-cover>
```


Ce que nous avons appris

- **Météocode: l'arbre qui cache la forêt...**
- **Quantité et variété de sources de données impressionnante à gérer!**



Sources de données



- Plusieurs modèles de prédiction (48h, ...)
- GRIB, nowcast, meteocode: formats différents

Ce que nous avons appris (suite)

Météocode: prédictions de **très haute qualité** (corrigées à la main par météorologue) **à faible résolution** (zone code).

GRIB: prédictions **brutes du modèle** à très haute résolution.

Nowcast: **observations + bonnes prédictions court terme** à **moyenne résolution** (stations météorologiques)

- Stockage de ces données dans base de donnée SQL: sans espoir!
- Développement d'un stockage et accès efficace (en temps et espace). Fait.

Visualiser les données grib, meteocode et nowcast pour différentes zones et/ou stations ainsi que des données grib partout

Accueil Visualiser Tester Tables Dataset Créer table Prétraiter Comparer Importer Créer dataset Admin

Cette page sert à obtenir et comparer différentes prévisions météo en créant deux graphiques. Le premier contient les données demandées et le deuxième contient les données qui on mener à la correction (jaune) de ce qui est appelé récent(vert). C'est deux informations sont liées à l'emplacement du marqueur vert sur la map. Toutes les autres informations sont liées à la station choisie et à la zone choisie qui y correspond. Soyez patient, la map est très lente du au grand nombre de station et la création des graphiques peut prendre quelques minutes.

Station: Aéroport int. de Mont
Zone: Montréal métropolitain
Zone 1: Vaudreuil - Soulange
Zone 2: Montréal métropolitain
Zone 3: Laurentides - r71.6
Caractéristique: air_temp
Date: 2010-08-27 10 UTC
Fenetre temporelle: 1 jour(s)
Delai de prevision: 0 heure(s)
Coordonnee géographique: 45.5694051,-73.8009789
Grib de la station
Grib00 de la station
Grib du marqueur vert
Meteocode de la zone
Nowcast de la station
Reference: meteocode
 Soumettre <- ->

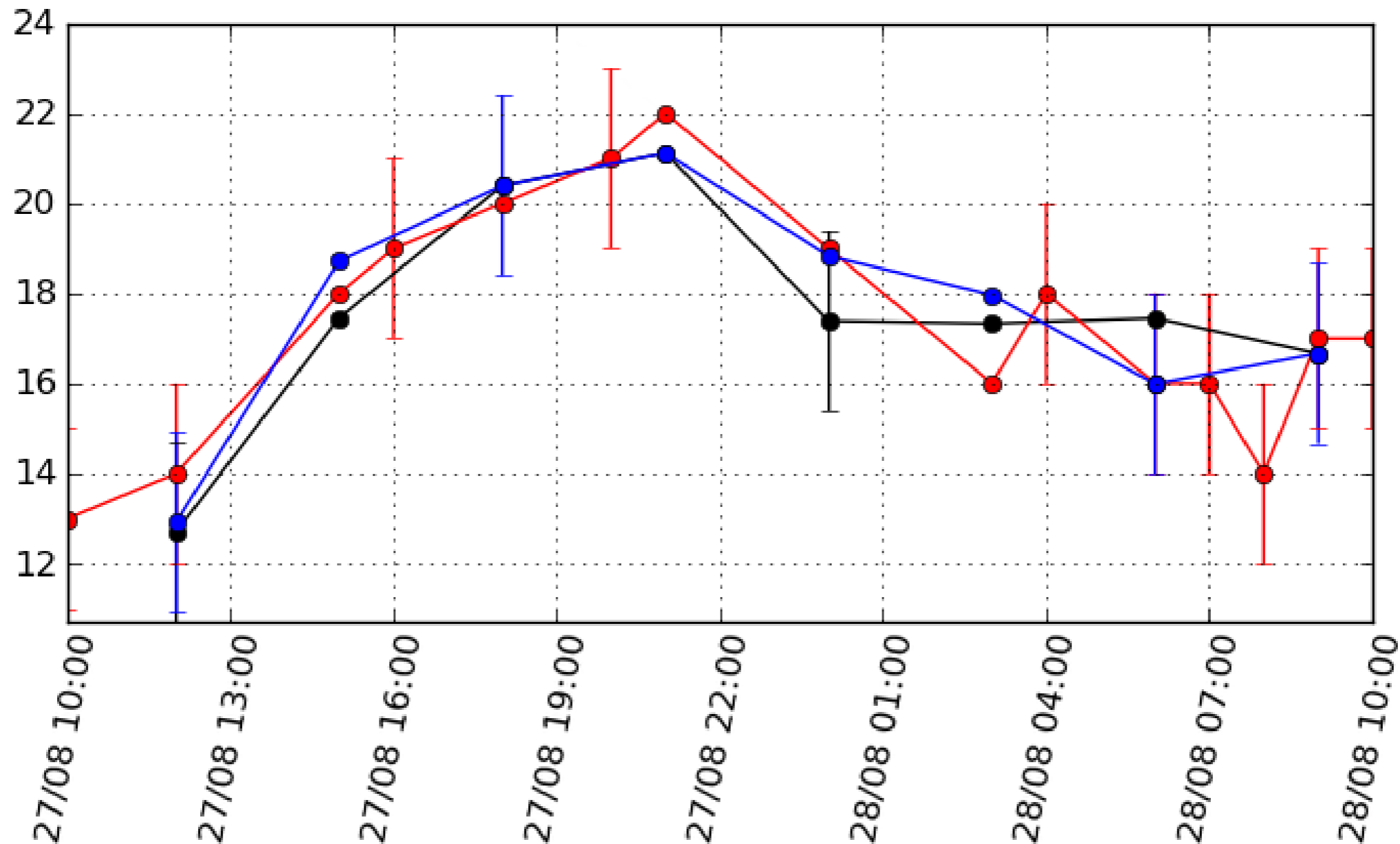
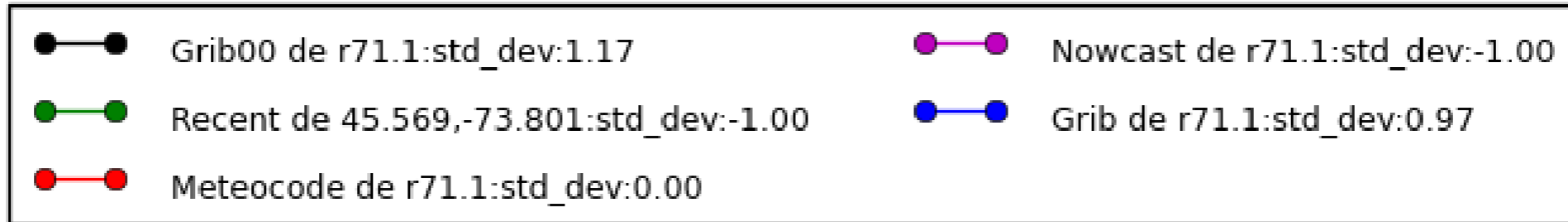
Distance(km) Stations les plus proche

8.8785814221	Sainte-Anne-de-Bellevue	WVQ
11.7584291808	Montréal-Trudeau Int'l Airport	YUL
21.0312318966	Montréal-Mirabel Int'l Airport	YMX

Plan Satellite Mixte Relief

Données cartographiques ©2011 Europa Technologies, Google - Conditions d'utilisation

Comparaison des prevision hautes resolution et meteocode



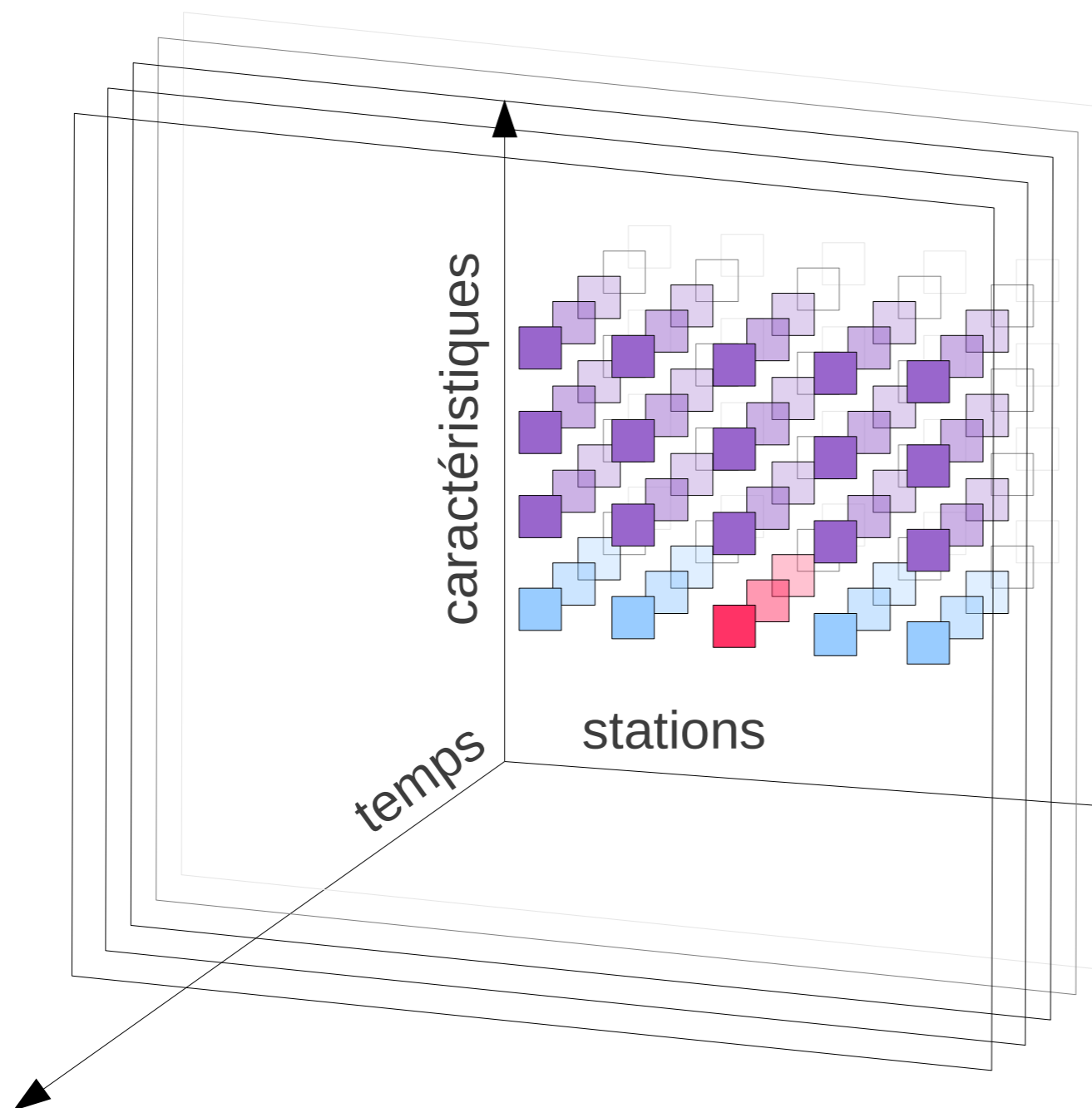
Ce que nous avons appris

(flot de traitement)

- Prévisions brutes d'un modèle «numérique»
(⇒ fichiers GRIB)
- Modèles de correction statistique
(⇒ UMOS Updatable Model Output Statistics)
- Corrections des météorologues
(⇒ Météocode XML)
- Comparaison aux observations des stations
(⇒ Nowcast?)
- Qui prédit mieux? **Métriques de comparaison?**

Perspectives

- Où mon expertise serait la plus pertinente:
- **Modèles de correction statistique**
(UMOS Updatable Model Output Statistics)
- Quels modèles? Données spécifiques utilisées pour entraîner ces modèles? Quelles métriques?
- Prise en compte des corrections des météorologues: se concentrer là où leur expertise est la plus utile.



- **Plusieurs** caractéristiques
- **Plusieurs** stations
- **Plusieurs** moments temporels

Étude de l'**évolution** des **caractéristiques** et des **corrections** apportées au cours du temps.

À suivre....

- Rencontrer (au CMC?) un spécialiste du UMOS
- Trouver un nouvel étudiant pour reprendre le flambeau....

Questions?