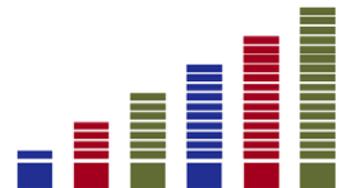


Improving building permit administrative data set for short term analysis

François LIMOUSIN, SOeS, France
Ministry of Environment – Ministry of Housing

- 1) What is the Sitadel2 database ?
- 2) Construction of the estimator of the number of authorized dwellings
- 3) Dealing with revisions



The Sitadel2 database : collecting data

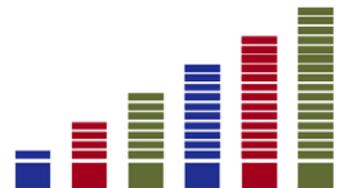
Building permit forms are registered in every 36 000 city hall

2000+ local authorities automatically send the collected information each month (using specific software) : the file format has been set in 2009.

Yesterday, 26 regional teams driven by a national supervisor,
Today one team in 4 places plus overseas

Structured ID for the permit number, numerous fields about who, where, what.

Follow-up for housing starts and end of works



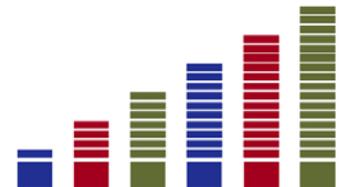
The Sitadel2 database : disseminating data

Monthly publication of the total number of dwellings and of the total area of non residential buildings : roughly 30 000 authorized dwellings each month.

Based on administrative data collection of the granted building permits

We used to sum the number of dwellings transmitted by the respondents during the month : DPC series which were used for short term analysis.

For long term studies, DR series were used because it is the actual authorisation figures.



The Sitadel2 database : Date issues

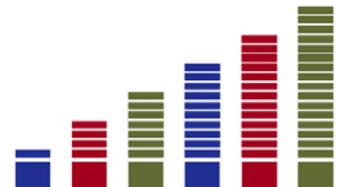
2 different dates for every event of a permit :

Date of collection (called DPC)

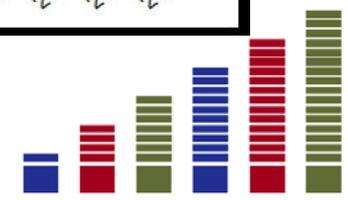
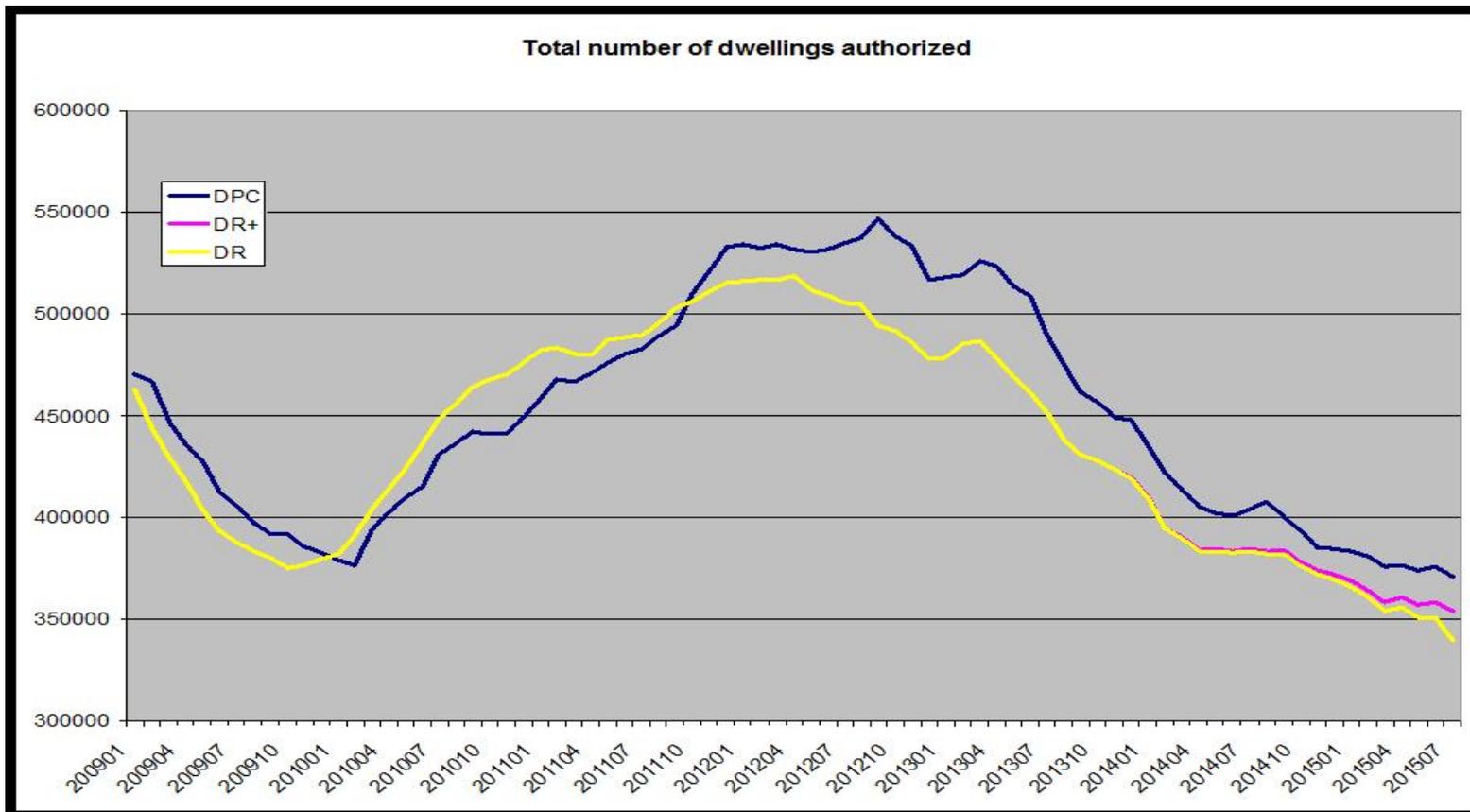
Real date (called DR)

$DPC \geq DR$, the difference is a delay due to collection

This delay is usually OK, but with changes in the regulatory framework (2007 mainly) or in the data collection organisation (2011-2012), some troubles appeared...



The Sitadel2 database : (12 month agregate)
2012/2011 : DR +8 %, DPC +20 %



The Sitadel2 database : Date issues

DPC estimates have been used since 1960s because :

- DPC is **easy** to compute

- ... at any geographical level

- ... is **never revised**

But DPC estimates depends on mean data collection delay

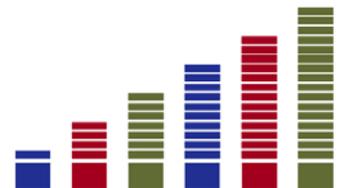
- ...which depends on institutional environment

- ...which is nowadays quite moving

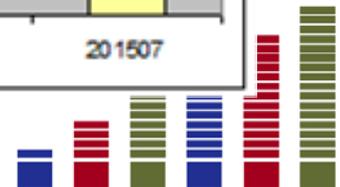
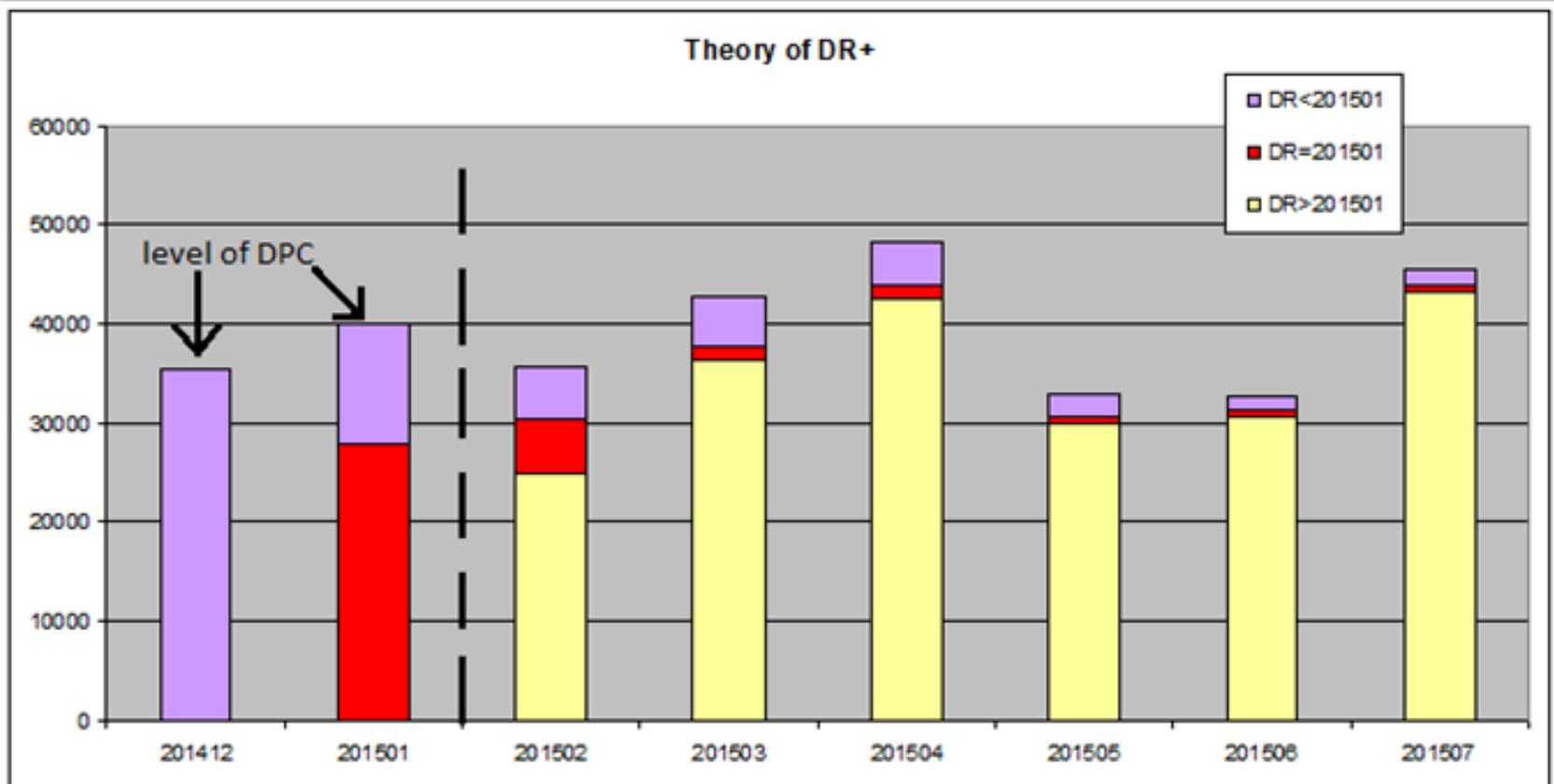
DPC is **always late** to identify turning points and can include old events that are said to occur at the time of issue

A wrong figure can never be changed

DPC is not consistent with reality at the finest geographical levels



Construction of DR+ estimates



Construction of DR+ estimates

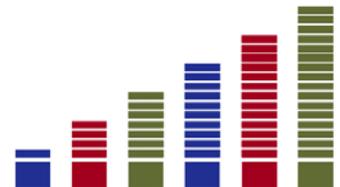
The forecast of the DPC level is made with an airline model

The forecast is done for 11 months, but the influence of the latest months is quite weak

It is not the actual DPC serie which is forecasted, but a truncated version, to avoid « catch-up » effects.

This leads to a DR timeseries, called DR_{12} , built with only 12 months of data.

In order to obtain the DR+ estimates, we use monthly evolution of DR_{12} at a stable DR level.



Construction of DR+ estimates

We consider the DR level to be stable at 24 month so

$$\mathbf{NDR+(t) = Nobs(t) \text{ where } t < m-24}$$

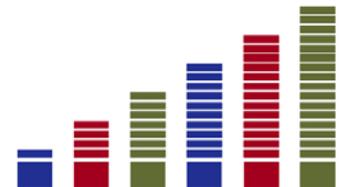
And for months between m and m-24 :

$$\mathbf{NDR+(t) = Nobs(m-24) * DR_{12}(t)/DR_{12}(m-24)}$$

Areas of authorized dwellings are computed using the formula :

$$\mathbf{ADR+(t) = Aobs(t) + (NDR+(t) - Nobs(t)) \times Amean(t)}$$

With A =areas and N= number of dwellings



Revisions

The main issue with this new estimate.

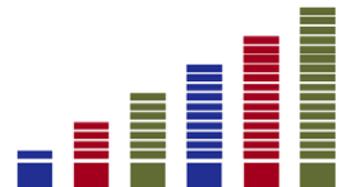
Several causes :

- Data collection for old permits (small impact)

- Each month, we update the timeseries we forecast : we can modify the first forecasted terms, which weights around 20%

- The use of a new m-24 month in chaining evolutions of DR₁₂

The entire set is revised every month, even though the oldest figures do not change much (only data collection)



Revisions

Since 201501 :

Average absolute revision $m \Rightarrow m+1$: 890

Average absolute revision $m+1 \Rightarrow m+2$: 550

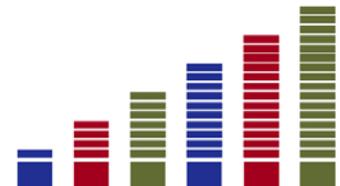
Average absolute revision $m+2 \Rightarrow m+3$: 460

Average absolute revision $m \Rightarrow m+2$: 840

Average absolute revision m (for 2015) \Rightarrow 201603 : 770

Those revisions were expected, with this order of magnitude that's why Eurostat received m at the end of the $m+1$ month

Not so much points to study, but we are focused on this matter as it is a strong issue for our users (who are fond of the wrong but unrevised DPC figures).



To sum up

Administrative data seems to be perfect for statistical analysis

But actually need a lot of work to meet the quality requirements

This method which consists in anticipating the data collection could be use for other sources.

Thank you for listening

