Cemagref

VEOLIA EAU

The impact of pipe segment length on break predictions in water distribution systems

Bernard BREMOND

eau - territoires - développement durable

LESAM 2007 - LISBON – 17th October

Context

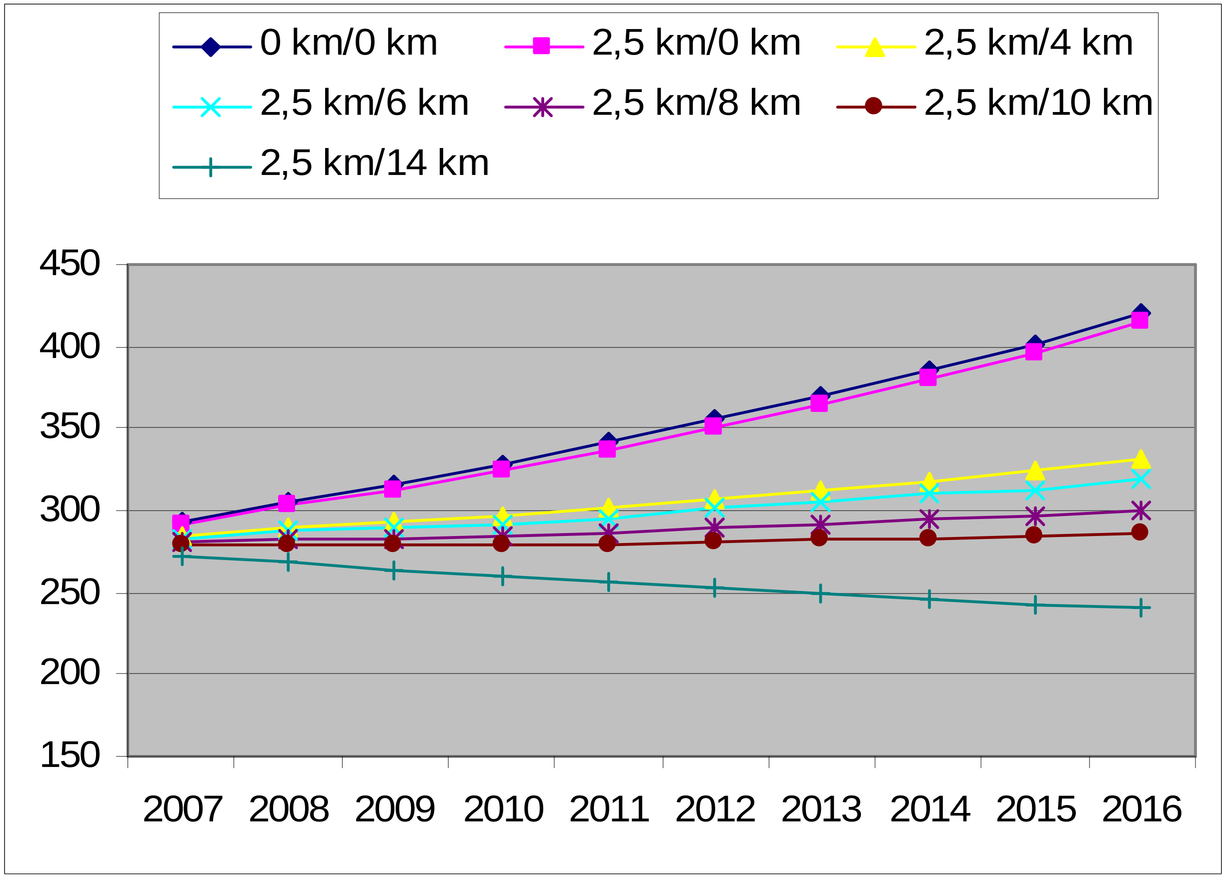
- ✓ Development of break prediction models for defining rehabilitation strategies in drinking water networks

- ✓ These models developed for predicting pipe breaks require the existence of a database that describes the network and an inventory of breaks recorded over several years

- ✓ The network is subdivided into pipe segments. The segment is the elementary object of the study.

- ✓ The principal of the calculations is to predict for each segment, the number of breaks likely during a given time period.

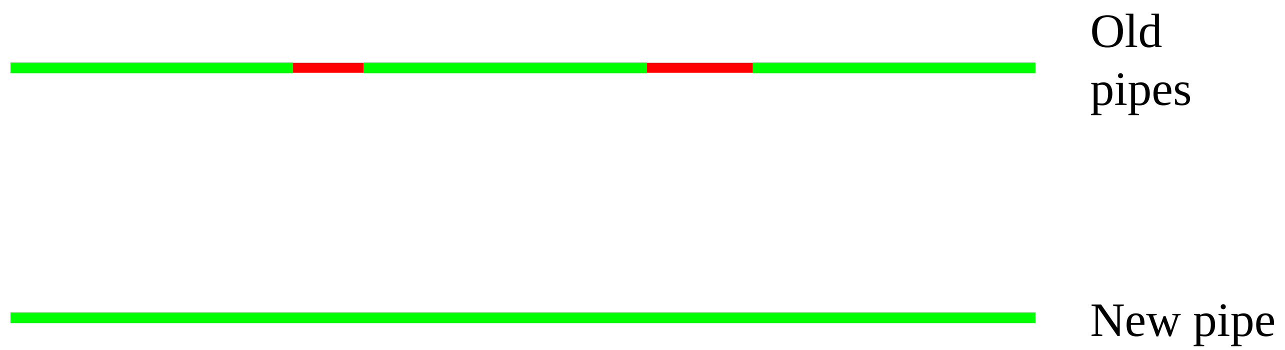
Example of break prediction model on a network of 2400 km



Context

- For each segment, the following information is necessary (or desirable):
  - Asset data – diameter, material, length, year laid, (joint type…)
  - Intervention data – identification of pipe concerned, date of intervention, type of incident, (reason for incident…)
  - Environmental data – (soil type, surface type, traffic level, water pressure…)
- segments registered in a geographic information system result from a splitting of a longer entity by the operator for functional purposes
- The object of this study is to determine the influence of this artificial splitting of pipe segments on the precision of break predictions

Case of sandwich pipe segments



Concatenation procedure

- One of the segments has a length less than the predetermined threshold
- The segments are of the same material
- The segments have the same diameter
- The segments were laid within five years of each other
- The segments don't need to be in the same road

Calculating the break predictions

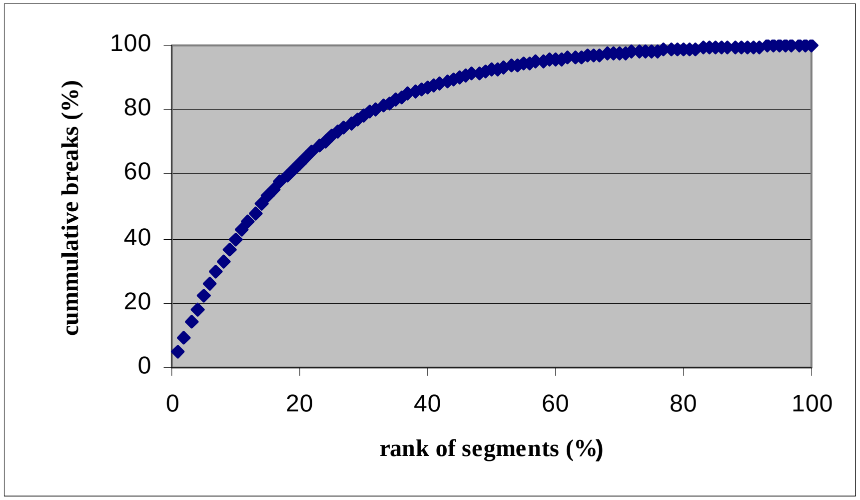
- ✓ **The calculations are performed using the statistical model LEYP (Linearly Extended Yule Process)**

- ✓ **The LEYP model is governed by an intensity function which is supposed to depend on:**
  - **The age of the segment considered**
  - **The number of previous events**
  - **The vector of covariates, Z**

- ✓ **The analytical form of the intensity function combines the product of:**
  - **The influence of previous events in a form derived from the Yule process (linear extension of this model)**
  - **The influence of age in the form of the Weibull model (power of time)**
  - **The influence of the covariates represented as in the Cox proportional hazards model**

Method for comparing results

- **Capacity of identifying the segments the most at risk**
  - **AI – Area under the curve of predicted performance**



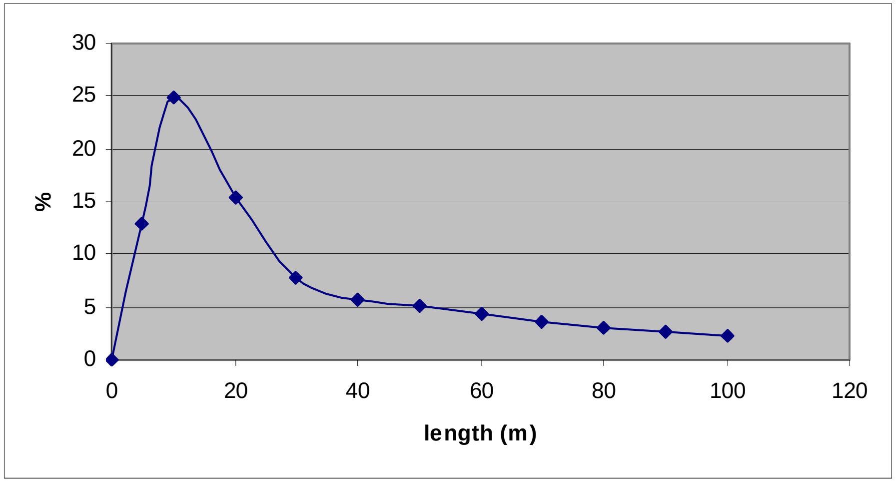
  - **CLx, – percentage of breaks really observed on the top x% of length of pipes sorted by decreasing predicted break rate, in other words, rehabilitating top x% of the pipes should avoid Clx % of the breaks**

- **The accuracy of the predictions**

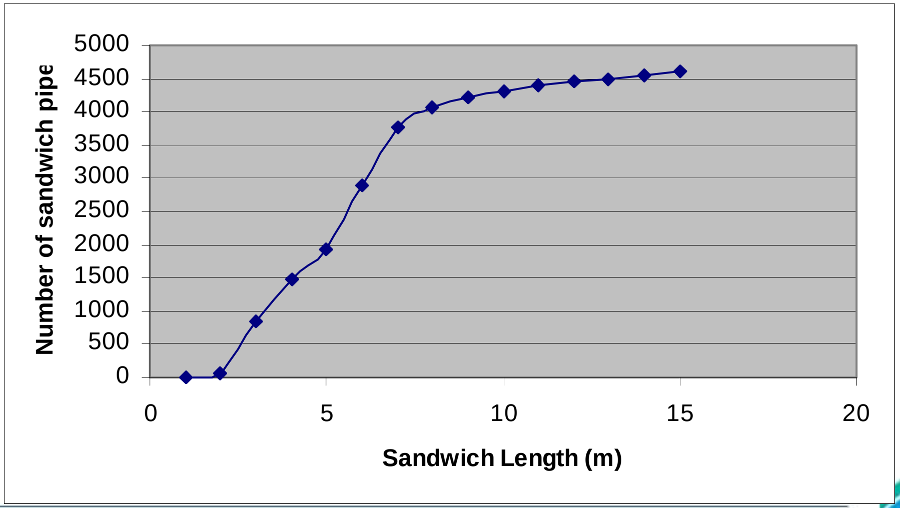
Study data Raw data

- ✓ **data supplied by Veolia Water, limited essentially to grey cast iron pipes: 3736 km, 79536 pipes**

Distribution of pipe segment lengths



Distribution function of sandwich length



Characteristics of data files

Length threshold (m)	Cast iron segments	Sandwich pipes	Orphan pipes	Maximum number of pipes
0 (original data)	79536	455	0	
0 (with sandwich pipes removed)	78941	0	0	19
5	68946	0	5785	7
10	57978	0	14595	7
20	51118	0	20637	10
50	45754	0	28040	15
100	41289	0	35274	19

Calculations and results

- ✓ **Calibration period: 9 years (01/01/1995 – 31/12/2003)**

  Total number of breaks during this period: 8228

- ✓ **The validation period: 2 years (01/01/2004 – 31/12/2005) Total number of breaks during this period was 1997**

Cemagref

parameters of interest

- α which controls the tendency of breaks to occur repeatedly on the same pipes;
- β_1 the regression coefficient of the natural logarithm of the length, *i.e.* the power of the length the failure rate is proportional to;
- β_2 the regression coefficient of the diameter category 40-60mm;
- β_3 the regression coefficient of the diameter category 70-80mm;
- β_4 the regression coefficient of the diameter category 90-100mm;
- β_5 the regression coefficient of the diameter category 110-150mm;
- β_6 the regression coefficient of the diameter category 160-250mm;
- β_7 the regression coefficient of the construction period 1850-1929 (pit cast iron);
- β_8 the regression coefficient of the construction period 1930-1945 (old spun cast iron).

consequences of concatenation on model parameters

- ✓ the sole removal of “sandwich” pipe segments does not significantly affect the model
- ✓ The model parameterised with raw data has a value close to 0.5 The concatenation makes decrease significantly, and makes the length factor consequently lose some explanation power.
- ✓ Another effect is the decrease in the covariate modulation, i.e. the decrease in the deviation between regression coefficients related to the modalities of a given qualitative factor; this holds for both diameter category and construction period
- ✓ the general consequence of concatenation lies in a decrease in the models ability to discriminate the failure risk and there is hence a fear that the model tends towards “averaging” of the predictions.

Cemagref

consequences of concatenation on predictive efficiency

Threshold Length (m)	Area under performance curve Al	CL0.5	CL1	CL5	Breaks observed	Breaks predicted
Original data	0.646	0.0225	0.0396	0.150	1997	1873
0	0.645	0.0225	0.0391	0.152	1997	1874
5	0.645	0.0235	0.0381	0.153	1997	1875
10	0.647	0.0210	0.0386	0.154	1997	1877
20	0.650	0.0215	0.0421	0.152	1997	1880
50	0.653	0.0200	0.0416	0.149	1997	1884
100	0.655	0.0185	0.0381	0.149	1997	1890

Conclusion

✓ Concatenation seems to dampen the model parameters and consequently weakens the model's ability to discriminate the break risk

✓ The model is not sensitive to datasets with numerous very short segments. If for file size or computation time reasons, one wishes to simplify the input data set, a 5 m threshold concatenation is however a reasonable option

✓The concatenation after the model has been calibrated deserves also to be studied in order to design practical replacement projects.This operation needs optimisation procedures to maximise the aggregated failure risk while matching operational constraints