MODELING OF THE SUBSIDENCE RISK IN METROPOLITAN FRANCE

Final presentation
Alice TOURRET
Master Statistics & Econometrics

INTRODUCTION

Natural disasters and the insurance sector

- Occurrence and severity
- Increase uncertainty in the insurance sector

Subsidence

- Second most important natural disaster in term of loss
- Consequence of drought
- Unique characteristics

→ Standard reserving methods

Model the subsidence risk in France and provide an estimate of the ultimate loss for the years of occurrence of our interest

OUTLINE

1.

WHAT IS SUBSIDENCE ?

SUBSIDENCE: THE DEFINITION

Subsidence is defined as the **displacement of the ground surface** due to withdrawal-shrinkage of the soil [4].

Consequences on constructions:
- Cracks in walls or ground
- Weakening of the foundations
- Collapse

Causes:
- Clay composition of the soil
- Nature of the building
- dry and hot weather

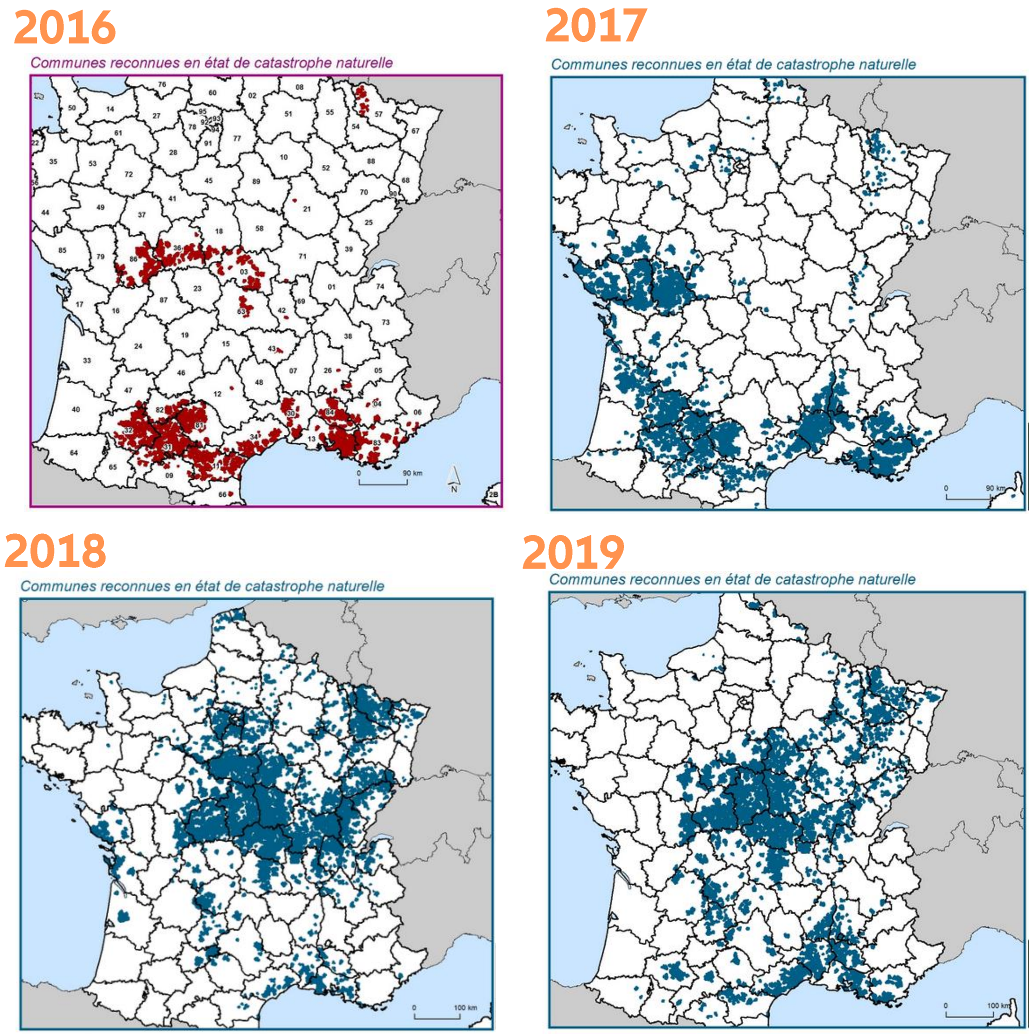
DROUGHTS: THE CONTEXT (2016-2020)

The period 2015-2020: severe for drought

Depending on the occurrence:

- Different **areas** affected
- Different **months** of occurrence
- Different **severity**, especially for **2018**

Year of occurrence	2016	2017	2018	2019
Number of municipalities	986	2 122	4 060	2 905



Distribution of the municipalities declared in a natural disaster state - Drought

SUBSIDENCE: A PARTICULAR NATURAL DISASTER

On average, for subsidence:

- **Duration of the event:** 50 days vs 5 days for other disasters
- **Report period :** 18 months VS 50 days for other disasters

Why ? **The Natural Disaster intergovernmental process (CatNat process)**

Municipality → **Interministerial commision**
- Geotechnical criterion
- Climatic criterion

→ Publication (or not) of a decree → Compensation of the victims by the insurer

ALLIANZ’S CLAIM DATA, TRIANGLE & CHAIN LADDER

What is Reserving ? Estimate the ultimate cost of the different P&C risks.

Year of development

Year of occurence

decl	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2001	140	65	9	6	7	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0
2002	114	688	766	159	26	3	28	2	3	0	1	1	0	0	0	0	0	0	0	
2003	1052	2857	3618	1986	216	74	11	5	8	1	2	0	0	0	1	0	0	1		
2004	404	48	10	3	43	9	6	0	0	0	0	0	0	0	0	0	0			
2005	226	68	10	1403	43	17	9	3	1	1	1	1	0	0	0	3				
2006	85	9	332	26	3	3	5	1	0	0	0	0	0	0	0					
2007	31	399	85	20	14	1	1	0	1	0	0	1	0	0						
2008	218	136	55	12	4	0	0	0	1	0	0	0	0							
2009	89	21	743	24	2	3	0	0	0	2	0	0								
2010	37	115	43	3	5	0	0	0	0	0	2									
2011	99	3315	546	50	11	5	3	2	2	0										
2012	345	726	114	12	8	3	3	2	0											
2013	78	37	7	2	0	2	0	0												
2014	21	39	23	5	0	0	0													
2015	58	270	166	23	4	0														
2016	115	2063	350	30	7															
2017	260	2981	496	55																
2018	480	5232	371																	
2019	664	2796																		
2020	535																			

Triangle - Uncumulated number of subsidence claims

- Year of development 4: triangle quite stable

Unsual shocks !
- CatNat process
- Chain Ladder asumptions and results

Try different ways to model subsidence

2.

MODELING THE SUBSIDENCE RISK

CONTEXT OF THE MODEL

GOAL

Predicting the number of claims at year of development 4 with external climatic data already available at the end of the first year of development

Characteristics:

- Training and optimization of the model on occurrences 2016-2017
- Modeling at the department level
- GLM and XGBoost

DATA SOURCES & DATA AGGREGATION

5 different data sources

Housing insurance contract portfolio

Allianz's claim data

Soil composition data from the French Bureau of Geological and Mining Research

Daily climatic data from the ECA&D

Monthly climatic index (SWI) from Météo France

- **Aggregation at the department level**
- **88 descriptive variables**

FITTING THE MODEL

“Best” model:

test-minimal RMSE

BUT:

- Relatively small sample
- Relatively few information
- unbalanced dataset

→

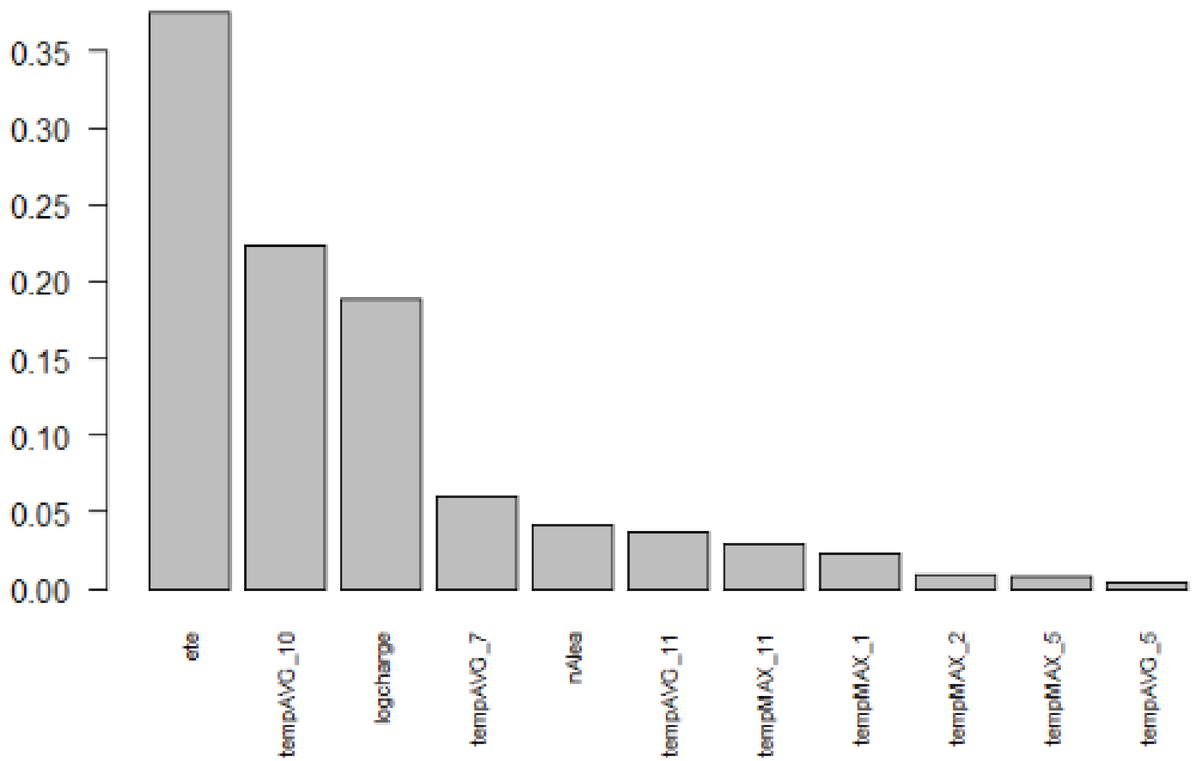
Easy to overfit !

Resampling and hyper parameter optimization (grid-search)

3.

RESULTS

XGBOOST: VARIABLE SELECTION AND IMPORTANCE



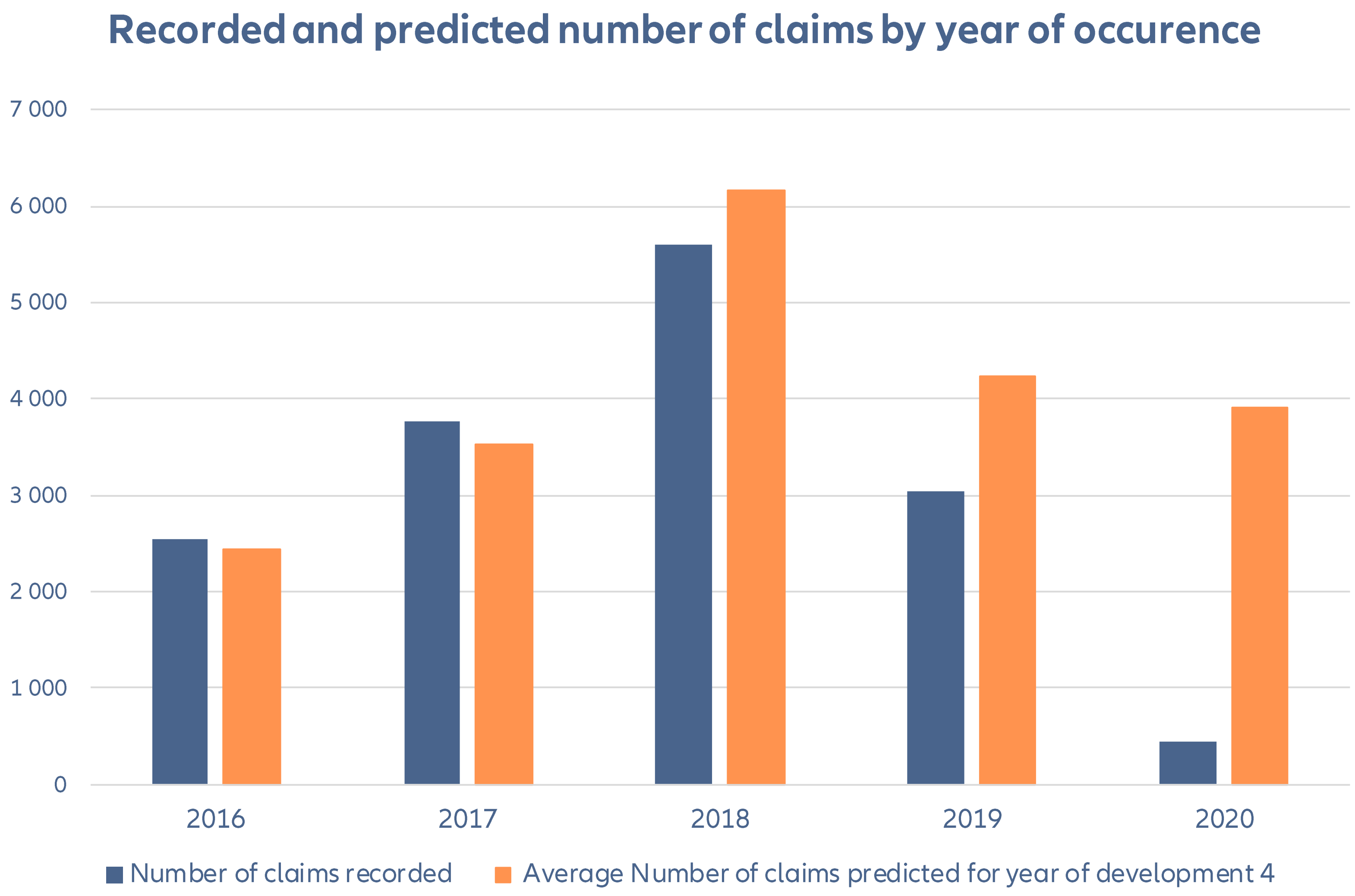
Mean gain of selected features of the tuned XGboost model

11 variables are selected:

- Claim variables
- Monthly max and average temperature
- Soil data

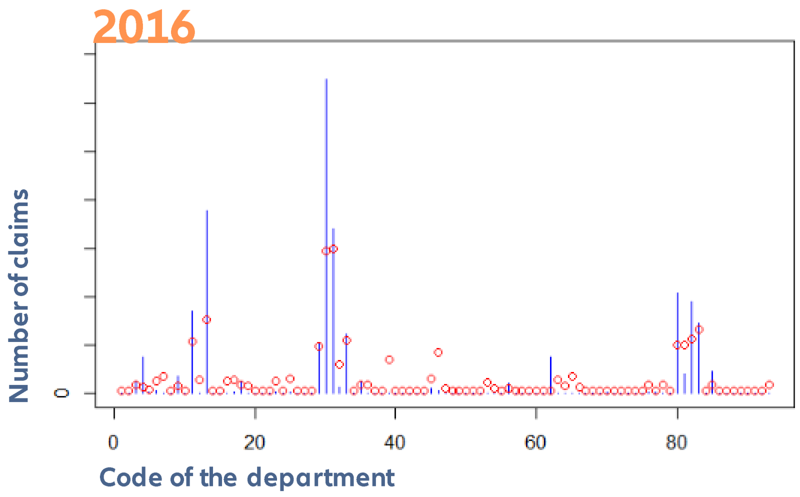
No precipitation data !

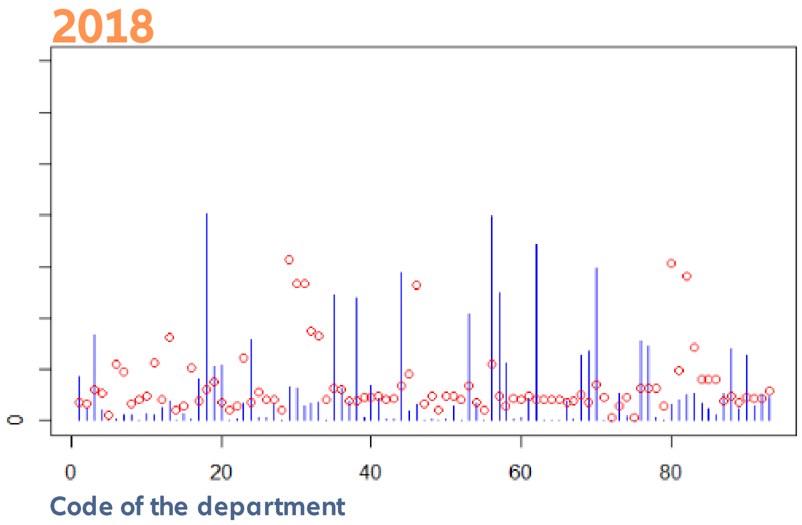
XGBOOST: OVERALL PREDICTIONS BY OCCURENCE

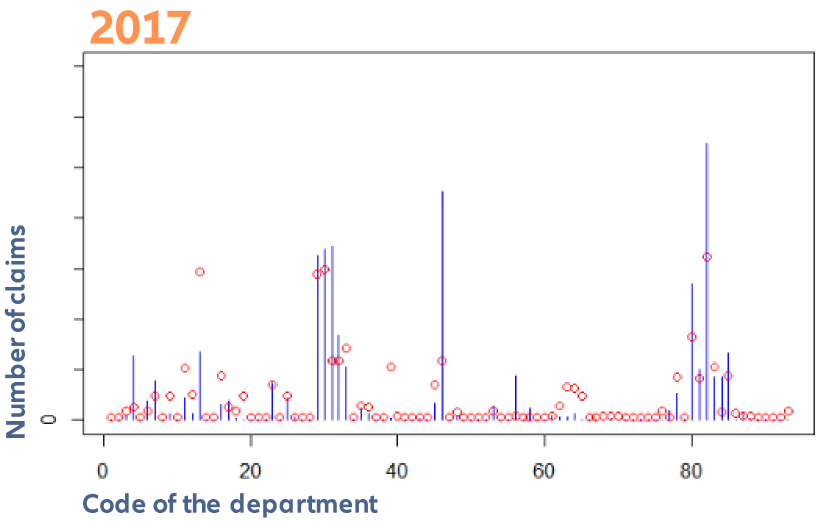


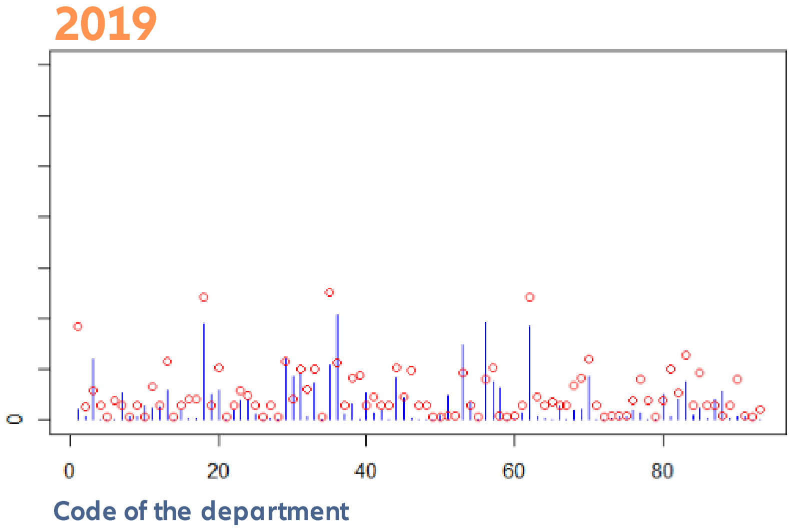
- Scale of **severity**
- **Early** estimate for 2020
- 2016,2017: model a bit **optimistic**

XGBOOST: NUMBER OF CLAIMS BY DEPARTMENT AND OCCURENCE









Recorded claims

Predicted claims

- 2019
- 2018: South of France predicted as highly affected by drought

- Atypical year
- Temperature patterns
- Model & precipitations

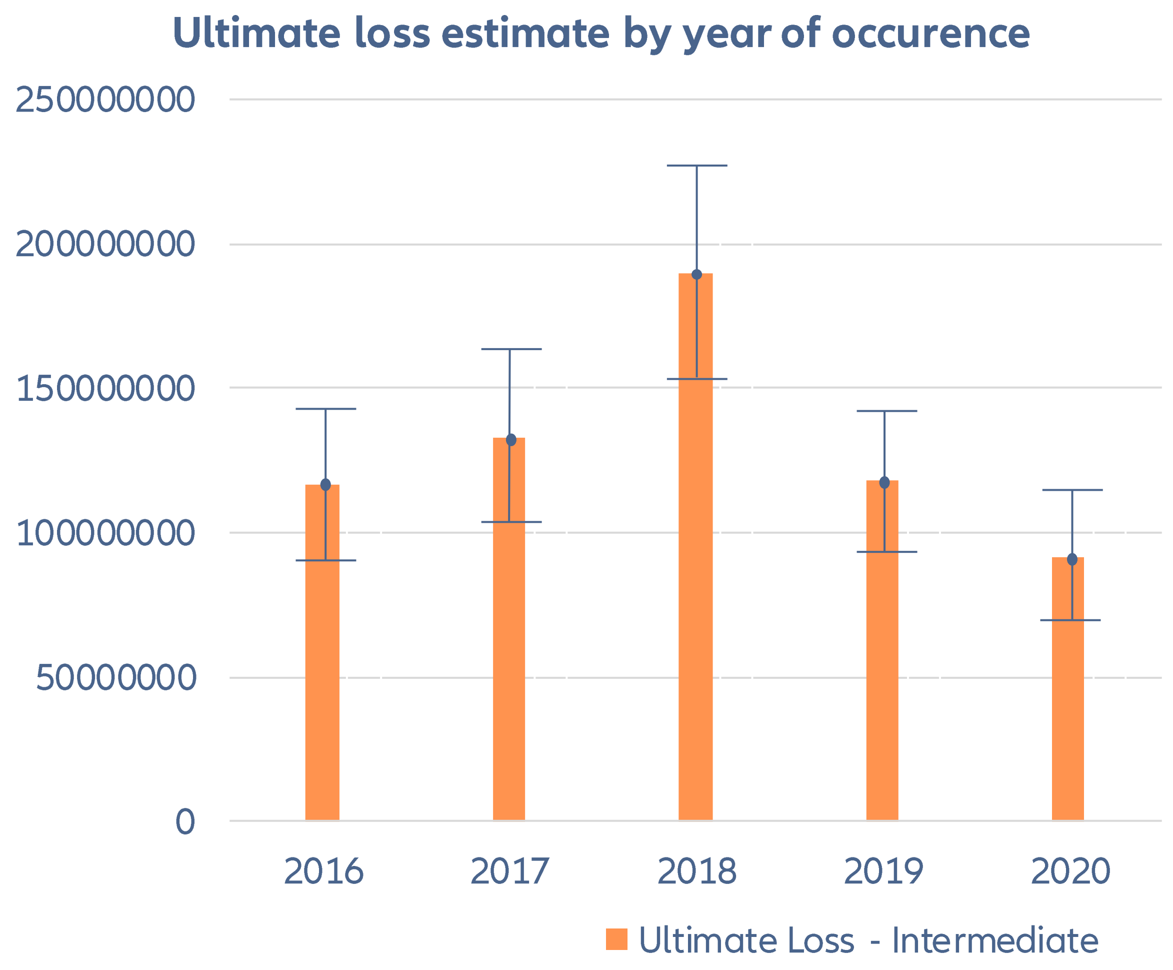
ULTIMATE LOSS

Ultimate loss computation:

- Ultimate number of claims: tail coefficient
- Average cost by claim

2018: 285 million predicted ultimate loss

An optimistic model/Not trained for atypical years



MODEL- CONCLUSION

The subsidence model can give:

- Reasonable first estimates of the overall number of claims by occurrence
- Fast results: by the end of the first year of development

But:

- Difficulties predicting atypical years
- A bit optimistic

MODEL - EXTENSION

Better results for the model updated with claim data at year of development 2

- Expand the training database
- Try implementing Canat process data
  - Catnat process criteria change rapidly