Bayesian Neural Network perspectives

Casualty Actuaries of Europe Meeting

Zurich – 07/10/2022

Aurelien COULOUMY - Chief Data & AI Officer
With the kind contributions of A. KAIS, E. LAVERGNE & A. BEN CHEIKH LEHOCINE

GROUPE
CAISSE CENTRALE DE RÉASSURANCE

1. Introduction

1. Introduction (1/2)

What is Wrong with common ML approaches

- **ML techniques are becoming standards** in many areas of the insurance industry and in actuarial science, with many successful implementations in terms of model performance, data understanding, process automation, etc.

- However, some **issues remain**, including:

Robustness	Trust	Continuity	Optimality

- Consequences are:

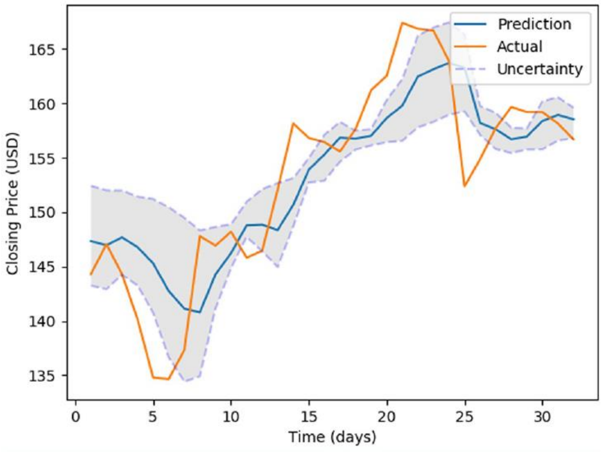
Lack of nuanced decisions and generalization	Difficulty to detect adversarial data and to interpret models	Stability and drift predictions through time	Limitation in algorithmic learning guidance, cost.

1. Introduction (2/2)

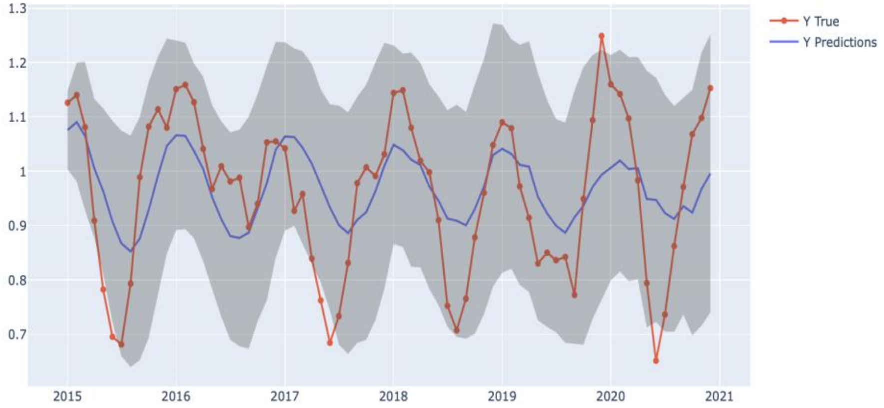
Uncertainty using BNNs

- Approaches that consider the **notion of “uncertainty”** could address such issues.
- **Bayesian Neural Networks (BNN)** are interesting candidates that allow to know when and what the model doesn’t know [1] and to give uncertainty estimations.
- This paradigm also fits well with actuarial science which is based on risk and uncertainty estimation:

Uncertainty quantification for stock price prediction [2]

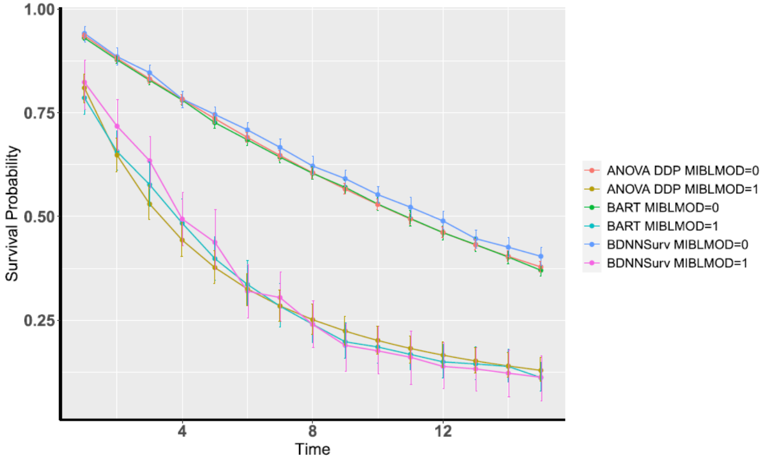


Forecasting of SWI indicators for drought severity prediction [3]



Predictions and confidence interval with Bayesian LSTM

Survival analysis prediction and uncertainty using pseudo values [4]

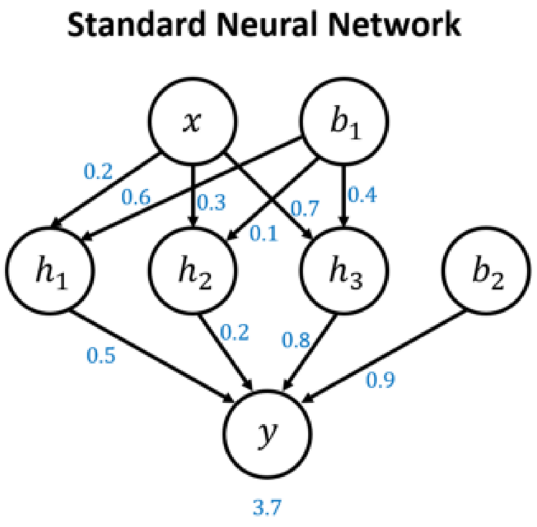


[1] Y Gal, (2016) Uncertainty in Deep Learning,http://www.cs.ox.ac.uk/people/yarin.gal/website//thesis/thesis.pdf,
[2] Chandra R, He Y, (2021) Bayesian neural networks for stock price forecasting before and during COVID-19 pandemic, https://doi.org/10.1371/journal.pone.0253217
[3] Internal CCR Group analysis, (2022) SWI indicators prediction
[4] D Feng, L Zhao, (2021) BDNNSurv: Bayesian deep neural networks for survival analysis using pseudo values, https://jds-online.org/journal/JDS/article/1244/info ,

2. What are BNNs and uncertainties

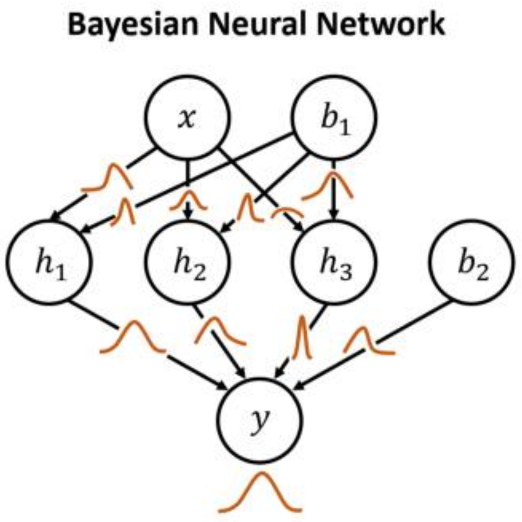
BNNs - Overview

- Classical ML approach: learn the most optimal **combinations of weights/parameters** minimizing a specified loss function. *



- Each weight has a single value referred as a **point estimation**.
- Use **differentiation** to find the optimal value such as gradient descent.

- Bayesian ML approach: learn the **a posteriori distribution** on the model parameters from Bayes' rule. * [5] [6] [7]



- Each weight is represented by an **optimal distribution**.
- Use **approximation** methods to draw the optimal posterior distribution.

* Formulas and decompositions available in EAA presentation https://www.actuview.com/media/downloadAttachment/key/526aea265200d7349c6f40a66f3afe41/maid/599

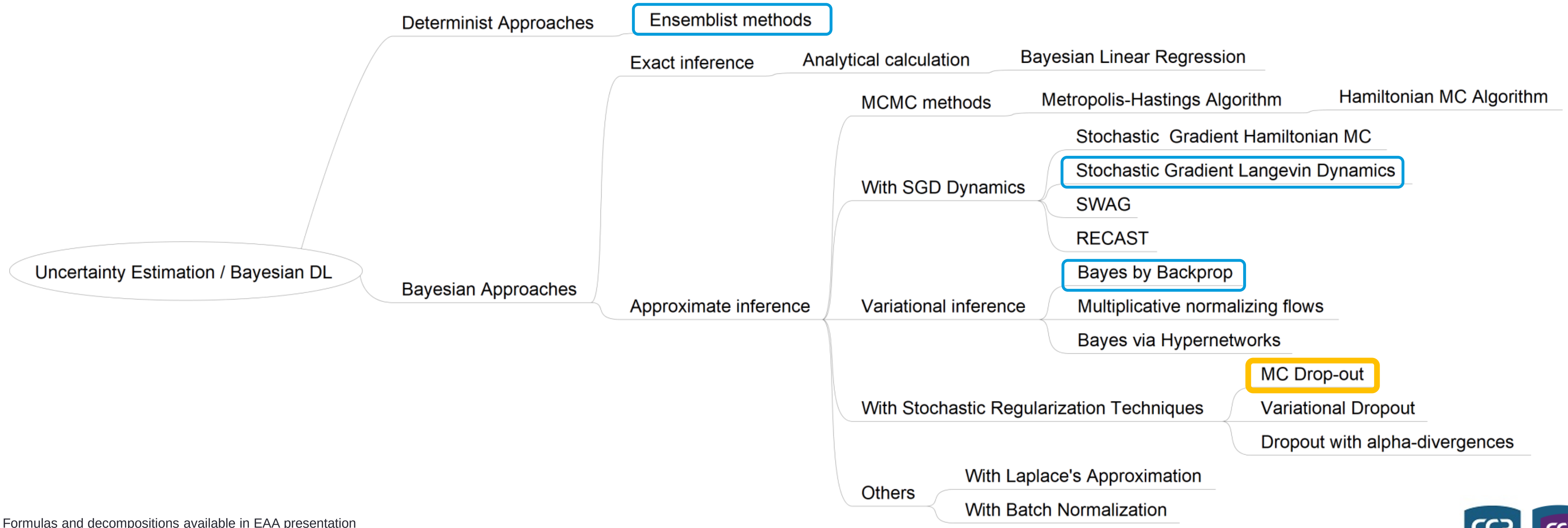
[5] N. G. Polson, V. Sokolov et al., (2017) Deep learning: a Bayesian perspective, Bayesian Analysis, vol. 12, no. 4, pp. 1275–1304
[6] J. Lampinen and A. Vehtari, (2001) Bayesian approach for neural networks—review and case studies, Neural Networks, vol. 14, no. 3, pp. 257 – 274,
[7] D. M. Titterington, (2004) Bayesian methods for neural networks and related models," Statist. Sci., vol. 19, no. 1, pp. 128–139, 02

2. BNNs and uncertainties (2/5)

BNNs - Approximations

- From a practical perspective, Bayesian inference using Neural Networks is not trivial:
  - **Impossible computation** of Bayes' rule analytically;
  - MCMC methods are **costly** both regarding computationally and memory.

- **Several approximation methods *** have emerged in recent years:

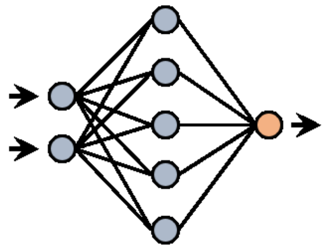


* Formulas and decompositions available in EAA presentation

2. BNNs and uncertainties (3/5)

BNNs - Monte-Carlo Dropout

- Dropout refers to **randomly dropping out units** (in our case nodes) during training.
- Monte Carlo Dropout [8] is currently one of the **most practical** methods available (because of its easiness)
- It allows to **reinterprets the dropout** as an approximation of the Bayesian approach.
- It continues to use the “stochasticity” of dropout **during the prediction/test phase** to get several credible models (weights from approximate posteriors).



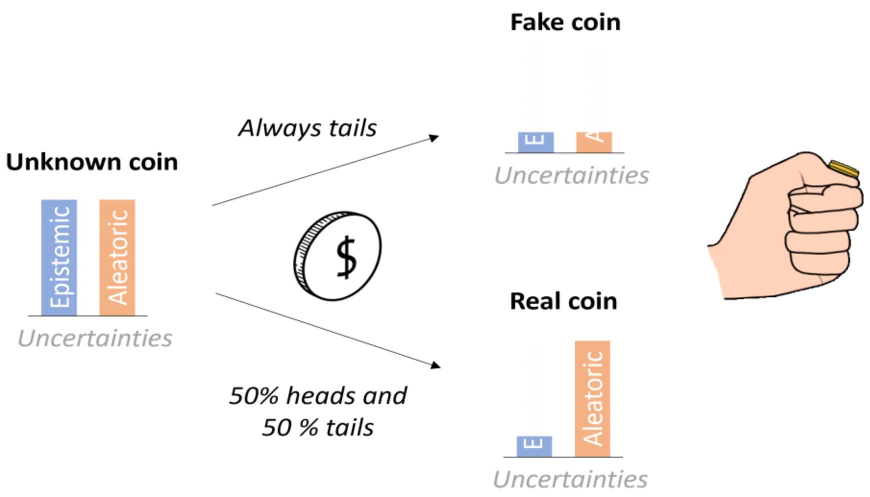
Monte Carlo Dropout

[8] G. Yarin and G. Zoubin. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. https://arxiv.org/abs/1506.02142

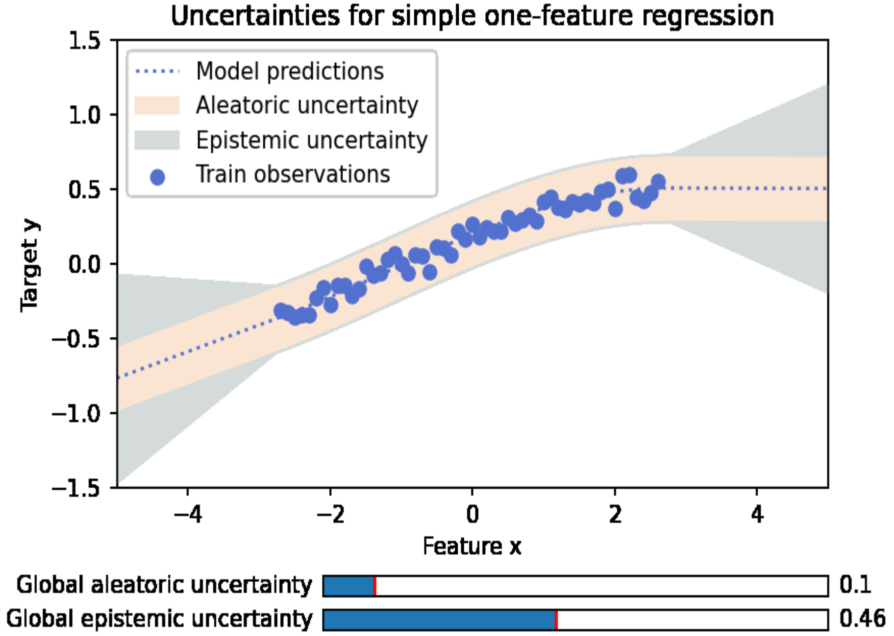
2. BNNs and uncertainties (4/5)

BNNs - Where does it come from?

- Predictive uncertainty reflects **how likely** a prediction is to be wrong on certain observations.
- Bayesian framework is **useful to estimate uncertainty** as it gives a range of credible predictions.
- Uncertainty can be decomposed [9] into:
  - **Aleatoric** uncertainty: noise in data
  - **Epistemic** uncertainty: model lack of knowledge



- One example:
  - Aleatoric uncertainty is high here in areas where the target variable **does not follow a deterministic** relationship with the feature variable
  - Epistemic uncertainty is high here in areas where there is **insufficient data**

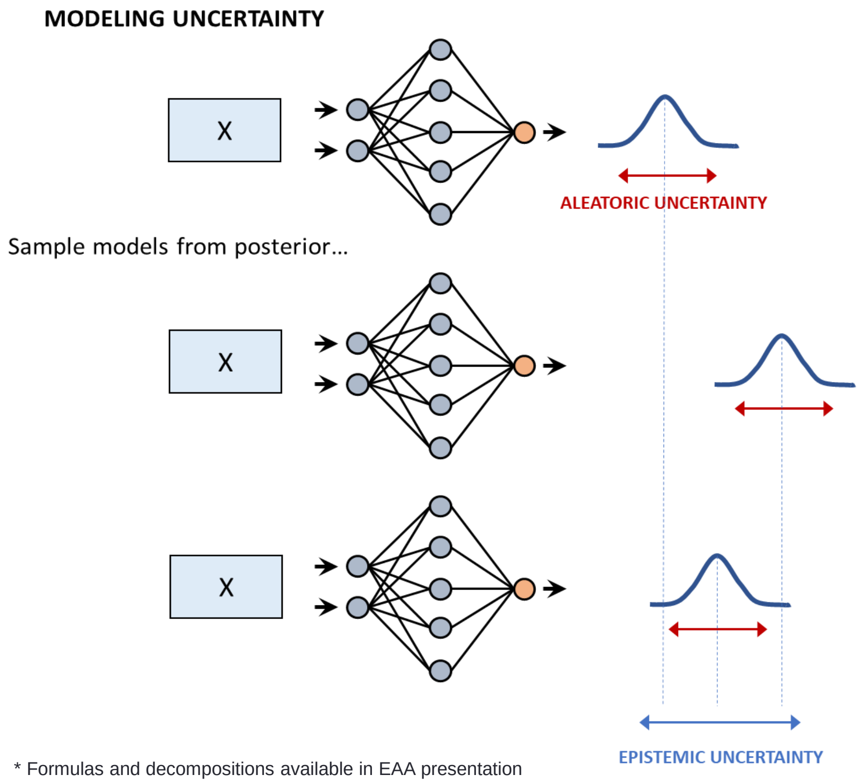


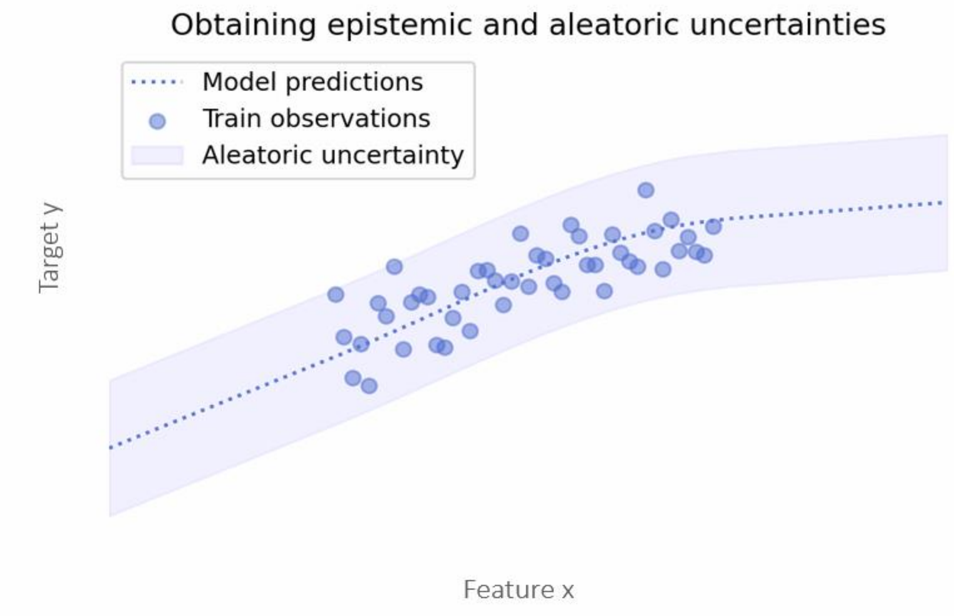
[9] A Der Kiureghian and O Ditlevsen. (2009) Aleatory or epistemic? does it matter? Structural Safety, 31 (2):105–112,

2. BNNs and uncertainties (5/5)

BNNs – How to estimate it?

- Epistemic uncertainty is **modelled with the Bayesian approach** by introducing a distribution on the parameters (posterior)
- Aleatoric uncertainty is modelled **using distribution on model output** (likelihood)





- For **classification** cases:
  - "Total predictive uncertainty can be measured by the predictive entropy, i.e. entropy of mean prediction" *
- For **regression** cases:
  - "Total predictive uncertainty can be measured by the total variance of the predictive distribution" *

* Formulas and decompositions available in EAA presentation

3. Application

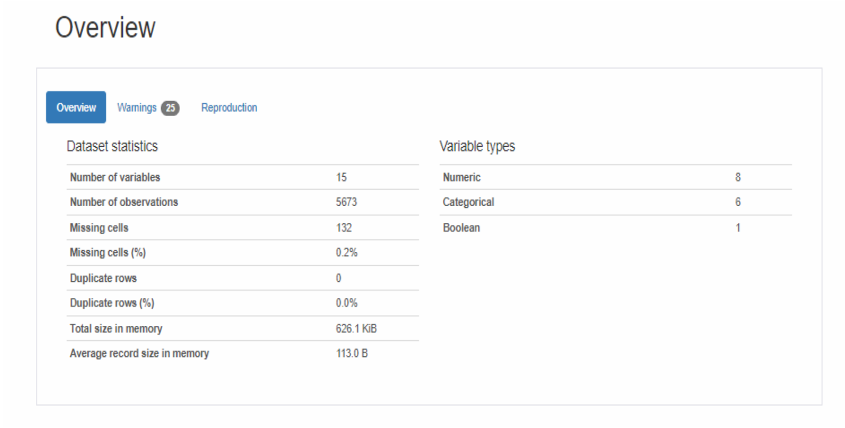
3. Application (1/7)

Context

- **French motor insurance** portfolio collected for reinsurance purpose. A first "manual" analysis was developed in 2019.

- **~2k severe bodily injury** claims from 1999 to 2021, reviewed annually.

- Updated prejudices charges with **~137k observations**.



Overview

Overview | Warnings 25 | Reproduction

Dataset statistics

Number of variables	15
Number of observations	5673
Missing cells	132
Missing cells (%)	0.2%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	626.1 KiB
Average record size in memory	113.0 B

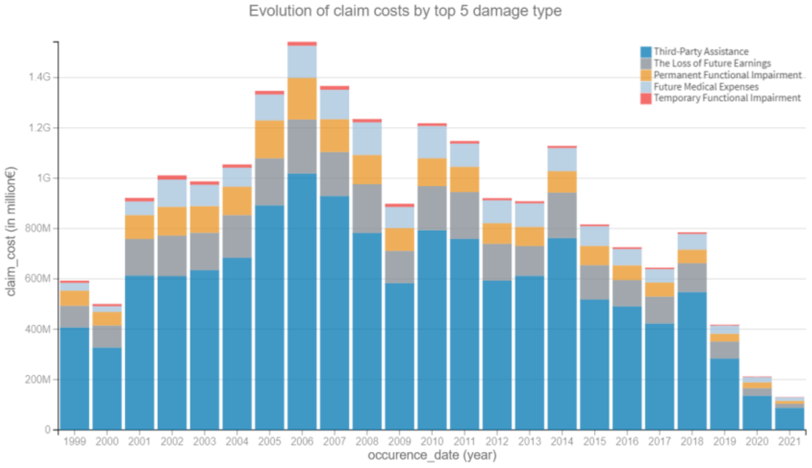
Variable types

Numeric	8
Categorical	6
Boolean	1

- **Key features** identified: age, sex and socio-professional category of the victim, type of injury, rate of permanent damage to physical integrity.

- Work will consist of **standard ML regression** with tabular data for predicting the severity of prejudice charges, globally and per type.

- About **45 prejudice types**. We focus on the top 3: Permanent functional deficit, Temporary functional deficit and Third party support.



[10]

[10] CCR Re Publication Third party liability compensation of severe motor bodily injury claims, (2019)
https://www.ccr-re.com/documents/20123/54390/Third+party+WhitePaper+-+CCR+Re+-+WEB.pdf

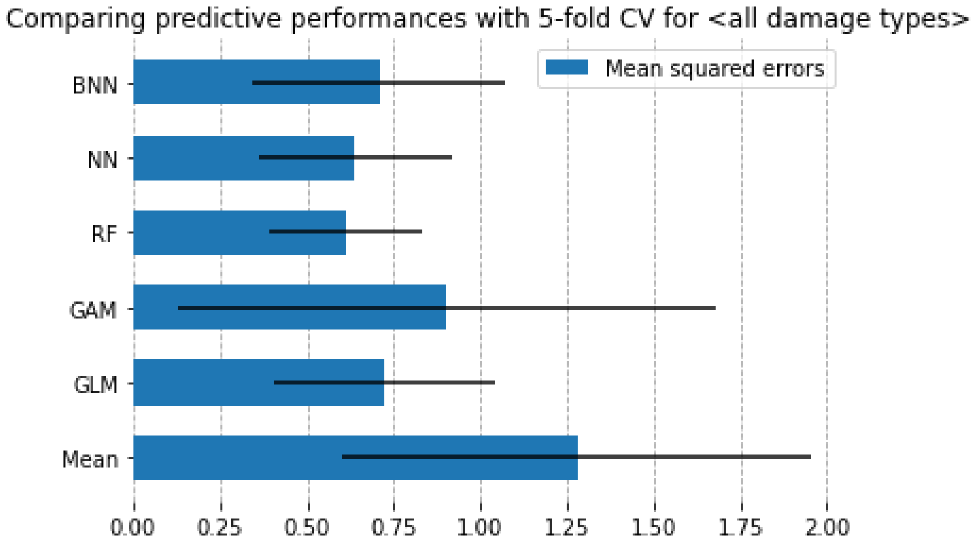
3. Application (2/7)

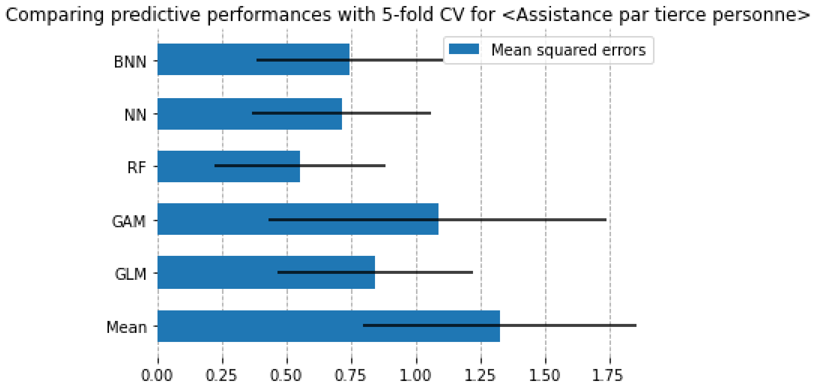
Robustness

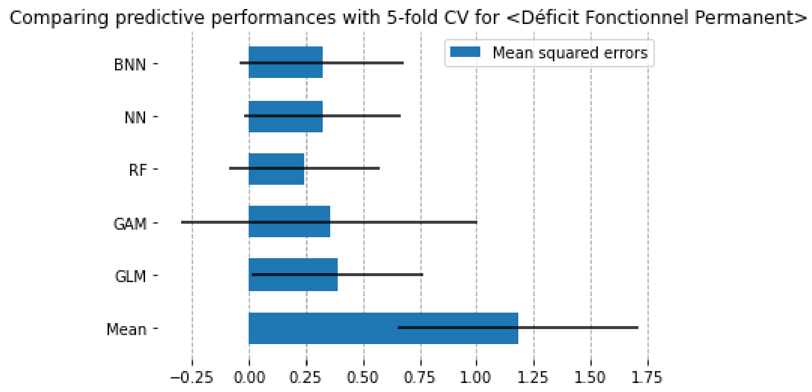
? **Are BNNs good enough** comparing to standard machine learning, neural networks or actuarial methods? *

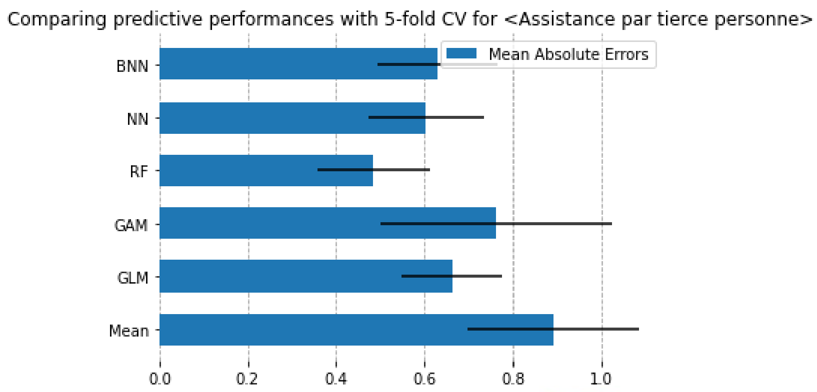
BNNs provide interesting results with limited volatility, most of the time with equal MSE compared to common NN.

RF still provide better results and common GLM (not specifically adapted) as well as GAM lag behind.









* Disclaimer: results are deeply correlated to data and use case specificities

3. Application (3/7)

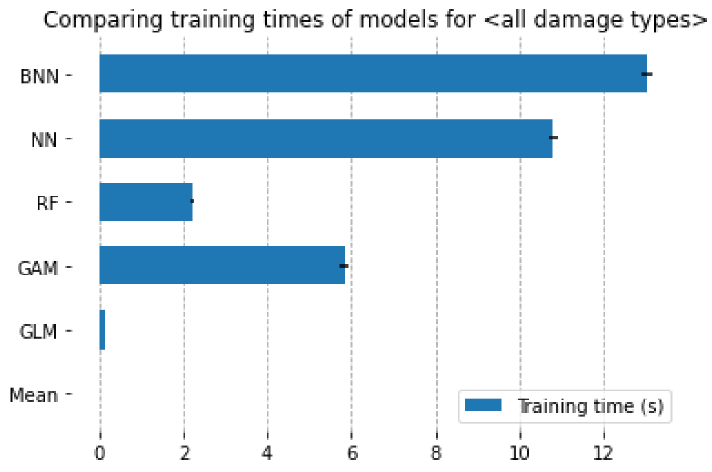
Robustness

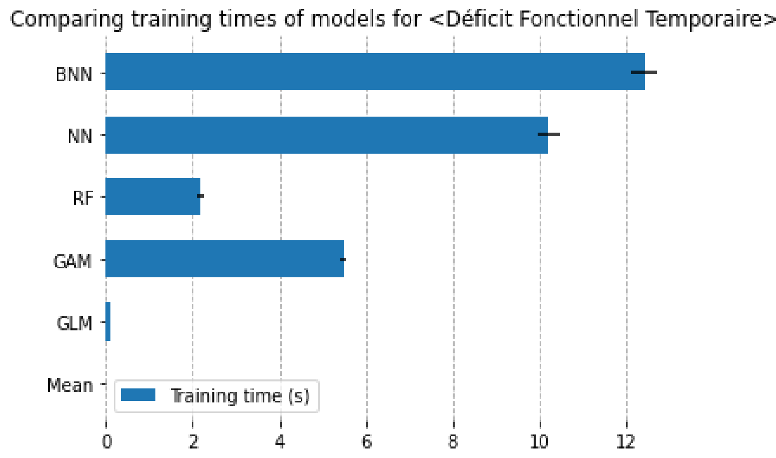
? **How fast are BNNs?** How to ensure that BNNs are viable for production run (regarding both training and inference time)? *

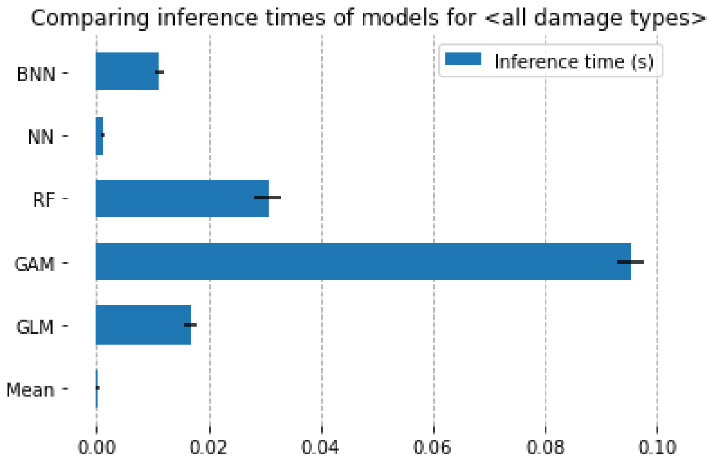
BNNs require a much longer time to converge for training

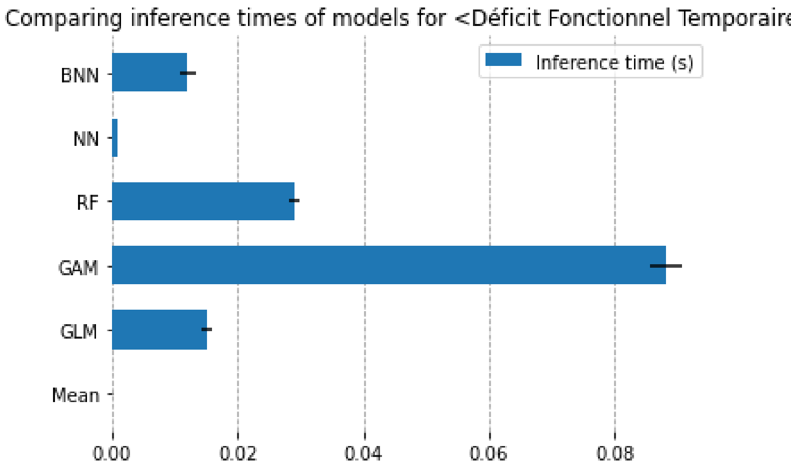
Inference time for BNNs on contrary is quite good, even compared to GLMs.

Results are not affected by prejudice type task









* Disclaimer: results are deeply correlated to data and use case specificities

3. Application (4/7)

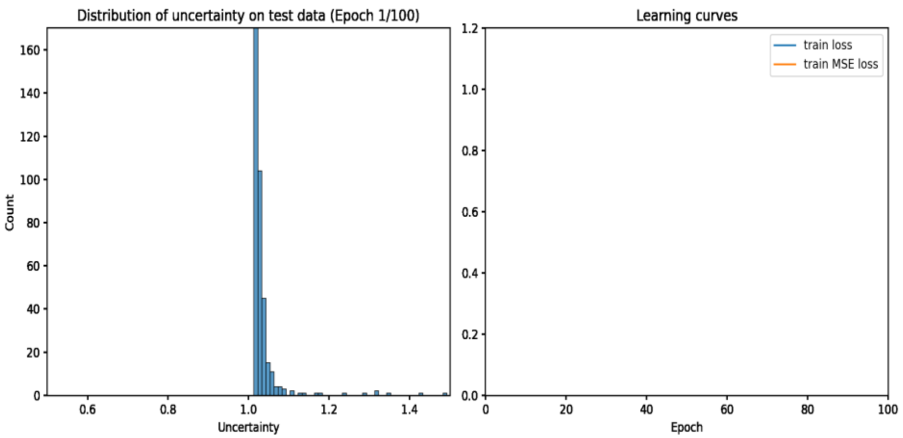
Trust

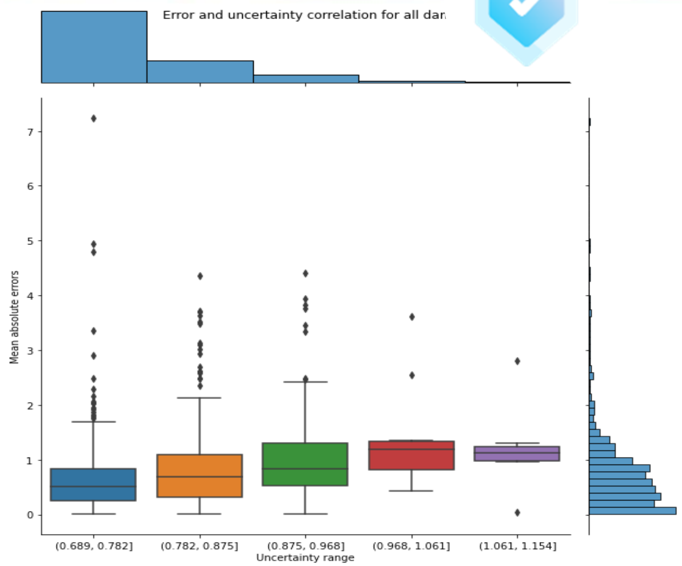
? Can we **profile uncertainty** over training time ? How related are **uncertainty and error** measures?

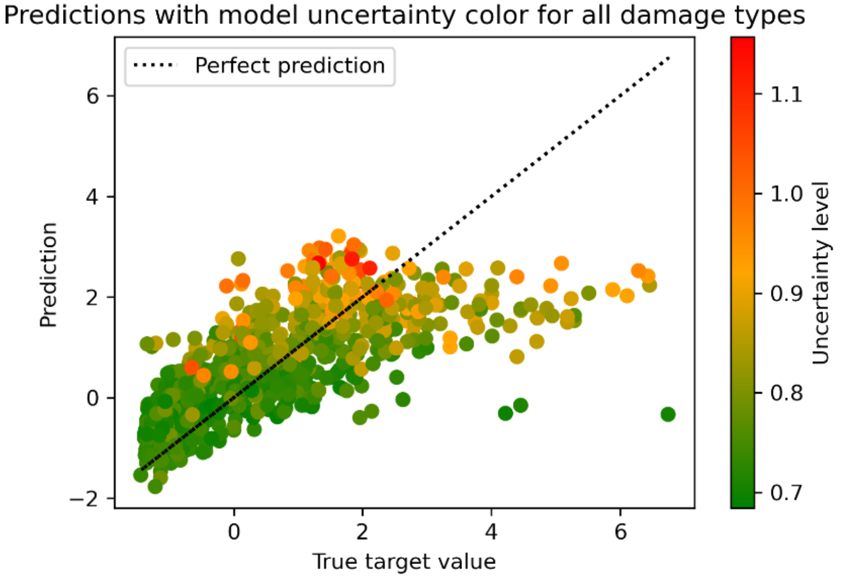
While loss is decreasing, we clearly observe uncertainty profile flatten.

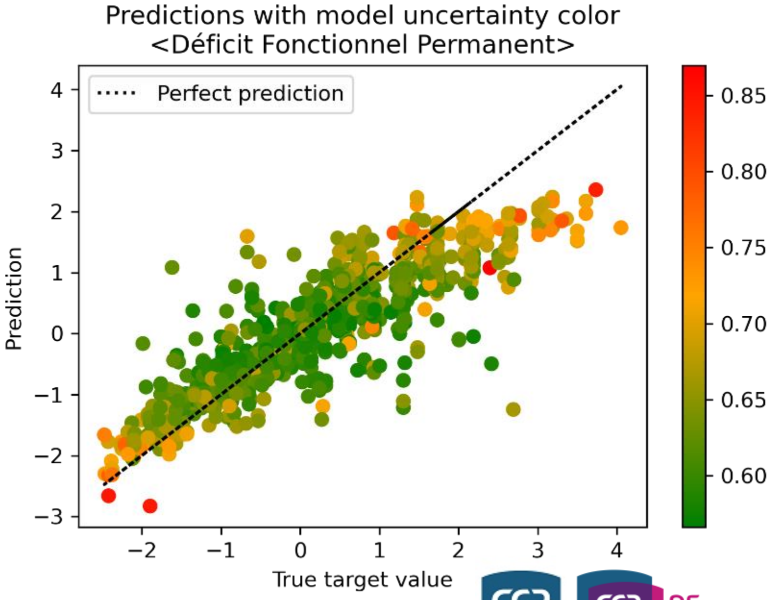
The more the error increases the more the uncertainty also increases and becomes more volatile.

Uncertainty is also observed for data at specific target ranges, with no evident errors.









CCR CCR RE

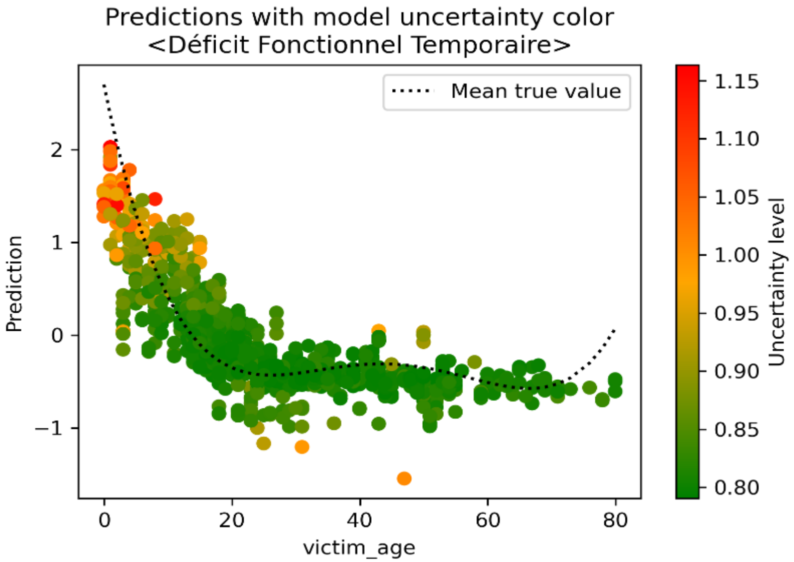
3. Application (5/7)

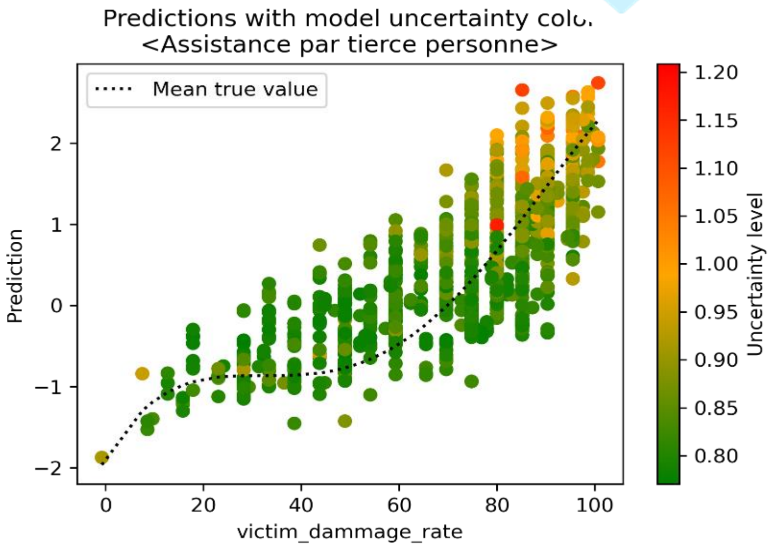
Trust

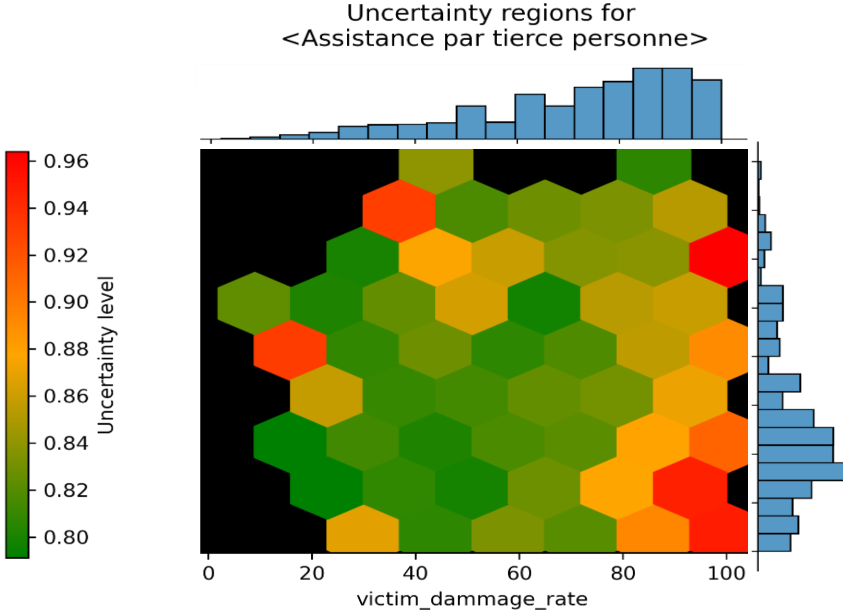
? How to formalize **links between uncertainty measure and features** or observations?

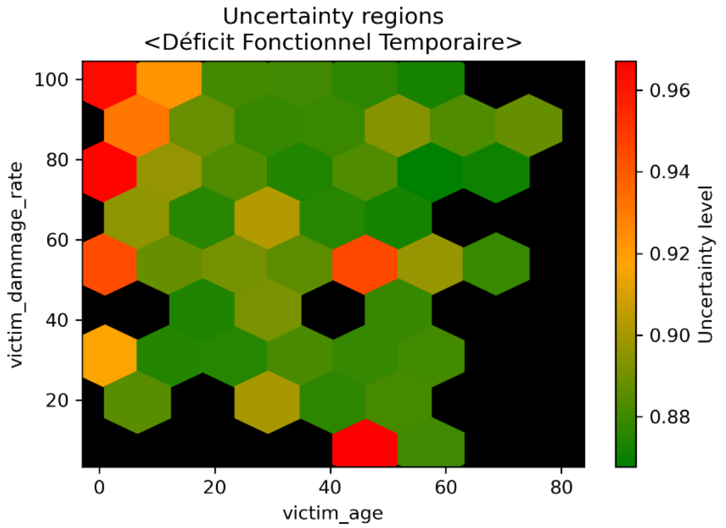
Using partial dependance plots with uncertainty we can analyze for some feature ranges unlikely predictions.

Multivariate analysis allows to highlights unknown combinations (missing observation profile).









3. Application (6/7)

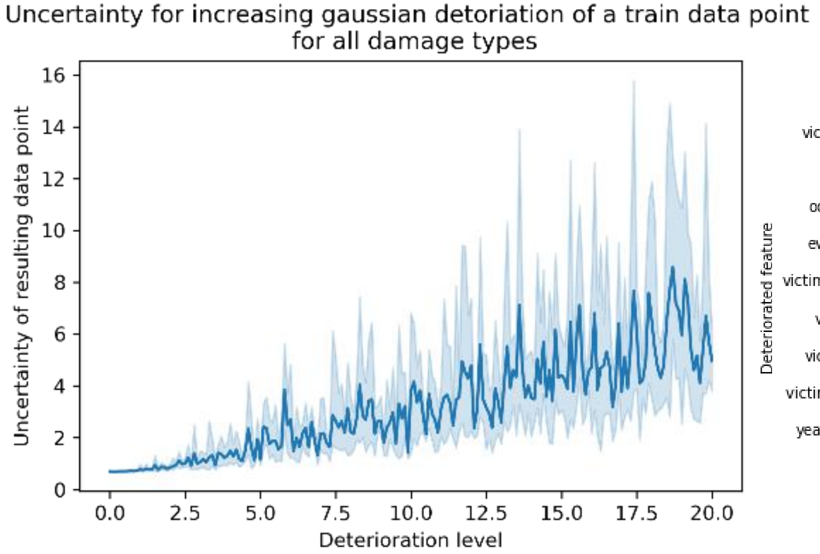
Continuity

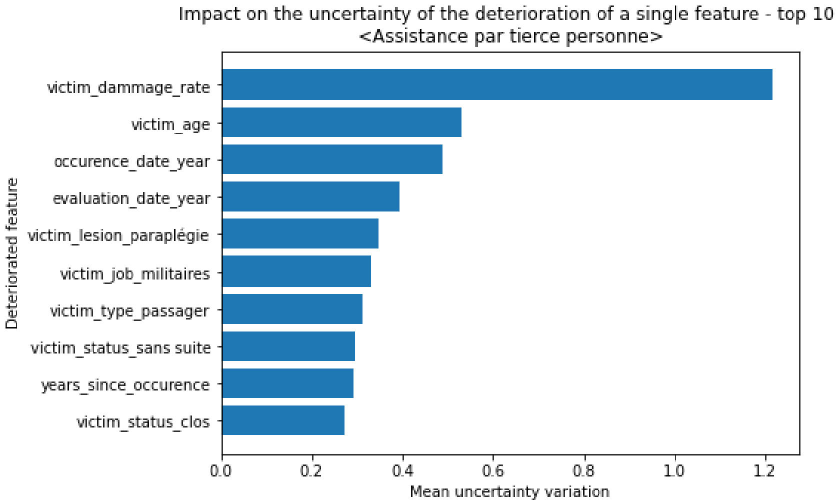
How BNNs can help regarding model or data **analysis through time**? How does it assist **drift** analysis?

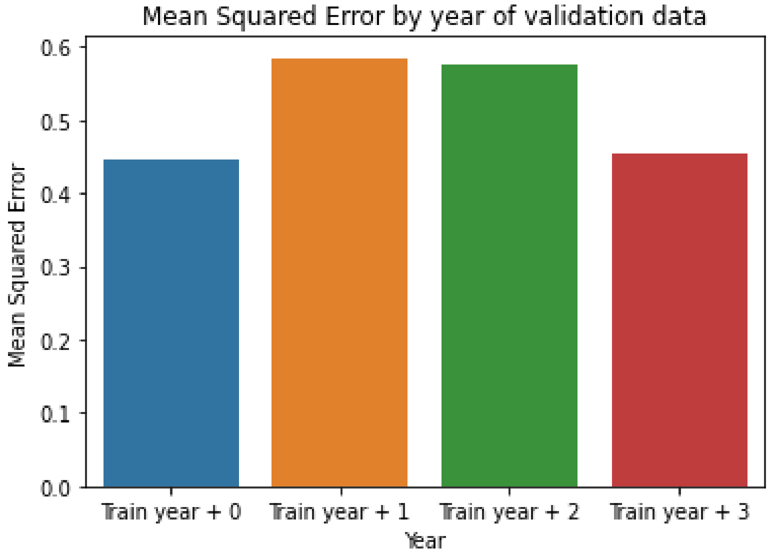
Deterioration function allow to demonstrate model adaptability to features changes.

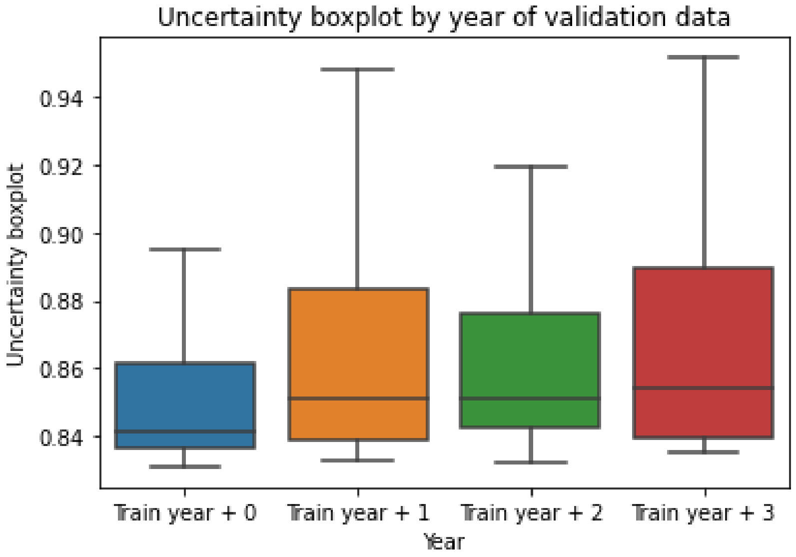
It appears helpful, in addition of importance feature analysis, to highlight key variables.

It is also a good complementary tool to follow model drift. We observe here stable MSE while uncertainty increases and becomes volatile after 3 years.









3. Application (7/7)

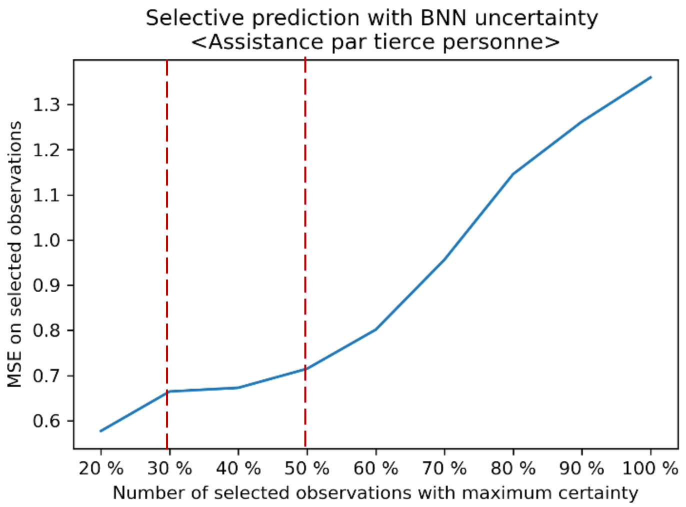
Optimality

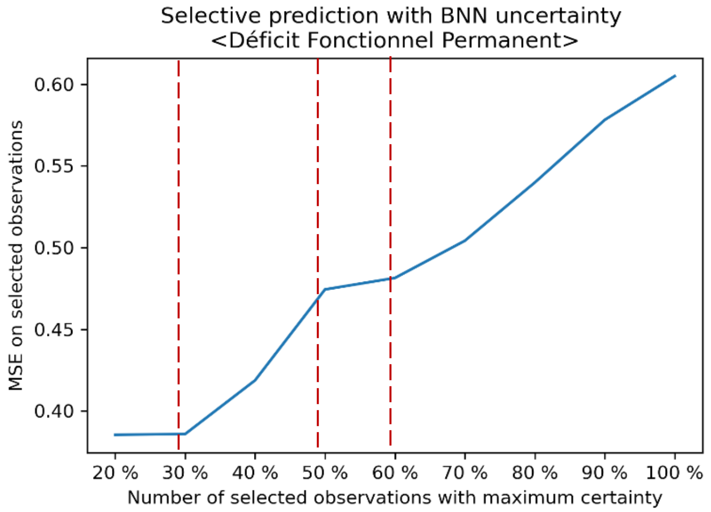
How can we benefit from BNNs and **optimize learning costs**, prediction quality, etc.?

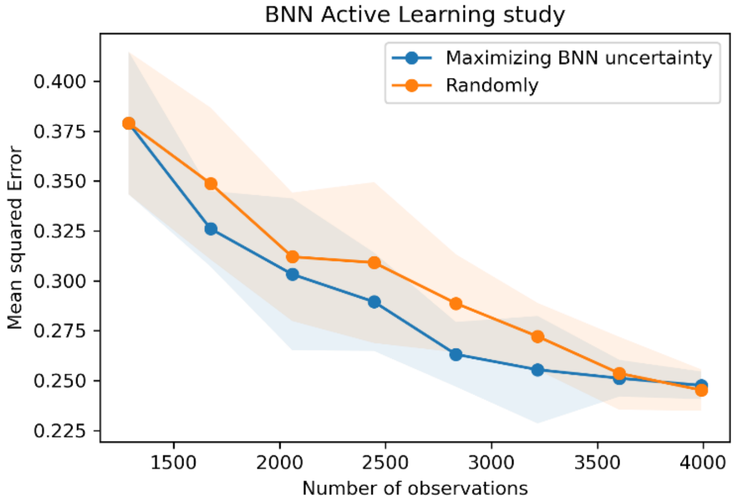
During inference, we can define uncertainty threshold to ensure MSE expectations.

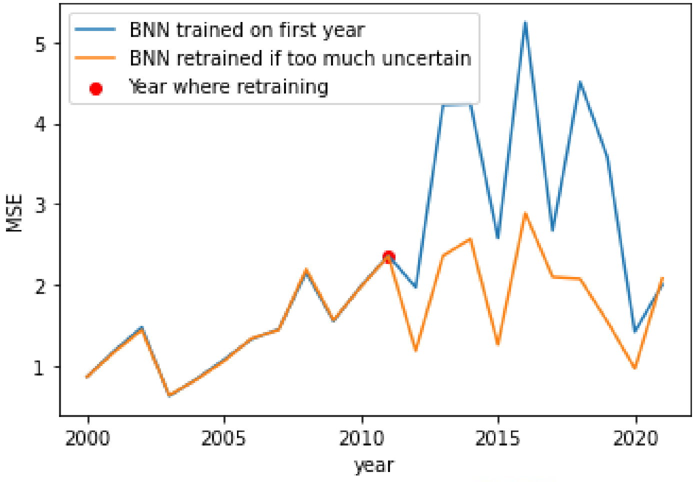
With active learning [11] approaches we can also minimize retraining costs while minimizing also MSE values.

Finally, we can mix both threshold and active learning to define retraining strategies.







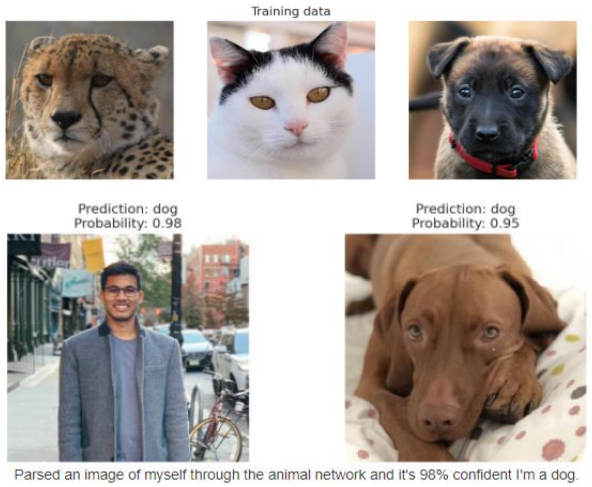


[11] B Settles, (2010), Active learning literature survey, http://burrsettles.com/pub/settles.activelearning.pdf

4. Conclusion & perspectives

Conclusion

- Despite **relative theoretical complexity**, BNNs can be developed to add uncertainty notions into standard actuarial / ML tasks.
- **Results are promising**, in terms of time inference, model quality, interpretability capabilities, continuity add-on, process optimization, etc.
- We observe **BNNs drawbacks**: training/test time, difficulty of training (choice of prior distribution), lack of interpretability chart baselines.
- At the end, there would be many **risks** [12] **not to consider BNNs and model uncertainty**:



Overconfident prediction of a dog [13]

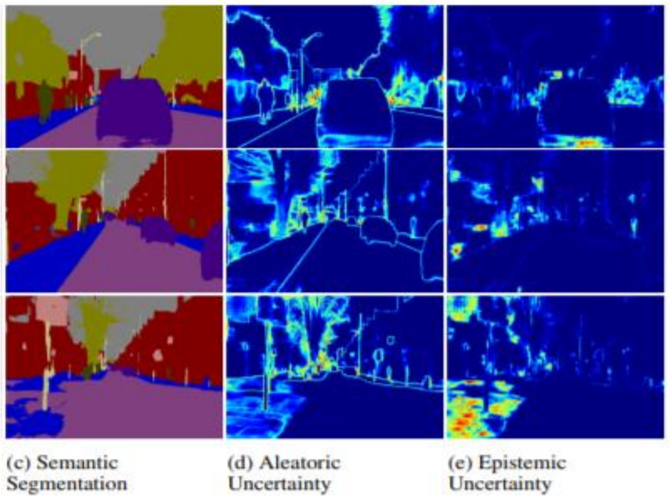
Most certain predictions

workclass	Private
fnlwgt	-1.22549
education	Doctorate
education-num	2.09553
marital-status	Married-civ-spouse
occupation	Prof-specialty
relationship	Husband
race	White
sex	Male
capital-gain	9.00439
capital-loss	-0.259806
hours-per-week	1.06957
native-country	United-States
salary	>=50k
education-num_na	False
Name: 6300, dtype: object,age	
workclass	Self-emp-inc
fnlwgt	0.159757
education	Doctorate
education-num	2.09553
marital-status	Married-civ-spouse
occupation	Prof-specialty
relationship	Husband
race	White
sex	Male
capital-gain	9.00439
capital-loss	-0.259806
hours-per-week	2.31517
native-country	United-States
salary	>=50k
education-num_na	False

Most uncertain predictions

workclass	Private
fnlwgt	-0.24207
education	Some-college
education-num	-0.227131
marital-status	Divorced
occupation	Transport-moving
relationship	Not-in-family
race	White
sex	Male
capital-gain	-0.201885
capital-loss	-0.259806
hours-per-week	0.654366
native-country	United-States
salary	<50k
education-num_na	False
Name: 13145, dtype: object,age	
workclass	Self-emp-not-inc
fnlwgt	1.20672
education	Some-college
education-num	-0.227131
marital-status	Married-civ-spouse
occupation	Farming-fishing
relationship	Husband
race	White
sex	Male
capital-gain	-0.201885
capital-loss	-0.259806
hours-per-week	3.47774
native-country	United-States
salary	<50k
education-num_na	False

Bias and Ethic in tabular data classification with Adults Income [14]



Capturing various uncertainty measures on computer vision tasks [15]

[12] A Nguyen, J. Yosinski, J. Clune, (2014), Deep Neural Networks are Easily Fooled https://arxiv.org/abs/1412.1897
[13] J Ramkissoon (2020) Dealing with Overconfidence in Neural Networks: Bayesian Approach, https://jramkiss.github.io/2020/07/29/overconfident-nn/
[14] D. Huynh (2019) Bayesian deep learning with Fastai,
[15] A Kendall, Y Gal, (2017) What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? https://arxiv.org/pdf/1703.04977.pdf

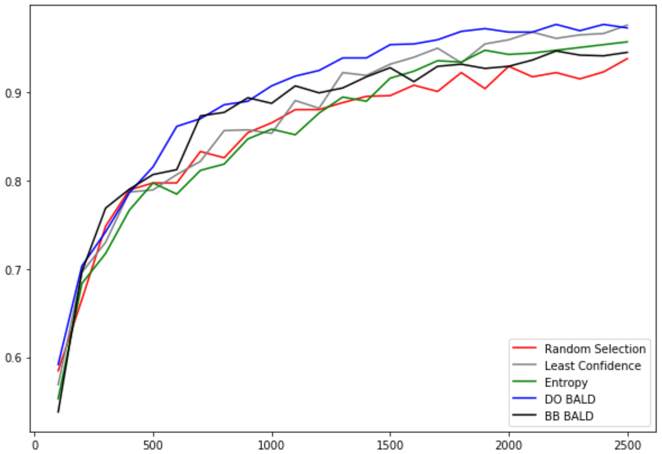
4. Conclusions & perspectives (2/3)

Perspectives

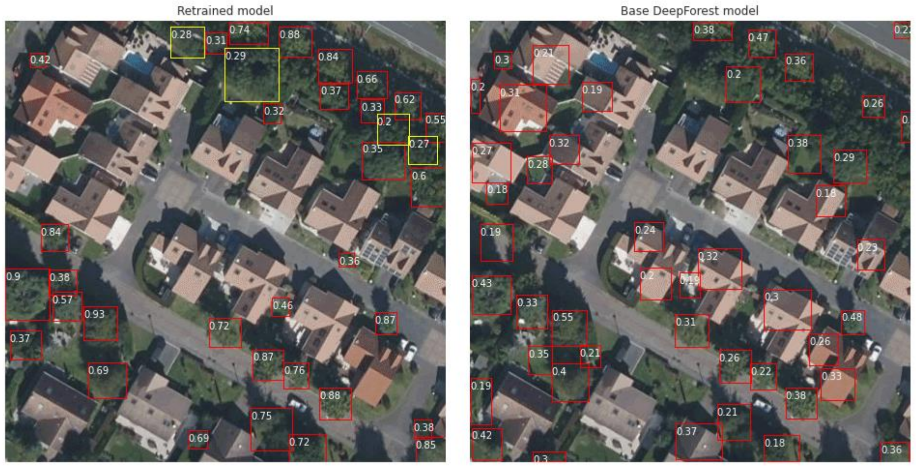
- **Several perspectives** can be discussed:
  - Deeper exploration of aleatoric or epistemic uncertainty measures relation and representation;
  - Integration of such uncertainty measures within daily processes (library?);
  - Exploration of out of domain data uncertainty;
  - Other examples in actuarial science (claim reserving, mortality rate prediction, ESG, BEL, etc.) or experienced in CCR Group (Cyber risk, SWI indicators for drought nat cat modelling, etc.);
  - Other insurance tasks such as NLPs (Custom NER Active learning and Clause classification outliering) or CV (for TreeDetection).

1998	3119574.000000	47184563.000000	129571818.000000	205210185.000000	250488154.000000	276516578.000000	289448982.000000	294634712.000000
1999	2611938.000000	48712676.000000	128109553.000000	199696111.000000	251226483.000000	278089709.000000	289787024.000000	296221318.000000
2000	3660623.000000	48578158.000000	131138360.000000	203536066.000000	248153147.000000	275513570.000000	290183175.000000	294503166.000000
2001	1969019.000000	46814727.000000	129874801.000000	200157921.000000	249622859.000000	277553530.000000	291323882.000000	296885103.000000
2002	4136072.000000	51272019.000000	133205433.000000	207331606.000000	250449602.000000	278628322.000000	291498846.000000	297631758.000000
2003	3592737.000000	49380700.000000	135461144.000000	203766029.000000	249377437.000000	279661188.000000	292897321.000000	298935320.000000
2004	1617625.000000	52367903.000000	127123345.000000	197583449.000000	243772955.000000	270937790.000000	284027624.000000	288476178.000000
2005	1617463.000000	54874170.000000	127918577.000000	199922166.000000	247481150.000000	271736199.000000	284764683.000000	290172197.000000
2006	4056397.000000	46620968.000000	128762059.000000	199871502.000000	246708416.000000	274025670.000000	288207447.000000	292632220.000000
2007	2094358.000000	46397031.000000	132602794.000000	199503974.000000	250744434.000000	276357306.000000	288067869.000000	293180137.000000
2008	3204310.000000	39314302.000000	127133116.000000	195618362.000000	242022989.000000	271386829.000000	285957669.000000	290641213.000000
2009	2281020.000000	48134704.000000	129443792.000000	207226593.000000	253528126.000000	275339116.000000	289163444.000000	294661052.000000
2010	3247383.000000	45355418.000000	130948545.000000	197259488.000000	245848248.000000	277067336.000000	288809927.000000	294575540.000000
2011	4023190.000000	52422934.000000	135085441.000000	213929806.000000	262775639.000000	289531575.000000	302093905.000000	304345060.600000
2012	4058287.000000	50069534.000000	142260025.000000	206867581.000000	257968282.000000	283492730.000000	298678830.800000	301681839.400000
2013	3335156.000000	52570292.000000	133763178.000000	206994774.000000	254275027.000000	278221412.100000	293279630.000000	295392654.300000
2014	1727421.000000	48142658.000000	129925423.000000	202030672.000000	247221887.000000	271516891.300000	287485484.600000	291799355.000000
2015	2707623.000000	44076323.000000	130639869.000000	202122480.400000	246644996.900000	269395892.600000	286703049.900000	291640882.900000
2016	3057485.000000	41569188.000000	130478630.000000	199761028.300000	244225817.400000	266442116.300000	284760652.100000	290886684.200000
2017	2458100.000000	47018050.100000	130123667.500000	201967398.100000	245349943.800000	268009981.800000	286374167.400000	292205569.000000

Individual claim reserving study example using Bayesian LTSM prediction



Active learning analysis to improve Custom NER annotation applied to reinsurance treaties analysis context [16] [17]



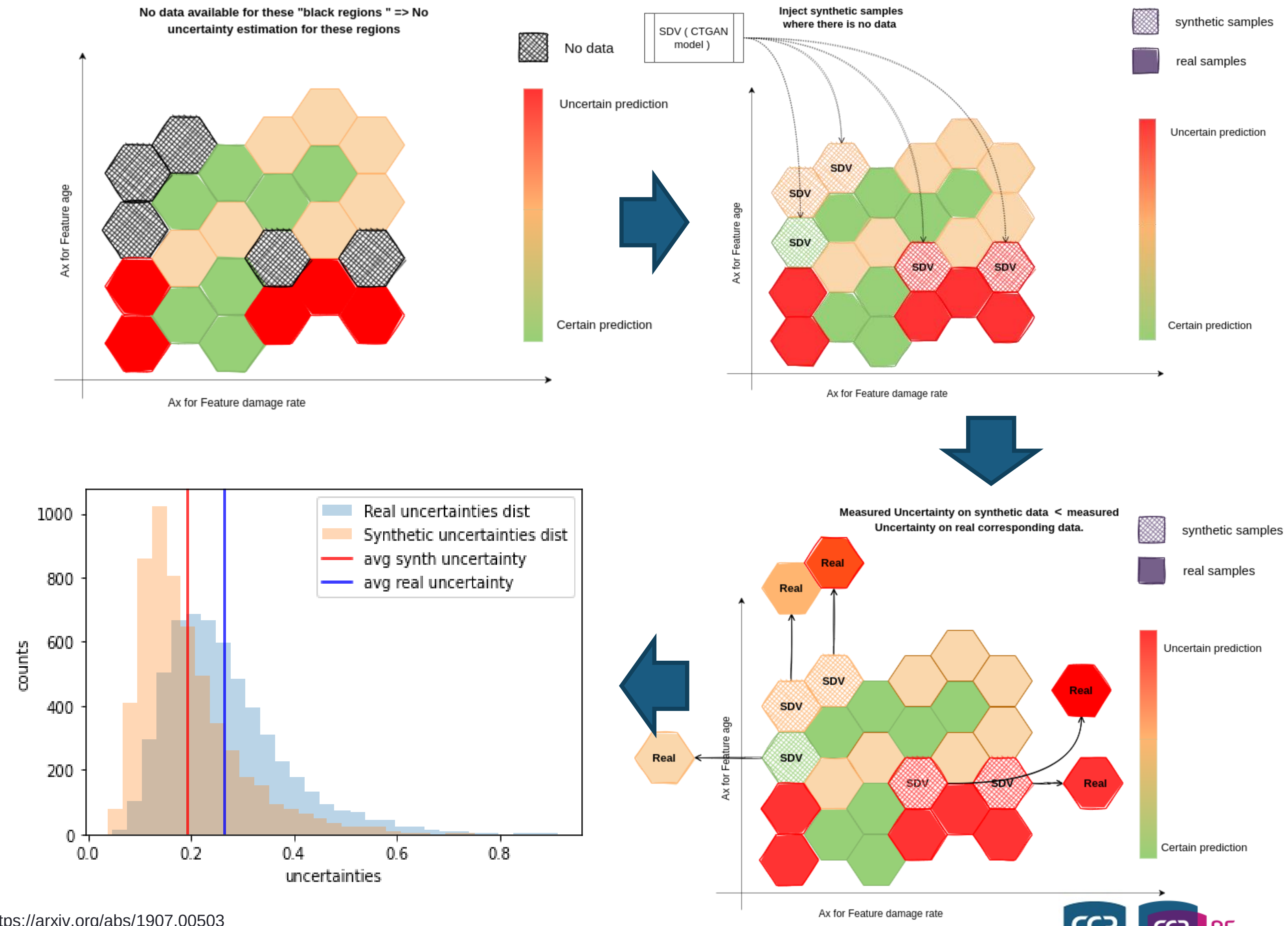
Custom DeepForest model inference on French aerial images study of softmax vs uncertainty.

[16] F. Planchet, C. Y Robert, (2019) Insurance Data Analytics, NLP methodological triggers to address Insurance domain issues, Economica
[17] A. Siddahant, Z. C. Lipton (2018), Deep Bayesian Active Learning for Natural Language Processing: Results of a Large-Scale Empirical Study, arXiv:1808.05697

4. Conclusions & perspectives (3/3)

Perspectives

- We have explored **Synthetic Data Vault (SDV)** to know better how models could react and be uncertain to rare or unknown events.
- We have used **CTGAN** [18] **on the dataset application**, by randomly dropping regions and training BNNs
- Then we use synthetic data for BNNs inference and we **compare to real data uncertainty**.
- We have observed that **average uncertainty estimation on synthetic data is a great lower bound** to its real valuation and tell us more on model understanding.



[18] Lei et al, Oct 2019. Modeling tabular data using conditional GAN. https://arxiv.org/abs/1907.00503

Contact

Thank you for your attention

Contact:

Aurélien COULOUMY
+33 6 26 13 09 97
Chief Data & AI Office – CCR Group – acouloumy@ccr.fr
Lecturer – Université Lyon 1 ISFA – aureliencouloumy@gmail.com

5. Appendix

5. Appendix

References

[1] Y Gal, (2016) Uncertainty in Deep Learning,http://www.cs.ox.ac.uk/people/yarin.gal/website//thesis/thesis.pdf,
[2] Chandra R, He Y, (2021) Bayesian neural networks for stock price forecasting before and during COVID-19 pandemic, https://doi.org/10.1371/journal.pone.0253217
[3] Internal CCR Group analysis, (2022) SWI indicators prediction
[4] D Feng, L Zhao, (2021) BDNNSurv: Bayesian deep neural networks for survival analysis using pseudo values, https://jds-online.org/journal/JDS/article/1244/info
[5] N. G. Polson, V. Sokolov et al., (2017) Deep learning: a Bayesian perspective, Bayesian Analysis, vol. 12, no. 4, pp. 1275–1304
[6] J. Lampinen and A. Vehtari, (2001) Bayesian approach for neural networks—review and case studies, Neural Networks, vol. 14, no. 3, pp. 257 – 274,
[7] D. M. Titterington, (2004) Bayesian methods for neural networks and related models," Statist. Sci., vol. 19, no. 1, pp. 128–139, 02
[8] G. Yarin and G. Zoubin. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. https://arxiv.org/abs/1506.02142
[9] A Der Kiureghian and O Ditlevsen. (2009) Aleatory or epistemic? does it matter? Structural Safety, 31 (2):105–112,
[10] CCR Re Publication Third party liability compensation of severe motor bodily injury claims, (2019)
https://www.ccr-re.com/documents/20123/54390/Third+party+WhitePaper+-+CCR+Re+-+WEB.pdf
[11] B Settles, (2010), Active learning literature survey, http://burrsettles.com/pub/settles.activelearning.pdf
[12] A Nguyen, J. Yosinski, J. Clune, (2014), Deep Neural Networks are Easily Fooled https://arxiv.org/abs/1412.1897
[13] J Ramkissoon (2020) Dealing with Overconfidence in Neural Networks: Bayesian Approach, https://jramkiss.github.io/2020/07/29/overconfident-nn/
[14] D. Huynh (2019) Bayesian deep learning with Fastai,
[15] A Kendall, Y Gal, (2017) What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? https://arxiv.org/pdf/1703.04977.pdf
[16] F. Planchet, C. Y Robert, (2019) Insurance Data Analytics, NLP methodological triggers to address Insurance domain issues, Economica
[17] A. Siddahant, Z. C. Lipton (2018), Deep Bayesian Active Learning for Natural Language Processing: Results of a Large-Scale Empirical Study, arXiv:1808.05697
[18] Lei et al, Oct 2019. Modeling tabular data using conditional GAN. https://arxiv.org/abs/1907.00503

CCR ™
Caisse Centrale de Réassurance
157 boulevard Haussmann 75008 Paris – France
Tél. : +33 1 44 31 00 – http://www.ccr.fr
SA au capital de 60 000 000 € - 388 202 533 RCS Paris

GROUPE
CAISSE CENTRALE DE RÉASSURANCE

CCR | CCR RE