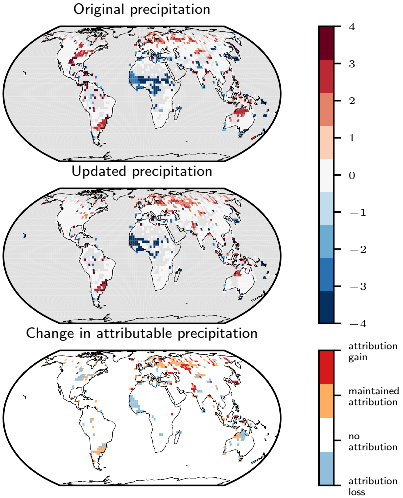
Machine learning based evidence and attribution mapping of 100,000 climate impact studies

Discussion points for resubmission

March 16, 2021

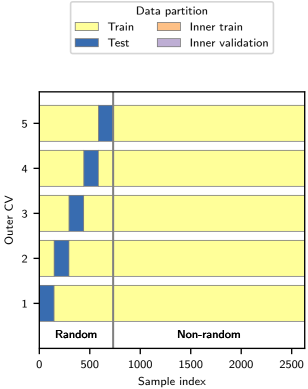
Discussion points

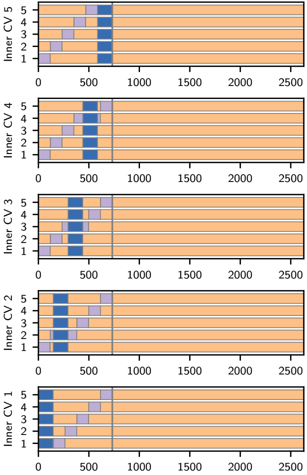
New results from Tom and Shruti



- Much greater area with attributable temperature trends (particularly N America and oceans)
- Some losses of cells with attributable precipitation (W Africa because of models predicting wrong sign, N America, Australia)
- Some additional cells with attributable precipitation in W and NW Asia

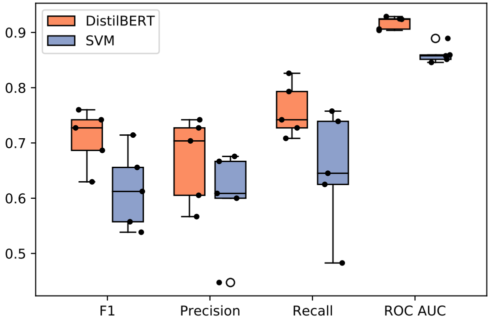
Nested cross-validation





- We no longer use non-representative samples validation
- Our previous validation procedure (solely cross-validation) was subject to overfitting, and did not systematically explore the parameter space
- **Nested cross-validation** separates the selection of hyperparameters from the validation of model performance

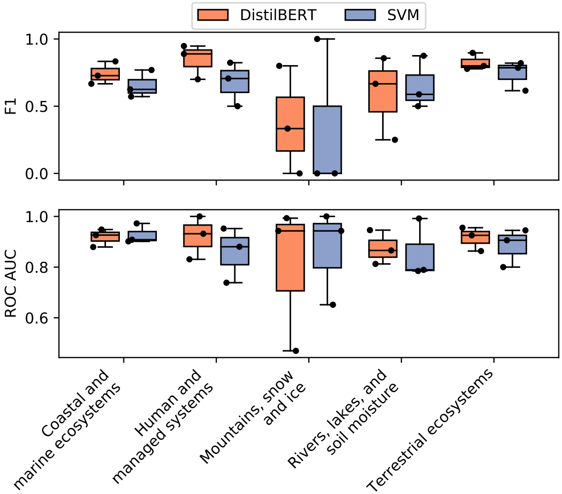
BERT



How to present these best?

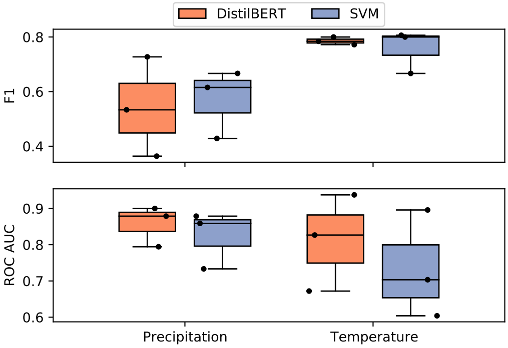
- Our machine learning approach (SVM) was no longer the most cutting edge
- BERT (Bidirectional Encoder Representations from Transformers) type models have in the last couple of years surpassed many benchmarks in NLP tasks.
- BERT are large language models trained on large corpora and can be fine-tuned to perform specific classification tasks
- DistilBERT (a smaller and faster version of BERT) outperforms SVM
- We have poor performance in mountains, snow and ice due to very few examples in the random sample

BERT



- Our machine learning approach (SVM) was no longer the most cutting edge
- BERT (Bidirectional Encoder Representations from Transformers) type models have in the last couple of years surpassed many benchmarks in NLP tasks.
- BERT are large language models trained on large corpora and can be fine-tuned to perform specific classification tasks
- DistilBERT (a smaller and faster version of BERT) outperforms SVM
- We have poor performance in mountains, snow and ice due to very few examples in the random sample

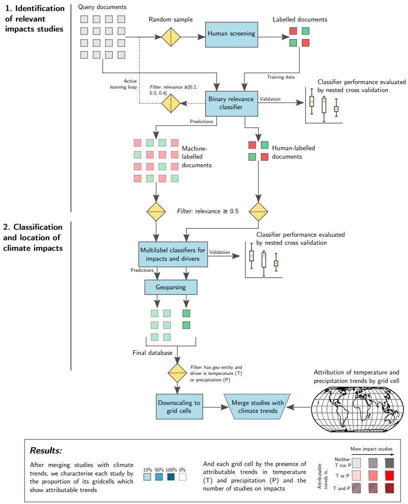
BERT



How to present these best?

- Our machine learning approach (SVM) was no longer the most cutting edge
- BERT (Bidirectional Encoder Representations from Transformers) type models have in the last couple of years surpassed many benchmarks in NLP tasks.
- BERT are large language models trained on large corpora and can be fine-tuned to perform specific classification tasks
- DistilBERT (a smaller and faster version of BERT) outperforms SVM
- We have poor performance in mountains, snow and ice due to very few examples in the random sample

Clarity of process



- We have a detailed diagram better describing the process
- The methods section is significantly rewritten

Abstracts

A major sticking point was the use of abstracts and the question of whether this biases our results

---

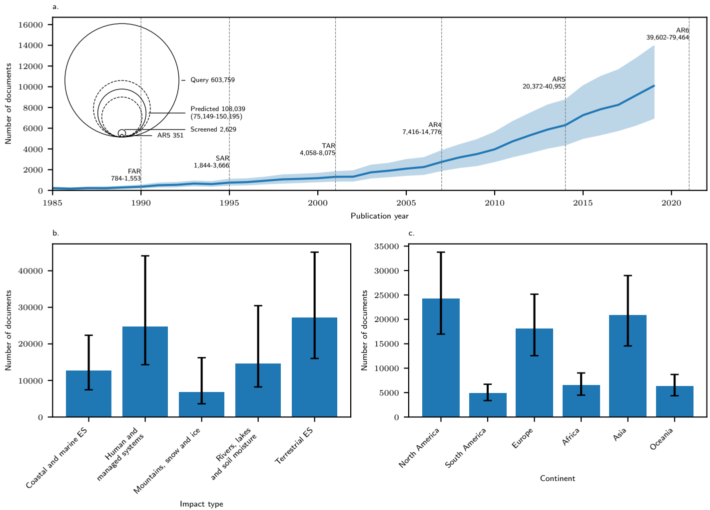
We respond that

- Abstracts contain the best summary of the topic of a document
- Within full texts you have many different types of texts, including reference to other results, which may increase the risk of false positives.
- Further, we simplify the information we want to extract from abstracts to
  - Does the article present evidence on climate impacts *broadly defined*?
  - What *broad* class of impacts does it describe?
  - Are impacts driven by temperature or precipitation or other variables?

  We no longer distinguish between the type of evidence provided (trends or variability) as it could be argued that abstracts are insufficient to answer this fully.

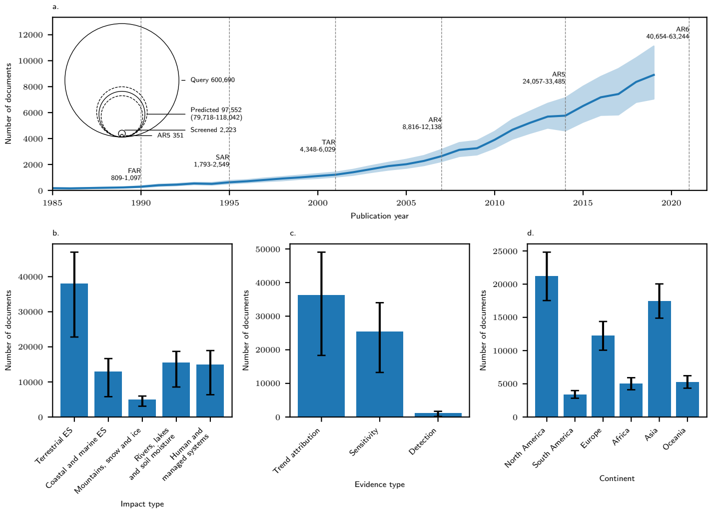
In summary, we may miss some documents which present evidence of climate impacts but do not discuss this in the abstract, but in doing so we avoid the (in our opinion likely more common) class of documents where impacts may be discussed in full texts but this is not the focus of the study.

Figure 1



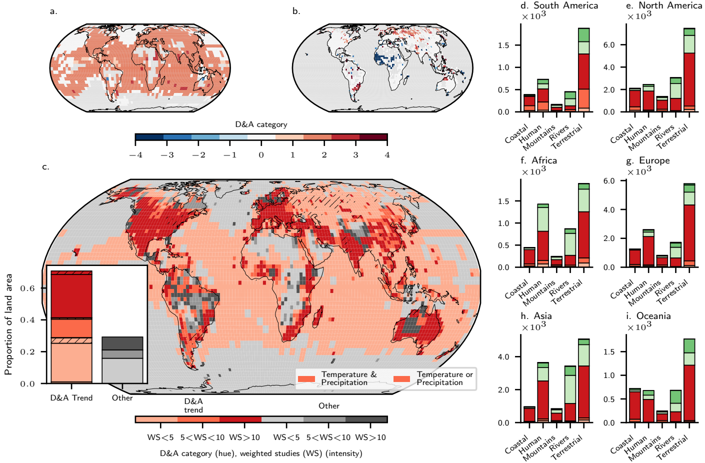
- The total number of predicted included results is larger (with more variation)
  - Previous model was optimized for marginal decisions and was overly conservative
- Qualitatively results remain similar
- Evidence type is removed

Figure 1



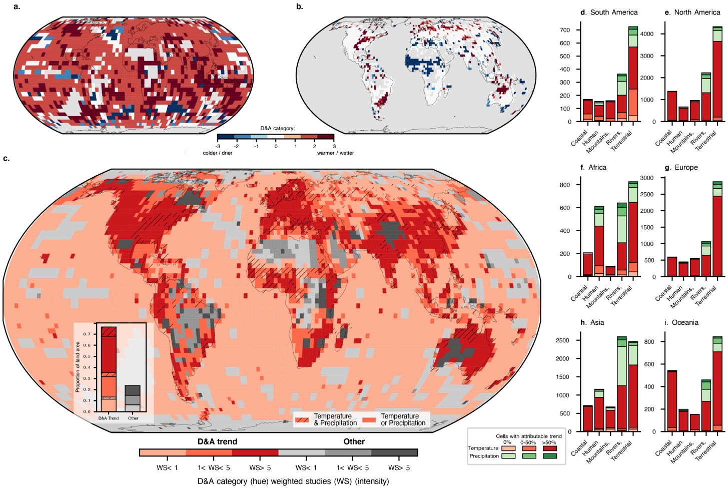
- The total number of predicted included results is larger (with more variation)
  - Previous model was optimized for marginal decisions and was overly conservative
- Qualitatively results remain similar
- Evidence type is removed

Figure 2



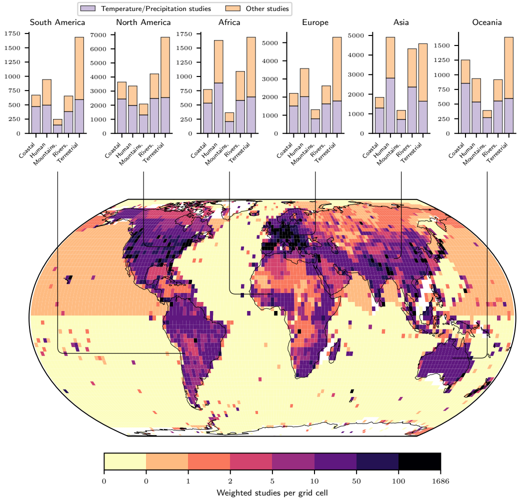
- Qualitative results are similar
- We reset thresholds (now we are including more evidence) to maintain distinction between qualitatively **lots** and **little** evidence
- Fewer attributable studies for precipitation

Figure 2



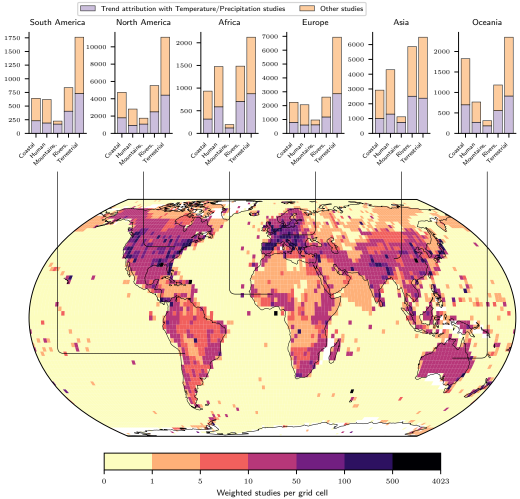
- Qualitative results are similar
- We reset thresholds (now we are including more evidence) to maintain distinction between qualitatively **lots** and **little** evidence
- Fewer attributable studies for precipitation

Figure 3



- Qualitative results are similar

Figure 3



- Qualitative results are similar

Next steps

- **Main text**: `https://docs.google.com/document/d/17NY4TB0yHsR5TsoTGiRUT6OkigT0LGEjTnN0Bb9UsiI/edit` major updates complete, and feedback from Tom mostly incorporated.
- **Methods and extended figures**: `https://docs.google.com/document/d/1mE3BD8WTQSNSPVtLYHHMVNqryctc8vypDvD5d8WBi30/edit` major updates complete.
- **Response to reviewers**: `https://docs.google.com/document/d/16DSrxjI3MbZVXX-E35RQYdwVzDjsWxPzF0W7zRWv0rM/edit?usp=sharing` needs work in parallel with final close edits to main text.