Assessment of Predictive Uncertainty of Data-Driven Environmental Models: An Argument-Based Approach

Benedikt Knüsel, Christoph Baumberger, Marius Zumwald, David N. Bresch, Reto Knutti

@beknuesel

benedikt.knuesel@usys.ethz.ch

EGU 2020

ITS4.3/AS5.2 I "Machine Learning for Earth System Modeling"

75 NRP Big Data National Research Programme

ETH zürich

Take-home messages

- New frameworks needed for a comprehensive uncertainty assessment for models built with machine learning
- Argument analysis offers a powerful approach for this task
- The focus lies on the justification of the assumption that the model is fit for the kind of prediction at hand
- Data-driven models are often applied to ill-understood problems (see Knüsel et al. 2019) and are often hard to interpret, thus, there can be substantial uncertainty in the uncertainty assessment ("second-order uncertainty")

Existing frameworks are not informative for data-driven environmental models

- Existing frameworks (e.g., Walker et al. 2003, Refsgaard et al. 2007) focus on several *locations of uncertainty*, including *model structure* or *model parameters*
- These locations are not informative for data-driven models:
  - For some machine learning models, it's unclear what the model structure would be
  - Some data-driven models are non-parametric
  - If defined, neither the model structure nor parameters can readily be interpreted in terms of the target system
  - Sometimes (but not always) we only care about good predictions of machine learning, not about the reasons for predictive success

Argument analysis

- Analyze the strength with which a proposition is justified
- Distinguish between propositions that are to be justified (*conclusion*) and those that do the justifying (*premises*)
- When evaluating the strength of the justification, consider
  - whether the premises are true (or close enough to the truth)
  - whether the premises provide good reasons to accept the truth of the conclusion

- Here, focus on assumption that a model is fit for the kind of prediction of interest and how this assumption can be justified

Framework

1. Reconstruct arguments that justify fitness-for-purpose assumption
2. Evaluate the strength of the justification
    1. Are premises true?
    2. Do they provide sufficiently good reasons to accept the conclusion?
3. Assess epistemic uncertainty based on argument analysis
    1. Epistemic first-order uncertainty arises to the extent that it cannot be conclusively justified that the model is maximally fit-for-purpose*, i.e.
        - if degree of fitness-for-purpose is lower than maximal
        - if some arguments for fitness-for-purpose are non-conclusive
    2. Epistemic second-order uncertainty arises due to factors that impair the assessment of first-order uncertainty
       (uncertainty of the uncertainty assessment)

* According to the framework presented here, a model is considered maximally fit-for-purpose if it reliably predicts the variable of interest up to some small error that depends only on the internal variability of the target system ("aleatory uncertainty")

Example: Long-term soil selenium projections by Jones et al. (2017)

- Three data-driven models trained with historical data for making projections of climate change impact on soil selenium
- For the models to be fit-for-purpose, they need to capture the broad-scale mechanisms
- Example (a full reconstruction is provided in the appendix):

P	*The models behave in consistency with background knowledge about soil selenium concentrations.*
C	*The models represent the most important mechanisms that drive soil selenium concentrations.*

Uncertainty:

- Justification is non-conclusive (first-order uncertainty)
- Processes are ill-understood → unclear how strong justification is (second-order uncertainty)

References

- Jones, Gerrad D., Boris Droz, Peter Greve, Pia Gottschalk, Deyan Poffet, Steve P. McGrath, Sonia I. Seneviratne, Pete Smith, and Lenny HE Winkel. "Selenium deficiency risk predicted to increase under future climate change." *Proceedings of the National Academy of Sciences* 114, no. 11 (2017): 2848-2853.
- Knüsel, Benedikt, Marius Zumwald, Christoph Baumberger, Gertrude Hirsch Hadorn, Erich M. Fischer, David N. Bresch, and Reto Knutti. "Applying big data beyond small problems in climate research." *Nature Climate Change* 9, no. 3 (2019): 196-202.
- Refsgaard, Jens Christian, Jeroen P. van der Sluijs, Anker Lajer Højberg, and Peter A. Vanrolleghem. "Uncertainty in the environmental modelling process–a framework and guidance." *Environmental modelling & software* 22, no. 11 (2007): 1543-1556.
- Walker, Warren E., Poul Harremoës, Jan Rotmans, Jeroen P. Van Der Sluijs, Marjolein BA Van Asselt, Peter Janssen, and Martin P. Krayer von Krauss. "Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support." *Integrated assessment* 4, no. 1 (2003): 5-17.

Appendix

Overview of appendix

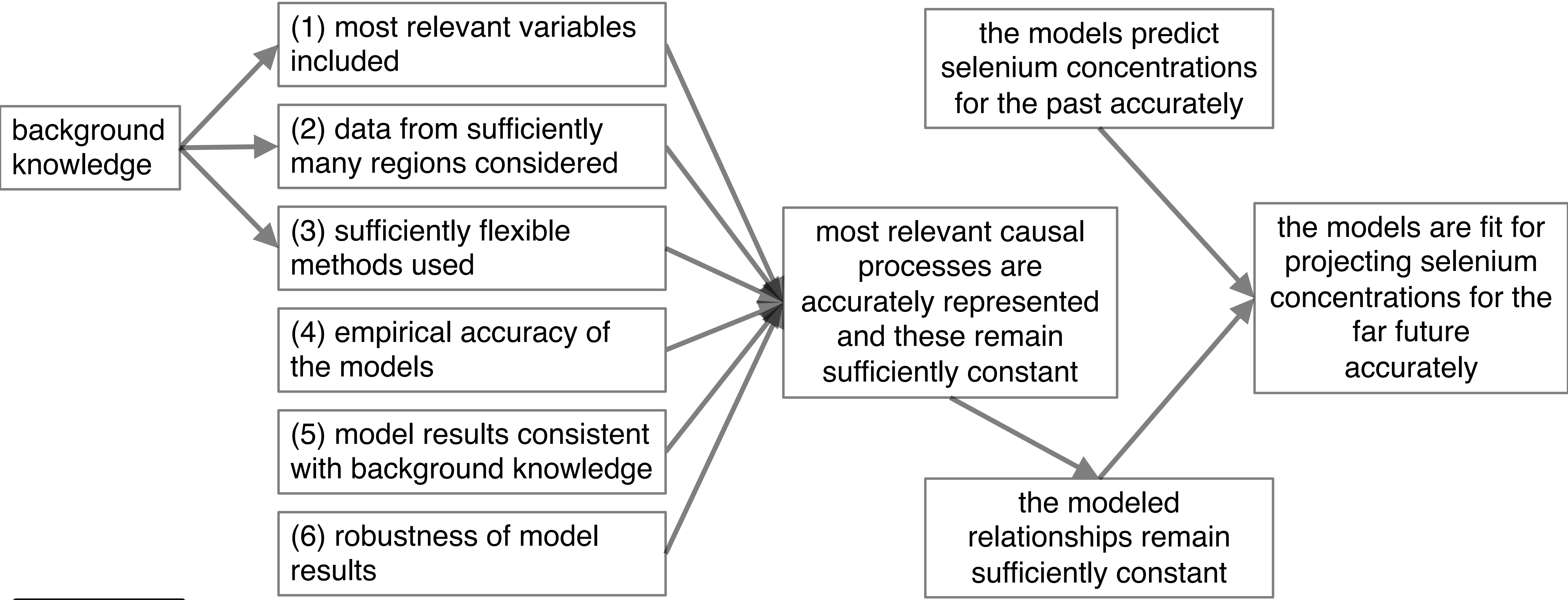
The following slides contain the application of the framework to the long-term soil selenium projection by Jones et al. (2017)

Slide 10 gives an overview of the justification

Slides 11 – 19 reconstruct the arguments explicitly, slide 20 presents them in an argument map

Slide 21 analyzes the uncertainty based on the previous slides

Reconstruction of Justification for Jones et al. (2017) PNAS



Reconstruction of Arguments (1)

Argument 1

P1.1	If a model has predicted many past instances of a phenomenon accurately and the modeled relationships remain sufficiently constant over time, that model is fit for predicting the phenomenon in the far future.
P1.2	M has predicted many past instances of S accurately.
P1.3	The modeled relationships in M remain sufficiently constant.
C1	M is fit for predicting S in the far future.

Reconstruction of Arguments (2)

	Argument 2
P2.1	If a model represents the most important causal processes producing a phenomenon accurately and these processes are unaffected by changing environmental conditions, the modeled relationships remain sufficiently constant.
P2.2	The causal processes represented in M are unaffected by changing environmental conditions.
P2.3	M accurately represents the important causal processes that produce S.
C2	The modeled relationships in M remain sufficiently constant. (= P1.3)

Reconstruction of Arguments (3)

	Argument 3
P3.1	M was constructed using data that represents sufficiently many configurations of S.
P3.2	M was constructed using the most important variables.
P3.3	M was constructed using sufficiently flexible methods while overfitting was avoided.
C3	M represents most important mechanisms that produce S. (=P2.3)

Reconstruction of Arguments (4)

	Argument 4
P4.1	M is empirically accurate with respect to the data from the past.
C4	M represents most important mechanisms that produce S. (= P2.3)

Reconstruction of Arguments (5)

	Argument 5
P5	M behaves in consistency with background knowledge about S.
C5	M represents most important mechanisms that produce S. (= P2.3)

Reconstruction of Arguments (6)

	Argument 6
P6	The predictions are only considered if the ensemble members of M agree on the sign of change of S.
C6	M represents most important mechanisms that produce S. (= P2.3)

Reconstruction of Arguments (7)

	Argument 7
P7	M was constructed using over 30.000 samples from different continents.
C7	M was constructed using data that represents sufficiently many configurations of S. (=P3.1)

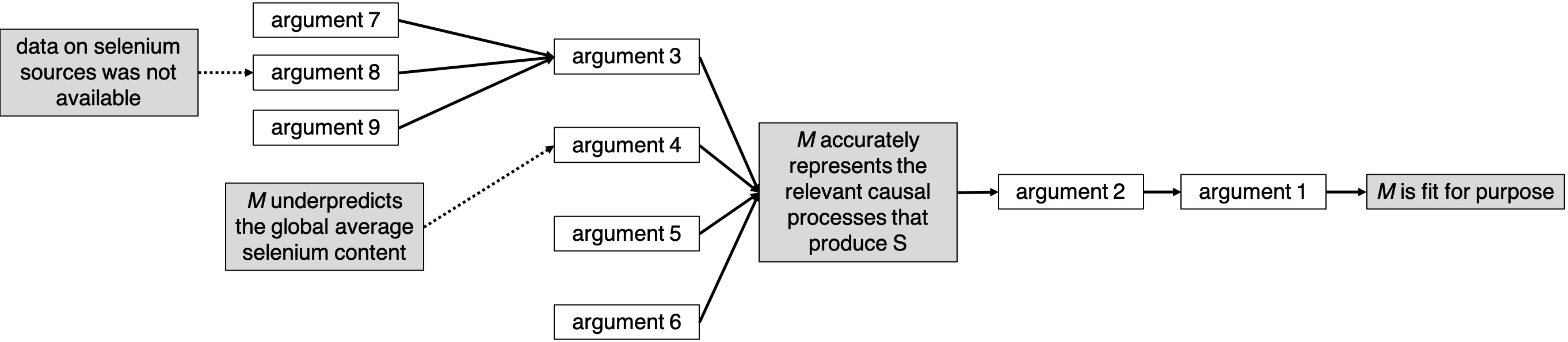
Reconstruction of Arguments (8)

	Argument 8
P8.1	M was constructed using seven variables chosen based on a variable selection procedure.
P8.2	Most potentially relevant variables were included in the variable selection procedure.
C8	M was constructed using the most important variables. (= P3.2)

Reconstruction of Arguments (6)

	Argument 9
P9.1	M was constructed using artificial neural networks and random forest.
P9.2	Measures were taken to avoid overfitting.
C9	M was constructed using sufficiently flexible methods while overfitting was avoided. (= P3.3)

Argument Map



Uncertainty

- There is substantial first-order uncertainty because
  - some of the arguments (specifically arguments 3 – 9) are non-conclusive
  - the actual degree of fitness-for-purpose is less-than maximal due to two theses that attack the justification
- There is substantial second-order uncertainty because
  - the strength of some of the arguments is difficult to assess due to limited background knowledge
  - the models are not transparent
  - the actual degree of fitness-for-purpose is unclear