COMMENT

To the brave scientists: Aren't we strong enough to stand (and profit from) uncertainty in Earth system measurement and modelling?

Hendrik Paasche1,2 | Matthias Gross3,4 | Jakob Lüttgau5 | David S. Greenberg6 | Tobias Weigel5

1Department Monitoring and Exploration Technologies, UFZ – Helmholtz Centre for Environmental Research, Leipzig, Germany
2Geological Survey of Norway (NGU), Trondheim, Norway
3Department of Urban and Environmental Sociology, UFZ – Helmholtz Centre for Environmental Research, Leipzig, Germany
4Institute of Sociology, University of Jena, Jena, Germany
5Helmholtz AI, DKRZ – Deutsches Klimarechenzentrum GmbH, Hamburg, Germany
6Model-Driven Machine Learning, Helmholtz AI, Helmholtz Centre Hereon, Geesthacht, Germany

Correspondence
Hendrik Paasche, Department Monitoring and Exploration Technologies, UFZ – Helmholtz Centre for Environmental Research, Permoserstr. 15, Leipzig 04318, Germany.
Email: hendrik.paasche@ufz.de

Funding information
Helmholtz Association’s Initiative and Networking Fund through Helmholtz AI, Grant/Award Number: ZT-I-PF-5-01; Uncertainty Quantification – From Data to Reliable Knowledge, Grant/Award Number: ZT-I-0029

Abstract
The current handling of data in earth observation, modelling and prediction measures gives cause for critical consideration, since we all too often carelessly ignore data uncertainty. We think that Earth scientists are generally aware of the importance of linking data to quantitative uncertainty measures. But we also think that uncertainty quantification of Earth observation data too often fails at very early stages. We claim that data acquisition without uncertainty quantification is not sustainable and machine learning and computational modelling cannot unfold their potential when analysing complex natural systems like the Earth. Current approaches such as stochastic perturbation of parameters or initial conditions cannot quantify uncertainty or bias arising from the choice of model, limiting scientific progress. We need incentives stimulating the honest treatment of uncertainty starting during data acquisition, continuing through analysis methodology and prediction results. Computational modellers and machine learning experts have a critical role, since they enjoy high esteem from stakeholders and their methodologies and their results critically depend on data uncertainty. If both want to advance their uncertainty assessment of models and predictions of complex systems like the Earth, they have a common problem to solve. Together, computational modellers and machine learners could develop new strategies for bias identification and uncertainty quantification offering a more all-embracing uncertainty quantification than any known methodology. But since it starts for computational modellers and machine learners with data and their uncertainty, the fundamental first step in such a development would be leveraging shareholder esteem to insistently advocate for reduction of ignorance when it comes to uncertainty quantification of data.

KEYWORDS
computational modelling, data uncertainty, machine learning, model uncertainty, uncertainty quantification
Earth scientists, like probably most scientists, are convenience-oriented ignoramuses. At least that is what one might think when glancing at the work of the last decades in Earth observation, modelling and prediction of environmental processes. Centuries-old awareness of the unavoidable existence of measurement errors and error propagation is not reflected by attempts to quantify the indeterminateness of Earth observation data and subsequent modelling and prediction efforts. The focus in Earth observation has apparently moved to the acquisition of large quantities of data in short time, following the idea of big data (Yang et al., 2019), rather than to invest into a thorough quantification of the trustworthiness of the measured data. Results of modelling and prediction of environmental processes are often presented deterministically (Uusitalo et al., 2015). Nevertheless, a number of methodologies have been suggested to assess the uncertainty of models and predictions (Beven, 2016; Refsgaard et al., 2007; Uusitalo et al., 2015) often limited to distinct types of uncertainty, coming with their own methodological uncertainty, and usually not starting at the beginning of the modelling endeavours, the data acquisition. Concealment of uncertainty when observing, modelling and predicting complex systems like the Earth provides a false certainty about the predictability of the consequences of adaptation decisions regarding social welfare for which stakeholders accept the responsibility. In a self-reinforcing loop, uncertainty concealment may be valued by stakeholders setting incentives allowing Earth scientists to (partly) ignore uncertainties related to their efforts in Earth observation and process modelling. However, since any modelling or prediction endeavour builds on data, working with uncertainty-concealed data fundamentally limits the trustworthiness of models and predictions of environmental processes, for example, climate, water balance, crop yield, pandemic spread, etc. (Falloon et al., 2014). Data uncertainty limits the accuracy machine learning models can achieve without overfitting the data, and it limits the certainty with which model parameters can be tuned in computational modelling approaches. Therefore, we should not rely on stimulation and incentives from outside the scientific community to advance uncertainty quantification in Earth observation and modelling. Whenever knowledge grows, the perception of ignorance does, too (Gross, 2010). When looking at the handling of data uncertainty in Earth observation, modelling and prediction efforts, we need to critically ask ourselves about the routines followed in modern Earth sciences. We claim that selective ignorance of uncertainty or aspects thereof in Earth observation, modelling and prediction prohibits the so-called ‘hard’ numerical sciences, that is, computational modelling and machine learning, from significantly advancing when analysing complex natural systems. We think that selective ignorance of data uncertainty represents threats to the general applicability and usefulness of any modelling and prediction approaches in Earth sciences. Since this problem has not been addressed sufficiently in the past, we call for the community to address this as it can also open up new data sources and increase the trust in data sources and reap their hidden insights with confidence. With this contribution, we aim to stimulate discussions among data set creators and data users and encourage new innovative approaches for data uncertainty quantification. While it is beyond our scope to summarize the state-of-the-art in uncertainty quantification or review previous empirical research, we highlight several examples to illustrate both the promises and limitations of current approaches.

2 | DATA UNCERTAINTY

While data uncertainty can hardly be discussed without modelling uncertainty due to its propagation into models, the two are not interchangeable. Earth scientists are aware of the importance of linking data to quantitative uncertainty measures and partly provide these with the data (Pérez-Díaz et al., 2020; Yang et al., 2013). Others consider uncertainty associated with data as noise and characterize it as unwanted information that complicates the certain extraction of important or useful information from data (Schmidt et al., 2020). Schmidt et al. (2020) state that noise associated with data never diminishes. This makes us believe that quantified knowledge about noise or information about the uncertainty of extracting important or useful information from data is important information on its own and essentially required (and therefore hopefully wanted) in the context of any data processing. However, it is worth considering whether attempts to realize generic and realistic uncertainty quantification of Earth observation data too often fail at very early stages. For example, commercial manufacturers of Earth observation systems might not reveal internal functionality, calibration details, etc., of their proprietary sensors and might lack a systematic sensor decomposition concept (Bumberger et al., 2015) for controlled assessment of all aspects of the measurement uncertainty of a sensor (Demuth et al., 2015).

Generally, all measurements suffer aleatory uncertainty (Kiureghian & Ditlevsen, 2009), that is, limited precision and accuracy linked to random and systematic measurement errors, respectively. Both error types can be quantified by choosing an appropriate experimental setup, but particularly the latter is rarely assessed, whereas the first is often only poorly assessed. Instead of investing in the quantification of aleatory uncertainty, geoscientific
research tends to promote experimental setups favouring, for example, the rapid acquisition of many readings at different locations or times. Alternatively, the experimental focus can be biased towards increasing the information return of experiments by acquiring data sets with low informational redundancy among the included readings (Stummer et al., 2004). This makes it practically more difficult to control the quality of the readings and identify readings with significantly different aleatory uncertainty than others (often referred to as ‘outliers’) on the basis that readings sharing a significant overlap in the information about the system under measurement should be similar. While the information content of such an optimized data set may indeed be higher and allow for better models in subsequent processing steps, a few readings associated with unrecognized high aleatory uncertainty may totally invalidate efforts to generate realistic models out of such a data set. Established practices focussing one-sidedly on acquisition of large data sets or low information redundancy within a data set are a base for concern that needs to be addressed.

Digital Earth observation builds on band-limited experiments causing epistemic uncertainty (Kiureghian & Ditlevsen, 2009) in data sets. Digitized experiments are discrete by nature, meaning that the sample density in time and space is limited resulting in uncertainty about aspects not recorded in the data due to limits in spatial and temporal resolution and coverage. In addition, experiments are physically band-limited, that is, by only sampling the ground within a bounded physical spectrum. This results in uncertainty about aspects of reality probably only recognizable beyond the limits of the probed spectrum, but not recorded by the data. They cannot be quantified without additional measurements. For example, the information of multiple data sets co-located in time and space can be fused to mutually fill information gaps. For each data set added, the overall epistemic uncertainty is reduced at the cost of adding aleatory uncertainty.

Experiments are often set up based on expectations hypothesized from existing theory and models, for example, what are relevant data sets to be considered, suitable spectrum selections, sample density, etc. Any flaws of these theories and models will bias the experiment and add ontological uncertainty (Lane & Maxfield, 2005) to the resultant data. Such methodological uncertainties usually remain unrecognized by experts routinely following state-of-the-art data acquisition and experimental setups. However, if such methodological limitations could be brought to light, they would be transformed from unrecognized measurement uncertainty into aleatory uncertainty of recorded data grounded in random and systematic errors of the experimental setup, which are generally quantifiable. Table 1 summarizes the characterization of different types of uncertainty. Raising awareness regarding these limitations is therefore a crucial first step in addressing and quantifying uncertainty.

### 3 | SCIENTIFIC METHODOLOGY FOR MODELLING AND PREDICTION OF PROCESSES IN COMPLEX NATURAL SYSTEMS

Natural data sets carrying information and uncertainty about processes in the Earth and related state variables are the basis for modelling complex systems and predicting future system states. Two mutually independent scientific methodologies exist for this purpose, currently named computational modelling and machine learning. Computational modelling builds on models once

| Uncertainty | Aleatory | Epistemic | Ontological |
|-------------|----------|-----------|-------------|
| Cause       | Limited observational accuracy and precision | Band-limited experiments | Inappropriate methodological choices |
| Quantifiable| Yes (Gellert et al., 1975) | No | No |
| Reduction by| Increased observational accuracy and precision | Increased experimental information return | better methodology |
| Convertible into aleatory uncertainty | By fusing information of multiple data sets co-located in time and space (Liggins et al., 2008). For each data set added, the overall epistemic uncertainty is reduced at the cost of adding aleatory uncertainty of the other data sets | By becoming aware of and varying methodological choices (Spradley, 1980). Usually, the identification of choices in the used methodology cannot be realized by people who are all bound to the same behaviour or state-of-the-art, for example, as learned by experiencing a common training procedure |
derived from natural data but now considered generic and transferable in space and time, although the need for calibration of model parameters to natural data acquired at the area of interest cannot always be ruled out. Traditionally, this approach builds on known physical processes, described through mathematical–analytical formulations of a relation between state variables identified as particularly suitable for describing processes in dynamic systems, for example, ratio equations as, for example, used in the law of the lever, or partial differential equations, such as Richards equation describing the movement of water in soils. A common feature of these physics-based models is that the nature of interactions between state variables is highly restricted by physical constraints, so that each variable is only affected by a specific set of others, and only according to existing physical laws. Critically, the laws and constraints used in these models are often derived from previous experiments and observations of a very different nature as a posterior hypothesis and then used as an a priori valid hypothesis of general applicability.

In contrast to computational modelling, machine learning (or other data-centric methods, e.g., statistical modelling) can learn a model from available data without considering a priori accepted domain knowledge about relevant mathematical–analytical process models or data sets. Instead, patterns are identified within and across data sets without asking for the ‘physics behind those patterns’, and without regard to patterns or rules derived from other data sets in the past. This makes machine learning particularly suitable for dealing with cross-disciplinary databases, leaving the restrictions of domain-specific mathematical–analytical model families behind. However, models learned by machine learning are only valid for the domain covered by the available data, but without the claim to honour the ‘generic physics behind recognized patterns’, its general transferability is highly uncertain.

Both approaches are methodologically independent but can build on the same natural data and strive to model the same phenomenon. Both approaches suffer from concealed data uncertainty and struggle to transfer it into model uncertainty. When the two approaches agree within the data-derived model uncertainty limits, this should increase our confidence in the results, but when they give different answers, then at least one approach must be biased in its methodology. The independence and possibility for mutual confirmation or bias identification are lost if both approaches are coupled by conscious implementation of architecture constraints of the employed machine learning algorithm (Beucler et al., 2019). A prominent example is physics-informed neural networks (Raissi et al., 2019) where the mathematical–analytical model becomes part of the objective function of a machine learning approach. Independence of the available scientific methods is also lost if both are subsequently applied to reach a modelling or prediction task, for example, if missing natural data shall be substituted with synthetic data generated by machine learning models trained with synthetic data fabricated by means of mathematical–analytical process simulation models (Kadow et al., 2020).

### 4 | LINKING DATA AND PREDICTION

Uncertainty associated with natural data goes into any processing methodology using the data. This uncertainty hinders the disambiguation of model candidates, as well as predictions derived from these models. When fitting a model to data, the uncertainty in model parameters can be reduced through previous knowledge, inductive biases and abundant data, but the same fitting procedure can introduce epistemic uncertainty when parameters are poorly constrained, and inevitably produces ontological uncertainty regarding the model itself. Thus, both physics- and machine learning-based approaches can filter or transform data uncertainty into other forms, but never reduce it. In practice, the uncertainty that arises will depend on the precise nature of modelling assumptions and data properties, as well as the specific problem to be solved.

For computational modelling, the uncertainty going along with the trust in the applicability of theory and models across time and space cannot be quantified. Note, this is ontological uncertainty linked to the modelling and its inherent choices, rather than the data acquisition. In this point, the belief could be regarded as a doctrine. This means that computational modelling by itself cannot quantify the ontological uncertainties inherent to its own assumptions and methodologies. This problem will remain as long as we do not have ‘a second physics’, which is independent of the established physics and its concepts and mathematical–analytical models formulated under its roof, but equally sound. For example, stochastic approaches for uncertainty quantification of models only vary the parameters of the model, but not the model itself. The comparison of different models, all building on computational modelling and rooted at some point in related theory can carry the same methodological bias which cannot be revealed by comparison of computational modelling results, albeit this is routine practice when assessing the trustworthiness of climate models (Pirtle et al., 2010; Williamson et al., 2013). Adding data or data uncertainty does not help here, since only model parameters are calibrated to data and data uncertainty, but the general model remains unchanged.
Machine learning can learn models of almost arbitrary complexity (Hornik et al., 1989) and can approach a perfect fit to data given enough parameters. Similarly, computational models with enough parameters and complexity can accommodate arbitrary data. However, knowledge of data uncertainty allows us to limit the complexity of these models and avoid overfitting on finite data sets. In the case of machine learning, complexity can be limited using architectural design choices that constrain the learned mapping of inputs to outputs, such as the continuity of neural networks or the partitioning of decision trees.

By concurrently employing machine learning methods with independent learning concepts, for example, continuity or partitioning as for example used in artificial neural networks or random forest algorithms, respectively, one could check whether the choice of the learning method and its settings has an impact on the applicability of the learned model. This would be controllable if the same data and the associated data uncertainty would be considered in the learning procedure – then results should coincide if the model, regardless of partitioning or continuity selection, was suitably chosen and not suffering inductive biases, such as overfitting the data, which essentially requires data uncertainty quantification. Opposite to computational modelling, where we lack a second physics for revealing ontological uncertainty in the model setup, machine learning offers here better capabilities, since two independent working concepts of algorithms can be formulated which should meet each other if correctly set up with regard to the information and uncertainty fractions of the considered data. However, machine learning models build on patterns and may fail to honour any physical limits when it comes to predictions away from the data used for learning the model.

5 CONCLUSIONS

Computational modelling and machine learning are in their fundamental methodological approaches independent from each other, but build both on (the same) natural data and can be used to model and predict the same phenomena of complex systems like the Earth. Computational modelling and machine learning approaches can complement each other, mutually reveal potential biases and thus strengthen the trustworthiness in their results. Although not established yet, we consider this complementary use as a promising way to enable uncertainty and bias assessment beyond established concepts, such as fuzziness or probability, which cannot incorporate ontological model uncertainty in their uncertainty quantification.

For being able to judge the coincidence of results of machine learning and computational modelling, it is essential to propagate data uncertainties via models into prediction results and incorporate model uncertainties arising in this data processing pipeline. Any compromises, for example, due to computational or economic reasons, avoiding the consideration of realistic data uncertainty beyond some aspects of aleatory uncertainty, for example, random errors, go along with degrading capabilities to assess the trustworthiness of models and derived predictions right from the beginning of all efforts. Ignorance of data uncertainty components will lead to artificially reduced uncertainty on prediction results with deterministic prediction as its most extreme case. Unreasonably small prediction uncertainty is deceptive and lowers the true prediction certainty of decisions striving to ensure social welfare and adaptation to a changing environment. Overall, we currently follow unsuitable incentives to consider data uncertainties in our experimental design since we focus on filling data gaps rather than quantifying the quality of the acquired data.

Without uncertainty quantification, data acquisition approaches are not sustainable, machine learning and computational modelling cannot unfold their potential and we fail to come in a position to quantify the uncertainty of our predictions and reveal biases in existing process models leading to significant scientific progress beyond established uncertainty assessment techniques not suitable to capture all aspects of uncertainty, like stochastic modelling varying model parameters only. We need incentives stimulating the honest treatment of uncertainty starting at the data acquisition, continuing through analysis methodology and prediction results. Computational modellers and machine learning experts have a critical role, since they enjoy high esteem from stakeholders and their methodologies and their results critically depend on data uncertainty. If both want to advance their uncertainty assessment of models and predictions of complex systems like the Earth, they have a common problem to solve. Together, computational modellers and machine learners could develop new strategies for bias identification and uncertainty quantification offering a more all-embracing uncertainty quantification than any known methodology. But since it starts for computational modellers and machine learners with natural data and their uncertainty, it is the fundamental first step in such a development to use the enjoyed esteem of stakeholders to insistently advocate for reduction of ignorance when it comes to uncertainty quantification of natural data.

In summary, we put forward first steps of action. Both, computational modelling and machine learning communities, need to come together as they follow two independent scientific methodologies essentially needed to mutually reveal biases in their workflows and to advance modelling and prediction of complex systems like the
Earth to new methodologies beyond centuries-old stochastic or fuzzy modelling concepts. The Earth sciences community should increase efforts to uncover and quantify uncertainty in data products, particularly in Earth observation, and make all-embracing uncertainty quantification an integral part of the data acquisition and provisioning processes. The Earth sciences community should also propagate data uncertainties via models into prediction results and add model uncertainties occurring during this processing chain. Computational modellers and machine learning experts should use the high esteem given to them by stakeholders to urge for incentives for advancement of data uncertainty quantification since all their modelling efforts critically depend on quantified data uncertainty.

ACKNOWLEDGEMENTS
David Greenberg, Jakob Lüttgau and Tobias Weigel were funded by Helmholtz Association’s Initiative and Networking Fund through Helmholtz AI. This work was supported by Helmholtz Association’s Initiative and Networking Fund through Helmholtz AI [grant number: ZT-I-PF-5-01] and Uncertainty Quantification – From Data to Reliable Knowledge [grant number: ZT-I-0029].

ORCID
Hendrik Paasche https://orcid.org/0000-0003-1814-0645

REFERENCES
Beucler, T., Rasp, S., Pritchard, M. & Gentine, P. (2019) Achieving conservation of energy in neural network emulators for climate modeling. Expanded abstract, ICML workshop 2019: climate change: how can AI help? Long Beach. arXiv:1906.06622v1.

Beven, K. (2016) Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. Hydrological Sciences Journal, 61, 1652–1665.

Bumberger, J., Paasche, H. & Dietrich, P. (2015) Systematic description of direct push sensor systems: a conceptual framework for system decomposition as the basis for the optimal sensor design. Journal of Applied Geophysics, 122, 210–217.

Demuth, D., Bumberger, J. & Paasche, H. (2015) Evaluation of direct push probes: sensor interface analysis of DC resistivity probes. Journal of Applied Geophysics, 122, 218–225.

Falloon, P., Challinor, A., Dessai, S., Hoang, L., Johnson, J. & Koehler, A.K. (2014) Ensembles and uncertainty in climate change impacts. Frontiers in Environmental Science, 2, 33.

Gellert, W., Gottwald, S., Hellwich, M., Kästner, H. & Küstner, H. (1975) Calculus of errors, adjustment of data, approximation theory. In: Gellert, W., Gottwald, S., Hellwich, M., Kästner, H. & Küstner, H. (Eds.) The VNR concise encyclopedia of mathematics. Dordrecht: Springer, pp. 607–630.

Gross, M. (2010) Ignorance and surprise: science, society, and ecological design. Cambridge, MA: MIT Press.

Hornik, K., Stinchcombe, M. & White, H. (1989) Multilayer feedforward networks are universal approximators. Neural Networks, 2, 359–366.

Kadow, C., Hall, D.M. & Ulbrich, U. (2020) Artificial intelligence re-constructs missing climate information. Nature Geoscience, 13, 408–413.

Kuurpleihan, A.D. & Ditlevson, O. (2009) Aleatory or epistemic? Does it matter? Structural Safety, 31, 105–112.

Lane, D.A. & Maxfield, R.R. (2005) Ontological uncertainty and innovation. Journal of Evolutionary Economics, 15, 3–50.

Liggins, M.E., Hall, D.M. & Llinas, J. (2008) Handbook of multisensor data fusion: theory and practice. Boca Raton, FL: CRC Press.

Paasche, H., Paasche, K. & Dietrich, P. (2020) Uncertainty as a driving force for geoscientific development. Nature and Culture, 15, 1–18.

Pérez-Díaz, L., Alcade, J. & Bond, C.E. (2020) Handling uncertainty in the geosciences: identification, mitigation and communication. Special Issue, Solid Earth, 11, 889–897. Available from: https://doi.org/10.5194/se-11-889-2020

Pirtle, Z., Meyer, R. & Hamilton, A. (2010) What does it mean when climate models agree? A case for assessing independence among general circulation models. Environmental Science and Policy, 13, 351–361.

Raisi, M., Perdikaris, P. & Karniadakis, G.E. (2019) Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378, 686–707.

Refsgaard, J.C., van der Sluijs, J.P., Hojberg, A.L. & Vanrolleghem, P.A. (2007) Uncertainty in the environmental modelling process – a framework and guidance. Environmental Modelling & Software, 22, 1543–1556.

Schmidt, A., Dabas, M. & Apostolos, S. (2020) Dreaming of perfect data: characterizing noise in archeo-geophysical measurements. Geosciences, 10, 382.

Spradley, J.P. (1980) Participant observation. Orlando, FL: Harcourt.

Stummer, P., Maurer, H. & Green, A. (2004) Experimental design: electrical resistivity data sets that provide optimum subsurface information. Geophysics, 69, 120–139.

Uusitalo, L., Lehikoinen, A., Helle, I. & Myrberg, K. (2015) An overview of methods to evaluate uncertainty of deterministic models in decision support. Environmental Modelling & Software, 63, 24–31.

Williamson, D., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L. et al. (2013) History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. Climate Dynamics, 41, 1703–1729.

Yang, C., Yu, M., Li, Y., Hu, F., Jiang, Y., Liu, Q. et al. (2019) Big Earth data analytics: a survey. Big Earth Data, 3, 83–107.

Yang, X., Blower, J.D., Bastin, L., Lush, V., Zabala, A., Maso, J. et al. (2013) An integrated view of data quality in Earth observation. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 371, 20120072.

How to cite this article: Paasche, H., Gross, M., Lüttingau, J., Greenberg, D.S. & Weigel, T. (2022) To the brave scientists: Aren’t we strong enough to stand (and profit from) uncertainty in Earth system measurement and modelling? Geoscience Data Journal, 9, 393–399. https://doi.org/10.1002/gdj3.132
APPENDIX

Glossary

Data: units of coded knowledge or information, either qualitative or quantitative.

Earth scientist: a person studying the Earth by means of scientific methodology rooted in the paradigms of Earth science domains, such as geophysics and geology, which all build on natural data and strive towards the retrieval of generic information about processes controlling the shaping of the Earth.

Information: refers to knowledge communication (e.g. education, instruction, etc.) and is intrinsically linked to data. Information must be coded for transmission and storage. Information can be retrieved by decoding data, which is usually context based.

Machine learning: scientific domain concerned with finding useful patterns in complex databases. Follows scientific methodology and builds on data but does not necessarily put data in a context of retrieving generic information about processes related to the database. If the database stores information about the Earth, machine learning can reveal and take advantage of data patterns which are intricately related to processes controlling the shaping of the Earth without asking for the processes itself or information about the processes controlling the shaping of the Earth.

Model: simplified image of reality that strives to describe aspects of reality that were deemed relevant by their inventors.

Natural data: data that were obtained through the observation of measurable factual aspects in our perception of natural systems.

Synthetic data: data generated on the basis of models, for example, a computer simulation, not obtained by measurements directly sampling natural systems.

Uncertainty: the level of imperfect or unknown information in epistemic situations, for example, with regard to natural data (data uncertainty) or models (model uncertainty).