A View on Model Misspecification in Uncertainty Quantification

Kato, Yuko; Tax, David M.J.; Loog, Marco

DOI
10.1007/978-3-031-39144-6_5

Publication date
2023

Document Version
Final published version

Published in
Artificial Intelligence and Machine Learning - 34th Joint Benelux Conference, BNAIC/Benelearn 2022, Revised Selected Papers

Citation (APA)
Kato, Y., Tax, D. M. J., & Loog, M. (2023). A View on Model Misspecification in Uncertainty Quantification. In T. Calders, B. Goethals, C. Vens, & J. Lijffijt (Eds.), *Artificial Intelligence and Machine Learning - 34th Joint Benelux Conference, BNAIC/Benelearn 2022, Revised Selected Papers* (pp. 65-77). (Communications in Computer and Information Science; Vol. 1805 CCIS). Springer. https://doi.org/10.1007/978-3-031-39144-6_5

A View on Model Misspecification in Uncertainty Quantification

Yuko Kato[1(✉)], David M. J. Tax[1], and Marco Loog[1,2]

[1] Delft University of Technology, Delft, The Netherlands
{y.kato,d.m.j.tax}@tudelft.nl
[2] Radboud University, Nijmegen, The Netherlands
marco.loog@ru.nl

Abstract. Estimating uncertainty of machine learning models is essential to assess the quality of the predictions that these models provide. However, there are several factors that influence the quality of uncertainty estimates, one of which is the amount of model misspecification. Model misspecification always exists as models are mere simplifications or approximations to reality. The question arises whether the estimated uncertainty under model misspecification is reliable or not. In this paper, we argue that model misspecification should receive more attention, by providing thought experiments and contextualizing these with relevant literature.

Keywords: Uncertainty quantification · Model misspecification · Epistemic and Aleatoric uncertainty

1 Introduction

In fields such as biology, chemistry, engineering, and medicine [1–4], having an accurate estimate of prediction uncertainty is of great importance to guarantee safety and prevent unnecessary costs. In this regard, uncertainty quantification (UQ) is a vital step in order to safely apply Machine Learning (ML) models to real world situations involving risk. It is well-known, however, that these ML models are generally poor at quantifying these uncertainties [5,6].

Generally, depending on the exact source, the type of uncertainty can be categorized as being epistemic or aleatoric [7,8]. Epistemic uncertainty refers to the uncertainty of the model and arises due to lack of data used to train the model and lack of domain knowledge. This type of uncertainty is considered to be reducible by an appropriate selection of the model and increasing data size. On the contrary, aleatoric uncertainty is a consequence of the random nature of data and is therefore considered to be irreducible [9].

Although total uncertainty (the combination of epistemic and aleatoric uncertainty) can be estimated using different methods, some articles have focused on the separate estimation of aleatoric uncertainty and epistemic uncertainty [10–14]. Given the fact that only epistemic uncertainty is reducible, separation

T. Calders et al. (Eds.): BNAIC/Benelearn 2022, CCIS 1805, pp. 65–77, 2023.
https://doi.org/10.1007/978-3-031-39144-6_5

into the two types of uncertainty can help to guide model development [12]. For example, during active learning, Nguyen and colleagues [15] concluded that quantifying epistemic uncertainty separately has the potential to provide useful information to learners of the model and can potentially improve the performance of the learning process.

Notwithstanding, due to model misspecification, a question remains whether the estimated aleatoric and epistemic uncertainty are reliable or not [16,17]. Model misspecification exists when the best model differs from the truth and the associated hypothesis space does not include the truth. It is important to realize that model misspecification always exists without any exceptions [18]. As there is no uniquely prescribed way to deal with model misspecification, it has been treated differently among researchers. Some do not mention it at all in their papers [11,14], while others consider it to be part of epistemic uncertainty [19]. These variable interpretations of model misspecification make it difficult to compare the different estimated epistemic uncertainties.

In this paper, we argue that model misspecification should receive more attention. We start by defining model misspecification and propose three possible ways to see model misspecification in relation to epistemic uncertainty. The paper proceeds with a brief investigation of how model misspecification is recognized among researchers and how they relate to our proposed views. In addition, we assess the possible consequences of mistreating model misspecification qualitatively and conclude with a brief discussion.

2 Model Misspecification

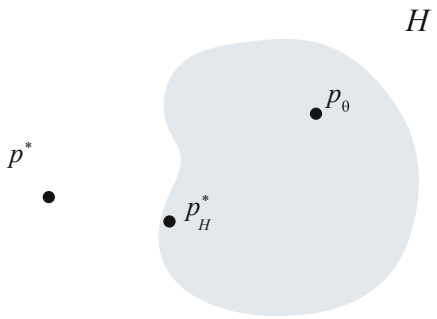

Fig. 1. Model misspecification: The truth p^* is not included in the hypothesis space H (shaded blue area) due to model misspecification. (Color figure online)

2.1 Definition of Model Misspecification

Whether the ML-method is applied in a regression or classification setting, the first step is to choose a model p_θ parameterized by θ. Let $H = \{p_\theta : \theta \in \Theta\}$

denote the hypothesis space defined by the chosen model class and the associated parameters θ in the set Θ. We denote the truth p^* and the best possible model in H by p^*_H. Now, model misspecification exists when the best model p^*_H differs from the truth p^* and $p^* \notin H$. This is illustrated in Fig. 1.

2.2 Example of Model Misspecification: Regression

Let us consider a simple regression problem using a dataset of n observations $\mathcal{D} = (x_i, y_i)$ with $i = 1, ..., n$ and additive noise. Let us assume, in addition, that the true additive noise follows a heavy-tailed distribution (see Fig. 2a). The typical aim is to approximate the unknown underlying true data distribution $p^*(y|x)$ at a new test point x. Now, a typical model choice $\hat{p}_n(y|x)$ is to assume the additive noise to be Gaussian. This means that the assumption on the noise distribution is misspecified. Since $p^*(y|x)$ is not in the hypothesis space H due to model misspecification (i.e., the assumed noise does not match the true one), we will never be able to reach the true distribution $p^*(y|x)$ even if we use an infinite amount of data to train the model $\hat{p}_\infty(y|x)$ (See Fig. 2b).

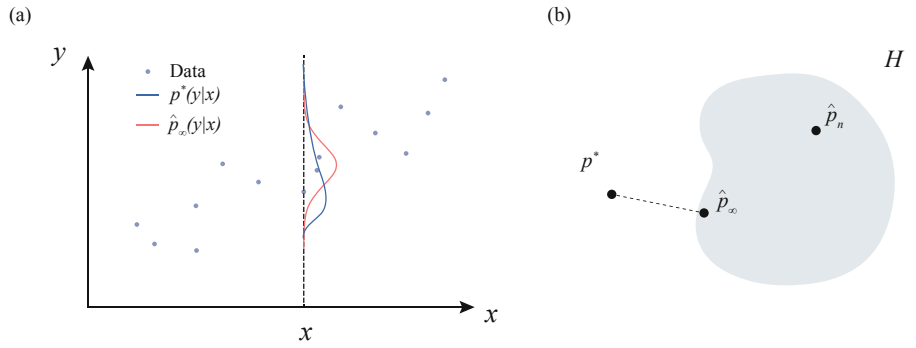

Fig. 2. Model misspecification in a regression example: (a) the true data distribution $p^*(y|x)$ (blue line) at a test point x and the estimated data distribution $\hat{p}_\infty(y|x)$ (red line) which is trained using an infinite amount of data at a test point x (b) $\hat{p}_\infty$ will never reach the truth p^* – even when optimizing for a proper loss, due to model misspecification. (Color figure online)

2.3 Contextualizing Model Misspecification

When looking at model misspecification, a key uncertainty we consider is epistemic uncertainty which is related to the model choice [18]. Figure 3 shows three perspectives on model misspecification and its relation to epistemic uncertainty for regression setting.

The first two scenarios (Fig. 3a and Fig. 3b) illustrate the most common perspective where model misspecification is ignored. Ignoring it could be acceptable

if the hypothesis space was sufficiently large and, therefore, model misspecification does not exist in principle (Fig. 3a). Even though most practitioners disregard model misspecification as illustrated in this Fig. 3a, in reality this is rarely the case. The most common circumstance in the literature is shown in Fig. 3b, where model misspecification exists despite being unintentionally ignored. As a result, for both scenarios, epistemic uncertainty does not contain model misspecification. In the third scenario (Fig. 3c), model misspecification is explicitly included as part of epistemic uncertainty. In this case, epistemic uncertainty consists of two parts: model misspecification (green line in Fig. 3c) and approximation uncertainty (blue line in Fig. 3c) as Hüllermeier and Waegeman [9] define it. In this scenario, model misspecification directly influences epistemic uncertainty estimates [20].

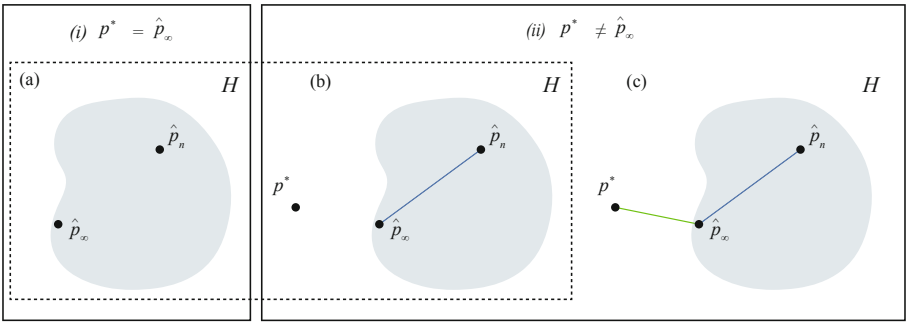

Fig. 3. Three views on model misspecification when a model class is assumed: In (a) and (b), model misspecification is ignored and epistemic uncertainty (blue line) is defined as a difference between $\hat{p}_\infty(y|x)$ and $\hat{p}_n$. In (c), model misspecification is treated as a part of epistemic uncertainty (green line). (Color figure online)

3 Effect of Model Misspecification on Uncertainty Estimates

We cover, in brief, the connection between model misspecification and epistemic uncertainty in different ML-methods and discuss possible consequences of model misspecification on uncertainty estimation and decision making. Although we mainly focus on neural network models in this section, the same principle applies to other methods as long as these models that require any distributional assumptions.

3.1 Model Misspecification and Uncertainty Estimates

Bayesian ML-methods are widely used for UQ [21]. Specifically, Bayesian neural networks (BNNs) are receiving a lot of attention due to their potential to

model both epistemic and aleatoric uncertainty [10,11,14,22]. The definition of uncertainty varies among researchers and BNN articles rarely include the effect of model misspecification. For instance, Gustafsson and colleagues [14] explain that epistemic uncertainty is uncertainty in the deep neural network model parameters, which ignores the fact that the model is not necessarily correctly specified. This definition of epistemic uncertainty is very common in the literature [9,10,23,24].

Without loss of generality, consider once again the regression setting from Subsect. 2.2. In BNNs, the total uncertainty is modelled by the posterior $p(\theta|\mathcal{D})$, where $\mathcal{D}$ symbolizes data. At test time, predictions are made via the posterior predictive distribution (PPD):

$$p(\mathbf{y}|\ \mathbf{x}, \mathcal{D}) = \int p(\mathbf{y}|\ \mathbf{x}, \theta) p(\theta|\mathcal{D}) d\theta. \tag{1}$$

In [14], both epistemic and aleatoric uncertainties are obtained assuming a Gaussian distribution with mean $\hat{\mu}$ and variance $\hat{\sigma^2}$ on PPD:

$$p(\mathbf{y}|\ \mathbf{x}, \mathcal{D}) \approx N(\mathbf{y}; \hat{\mu}(\mathbf{x}), \hat{\sigma}^2(\mathbf{x})). \tag{2}$$

It is safe to say that no true distribution is exactly Gaussian. The possible effects on UQ of such assumption are neither quantified nor discussed in this work; not as part of epistemic uncertainty itself, nor as a separate uncertainty. In the literature, the (implicit) assumption of a PPD with a Gaussian distribution seems to be made more generally when creating Bayesian ML-models [9,10,23,24].

Another popular approach for UQ is the use of ensemble methods, such as Monte Carlo dropout (MC-dropout) [24], where predictive uncertainty is estimated using Dropout [24] at test time, and Deep ensembles [5] which rely on retraining the same network many times with different weight initializations. Both methods can be considered a simple alternative to Bayesian methods. It is known that ensemble methods can provide an estimation of the (epistemic) uncertainty of a prediction [9], meaning that the variance of the predictions can be used to estimate epistemic uncertainty. By increasing the number of ensemble members, improved estimation of epistemic uncertainty is possible [9,25,26].

Liu and colleagues pointed out two issues when performing uncertainty estimation (both epistemic and aleatoric) using ensemble methods [27]. Similar to Bayesian methods, currently existing ensemble methods typically assume that the ground-truth data distribution $p(y|x)$ follows a Gaussian distribution [5]. Furthermore, ensemble methods perform uncertainty estimation using base models of the same class, meaning in the same hypothesis space H [5]. The consequence is that the creation of these models (that are all potentially misspecified) might result in a hypothesis space that still does not include the true distribution. Therefore epistemic uncertainty estimates from ensemble methods do not include model misspecification.

3.2 Consequences of Ignoring Model Misspecification

The central question that remains is what are the consequences of not taking model misspecification into account?

If we do not include model misspecification into epistemic uncertainty, we can only trust the estimated epistemic uncertainty when the model is correctly specified. As a result, we may significantly underestimate total uncertainty, which can be problematic in risk-involving tasks. Therefore, ignoring model misspecification usually leads to the scenario in Fig. 3b which can overestimate or underestimate the total uncertainty [16]. Specifically for Bayesian methods, when the distributional assumption on PPD is not correctly specified, the probability of the true distribution lying outside hypothesis space increases [16,28–32]. The consequences of ignoring model misspecification in real-world tasks are illustrated in Fig. 4, which is based on an example from [16].

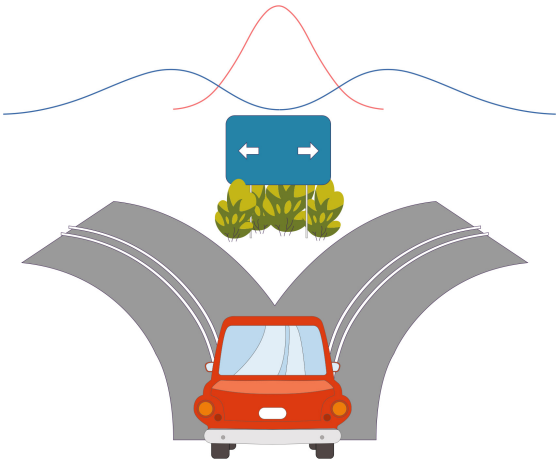

Fig. 4. A real-world example of model misspecification: In a decision making process at a split road, assuming a Gaussian distribution on the output distribution (red line) can cause a serious accident (Color figure online)

In Fig. 4, a lane splits in two directions, i.e., we have two outputs. In order to capture the multimodality, our output distribution has to follow a bi-modal distribution. However, by assuming a Gaussian distribution on the output distribution instead, the model predicts the mean of the two possible outcomes and the car goes in between lanes [33]. Although this can lead to a fatal accident, this possibility cannot be captured by estimating epistemic uncertainty with such model.

4 Discussion

In this paper, we highlighted the importance of considering model misspecification when performing UQ. Reviewing the literature, it should be noted that model misspecification is often not explicitly described or strong assumptions are imposed on the underlying data distributions. This can lead, in turn, to models that provide unreliable uncertainty estimates. These findings raise an important question: how should we handle model misspecification? As for the definition of different uncertainty types (i.e., aleatoric and epistemic uncertainty), the question remains whether epistemic uncertainty should contain model misspecification or not. These issues are discussed in the following Subsect. 4.1 and Subsect. 4.2.

4.1 How to Handle Model Misspecification

It is not possible to entirely avoid model misspecification when modeling real-world phenomena, mainly due to the imposed model assumptions. This means that we have to consider the impact of model misspecification and to explore ways to deal with it in various scenarios.

In principle, model misspecification can be reduced by expanding the initial hypothesis space H_1 (i.e., changing the associated model) (Fig. 5a). As a result, the impact of model misspecification can be reduced subsequently. Expanding the hypothesis space can be done, for instance, by easing the imposed assumptions. In the most favorable situation, the expanded hypothesis space H_3 includes the truth p^*. However, there is a possibility that we increase model misspecification when changing hypothesis space (Fig. 5b). As a result, we would never be able to reach the p^* in H_3. This situation can be avoided by assigning a hierarchy to the expanded hypotheses spaces as follows, $H_1 \subset H_2 \subset H_3$. Under this assumption, structural risk minimization (SRM) [34], which is strongly universally consistent [35], can be arbitrary close to the p^* in H_3. SRM uses the size of hypothesis space as a variable and tries to minimize the guaranteed risk over each hypothesis space [36,37]. This can reduce model misspecification.

Therefore, in either way (Fig. 5a or Fig. 5b), the impact of model misspecification can be minimized. This fact is important since it is not always clear if we change our hypothesis space inclusively (Fig. 5a) or exclusively (Fig. 5b).

Additionally, it is theoretically possible to reduce model misspecification completely according to the universal approximation theorem [38,39]. The theorem guarantees a neural network to represent any function (e.g., input-output relationship in regression) generically on its associated function space [40]. However, the theorem cannot be applicable to most practical situations due to the following reasons. Firstly, the theorem does not guarantee the network to learn the model [41]. Therefore, the chosen network can overfit to the training data, resulting in a poor generalization. Secondly, the network can require an exponential number of hidden units, which are often not desirable in practical situations [41]. These concerns lead us to the next question; What are the consequences when we attempt to reduce model misspecification during UQ?

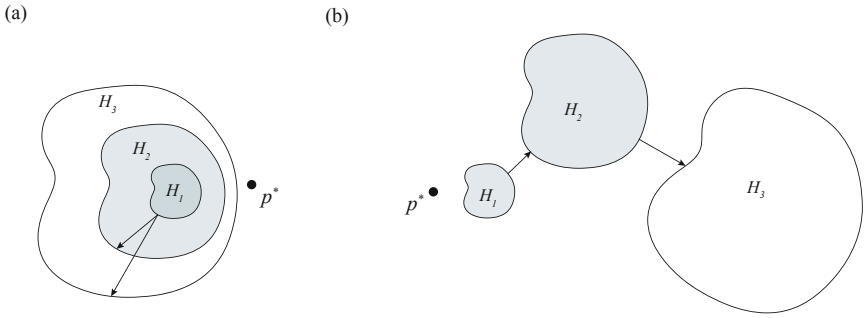

Fig. 5. Reducing model misspecification: By changing the initial hypothesis space H_1, model misspecification is either (a) reduced by including the truth p^* in the new hypothesis space H_3 or (b) increased.

In order to consider the question, let us go back to the scenario where model misspecification is considered to be part of epistemic uncertainty (Fig. 3c). Lahlou and colleagues [19] state that model misspecification can be considered as a form of bias while approximation uncertainty can be a form of variance. With the use of this concept, we can think about a bias-variance tradeoff with respect to the size of hypothesis space, which is illustrated in Fig. 6. Although bias-variance decomposition has been studied for a long time, to our best knowledge, no paper has considered this bias-variance trade-off with respect to uncertainty quantification. It is important to emphasize the following points: 1) by considering this bias-variance trade-off, we cannot guarantee to choose the model with the lowest total error in contrast to the original concept (i.e. the model class with the lowest total uncertainty does not necessarily provide the lowest total error.), 2) increasing the size of the hypothesis space does not make the model more complex (instead, it would be less complex with fewer assumptions).

In Fig. 6, we can see the effect of model misspecification in relation to total uncertainty. Assume that we have an infinite amount of data, then every hypothesis space shares the same amount of aleatoric uncertainty (red shaded area in Fig. 6). Depending on the model choice, the size of hypothesis space can vary. If we choose the initial hypothesis space H_1, there is a high possibility that p^* is not included in the hypothesis space, potentially increasing model misspecification and therefore bias of the model. On the contrary, if we choose an expanded hypothesis space H_3, model misspecification can be decreased due to the fact there is a higher possibility that p^* lies in H_3. However, at the same time, it will be harder to find an optimal model in an expanded hypothesis space. Therefore, the variance which represents approximation uncertainty increases when hypothesis space expands. This means that there is trade-off between bias (model misspecification) and variance (approximation error). In this example, hypothesis space H_2 has an optimal size. In this case, H_2 results in the best trade-off between model misspecification and its size. Using this concept, we can choose the best possible model and the associated hypothesis space, even if we do not know p^*.

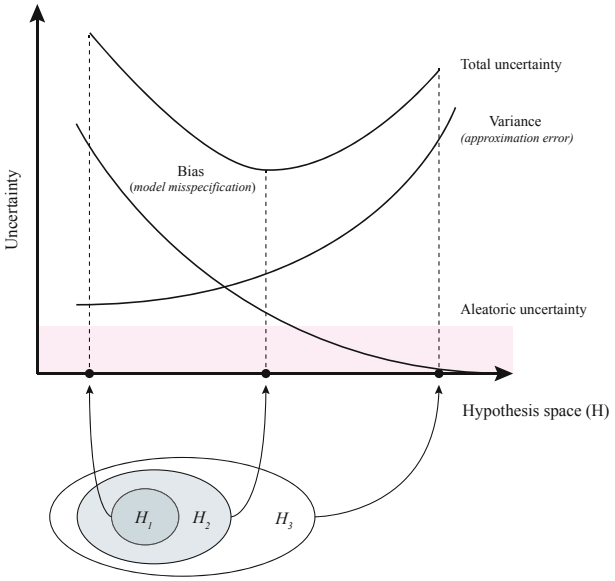

Fig. 6. Intuition for bias (model misspecification) variance (approximation error) trade-off with respect to the size of hypothesis space. Red shaded area represents aleatoric uncertainty. (Color figure online)

However, whether the approximation error and model misspecification can be estimated in practical scenarios is not entirely clear yet. This means that how to handle model misspecification still remains an open question.

4.2 Conflicting Views on Uncertainty Definitions

There are conflicting views and significant terminology diversity about the different types of uncertainty in the literature. In particular, model misspecification has been included as part of epistemic uncertainty by some authors while others do not [27,28]. This is problematic since it prevents a direct comparison regarding the performance of different methods. A uniform view on how to treat model misspecification (either being part of epistemic uncertainty or as a separate uncertainty type) should be made to make such comparison possible. Furthermore, it is not very clear which type of uncertainty is estimated in some situations. For example, Valdenegro-Toro and colleagues [30] claim that there is a clear connection between epistemic and aleatoric uncertainty. This would mean that underestimating or overestimating epistemic uncertainty can have an effect on the quality of aleatoric uncertainty. Therefore, a true possibility for such a separation can be doubted in this case. Additionally, there is another issue related to the definition of uncertainty, which is inconsistent naming of uncertainties. This variability in naming of epistemic uncertainty is illustrated

in Fig. 7. Note that this figure does not reflect an exhaustive literature search, so there is even more diversity in terminology than what is illustrated there.

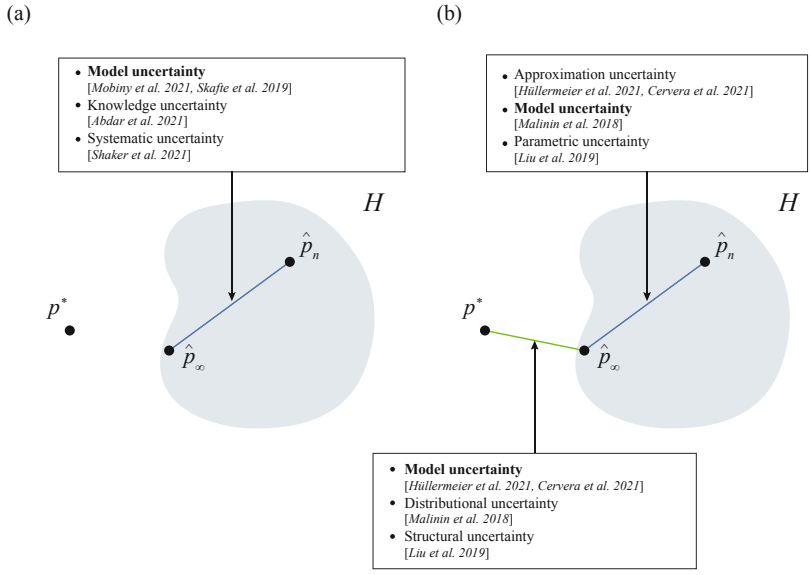

Fig. 7. Example of variability in naming of epistemic uncertainty: (a) Model misspecification is ignored. (b) Model misspecification is considered as part of epistemic uncertainty (a green line). Model uncertainty represents different concepts depending on literature. (Color figure online)

Epistemic uncertainty is called as, for instance, model uncertainty [23,42], knowledge uncertainty [8], systematic uncertainty [43] (Fig. 7a). In other work, model uncertainty is considered to be part of epistemic uncertainty (Fig. 7b), representing model misspecification [9,16]. Malinin and colleagues [44] call model misspecification as distributional uncertainty. This inconsistency can be confusing and leading to wrong interpretation of published work.

5 Conclusion

There are a number of concerns when it comes to model misspecification in relation to uncertainty estimation that can considerably influence the accuracy by which uncertainty estimation can be performed. No general definition of both epistemic and aleatoric uncertainty currently exists. Additionally, researchers treat model misspecification differently. Since model misspecification influences the reliability of uncertainty estimates, we would argue that model misspecification, and the ways to measure and control it, should receive more attention

in the current UQ research. Furthermore, we propose directions for future work to minimize model misspecification, for instance, by looking into bias-variance trade-offs.

References

1. Hie, B., Bryson, B.D., Berger, B.: Leveraging uncertainty in machine learning accelerates biological discovery and design. Cell Syst. **11**, 461–477 (2020)
2. Vishwakarma, G., Sonpal, A., Hachmann, J.: Metrics for benchmarking and uncertainty quantification: quality, applicability, and best practices for machine learning in chemistry. Trends Chem. **3**, 146–156 (2021)
3. Begoli, E., Bhattacharya, T., Kusnezov, D.: The need for uncertainty quantification in machine-assisted medical decision making. Nat. Mach. Intell. **1**, 20–23 (2019)
4. Michelmore, R., Kwiatkowska, M., Gal, Y.: Evaluating Uncertainty Quantification in End-to-End Autonomous Driving Control. arXiv: 1811.06817 (2018)
5. Lakshminarayanan, B., Pritzel, A., Blundell, C.: Simple and scalable predictive uncertainty estimation using deep ensembles. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
6. Maddox, W.J., Garipov, T., Izmailov, P., Vetrov, D., Wilson, A.G.: in Proceedings of the 33rd International Conference on Neural Information Processing Systems, pp. 13153–13164. Curran Associates Inc., Red Hook (2019)
7. Kiureghian, A.D., Ditlevsen, O.: Aleatory or epistemic? Does it matter? Struct. Saf. **31**, 105–112 (2009)
8. Abdar, M., et al.: A review of uncertainty quantification in deep learning: techniques, applications and challenges. Inf. Fusion **76**, 243–297 (2021)
9. Hüllermeier, E., Waegeman, W.: Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. Mach. Learn. **110**, 457–506 (2021)
10. Depeweg, S., Hernandez-Lobato, J.-M., Doshi-Velez, F., Udluft, S.: Decomposition of uncertainty in Bayesian deep learning for efficient and risk-sensitive learning. In: International Conference on Machine Learning, pp. 1184–1193 (2018)
11. Kendall, A., Gal, Y.: What uncertainties do we need in Bayesian deep learning for computer vision? In: Guyon, I., et al. (eds.) Advances in Neural Information Processing Systems, vol. 30. Curran Associates Inc. (2017)
12. Senge, R., et al.: Reliable classification: learning classifiers that distinguish aleatoric and epistemic uncertainty. Inf. Sci. **255**, 16–29 (2014)
13. Prado, A., Kausik, R., Venkataramanan, L.: Dual Neural Network Architecture for Determining Epistemic and Aleatoric Uncertainties. arXiv: 1910.06153 (2019)
14. Gustafsson, F.K., Danelljan, M., Schon, T.B.: Evaluating scalable Bayesian deep learning methods for robust computer vision. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1289–1298 (2020)
15. Nguyen, V.-L., Shaker, M.H., Hüllermeier, E.: How to measure uncertainty in uncertainty sampling for active learning. Mach. Learn. **111**, 89–122 (2022)
16. Cervera, M.R., et al.: Uncertainty estimation under model misspecification in neural network regression. arXiv: 2111.11763 (2021)
17. Lv, J., Liu, J.S.: Model selection principles in misspecified models. J. R. Stat. Soc. Series B Stat. Methodol. **76**, 141–167 (2014)

18. Aydogan, I., Berger, L., Bosetti, V., Ning, L.I.U.: Three Layers of Uncertainty and the Role of Model Misspecification. Working Papers (2020)
19. Lahlou, S., et al.: DEUP: Direct Epistemic Uncertainty Prediction. arXiv:2102.08501 (2021)
20. Xu, A., Raginsky, M.: Minimum excess risk in Bayesian learning. IEEE Trans. Inf. Theory **68**(12), 7935–7955 (2022)
21. Tipping, M.E.: Bayesian inference: an introduction to principles and practice in machine learning. In: Bousquet, O., von Luxburg, U., Rätsch, G. (eds.) ML -2003. LNCS (LNAI), vol. 3176, pp. 41–62. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-28650-9_3
22. Scalia, G., Grambow, C.A., Pernici, B., Li, Y.-P., Green, W.H.: Evaluating scalable uncertainty estimation methods for deep learning-based molecular property prediction. J. Chem. Inf. Model. **60**, 2697–2717 (2020)
23. Mobiny, A., et al.: DropConnect is effective in modeling uncertainty of Bayesian deep networks. Sci. Rep. **11**, 5458 (2021)
24. Gal, Y., Ghahramani, Z.: Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: Balcan, M.F., Weinberger, K.Q. (eds.) Proceedings of the 33rd International Conference on Machine Learning, vol. 48, pp. 1050–1059. PMLR, New York (2016)
25. Charpentier, B., Senanayake, R., Kochenderfer, M., Günnemann, S.: Disentangling Epistemic and Aleatoric Uncertainty in Reinforcement Learning. arXiv: 2206.01558 (2022)
26. Caldeira, J., Nord, B.: Deeply uncertain: comparing methods of uncertainty quantification in deep learning algorithms. Mach. Learn. Sci. Technol. **2**, 015002 (2020)
27. Liu, J.Z., Paisley, J., Kioumourtzoglou, M.-A., Coull, B.A.: in Proceedings of the 33rd International Conference on Neural Information Processing Systems, pp. 8952–8963. Curran Associates Inc., Red Hook (2019)
28. Masegosa, A.: Learning under model misspecification: applications to variational and ensemble methods. In: Advances in Neural Information Processing Systems (2020)
29. Jewson, J., Smith, J.Q., Holmes, C.: Principles of Bayesian inference using general divergence criteria. Entropy **20**, 442 (2018)
30. Valdenegro-Toro, M., Mori, D.S.: A deeper look into aleatoric and epistemic uncertainty disentanglement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1509–1517 (2022)
31. Hansen, L.P., Sargent, T.J.: Structured uncertainty and model misspecification. University of Chicago, Becker (2019)
32. Ramamoorthi, R.V., Sriram, K., Martin, R.: On Posterior Concentration in Misspecified Models, vol. 10, pp. 759–789 (2015)
33. Cerreia-Vioglio, S., Hansen, L.P., Maccheroni, F., Marinacci, M.: Making Decisions under Model Misspecification (2020)
34. Guyon, I., Vapnik, V., Boser, B., Bottou, L., Solla, S.A.: Structural risk minimization for character recognition. In: Moody, J., Hanson, S., Lippmann, R.P. (eds.) Advances in Neural Information Processing Systems, vol. 4. Morgan-Kaufmann (1991)
35. Lugosi, G., Zeger, K.: Concept learning using complexity regularization. In: Proceedings of 1995 IEEE International Symposium on Information Theory, Whistler, BC, Canada. IEEE (2002)
36. Corani, G., Gatto, M.: Structural risk minimization: a robust method for density-dependence detection and model selection. Ecography **30**, 400–416 (2007)

37. Zhang, X.: Structural risk minimization. In: Sammut, C., Webb, G.I. (eds.) Encyclopedia of Machine Learning, pp. 929–930. Springer, Boston (2010). https://doi.org/10.1007/978-0-387-30164-8_793
38. Hornik, K., Stinchcombe, M., White, H.: Multilayer feedforward networks are universal approximators. Neural Netw. **2**, 359–366 (1989)
39. Cybenko, G.: Approximation by superpositions of a sigmoidal function. Math. Control Signals Syst. **2**, 303–314 (1989)
40. Kratsios, A.: The universal approximation property. Ann. Math. Artif. Intell. **89**, 435–469 (2021)
41. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge (2016)
42. Skafte, N., Jo rgensen, M., Hauberg, S.R.: Reliable training and estimation of variance networks. In: Wallach, H., et al. (eds.) Advances in Neural Information Processing Systems, vol. 32. Curran Associates Inc. (2019)
43. Shaker, M.H., Hüllermeier, E.: Ensemble-based Uncertainty Quantification: Bayesian versus Credal Inference. arXiv: 2107.10384 (2021)
44. Malinin, A., Gales, M.: Predictive uncertainty estimation via prior networks. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems, Montréal, Canada, pp. 7047–7058. Curran Associates Inc. (2018)