Panel Debate on the integration of AI and machine learning with geostatistics

Abstract On the afternoon of August 10, 2023, in the Natural Science Building of the Norwegian University of Science and Technology in Trondheim, it took place a panel discussion on "How can the integration of AI and machine learning approaches with geostatistics be achieved without compromising the rigorous statistical foundations of traditional geostatistical methods?" The discussion was part of the closure session of the 22nd Annual Conference of the International Association for Mathematical Geosciences. The invited panelists were Jennifer McKinley (Queens University Belfast), Shaunna Morrison (Carnegie Institution for Science), Lingchen Zhu (Schlumberger), Thomas Mejer Hansen (Aarhus University), Odd Kolbjørnsen (University of Oslo and Aker BP), Gerald v.d. Boogart (Techinical University Freiberg), and the moderator was Richard Sinding-Larsen (Norwegian University of Science and Technology).
The debate explored the transformative impact of Machine Learning (ML) and AI on geostatistical practices, scrutinizing the most beneficial integration of these technologies within a geostatistical framework. Reflecting on the developments since the 1999 IAMG event, the panelists commented upon how AI and ML should be leveraged within a geostatistical context, and the role of geoscientists versus computer scientists in driving these advancements.
Discussions emphasized the historical integration of machine learning in geoscience, the significance of combining domain knowledge with mathematical algorithms, and the practical application of AI in geoscientific problem-solving, highlighting the need to quantify uncertainty through Bayesian approaches.
Audience insights underscored the importance of cross-disciplinary collaboration, open code policy, and a profound understanding of underlying assumptions.
The debate revealed AI's transformative potential in geostatistical applications when coupled with geoscientific insights.
The panel debate was succeeded by the announcement of a forthcoming special issue in Mathematical Geoscience on how AI and ML could be leveraged within a geostatistical framework. It aims to present a visionary perspective on the impact of collaborative advancements of integrating Machine Learning (ML) and AI into geostatistical practices.

Introduction

On the afternoon of August 10, 2023, in the Natural Science Building of the Norwegian

University of Science and Technology in Trondheim, it took place a panel discussion on "How can the integration of AI and machine learning approaches with geostatistics be achieved without compromising the rigorous statistical foundations of traditional geostatistical methods?" The discussion was part of the 22nd Annual Conference of the International Association for Mathematical Geosciences. The invited panelists were Jennifer McKinley (Queens University Belfast), Shaunna Morrison (Carnegie Institution for Science), Lingchen Zhu (Schlumberger), Thomas Mejer Hansen (Aarhus University), Odd Kolbjørnsen (University of Oslo and Aker BP), Gerald v.d. Boogart (Techinical University Freiberg), and the moderator was Richard Sinding-Larsen (Norwegian University of Science and Technology). The panel debate was recorded and automatically transcribed in the conference system Panopto. Some parts of the audio file were less audible. So, there are some gaps and some guesswork in the transcription. We have made only moderate adjustments, in order to preserve the oral, spontaneous character of the interventions.

Debate

Richard. *I want to extend a warm thank you for being here and joining us for this panel debate. It's noteworthy that the organizing committee made a decision to hold this panel debate, which actually breaks away from the usual tradition of IAMG meetings. But it's always valuable to explore new opportunities, and this one seemed particularly promising. The reason behind this opportunity, of course, stems from the remarkable advancements and progress within the realm of AI. Looking back to 1999, when the last IAMG meeting took place in Trondheim, none of these possibilities were within reach. Yet, even back then, we engaged in vibrant discussions around multipoint statistics, which was quite a hot topic at the time. Many of the items on the agenda then resonate with those on our agenda today. However, the landscape has shifted with the emergence of new advanced tools. This prompts us to question: how can we leverage these tools for the benefit of geoscientists? How can geostatisticians feel empowered? To achieve this, we need to consider these tools through a geostatistical lens.*
So, the central question becomes: who should flex? Who should adjust? Should it be the burgeoning AI technology adapting to geoscientists, or the other way around? Perhaps it's a symbiotic relationship. These are the fascinating dimensions we're about to delve into. Amidst these discussions, a recurring theme is the trustworthiness of results and the demand for explicable AI. AI is promising as long as we comprehend its workings. If we can't grasp its processes, there's a risk that many might shy away from embracing these technologies. It's akin to administering an untested, new medicine to a patient, only to witness unfortunate consequences. Hence, the critical inquiry becomes: how can mathematical geoscientists contribute to enhance the fruitful integration of AI and machine learning with geostatistics? The goal is to ensure that everyone perceives it as a benefit to the geostatistical community rather than a distracting deviation.
So, with this question, I'm eager to introduce our panelists. We're fortunate to have a diverse blend of expertise and experiences, spanning spatial domains from the vastness of outer space to the minutiae of the tiniest mineral grains.

Our first panelist is ***Jenny McKinley****, a Professor of Geoscience based in Geography in the School of Natural and Built Environment, Queen's University Belfast, UK. Her* *research is focused on the application of spatial analysis techniques, including geostatistics, compositional data analysis and machine learning in the natural environment. The scope and applications of her geospatial research have been varied and span: soil geochemistry; environmental and criminal forensics; natural stone heritage weathering and conversation; human health; ground instability; renewable energy, airborne geophysics and peatland conservation and nature-based solutions in the urban environment. She is passionate about how mathematical geoscience can be meaningful to addressing the real-world relevance and impact of research. A key focus of hercurrent research seeks to gain a greater understanding of the link between health, the environment and climatic impacts. Please welcome Jenny McKinley!*

Our next panelist is ***Shaunna Morrison*** *(Carnegie Research Scientist at the Earth and Planets Laboratory of the Carnegie Institution for Science). She is currently focused on the data-driven exploration of Earth and planetary systems and has had experience in mineralogy, crystallography, planetary science (NASA Curiosity rover mission), and the application of data science techniques to mineralogical systems. She is passionate about integrating across scientific disciplines to answer outstanding questions in Earth and planetary science, utilizing data science as the shared language between fields. Her interests include understanding the deep-time evolution of Earth and planetary bodies, the coevolution of Earth materials and system with life, and identifying biosignatures on other planets. Please welcome Shaunna Morrison!*

Our next panelist is ***Lingchen Zhu****, a Senior Research Scientist at Schlumberger-Doll Research Center at Cambridge, Massachusetts, USA. He is currently focused on reservoir forward modeling and simulation, high-performance computing and building generative AI products for high-fidelity 3D subsurface modeling, and has worked with a range of/had experience in signal processing algorithms, deterministic and probabilistic deep learning methods, dimensionality reduction, and their applications on well integrity evaluation and full-waveform inversion. He is passionate about physics-informed deep learning methods, generative AI for large-scale 3D geological models and accurate geostatistics as well as building robust and reliable products leveraging machine learning algorithms and MLOps practices. Please welcome Lingchen Zhu!”*

Our next panelist is ***Thomas Mejer Hansen*** *(Associate Professor, Dept. of Geoscience, Aarhus University). He is currently focused on allowing the use of realistic (spatial) prior knowledge in the solution of probabilistic inverse problems, and has worked in geoscience and medical imaging. He is passionate about probabilistic models, and utilizing the potential of machine learning in that field. Please welcome Thomas Mejer Hansen!*

Our next panelist is ***Odd Kolbjørnsen****, an Associate Professor in Statistics and Data Science at the Department of mathematics, University of Oslo and a Value stream manager for subsurface digital twin, Aker BP. He is currently focused on Bayesian methods for utilizing geoscience data in reservoir characterization. Main focus on integrating seismic data, and production history. He was an associate editor of Mathematical geosciences 2010-2015. He is passionate about transparency, clarity, and* *accountability in the decision-making process when using data and computational models. Please welcome Odd Kolbjørnsen!*

Our last panelist ***Gerald van den Boogaart,*** *leads the department of Modelling and Valuation at the Helmholtz Institute Freiberg for Resource technology and is a Professor of Applied Stochastics at the TU Bergakademie Freiberg. He is currently focused on Optimized adaptive Processing for Primary and Secondary Raw Materials and has worked in Geostatistics, Microstructure Modelling, Compositional data, Bayes Space Theory and other fields of stochastic inference. He is passionate about not mistaking an example with a small prediction error for good science. Please welcome Gerald van den Boogaart!*

Richard*. From what you've just heard in these introductions, there's a wide range of interests and, of course, expertise. So, we're really eager to hear your thoughts on how a mathematical geoscientist can contribute to and enhance insight into the fusion of AI and machine learning with geostatistics. I would like to kick off the debate with you, Jenny, what are your opinions on this topic?*

Jenny. First of all, I would like to thank the panelists for agreeing to be part of this discussion on how mathematical geoscience can enhance the current work on AI in geostatistics. I think most people know that my expertise in not in machine learning, but I have published with colleagues using machine learning techniques. I would like to make three points, if I may. Speaking as a voice of experience in the IAMG, I've been a member of the IAMG Council since 2012. So I've been involved in planning for the IAMG conferences during this time. This involved thinking about the conference sessions and topics and thinking more strategically with council members about the future direction of the IAMG. Steinar Ellefmo mentioned this morning in his opening welcome about geology being a diverse subject and geoscience as being a diverse subject. The IAMG has always been a diverse community. Mathematical geoscience is diverse and has always included a diverse range of areas. That's why I believe as a community and association, we, as the IAMG, are unique. Soin thinking about the IAMG conference sessions in the past, and I'll go back to 1999 when the IAMG annual conference was held in Trondheim, there were sessions and talks on machine learning I'm going to acknowledge Vasily Demyanov and his colleagues, in organising at IAMG1999 in Trondheim which specifically explored issues on how to integrate ML techniques with geostatistics. So the IAMG has a history in discussing ML in mathematical geoscience. This is not new, and Michael Pyrcz in his excellent keynote at this conference raised this very point. I think we need to acknowledge all our IAMG colleagues including Mikhail Kanevsky, Vasily Demyanov and colleagues for their pioneering work, trying to develop and integrate ML into mathematical geoscience. The he poster by Marshall Ma looks at the development of the IAMG through paper subject matter in the IAMG journals, and we can see a consistency in the occurrence of geostatistics, Kriging, with an increase in GIS which comes down again. We can see that machine learning is consistently represented in the journals, whereas referfence to. AI is more recent. That's really interesting, and I think, again, it shows the diversity of mathematical geoscience. My second point is the reason why as a diverse community we have to work together is because there are special characteristics of geoscience data that we have to acknowledge. One that has been mentioned throughout the conference, is the importance of spatial dependence. There are also characteristics of different geoscience data, such as geochemical data that are proportional or compositional.There is excellent work being done and more papers being published which combinemachine learning and AI techniques with compositional data analysis. There is so much more to do in that space. My last point is linked to this. I'm a geoscientist and I work in applied fields. I love hearing about new approaches, and I always think, what can I use that for? Is that going to be applicable to what I do? You've got to think about how any results can be interpreted by geoscientists. There is no point in people working in silos if for example a computer scientist comes up with an AI solution, and presents it to me saying there's your result, and I think yes, but it's meaningless to me. Why didn't we do this together, and then we could have produced a really important and insightful output? I'm very aware that my wonderful colleague sitting beside me, keynote's title was Data-driven discoveries. When I hear data-driven, I think, actually, it's not, because we always bring our knowledge. So, yes, data can provide so many insights, but it's not data-driven, because we're the ones who are analysing the data and as geoscientists our knowledge informs all we do. So I would stress in this discussion that we lose the importance of our expertise and our geoscience understanding at our peril.

Richard. *Well, thank you very much. I think that's a very good start for this debate, and we are then welcoming the cosmic perspective on what you said, Jennifer.*

Shaunna. Thank you - You set me up very well. Actually, I completely agree with everything that you said, including the last point. And I also have three points to make, but one really big point, and it's kind of the umbrella of everything that I want to say. Certainly, you can't lose the geo component. You have to have domain scientists in the loop. I'm not a geo-statistician. I'm learning so much from all of you in this room. I'm a geologist. I'm a planetary scientist. I'm a mineralogist. My statistical background was very weak coming into geology, and that's actually what led me into a more data science or data-driven approach. The main point for me or, really, everyone (especially anyone who doesn't have a strong machine learning background) is that it's very important to partner with people who actually understand the mathematics and the algorithms that are underlying these methods. Alone, I would never have been able to do any of the things that I do – well, maybe I could have done them by myself, but I don't know how much I would trust those results, and that really comes back to a lot of what we've been talking about with explainability. If you talk to any of my colleagues, they are forced to explain everything to me, and so that's usually our conversation with my data science colleagues or my statistician colleagues is, "wait, but why did it do that? But what is this based on?" and that discourse between us is really what has – first of all, not only has it helped us solve the problem at hand, but it's actually created a lot of new ideas and new opportunities both within geoscience but also in data science. I think the most important thing that we could be doing is collaborating with pure computer scientists, pure data scientists, pure mathematicians who really think about the fundamentals of the math and data science, where we bring the geo angle to the conversation. Sometimes it can be hard to reach across these silos. It can be hard to find these people. It can be hard to engage with them. Any of us who are at a university likely have access to computer science departments and statistics departments, and I think we offer them pretty cool problems. These are really cool questions. I'm looking at Grethe Hystad right now. Greta's a statistician, and in working with us, became a geostatistician. We pulled her in. And the same thing, Anirudh Prabhu sitting right behind her. He's a data scientist, and he works in many different domains, not just earth science. Because our problems were so cool and interesting, he's here at this conference, and he's at AGU, and he works with me very closely because we have incredibly cool micro- to cosmic-scale questions. And not only do we offer them really cool problems, but we also offer them opportunities to make advances in their own fields. They are still mathematicians and computer scientists, and it's important that we give them the opportunities to make new algorithms, to apply new algorithms in different settings, and to get the papers that they also need for their tenure and advancement. And it's doable. We do that on a regular basis. Some portion of the papers that we publish are actually in data science journals. I think those are the two key components of that relationship. The other point I would like to make is: borrowing from other fields. A lot of fields have encountered a problem before we came to it. Biology is one that I borrow from a lot in my work. They've been dealing with big data issues long before mineralogy. In truth, mineralogy still doesn’t have “big data.” I have very small data. But they have encountered a lot of issues before, and they have a lot of methods that can be suitable for my problems. Keeping that outward look, going into the sessions, and even the conferences and domains that are not your own, is really, really important. And forming those relationships. I've mainly talked about the relationship between the earth scientist and the data scientist up until this point, but making those relationships with other fields that can, or maybe you think can't, influence your work - They might actually be able to. You might be able to do some cool stuff together. So, I think relationships with other scientists are very important. The third point that I wanted to make was: make sure we're really trying everything that is available out there to us and we're not sticking to one method because we know it and we're comfortable with it. We should try those methods that we know and are comfortable with, but we should be trying everything else too. Maybe not *everything*. We all have limited hours in the day. But trying different methods and learning about different techniques and trying to see what fits our data and our problem the best. Because it's not always what you expect - it's often in my experience actually not even the newest, fanciest, biggest tool that functions best. It's sometimes something sort of in between the most basic and the most complex or newest. So, I trying different methods is important. In the end, my main takeway is collaboration.

***Richard**. Thank you so much. That was really a very stimulating presentation, especially for the younger folks. You're basically highlighting that there's an incredible array of avenues to explore. So, the younger generation need not worry about job prospects, as there's an abundance of fascinating challenges to dive into. And then to Lingchen.*

Lingchen. Thank you very much, Richard. And actually I think I may be the only one who is not from a university background here. I am the only one from industry, from

Schlumberger or SLB right now. And, yeah, so I can think a little bit about how the AI is reflecting our AI in the industry and especially in a company. Also, if you speak about data-driven, which is part of my Ph.D. dissertation title, that is about data-driven seismic and noising, data-driven full-waveform inversion, that about 10 years before I was getting into the machine learning period, though at that time I think deep learning may not be the very real top year because of the limitation of the GPUs at that time. But still we're getting to that. And speaking of the connection between the academic and industry, I think it's quite intertwined that even think about, say, Google. Now SLB is a big customer of the Google Cloud and the Azure Cloud and AWS. We use all of their services to process our data. And either in the – not only in the forward modeling, but also in machine learning methods. So that means that it's tightly very close to each other. And so the big companies provide resources to industries to let them test their ideas on the geoinformation and support us to process our workflows, no matter on both scenarios, on traditional numerical methods or on the data-driven methods. And so I think it's definitely AI – incorporating AI into the geoinformation field is definitely a good point for these companies. And also speaking of the latest AI algorithms, as Shona and Jennifer said a lot, lots of the algorithms we pick, not only because they're fancy, they're new, but also they can solve our problems. And we can even find their limitations of the algorithms and we give feedbacks to them. And it's kind of like a loop, a beneficial loop, that from the pure CS world to the geoscience world. And so I think it's – we don't say it's a split of the two disciplines, but it's actually a cross-discipline effort that both of the fields can be favoring from there. And speaking of me myself, actually it's my first time to join an IAMG conference. I'm really grateful that Richard can invite me to this panel debate. I'm really grateful to sit between all these professors. And actually my background is from signal processing, but signal processing is actually – not everyone in the signal processing, in every ECE department or CS department is now working on the deep learning problems, AI problems for different disciplines. Not only for the bioinfo or geoinfo or astronomy or the planetary science, lots of things. So I think that AI is just like – my opinion is AI and computer science is just like a basic skill of everyone in the world. They have to know something about that. But it's not the only skill you want to learn. If you want to solve real problems, you need to find a problem that fits your needs. I mean, as everyone has introduced and all of the practitioners from this conference, that these problems are very interesting and very beneficial to the society and as far as humankind, that they are very interesting and they can be beneficial for all the researchers that we are involved. And even if you are devoting yourself, your time, your efforts into this research area, it's also very helpful for your own personal skills development. And say for us, we have so many employees from so many different backgrounds of the different talented universities worldwide and some of them, they work for a long time and because they found our problem so interesting, they want to keep spending time on there. And some of the employees may even go to Google, Amazon, these big companies, big tech companies. So I think from my own point of view that AI is a common recipe that can be worked into for every industry. And now our topic here today is about explainable AI, which is also a very real hot topic on the AI research itself, that we have so many packages already in the Python community that can do something about it, like the LIME, the Shapley, and other stuff that I may have heard of. And I use some of that and I think it definitely is helpful to explain a little bit about our results. And these results, sometimes we can feedback from them and submit GitHub issues and to

explain what we have found on something that they cannot explain. And it's kind of a very helpful iteration for the economic and industry to push this research ahead. So I think my point here is, we all benefit a lot and it's a cycle that both industries can benefit from that. And not only AI helps out the geoinformation discipline, but geoinformation can also help the AI research as well.
Richard. *Thank you. I found your morning presentation really captivating, especially the industrial angle you brought in. It seems Schlumberger is on a path to reshape things by essentially forward modeling a new digital version of the geoscience reality. And from what I gathered, you mentioned that by 2030, most of the training will rely on simulated reality – quite a thought-provoking shift. There's something that caught me off guard in your talk. You stated that the beauty of forward modeling is that we can have a version of reality with no noise.*

Lingchen. We can introduce as much noise as we want, in whatever form we prefer. It's similar to what Tesla did – they must have developed their own real-world engine to mimic various road scenarios. This way, they could train their system using that simulated data instead of relying solely on real-world data collection.

Richard. *Thank you again. Now, we're curious to learn how the Danish wisdom you sheared from my anecdote on Niels Bohr yesterday ties into this. And seeing that you Thomas come from the same background adds an interesting dimension.*

Thomas. If you put the bar that high, there's this telling about Niels Bohr that he went to make a presentation at a fine meeting like this, and he couldn't figure out how to adjust the projector so that you could actually see something on the screen. So, he completely messed up that part. I hope the story true, and that he was ‘human’ too.
A comment about this AI term. I realized over the last couple of days that I probably don't apply AI at all, but instead machine learning. The thing is, when I see something like an AI based geological model, it may suggest a solution, but how did that happen? How did it make that decision? That's probably AI. You know, when I program, I love to use AI such as ChatGPT. It helps me and I'm kind of amazed what it does. Sometimes it's a little off but it is certainly useful. But when I work on models and taking decisions, I am not satisfied being a little off. It's super difficult for me to advocate the use of that for anybody if you cannot really explain where information comes from. So, I mostly do machine learning where we can most often understand the model.
I noticed over the first few days that in this community, there's at least two types of machine learning. One of them which we really do understand completely. There's no black box at all. I think it was Pyrcz who said in his keynote that machine learning is all about mapping; mapping from one domain to another. All I've been doing the last 20 years is going from the data space to the model space, or data with noise to something related to the model space. And that is all just a mapping. And what we can do in this community, we can simulate everything, all the information that goes into this mapping. We're really, really good at that. And so we can construct amazing training data set. And then using just a simple neural network, by choosing the loss function, we choose exactly how we understand the output of that model. So, there's no black box there. It's like when I use a Metropolis

algorithm, I know that it's guaranteed to sample the correct distribution within finite time. Sometimes it's practically unsolvable actually, but in principle it works. The same with these networks when you design them. It's only a matter of having a large enough training data set and a complex enough structure of network, then we can actually use it to represent any mapping. And then it's just a tool like anything else. And then there's other types of machine learning that I don't really understand. And maybe we don't really understand yet some of these methods that generate realizations of spatial fields. And then I see in this community what we do here that I really like. We spend so much time analyzing what comes out of these models. So I think this community is the place that can actually make sure that these kind of models actually make sense to use for making predictions.
And then just finally also a story from this IAMG meeting, related to Lukas Mosser's presentation. In my mental baggage I carry some problems. I understand everything from a perspective of inverse problems. I look for solutions that can help me. And then sometimes somebody comes with a solution to something. Maybe based on machine learning as in this case. That solves some problem. And as a byproduct it just makes something feasible, a problem I have had, that yesterday I didn't spend much time on it because it was unsolvable. But now if, I understand correctly, in principle it's a trivial problem. And that's the potential of machine learning in these years. Some problems appear impossible; so we don't really think of going that way because we practically cannot. And then something comes along that is so much faster. And suddenly we can do it. So that's super exciting. I had a little highlight there. That was super nice. So just embrace those machine learning methods. And some of them are definitely not black boxes. We know exactly what we do. So they are easy to explain. Thank you.

Richard. *Thanks, Thomas. I must say, those points really hit the mark. Your reference to machine learning as a mapping from one domain to another between data and model was quite insightful. If I've got it right, having an accurate model is like having an unlimited dataset. So, in essence, there's a direct link between a well-chosen model and the data. Thanks a lot. Now, let's turn our ears to what Kolbjørnsen has to share.*

***Odd*.** This adds into how I think about this. These methods are extra tools that we can utilize. We really need to understand how we can utilize this. I'm also from the industry I work is in AKBP, and I have a 20% position at the University of Oslo. So I was asked also to bring the industry perspective into AI and ML. When we think of AI, we think of turning data into decisions. Sometimes it's advisable to try to think as the person that's taking the decision, But in our case when we talk about explainable AI. We could imagine ourself explaining to this person why your method failed.. All the things I would like to say to him when I try to explain why it failed, I should have told him that up front. We need to have a basic understanding of what the method does, what are the weaknesses. This also goes into how we think about as ML, which is turning data into knowledge. Why are we not eager to use these new methods in the industry? It's because these are new methods, and new methods have also new type of errors. It's much easier for me to say when analyzing the data. These are known problems, and these are issues that are coming. Others that did the analysis did all of these mistakes yesterday, so I'm likely to do some of those mistakes today as well. When I use a new method, I put my head on the block,. I really need the data scientists, geoscientists to help me explain what are the risks of using this method to go forward. . So I'm really excited about these new possibilities we are offered in our time, and I've seen many possible usages here at this conference I'm really looking forward to see where this will go in the future.

Richard. *Thanks. I really like how you're emphasizing that adopting new methods demand courage to put your head on the block and stand by your promises. So, having this mindset of not risking your reputation unless you're really confident in the outcomes is crucial. We need to use that motivation to ensure that what we're adopting is valid. Many others have also highlighted these points, and I think they're really important.. Now, Gerald, we're all ears for your seasoned IAMG perspective.*

Gerald. To me it is about the science. So where do we find the science? And there is methodological science in AI or machine learning research. There is geoscience. And there is engineering science. So using that, for example, for decision, an optimal way of decision. And the first thing about when it comes to where is the science is the methodological science to me as a mathematician. So and there I saw at least often what he (Thomas) said, publications where a method was proposed and then proven by example. And I understand that we want to accept these types of papers because obviously we can prove with one example that a method might work perfectly well. And that is an interesting method. But we shouldn't confuse that, that the method works in one well-chosen example very well with the understanding that the method is a good method for my different approach or my different question. In my different question, there might, for example, be the problem of causality. So one of the things I put in or want to put in, I want to take a decision, leads to different conditional distribution than an observation. We had (in this conference) a thing about road traffic here where they learned how much the road deteriorates. And obviously the roads deteriorated most in the areas with the best road type because that was obviously in the areas with the most traffic. So the better road type causes the road to stay longer while you will observe that it deteriorates faster because it's more traffic on it. So if we use something for decisions, it will be different and we need to understand that causality. We also need to understand representation and representativity of data. So we learned with one thing, for example, with our error-free simulation model, and then we want to apply it to reality. It works perfectly on that simulation model. And one of the first applications of AI was, I think, a tank trained on Russian tanks, American tank, to fire on Russian tanks. And it fired on the first American tank it saw simply because it had photos of the Russian tanks which were bad weather photos and good weather photos of American tanks. And then it fired on anything that looked real. So it does things wrong, and we need to understand we are not the experts for the method. We are not the experts for how to develop machine learning, but we need to understand what machine learning does in our reality, in our geoscientific reality. For example, we don't have independent data. And all these things Google does is about independent data. We work with representative data sets. Even the idea of cross-validation, drop 20% of the points, is wrong when you understand that you have spatial correlation in your data set. You remove 20% of that mountain range, and you try to cross-validate that you get that mountain range correctly, which is still an in-sample learning. Just because you took the nearby point doesn't mean it's very different from the one point you have still in the data set. And I have never in that conference or in the last conference gotten that even mentioned that this is a problem in our things. So we need to learn as geoscientists to apply that in the right way, and that is methodological research, how we do apply it. So the first thing, methodological research, and we need to learn how these things apply to our reality. The second thing is I saw a lot of engineering applications where I totally understand, let's do it this way. This gives a better predictor. It shows the real subsurface structure as good as possible. But let us think of that in terms of geosciences. So when we start to infer what the subsurface structure is from machine learning tool based on our simulated data and our simulated subsurface microstructure, have we learned anything about reality? Or have we just created our, reinforced our beliefs in how the subsurface looks like? It's perfect from an engineering point of view. We predict the correct microstructure in 97.5% of all cases perfectly, and in 100% of our simulated examples it works perfectly. But we need to understand that implication of doing so in what it does to people learning from our microstructure models. The same for microstructures. In geometallurgy I often think about the microstructures. So we need to learn how geoscience works in the presence of these very, very awesome tools which we now have. Where suddenly whole branches of science might get obsolete. I don't know. Are we going to do geostatistics in 10 years or are we just having an algorithm that does it for us? And one of these aspects is we learned in geostatistics a lot of things. Like, for example, that the simulation, the quantification of uncertainty is the key thing. For example, for decision making in geometallurgy. For groundwater flow modeling. For nuclear safety. We learned that, but we are now in the machine learning there that we do simulations, but we don't understand, I think we don't understand yet, how well they capture our uncertainty. We have no theory about that. So we need to understand how to use our tools. Not to repeat the errors of geostatistics. For example, that we just predicted the ore grade and then wondered why the production is always below, systematically below what we predicted. Because those blocks that were higher in grade, that's fine. But if they're lower in grade, we have them in the sample. But if we underestimated them, we don't have them in the sample. So just these errors, these small little things, need to be learned again for the new tool of machine learning.

Richard. *Thank you so much. You've really sparked some great suggestions here, and I particularly appreciate the focus on how uncertainty is a major concern for geoscientists. Maybe we have a deeper interest in uncertainty than many other fields. Now, let me bring up a point from this conference. We've been diving into spatial dependency, and Professor Bardossy's insightful talk on copulas highlighted this aspect. It's clear we're trying to appraise spatial uncertainty in various ways – Variograms capture some of it, but it is often more intricate. So, the crux is, how to consider the right level of spatial dependency for a given problem, whether it's of higher or lower order, but most importantly, what's relevant. That brings us to a crucial question: Do we possess the necessary tools for this? How well do we comprehend how the new era of AI handles these intricate aspects? These are the questions that have emerged during the conference, and I'd like to throw this to the panel. And now any of you could answer. What are your thinking about the* *challenge to try to capture more of the uncertainty than we currently do?*

Lingchen. Can I put a little bit of a point on the uncertainty? I think uncertainty quantification is very important on AI methods. From this conference I see many people are working with uncertainty quantification with Bayesian methods. Which I think is something very important. And because you're now no longer just giving a point-wise result, but now you are giving a probability. So if you're training a classifier, but your training samples are biased, not well sampled from different categories, then when you're giving test data that may be similar to some class that was under-sampled, then the output might have a very large standard deviation error. And then you can think, oh, this one is untrustful. Then it gives you some hint that this result is no longer trustful. You might not trust this result. So one of the ideas is Bayesian methods. I think it might be helpful for any geoscience problem that uses AI solutions.

Richard. *Thank you. But then the question is, – in your opinion, do you believe our current Bayesian methods are truly equipped to grasp the spatial dependency that we should be encompassing, rather than the ones which we currently capture? So, maybe, Odd would start.*

Odd. The point is we really don't know the models that we haven't tested yet. As a Bayesian, we really want to test many models, but due to computational limits, we have not been able to do that. What I liked with what Thomas said, is that all of a sudden he has a possibility to test one of these models that previously has been in the dark. We can't tell whether it's important or not before we have tested it. Now we have this opportunity, and we need to investigate does this make an impact or is it just as good or bad as what we had earlier? Like, you know, there's always a problem with the two-point statistics. Multi-points doesn't always look like the training image. There's lots of issues that we have in the methodologies that we are using. Perhaps we now can have that will work a bit better.

Richard. *Thank you for that. It's actually a suggestion for all of you to reflect on – whether you're truly capturing what you're aiming for, and whether the models you're now able to test proves to be important. I'd like to circle back to a point Gerald brought up – the concept of causality. During the war in Norway, the Germans confiscated all radios to prevent exposure to London's propaganda. Post-war, some statisticians discovered a surprisingly strong correlation between improved tooth health in the population from 1940 to 1945 and the lack of radio listeners. It just goes to show that correlations don't necessarily equate to causality, and that's an essential consideration we should always keep in mind. I'm curious to hear from you, Thomas, since you were mentioned. Could you perhaps dig a bit deeper into this puzzle – the dilemma of whether we're actually addressing the problem we ought to be tackling, or if we're merely tackling the issue that appears to be at hand?*

Thomas. I have a comment for something else, maybe not specifically for that, but what those two gentlemen said, you know, we should split it up because in this room we kind of have the knowledge to build proper models. That has nothing to do with machine learning. If we have a very good machine learning method, give it some bad data, it would probably give us bad results. And also, I have no hope to use machine learning on, for example, a Bayesian formulation of a problem and then get better results. That would really make me worried, right? I would like to get exactly the same result, exactly the results that I'm looking for, but that is just not possible today because it's too slow, right? I just want to get it faster and hopefully more precise, but not better.

***Richard**. Shaunna or Jennifer, I'm curious about your thoughts on this. Are the tools we employ finely tuned to match the specific problem? Or are we shaping the problem to fit the tools we've got? Or, on a positive note, are we effectively adapting the tools to suit the problem at hand?*

Jenny. I'm not sure I'm going to answer the question that you posed, but there is a niggling thought in the back of my mind. I'll just say one thing. We used to say the best geologists are the ones who have seen the most rocks. I'm a bit concerned that it's become the best mathematical geoscientists are the ones that have seen the the most models or simulations, but are we always looking for the model that we haven't produced yet. I really hope that for the IAMG conference in 10 years' time, geologists still feel that it's relevant for them to come. And that they will feel that there's something they are hearing that makes sense to them. So the simulations are really great but, I always love the talks where there's a real life case study. Then I think, great now it’s time to test the model and see how it actually acts and works.

***Richard.** Thanks a lot, and now we're ready for some questions from the audience. We'll have to stick to one question per person, just to keep things on track with our tight schedule. So, who's got a burning question or a comment for any of our panelists? It's a first-come, first-served basis.*

First speaker from the floor (Mathieu Gravey?). We hear that a lot of mistakes were done when people are doing some tests and whatever, and totally agree. And other fields have the same problem. For example, in remote sensing, they turn machine learning people to do like splitting randomly as well. It doesn't make sense, no spatial variability. And on the other side, we hear like we should collaborate more with people who are experts in this field. And my feeling is I'm sorry, no, we need to learn their field. In fact, we need to double check everything they do, and they need to double check everything we do. In fact, we need to do the job twice. We cannot collaborate with them. We need to do exactly the same both. It doesn't work. And my opinion, my question is, what's the opinion of the others? Can we really collaborate with experts on another field? Or should we just locally merge our fields together to make a temporary new one?

***Richard.** Thank you. Any other comments?*

Second speaker from the floor (N.N.). Coming back to the question about can we capture the dependence in the phenomenon we are talking about. Well, today we got two experts. Today, Professor Bardossy showed us the copulas. And from the, sorry, from the copulas, where he separates marginals from dependency. I asked him, can I test it? And he said, well, I am doing my crazy code, and I still cannot. There is some code over there. And then we hear Professor Emery, Javier Emery. He gave us a recipe of what he called a pure and heavy geostatistics. I guess everyone hears his presentation. He gave us a whole recipe of different type of covariances that he was very proud telling me that he fits. He fits his simulation. We're fitting very well. But also we got Michael. He gave us a data, a machine learning techniques that could collaborate and could provide tools that allow us to do that much easier. I don't know. That's what I think.

Richard. Any other comments?

Third speaker from the floor (Lukas Mosser). Really interesting points that you all made. I have a comment to Gerald. You spoke about the need to understand ML techniques. And we need to inquire and check if they really work - we have the tools to do that. And then you also mentioned that people sort of do this random splitting, and it's not valid. I think a lot of that has to do with education. You have a lot of tutorials on the web. People are getting into these fields with no background knowledge on the assumptions. They don't understand what is really underlying the problems and methods that people are making. And I would like to make two points: One is, do you think we have to essentially understand the machine learning domain so well that we are able to identify where they fail? Which means, are we becoming machine learning researchers ourselves in that process? And the other is, where do we start in education? What can we do to make this better, the required knowledge more widespread? I think it's essential that we do that.

Gerald. When I teach statistics, I teach such things about the classical methods. I don't expect my geoscience students to become experts in statistical method development. But I expect them to become experts in using the methods. And I think this is what we need to reach. And when it comes to teaching, yeah, we need to introduce that obviously in the teaching at the university. But first, we need to understand it. And I'm not yet there that I understand it. So we need, as a society, to grow into that and understand these things. How to use machine learning in our field.

Richard: We have one comment here. Please, start.

Fourth speaker from the floor (NN). Yeah, I would like to follow up on Lukas's comment. Because I think in his talk today, he actually was one of the few attempts to put together a holistic sort of framework. And this is what ideally would be bringing together the core geostats where we are coming from with what we are borrowing from other sciences. And this is very important. But I want to take it even further to where I think it's very important to understand that machine learning... I'm very grateful to see so many applications of data science. And I've been jumping from sessions to sessions because I couldn't cover all the data science of my interest while my talk in 1999 was odd one out. So that's great that the community has accepted the value of data sciences here. What is... That's my subjective opinion. It's not there yet. I think there's a lot of view of data submission learning AI as a silver bullet. And I know I'm guilty as many that we all like to show good results where it works. But I think it's equally important to show where it breaks. And that's I think there's not enough in our conference to people coming and saying, look, this is where the limitation is. And I think, again, computer scientists, with all due respect, I don't know how many computer scientists are here or data scientists. Not many, right. I think it's still responsibility of the geoscientists, which are the majority here, to show where things break. And things may break because they're abused. And this is, again, we need to do the homework and apply them properly. But that's, again, their textbooks. And IAMG as an educational organization can do lots more. And we've got distinguished lecturers here and keynotes. They're doing great stuff. But also any method has a limitation. So I would like to see more things where there's a limit. And here it comes. Okay, again, I believe, and I've seen examples of this, and I hope to see more, that actually some keys for solutions to improve the computer science methods actually sits with the domain context. So to apply, so we can basically improve the way some top computer science algorithms, which we do not develop, we just borrow. But we can adapt them in a way that they perform better to our great problems, which are of interest to us, but also beyond the area, which Shonna quite rightly said.

Richard. *Thank you. I think we have to move on now to the next speaker.*

Fifth speaker from the floor (Vera Pawlowsky-Glahn). I just wanted to continue with one of the comments from the third speaker from the floor. The importance of the underlying assumptions. When I hear artificial intelligence and data-driven approaches, I get scared, because I wonder if really the underlying assumptions are present in the mind of those who implement the methods, the algorithms whatsoever. I mean, many here know that I work in compositional data analysis. And one thing that I have learned is that you have to look at the sample space. You have to define clearly the difference between real random variables and, for instance, compositional variables. They are not the same. It's not that they are not real numbers. But they don't cover the real space, and they have a different structure. So this is something that I doubt that is easy to incorporate into the models, even if it's easy to implement. But to check which is the right sample space to work in, this is, from my point of view, something that is essential.

Richard. *Thanks, Vera. I find that point particularly crucial, given your extensive involvement in IAMG over the years. You were among those who organized the compositional data workshop back in 1999 here in Trondheim. So, big thanks to you..*

Sixth speaker from the floor (Anirudh Prabhu). I thank you so much for all your introductory statements. And I find myself agreeing with almost everything that you said. A theme that emerges for me is, you know, we each spend our lives getting expertise on so many things and gaining expertise. And, you know, Shaunna mentioned collaboration. Jennifer mentioned collaboration. And I completely agree. But what's emerging in this discussion in this room for me was a bit, you know, like, we're all scientists. We've all gained expertise on something. But yet there's this distrust. I don't trust this algorithm. I don't trust this result, which I agree. There has to be a certain kind of skepticism that we're looking for. And, you know, I think we should all check each other. But I think the educational aspect there is not necessarily to learn every other person's domain because where do we stop there? Rather, can we learn how to collaborate? The educational aspect of interdisciplinary collaboration is what will make us work together, leverage each other's expertise, and ask the right questions. Annoy your collaborator until you are completely clear that everything is working fine from both parties. But that seems like an easier, more pragmatic approach in next steps than saying that, well, I'm a geochemist or I'm a geophysicist, I'm a space scientist, and I'm going to learn everything there is to know about machine learning. Or I'm a data scientist, I'm a computer scientist, and I need to know everything there is to know about every domain I apply myself in. So it just seems like we have, and this group is perfect. We are mathematical geoscientists. We have our foot enough in both doors that we can be the spokespeople for people to collaborate better because we've been doing this for a long time. So that's what I wanted to say. Thank you.

Richard. *Thanks. I found the suggestion of geoscientists as go-betweens really intriguing – the notion that in order to collaborate effectively, you need to be able to communicate and understand each other's domain language. It's all about being clear on the actual problem at hand, without any communication barriers holding you back. This reminds me of a young man who tried to delicately express to his girlfriend that their relationship might be going through a rough patch. He chose his words carefully, mentioning that their connection seemed a bit strained lately. Her response? "Oh, you mean we've been hitting the gym too hard." So, with that tale in mind, I'd like to invite the panelists to wrap things up with some closing thoughts on where we stand and how we've experienced this panel debate. Let's kick off with Gerald, just to keep the panelists interventions symmetrical.*

Gerald. I would like to comment on Matthieu's suggestion that we need to develop that feel, the new science of being both geoscientist and a mathematician. And I think that's what the IAMG is about and has always been.

Sixth speaker from the floor (Anirudh Prabhu). It's more, I would say, about AI than mathematics.

Gerald. Yeah, but that extends, obviously, to mathematics now.

Odd: Some closing remarks, where are we in like 10 years? Are we all in machine learning or are some of us still discussing variogram ranges? I think we always will have a need for the simple models because they are simple, understandable, but I think we will also see very good examples where we have machine learning that has helped us with problems we previously couldn't solve. I have a great belief in this, but we need to acknowledge the special features of spatial statistics that are not in like image processing. Because images, there's 256 colors, three channels, our problems are different, An example in image analysis it's the contrast in the images that's important. In many geo problems, the absolute values are important. So there's real difference in what is important in the problems we are trying to solve. We need to have focus on the important aspects in our problems.

Thomas: Maybe we don't need to be domain experts in all field, but we need to communicate. We have a language to talk to each other, that's statistics. So, we have a formulation, Bayesian statistics, probabilistic modeling, where we can work together without being expert in each other's fields. And to you, Mathieu, can we actually work together? Yes! As when I work with geologists, and set up a model. The moment I show a realization to people who know something about that model, they will have an opinion. So it's actually super easy. And I think in this geostatistical community, we have those tools that allow us to express visually what it is that we assume. So, I'm pretty sure we can do that. And in 10 years' time, we'll still be doing that, I think, with a core in Bayesian statistics. But we'll just use machine learning to achieve the goals.

Lingchen. I think I would like to give some comments about education and collaboration. And especially to Lukas' comments about education from online resources and how do we learn about them. But the people who go to learn from resources may just, because they are unaware of the knowledge, they may accept everything. But they've just ignored the traditional assumption on the geological information processing. So, yeah, this is something that it will be greater for this entire community, including us, to publish more tutorials online and more online courses to the people who are in both industries. Not only in the geoprocessing, but also for students who have the love for the AI domain. And it will be even better for those more broadly and general tutorial online. It will be even better because I have so many experiences on MOOCs, on these online courses. I know their intention was to educate people, but normal people, not for the specialized people on the specialized industries. So it will be even greater to have these resources published online on YouTube, on TikTok, on different online sources. And speaking of the collaboration between different disciplines, I totally agree that we need to learn from each other. I am a data scientist. My background is about signal processing, image processing, right beginning later to the acoustic data processing. And then I go to the geological facies modeling information, which is like still the 256 images, but with more than three channels like Thomas had mentioned. So we are learning every day. And I'm learning every day from my colleagues about what they are doing on their domain. And I'm translating myself on the background to know what this is really about. Is this really a still image of a pixel or any uncertainty was involved over there? And so I have to understand their domain and start to think about it with my own way. Meanwhile, I try to explain what the data science is, what this model is doing, and how do we tune the loss function with long-term physically is about explaining to them. And then we finally reach a consensus. So this can benefit us a lot about that. So I think also another idea would be about so it's for the industry to embrace more about data science. Maybe it's more proposals to submit to the big companies, big IT companies with strong GPU computational resources to make them feel interested and satisfied with our work. And they will sponsor us even more about giving us resources on the GPUs, on the computational cloud nodes, so then we can accelerate our research even faster. So then because our interest is no less interesting than those commercial interesting problems like the TikTok recognition system or the other. We have our own interesting problems and also with a very large, vast amount of data that's already there. So as long as this big industry can help us more sponsoring our resources for academy, not only academy but also the industry as well, so we can definitely have better improvement in the following years.

Shaunna. Thanks to my fellow panelists for what they said because I think they've covered a lot of what I would like to say. So, I'll actually speak to a little bit of the future. You're asking about what things are going to be like in 10 years and we were talking about the difficulty of the common language between different fields. Helping people understand the geologic jargon is always going to be a problem because unfortunately that's not a huge part of most education systems (although, I can really only speak to the American system). We're always going to have to teach folks what we mean when we use rock names and mineral names and all of these many specific geologic terms. However, I see that computer science language, which of course is underlain by statistics, is making its way into primary school vernacular and into the vernacular of children now. It's really bringing math back into something that is being talked about in a way in which it's cool and it's exciting. I think what we're going to find is, whereas I had to learn a lot of these terms older as an adult in an effort to try and improve my science and ask the questions I want to ask, children are going to know these terms and methods. Our young college students already have that language to be able to have those conversations. And so in the future we're going to be able to talk to each other much more easily than we do now. There's going to be a less steep learning curve for you to have the language needed in order to understand what your colleagues are doing, which is critical to any method. You have to be able to understand it and interpret it and recognize when there is a problem with it. So, I think in the future it's going to be easier than it is today.

Jenny. Well, I think to look forward, I'll reflect on why I became part of the mathematical geoscience community. As a geoscientist, as a geologist, there were things that I could not address or solve by myself, geological problems. There are still geological problems, they may have shifted in terms of focus but geology is key here. Understanding geoscience is key to all of the issues that we face. And therefore, mathematical geoscience, I believe, is even more relevant to addressing current global issues which involve geoscience. I hope that AI and machine learning remain part of the mathematical geoscience toolkit. If we reflect in 10 years' time on Marshall's poster approach to observing the ebbs and flows in mathematical geoscience, that AI and ML will be a consistently used approach along with geostatistics and and other mathematical geoscience approaches. For example the use of GIS has become commonplace, so much so that ewe don't even mention how it is used sometimes in presentations. I suppose the final thing I would say is, as a teacher, I know that if I have to explain something to a student or to a colleague, I get a better understanding myself by doing that. So I think when we have to explain what we think our problems or the issues are that we're facing, whether it's to a geoscientist,mathematician,computer scientist or statistician, actually it helps us all to understand the issue and hopefully find a solution. If we are honest about the limitations and explains it in a way that it's not ‘a black box’, then I think that this is key. I'm not going to become a mathematician, or an AI expert, but I can help you, as you explain to me how that approach might actually address the geological problem that I'm trying to solve. So I think the outlook is very good for us. I'm delighted that there are just so many people here from various expertise backgrounds and that everyone feels welcome at this conference. I hope that this continues to be the case.

Richard. *Well, a big thank you indeed. Let's give the panelists a round of applause. Their insights have certainly given us food for thought.*
I'd like to take this moment to announce something exciting. The panelists have met with Roussos Dimitrakopoulos, Editor-in-Chief of Mathematical Geosciences, and

we're gearing up for a special issue on the integration of AI with geostatistics – a topic we've dived into here. Russos has extended an invitation to those interested to submit a paper for this special issue. The papers will go through peer review, and upon acceptance, they'll be available online right away while being prepped for the print volume. As for the timeline, the paper submission deadline is set for January 31st, and we're looking at having the issue in print by Easter, adhering to standard procedures. So, watch out for a call for papers post-conference. My co-chairs for this initiative, Jo Eidsvik and Michael J. Pyrcz, have already agreed to help realize this endeavor with additional help from Steinar Ellefmo. Our hope is that this special issue will gather papers that steer us towards the direction we've been talking about – how geostatistics in the age of AI can hold onto its essence while gaining more prowess. In essence, we're aiming for a leap forward, even a quantum leap. So, a huge thanks to our panelists for sharing their invaluable insights, and to the audience for joining in on this engaging discussion.

With that, let's consider this panel officially closed.

.