Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-61-AC2, 2018

Geoscientific Model Development Discussions

Interactive comment on "GemPy 1.0: open-source stochastic geological modeling and inversion" *by* Miguel de la Varga et al.

Miguel de la Varga et al.

miguel.de.la.varga@rwth-aachen.de

Received and published: 6 July 2018

Answer Review 2:

Thank you very much for the detailed suggestions and comments and questions, as well as the positive feedback. After incorporating your input, the paper has become much more polished and easier to understand.

Despite agreeing with most of the recommendations about expanding the explanations about the probabilistic graphical models and fault networks, we are the opinion that

the extent of the paper is already too large to incorporate this information. Therefore sections that may be self contained in further publications—i.e. Bayesian network, fault network—or that have been already published (although sadly quite sparsely) have been reworked but not extended. Hopefully all of this will come together in the first's author PhD thesis.

Regarding the Bayesian model, the scope of this particular paper was on the generative model of the Bayesian inference—i.e. the mathematical formulation that connects the model parameters with different datasets. We have improved the clarity of our description and added the probabilistic graphical model into the main text.

The definition of faults drift functions are, again, too broad to be treated in an already too large paper. So far the drift function is manually defined and in the authors view the best results would be yielded by optimization. We are not aware of work done in this direction and we have just started to explore this option. Regarding their definition in the GemPy code, we have added some further explanation in the manuscript but, in short, it is done by categorizing the input data. Although in the paper we aim to give a flavour of the use of GemPy, for a good understanding of how to create models, we recommend the online documentation and tutorials (an endless work-in-progress). We will try to provide an example of fault networks sooner rather than later.

Finally, we took the decision of writing this paper as a compendium of the necessary algebra with the correspondent open-source code to generate implicit geological models. We are aware that a more extensive explanation of the kriging system and the Appendix C would be useful for the community but our mathematical knowledge is limited and we are not convinced of being able to add much value to the already published work in this matter. Therefore we have opted for a more descriptive approach referencing the original sources.

Once again, we highly appreciate all comments and we accepted most suggestion not only for this paper but also for future work.

RESPONSES

Answer to pdf comments:

[73] reword

Authors response: reworded.

Change in manuscript: *The method was first introduce by \ cite{ Lajaunie.1997} and it is grounded on the mathematical principles of Universal CoKriging*.

[102] I would also say to provide an environment/ecosystem to improve/enhance/add existing methodologies.

Authors response: Accepted suggestion

Change in manuscript: *. . . , but to provide an environment to enhance existing methodologies as well as give access to an advanced modeling algorithm for scientific experiments in the field of geomodeling.*

[135] I wouldn't refer to this as anisotropy. More precisely, it describes the planar

orientation of the stratigraphic structure throughout the volume.

Authors response: Accepted suggestion. We agree, it is less confusing now

Change in manuscript: *The gradient of the scalar field will follow the planar orientation of the stratigraphic structure throughout the volume or, in other words. . .*

[143] reword

Authors response: Reword:

Change in manuscript: *...as it is not possible to obtain remotely accurate estimations*

[168] why 0 subscript? shouldn't the function 's variable be a general point x?

Authors response: Further explanation and added citation. x_0 is the terminology used by Chiles on his book Geostatistcs Modeling Spatial Uncertainty and the sub index 0 is used to differentiate the interpolated points x_0 to the input data x_alpha.

Change in manuscript: *where $ { \ bf{ x} } _{ 0} $ refers to the estimated quantity for some integrable measure $ p_0$ . . . . . we will try to be especially verbose regarding the mathematical terminology based primarily on \ cite{ Chiles.2004}*

[173] ?; unclear (and can lead to confusion) and not needed.

Authors response: Reword but not deleted. Despite we agree with the reviewer that the parameter p may lead to confusion we have decided to keep the original sentence for two reason:

1. Keep the nomenclature consistent with Chiles description
2. To point out that the method is multivariate and there is no limitations to extend the current mathematical formulation

[184] keep the terminology consistent, change to scalar field

Authors response: Potential field naming fixed, p substituted by rho and u defined as any unit vector. Everything here was carelessness from our side. Thank you to the reviewer for pointing them out.

[188] unclear

Authors response: Reword.

Change in manuscript: *Note that in this context the scalar field property* $ \ *alpha*$ *is dimensionless. The only limitation is that the value must increase in the direction of the gradient which in turn describes the stratigraphic deposition.*

[195] Mention here that the choice of the reference point does not matter - choosing different reference points has no effect on the results.

Authors response: Added suggestion about reference point.

Change in manuscript: *It is important to mention that the choice of the reference points* $ { \ *bf*{ *x*} } _{ \ *alpha*_\ *, 0*} ^ *k*$ *has no effect on the results.*

[200] most of this is unclear

Authors response: Improved explanation

Change in manuscript: *The advantage of this mathematical construction is that by not fixing the values of each interface* $ { { \ *bf*{ *Z*} } *(* \ *bf*{ *x*} } _{ \ *alpha*} ^ *k)*$ *, the compression of layers—i.e. the rate of change of the scalar field—will be only defined by the gradients* \ *(*\ *partial* { \ *bf*{ *Z*} } */* \ *partial u*\ *). This allows to propagate the effect of each gradient beyond the surrounding interfaces creating smoother formations.*

[209] define after this equation.

Authors response: Extended eq 3 definition

[210] Although this system is indeed correct. I am concerned that for most readers this system will be confusing. I would strongly suggest that appendix section C2 be expanded to explain in detail the meaning of this system and its clear consequences. Then from that derive separate linear systems for both the scalar field and the gradient of the scalar field. In addition, refer in lines 210-215 to that appendix section for full details.

Authors response: After long consideration. We have decided to leave it as system for two reason: (i) a proper explanation of a CoKriging system would need to be too long and we may not have the mathematical background to explain error-free, and (ii) the matricial form is more consistent—at least to us—with Chiles' book explanation.

[217] on the right you have a 3x2 matrix

Authors response: Fixed terminology

[220] this description is insufficient. e.g. f10 = f1(x0) and you never defined what x0 is. same goes for f20. In connection with my comment about eqn 4 you should show or say why f10 reduces to zero in the appendix. It would tremendously help new researchers fully understand the mathematics for potential enhancements - just as you have done for most of the kriging system.

Authors response: Added suggestions

[300] How do you ensure that the modelled conformable stratigraphic layers respect the specified sequence for the stratigraphic pile? e.g. the interpolation constraints for interface pts (the increment points) is that the scalar field at these interface pts is the same - the is no method of inputting the sequence via the increments since their particular value has no effect. This problem usually only presents itself when interface data is characterize by strong horizontally sampling bias.

Authors response: Added explanation. If we are not mistaken the reviewer refers here to the order of the formations within each series. Indeed this only depends on the

geometry and now we added a sentence clarifying this point

Change in manuscript: *It is interesting to point out that the the sequential pile only controls the order of each individual series. Within each series, the stratigraphic sequence is strictly determined by the geometry and the interpolation algorithm.*

[345] why is there no input data for the fault and the unconformity?

Added explanation. The separation between the different series/faults is done by labeling the input data. As the paper only contains a small snippet of the code this was indeed misleading. We increased the comments on the listing hoping to clarify this point.

[370] spelling

Authors response: Corrected

[376] which rules?

Reworded. We are aware that the fault section is a bit too vague but again the paper is already too long and the faulting algorithms may fit in future work.

[379] how does one choose the perfect drift function? trial and error?

Authors response: Essentially yes, i.e. optimization. Nobody so far has came out

with a better idea. Probably our framework would be the ideal place to test some of this since with AD optimizations work better.

[391-392] ?; unclear explicitly state what you mean by input parameters

Authors response: Reworded

Change in manuscript: *variables*

[401] explicitly define graph in the current context

Authors response: Reworded and added explanation of symbolic graph

Change in manuscript: *Writing symbolically requires a priori declaration of all algebra, from variables which will behave as latent parameters—i.e. the parameters we try to tune for optimization or*

uncertainty quantification—to all involved constants and the specific mathematical functions that relates them. This statements generate a so called graph that encapsulates symbolically all the logic what enables to perform further analysis on the logic itself (e.g. differentiation or optimization).

[455] colloquial. please reword. can a method have intuitions?

Authors response: Reworded

Change in manuscript: *There is extensive literature explaining in detail the method and its related intuitions since it*

[498] these have not yet been defined yet - as it relates to implicit modelling

Authors response: Added reference to chapter

[FIGURE 6] Very confusing figure

Authors response: Reworked in 3D to improve clarity

[575] what is this variable?

Authors response: Added exaplanation

Change in manuscript: $ *t_z*$ *—i.e. the distance dependent side of Equation* \ *ref*{ *eq:grav_0*} *—*

[643] how is this done?

Authors response: Added explanation to paragraph

Change in manuscript: *We multiply the binary fault array (0 for foot wall, 1 for hanging wall) with the maximum lithology value incremented by one. We then add it to the lithology array to make sure that layers that are in contact across faults are assigned a*

unique integer in the resulting array.

[720] Is there a best practices guide for PGM construction?

Authors response: Added some references. Sadly making Bayesian models nowadays is pretty much an art. The variability on topics datasets and complexity makes very difficult to give a close set of rules to construct the models. Probably the best reference would be Bayesian methods for hackers but seems to generalistic for the paper scope. Our past work and but in especial our future work will try to address this problem in a comprehensive manner since we agree that there is a lack guidelines especially in geological modeling

Relevant citations with bibkeys:

- { *bishop2013model*} *Bishop, C. M. (2013). Model-based machine learning. Phil. Trans. R. Soc. A, 371(1984), 20120222.*
- { *sucar2015probabilistic*} *Sucar, L. E. (2015). Probabilistic Graphical Models. Advances in Computer Vision and Pattern Recognition. London: Springer London. doi, 10, 978-1.*
- { *Patil:FseZoIYV*} *Patil, A., Huard, D., & Fonnesbeck, C. J. (2010). PyMC: Bayesian stochastic modelling in Python. Journal of statistical software, 35(4), 1.*
- { *Koller:2009wk*} *Koller, D., & Friedman, N. (2009). Probabilistic graphical models: principles and techniques. MIT press.*

[734] Why just the Z axis? It should be perturbed along each of the 3 axis. Does just choosing the one axis easier for generating results?

Authors response: Added explanation

The choice of perturbing only the Z axis is merely due to computational limitations. Uncertainty tends to be higher in this direction (e.g. wells data or seismic velocity), however there is a lot of room for further research on the definition of prior data—i.e. its choice and probabilistic description—on both directions, to ensure that we properly explore the space of feasible models and to generate a parametric space as close as possible to the posterior.

[744] reword

Authors response: Reworded

[765] what is the meaning of this variable?

Authors response: Removed . This is a pymc 2 dummy value to initialize the object. I just deleted it to avoid confusion.

[FIGURE 8 CAPTION] is this the number of realizations?; says probability of layer 2 in the figure! not layer 1

Authors response: Increased verbosity, fixed typo

[850] should change to p_i represents the probability of a layer at the i-th cell.

Authors response: Fixed typo

[851] compress? you mean quantity?

Authors response: Reworded

Change in manuscript: *we can use information entropy to reduce the dimensionality of probability fields into a single value at each voxel as an indication of uncertainty,*

[856] ? comparison

Authors response: Reworded

[881] what are these parameters? you also used alpha and beta elsewhere for something else

Authors response: Change in manuscript to specify parameters and how the topology is evaluated in the PGM:

Change in manuscript: *To evaluate the likelihood of the simulated model topology we use a factor potential with a half-Cauchy parametrization (shape parameter $ \ alpha=0$ and rate parameter $ \ beta==10^ { -3} $) to constrain our model using the "soft data" of our topological knowledge \ citep{ lauritzen1990independence, jor-*

dan1998learning, christakos2002assimilation} . This specific parametrization was chosen due to empirical evidence from different model runs to allow for effective parameter space exploration in the used MCMC scheme and due to the Cauchy distribution being more robust to outliers than parameter estimates based on, say, Gaussian likelihoods.

[948] reword

Authors response: Reworded

To sample from the posterior we use adaptive Metropolis

[958-961] reword; Also which example ? from Fig 8's posterior models?

Authors response: Specified figure and reworded explanation

[1018] spelling

Authors response: Fixed

[1036] reword

Authors response: Fixed

Change in manuscript: *up to now*

[1037] reword

Authors response: Fixed

Change in manuscript: *able to construct*

[1067] general?

Authors response: Fixed

Change in manuscript: *general*

[FIGURE 9] Why are you presenting this graph? It does not add any insight. Please remove

Authors response: Reworked figures and equations. The reason to add this figures was to clarify the different distances that goes to each covariance functions since for us were one of the main challenges. However we agree that the implementation was not very clear. Hopefully after the rework the use of the figures is more obvious

[1159] please define hx, hy. shouldn't this be r = sqrt (hxˆ 2 + hyˆ 2 + hzˆ 2) with hx = xi - xj, hy = yi - yj, hz = zi - zj

Authors response: Added suggestion and reworded.

Change in manuscript:

\ *begin*{ *equation*}

r = \ *sqrt*{ *hˆ 2_x+hˆ 2_y+hˆ 2_z*}

\ *end*{ *equation*}

and $ *h_u*$ *as the distance* $ *u_i - u_j*$ *in the given direction (usually Cartesian directions). Therefore, since we aim to derive* $ *C_*{ *Z*} *(r)*$ *respect an arbitrary direction* $ *u*$ *we must apply the* \ *textit*{ *directional derivative*} *rules as follows:*

[1164] repeat wrt lhs

Authors response: Fixed

[1175] reword; its related to the smoothness of a function

Authors response: Reworded and further explanation

Change in manuscript: *This derivation is independent to the covariance function of choice. However, some covariances may lead to mathematical indeterminations if they are not sufficiently differentiable.*

[FIGURE 10] Why are you presenting this graph? It does not add any insight. what co the different colored dots, lines, and arrows represent?

Authors response: same as Figure 9

[1186] keep terminology consistent throughout manuscript. change to scalar

Fixed

[FIGURE 11] Again, its unclear what the purpose of these graphs are trying to convey.

Authors response: same as figure 9

[1194] this eqn is incorrect. you have to be very careful with the notation. to properly show this you will have to restructure this.

Authors response: Fixed. Thank you for realizing. This was an important mistake.

[1195] ???

Authors response: Reworded.

Change in manuscript: *As the interfaces are relative to a reference point per later*$ { \ *bf*{ *x*} } _{ \ *alpha*_\ *, 0*} ^ *k*$
the value of the covariance function will be the difference between this point and the rest on the same layer:

[1196] formatting problem

Authors response: Fixed formatting

[1214] formatting errors. ; fii should be beta I assume

Authors response: Fixed formatting

[1220] Is there another reference for this? Matheron, 1981 is cryptic and written in a time (type writers) where there are no matrices in their explicit form. This reference is what is given in Lajaunie, but if there is a better reference can you add that?

Authors response: Added Chiles and Delfiner 2009 book reference

[1235] ??; these functions also have limits - especially so in scenarios where data sample highly variant geological structures.

[1238] before you indicated that C_0 is the value for the nugget effect.

Authors response: Fixed

[1245-46] grammar; not clear how one inputs the nugget effect given D1

Authors response: Added short explanation about nugget effect.

Change in manuscript: *The implementation of nugget effects in covariance matrices are done by adding the value to the diagonal*.

[FIGURE 13] Explain in more depth this figure. what are the "modifiers", why are there

11 of them?

Authors response: Whole PGM has been reworked manually and added to the main text.

Please also note the supplement to this comment:
https://www.geosci-model-dev-discuss.net/gmd-2018-61/gmd-2018-61-AC2-supplement.pdf

---

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-61, 2018.