Review of the Phd thesis of Filip Seitl

Modeling and statistics of random tessellations with applications to the study of microstructure of polycrystallic materials

The main aim of this PhD thesis is to model 3D microstructure using the Laguerre tessellation, in order to obtain realistic realizations of microstructures without expansive data collection. To fulfil this task, the author uses two different modelling approaches. The first, described in Chapter 3, relies on modelling the full model, i.e. tesseletion represented by 3D marked point pattern, via Gibbs-Laguerre tessellation. The author is not satisfied by simulations from this model and therefore he concentrates further on reconstruction methods based on the tessellation characteristics moments or the whole histograms of the tessellation characteristics.
The second approach, described in Chapter 4, uses the hierarchical model, where the points are modelled by multiscale point process and subsequently the marks conditioned on points positions are modelled by an exponential density, again based on certain tesseletation characteristics. Author here also studies the dependence between points and marks of the tessellation representation.
Chapter 3 also contains theoretical result for the existence of Gibbs-Laguerre models.
All the methodology is illustrated on two real data polycrystallic materials introduced in the first chapter. Finally, Chapter 2 contains the detailed theory of point processes, marked point processes, tessellation, simulation algorithms and statistical methods which is used throughout the work.
I would like to highlight especially the hierarchical method for its practicality. In fact, it enables to select a model for the microstructure data from the simplest to the very complex. Using the subsequent testing it enables to select the least complex model which pass the goodness-of-fit testing procedure.
The topic of the work is current. The work is very well written and organized. It is easy to read and well described. The formal side of the PhD thesis is also good. The work is based on 4 papers published in journals Kybernetika, Methodology and Computing in Applied Probability, Australian & New Zealand Journal of Statistics and Spatial Statistics. The author is either the first author or he had a non-negligible role in their writing. Also the codes for estimation and simulation of both approaches were made publically available through C++ package, available on github.
All together the PhD thesis fulfils the requirements and it demonstrates the author's prerequisites for independent creative work, therefore I definitely recommend it for the defense.

In the following I would like to mention several comments or questions, which are not meant as critique of the work but rather the points for discussion at the defense or further work.

1) Page 33: Are there any possibilities for periodic boundary condition for irregular windows? Or do we have to stick on block windows?
2) Page 72: Since the parameters α, β, B are estimated from the finite population, the unbiased estimates of these parameters from the finite population should be used.
3) The reconstruction methods were preferred in front of the simulations from the fitted model. In reconstruction one specifies a set of realizations with respect to the chosen characteristic which is used in the energy. Apparently, the other characteristics, which are not used in the energy, differ from the target data characteristic. So, the reconstructed data differs from the target data in everything what is not contained in the energy. Thus, what is the reason to prefer reconstruction with respect to the simulation from the fitted model? Table 3.8 specifies that only the characteristic z (connected with number of

cells) differs (not significantly) from the target data when simulating from the fitted model. I do not see, why the MPLE is so bad.

4) Page 86: The reconstruction algorithms are based on the parameters $\theta_{1,}\theta_n$ which specifies the importance of certain energy. These parameters are set for reconstruction by the user. Could this be estimated from data?
5) Page 90: The authors does not specify, for which purpose these two simulations are made. Why it is important to simulate tesselletation with extreme number of faces?
6) Page 96: How did the author determined what is too smooth and what is too wiggly, when the truth is not known?
7) Page 98: If we delete the points, with small mark, because they are too close to a point with big mark, then we create a certain dependence among the close marks artificially. If they are close to each other, they are forced to be small. If these points were not deleted, would they be independent?
8) The author shows the dependence between marks for close points. This rejects the independent marking model, but it does not reject geostatistical marking model. Is there a test which would be able to test geostatistical marking assumption?
9) Page 100: The differences between characteristics computed under marks permutations and data themselves are not significant. So nothing is proven by Table 4.1
10) Page 101: When the composite hypothesis is tested and the adjusted test is too computational it is recommended to apply principally different summaries for estimation and testing. (The composite test can be even liberal, not only conservative.)
11) Page 105: It is not a confidence region but rather the critical range.
12) Quantification of usefulness of all methods for generating the new data would be rather interesting. E.g. one can use a certain characteristic for estimation or reconstruction and then study the distance of the generated data and target data with respect to a statistic of interest which is not included in any energy.
13) Page 110: Why it was chosen to use the more complex model when the simpler model with beta and dvol also passed the goodness-of-fit test?

doc. RNDr. Tomáš Mrkvička, Ph.D.

In České Budějovice, 26. 10. 2022