The paper by Smith et al. argues that, for a given airborne lidar dataset, there exists an optimal resolution which minimizes the impact of both gridding errors and any uncertainty in the DEM in the calculation of topographic metrics.

I think this paper should eventually be published but I have serious concerns about the analysis that I urge the authors to consider prior to publication.

1) I am concerned with how the authors created their DEMs. When I create a DEM of a mathematical function I sample the function at regularly spaced points. The resulting DEM is an incomplete representation of the surface but it is an accurate representation where the function is sampled. When I compute slope and aspect using DEMs created this way there are errors associated with discretization, but they are small and converge to zero as the pixel size becomes small.

As such, I was surprised by magnitude and types of errors computed by Smith et al. for their synthetic cases. When I looked more closely, I saw the reason for this discrepancy. If I am interpreting the code https://github.com/UP-RS-ESP/TopoMetricUncertainty/blob/master/gaussian_hill_example.py correctly, Smith et al. have generated their DEMs of Gaussian hills by randomly sampling a Gaussian function and then computing the mean elevation of those random samples within each domain. None of this is explained in the manuscript in the section on synthetic data analysis. Since no information is given, I have no idea how many random samples were used to compute the mean value, or why the authors choose to use the mean value. When scientists grid data from a point cloud, they generally use Independent Distance Weighting (IDW) because this method weighs measurement points close to the sample location more heavily than points father from the sample. The author's approach is not an interpolation of any kind – it treats measurements points far from the DEM grid point location (x_i, y_j) equal to those close to the grid point location. DEM values are supposed to represent the elevation at each point on the surface. DEMs are never supposed to represent the mean elevation within some square domain. Yet, that is how they have been created in this manuscript and I believe that much of the error that the authors are studying is due to the nonstandard way that they have created their DEMs. To address this issue, I urge the authors to explain how their DEMs are created and use IDW to create the DEMs from the point cloud. I would further urge the authors not to assume a random sample in their synthetic DEMs, since lidar data are not a random sample. Before this error is fixed it is difficult to even fully review the paper. However, I will do the best I can, recognizing that this can only be a preliminary review until the DEMs are properly computed.

2) I am having difficulty understanding the error equations. In eqns 4 and 5 there is a partial derivative ∂ with another partial derivative $\partial z/\partial x$ as a subscript. I have never seen this syntax before. What does it mean? I am also confused by the reference to epsilons as uncertainties. In the paper, the epsilon values used to compute TE are computed using the standard deviations within each pixel, which is not the same as uncertainty. The uncertainty of a mean value can be quantified using a standard deviation, but only after being divided by the square root of the number of samples used to compute the mean. I am similarly confused by the use of standard deviations without any scaling by the number of samples in the PEU calculations.

3) One of the metrics used by the authors, the truncation error is, according to the authors, "uncertainty associated with the representation of a continuous surface as a grid." However, since landscapes have roughness at all scales (i.e., they are not differentiable and, more broadly, any increase in DEM resolution almost always results in additional real features being resolved in the topography), it is not necessarily the case that a polynomial is a better approximation of the surface than a straight line, as implied by the truncation error and the associated assumption that minimizing TE leads to a better result. I can see how TE would be a useful measure if topography was smooth at small scales, but I don't think this is supported

by observations of actual topography. To address this issue, the authors could explore and defend their choice of TE in landscapes with microtopography present (i.e., nearly all landscapes) or they could perform their analysis without using TE.

4) I don't understand Figure 4. Part B illustrates conceptually how aspect values are pushed away from and towards certain angles. That is not what part A shows. Part A shows that the probability density of 91 degrees is anomalously high and that of 89 degrees is anomalously low. The same bias towards larger values just above angles that are multiples of 45 degrees applies to all other values. I don't understand how this bias occurs but it is certainly not the result of a tendency of the algorithm to result in higher values at angles that are multiples of 45 degrees, as implied by part B.

5) The most common method for determining the appropriate scale for computing slopes and curvatures that reflect landscape-scale attributes is to plot curvature as a function of scale and identify the scaling break following Roering et al., 2010, Evidence for biotic controls on topography and soil production. I think this alternative should be mentioned. At present the paper isn't referenced.