

ASSESSING THE SIMILARITIES OF 3D SIMULATION MODEL OUTCOMES

Alex K. Smith, Suzana Dragičević

Spatial Analysis and Modeling Laboratory, Department of Geography, Simon Fraser University (SFU), Burnaby, BC, Canada

Commission IV, WG IV/8

KEY WORDS: 3D Kappa, 3D Accuracy, 3D Agent-Based Model, 3D Simulation Outcome Comparisons, 3D Map Comparison

ABSTRACT:

The recent advancement of simulation modeling to represent phenomena in three spatial dimensions (3D) requires the development of techniques that will allow comparison of the modeling outputs in multiple dimensions. However, many existing techniques for map comparison in two spatial dimensions (2D) have been developed from non-spatial method such Cohen's Kappa. These techniques are not yet fully extended to deal with 3D map data or simulation outcomes. Therefore, the main objective of this study is to investigate the use of the 3D Accuracy and 3D Cohen's Kappa coefficients to compare simulation model outputs in 3D. An existing agent-based model (ABM) of forest-fire smoke propagation was used to generate multiple scenarios for the purpose of comparing 3D simulation outputs. The results for 3D Accuracy and 3D Cohen's Kappa produces meaningful values when comparing several scenarios with different 3D ABM outputs. This study emphasizes the need for the development of more advanced simulation output comparison techniques that operate in 3D and potentially over time (4D).

1. INTRODUCTION

The multidimensional characteristics of geospatial phenomena and the increase in three-dimensional (3D) and four-dimensional (4D) simulation modeling (Eshraghi et al. 2012; Gobron et al. 2011; Jjumba and Dragičević 2015; Narteau et al. 2009) indicates the necessity to develop improved methods for simulation model output comparisons. With spatial analysis and modeling approaches advancing from two-dimensional (2D) into 3D space, and even 4D with space-time considerations, the methods required for map comparisons must evolve with this advance into 3D and 4D. However, non-spatial techniques from concepts independent of the spatial dimensions can be applied to spatial scenarios including 3D and 4D. Potentially, the most common map comparison method is the simple overall map accuracy where two maps are compared cell-by-cell to calculate the percent agreement (Van Genderen et al. 1978). Accuracy does not account for chance agreement in its calculation where the two maps being compared have some degree of similarity expected due to the randomness of cell values. Cohen's Kappa coefficient (Cohen 1960, 1968), or simply Kappa, was developed to account for the chance agreement between two datasets although it was originally developed for the field of psychology. Previous to the Kappa, the mathematically identical Heidke Skill Score was developed for meteorology (Heidke 1926). However, Kappa is the commonly used name in Remote Sensing and GIScience studies when dealing with image or map comparisons. Since its development, Kappa has been widely adopted for use in remote sensing (Congalton and Mead 1986; Hudson 1987) and raster GIS for the purpose of comparison of raster-based images using paired cells, also known as pixels or raster, from two images or GIS data layers representing maps. The Kappa coefficient has been further enhanced to multiple variants such as Fleiss' Kappa (Fleiss 1971) which has seen use in the comparison of more than two maps (Rogers et al. 2014), Fuzzy Kappa which introduces spatial fuzzy logic into the calculation of Kappa (Hagen 2003), and many others specifically for land-use change comparisons including Kappa Histogram, Kappa Location, Kappa Simulation, Kappa Transition, and Kappa Transition Location (van Vliet et al. 2011). The hierarchy of the various Kappa variants are shown in

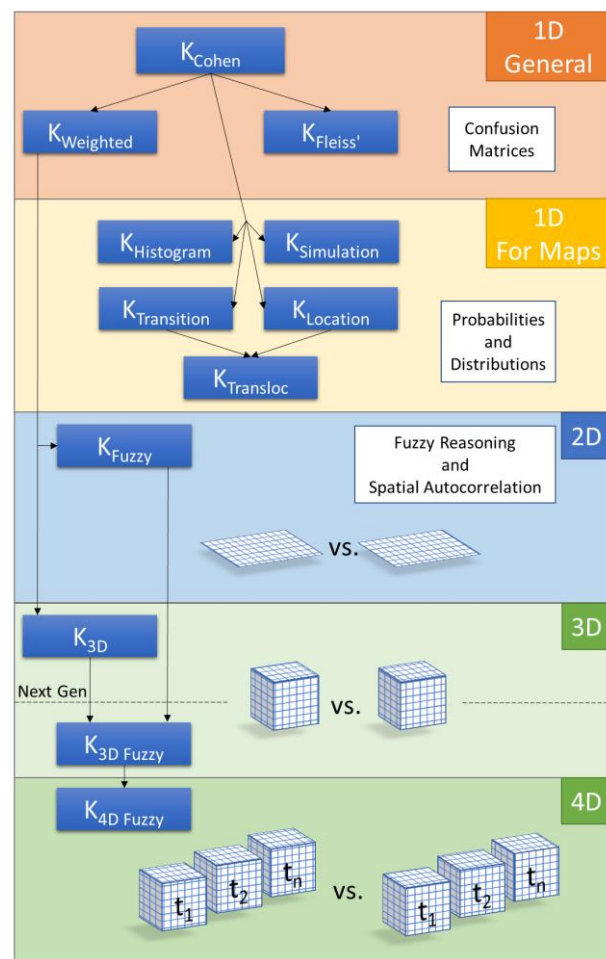


Figure 1. Hierarchy of Kappa approaches originating from Cohen's Kappa

Figure 1 and includes new generations of metrics, 3D Kappa, 3D Fuzzy Kappa and 4D Fuzzy Kappa. The techniques used by 2D map comparison methods such as Fuzzy Kappa are incompatible with 3D applications and therefore require major redevelopment before they can be used for comparisons of simulation outcomes in multiple spatial dimensions.

Both accuracy and Kappa measures are used for the comparison of 3D maps, typically generated from LiDAR data (Roberts et al. 2019; Wang et al. 2018) however were not as yet applied for the needs of geosimulation modeling. Therefore, the main objective of this study is to compare the similarity of 3D simulation modeling outcomes using 3D Accuracy and 3D Kappa, and then to theoretically extend the analysis towards possible spatial extension of Kappa to 3D and 4D metrics. A previously developed 4D agent-based model (ABM) for forest fire smoke propagation (Smith and Dragičević 2019) was used to generate the simulation outcomes used as a case study for comparison of 3D simulation outcomes using 3D Accuracy and 3D Kappa measures.

2. METHODOLOGY

Similar to the resel or raster concept in 2D GIS (Tobler 1995), voxels are the smallest representation of 3D space, divided into a regular 3D lattice, typically forming cubes (Greene 1989; Jjumba and Dragičević 2016). Similar to raster GIS data layer, representing array of equal-size uniform square cells, a voxel data layer represents the equal-size cubes. Each voxel contains a value, such as a class, that represents an attribute characterizing the 3D space that the voxel occupies, empty voxels must be represented by null values. In this study the 3D Kappa approach (Smith and Dragičević in press) has been used to compare simulation outcomes in 3D. The 3D Accuracy and 3D Kappa measures are designed for use with voxel representation of model simulation outputs. The 3D Accuracy is calculated by comparing two voxel datasets voxel-by-voxel and calculating the proportion of agreement between the pairs of voxels. The name 3D Accuracy is given to distinguish it from the cell-by-cell methods used by Accuracy in 2D map comparison. Cohen's Kappa is extended to be applied to the 3D voxel datasets through 3D Kappa (K_{3D}) and can be defined by the following equation:

$$K_{3D} = \frac{P_o - P_e}{1 - P_e} \quad (1)$$

where P_o represents the observed proportion of agreement, or observed similarity, among voxels in two 3D maps or simulation outcomes S , and P_e represents the expected proportion of agreement, or expected similarity, if voxel classes were assigned randomly to the voxels of both datasets. P_o is equal to the 3D Accuracy of the two voxel datasets. The expected similarity P_e is defined by:

$$P_e = \sum_{i=0}^n P_{iA} * P_{iB} \quad (2)$$

where P_{iA} and P_{iB} represent the proportion of class i in simulation outcome S_A and outcome S_B respectively, for all n classes in the two 3D maps or simulation outcomes. To implement both 3D Accuracy and 3D Kappa for the 3D modeling outputs, a custom program in Python 2.7 was developed. This program accepts two voxel datasets converted into 3D arrays of the same extent. It synchronously iterates over each entry of the arrays, recording relevant information,

including the counts of each class per array, the number of matches, and the total number of voxels with data. Equations (1) and (2) are calculated and 3D Kappa is returned alongside P_o as 3D Accuracy.

This research study uses an existing ABM to simulate the propagation of forest fire smoke (Smith and Dragičević 2019) based on a hypothetical fire. The ABM operates with two primary agent types - fire agents and smoke agents. The fire agents represent fire locations, and they emit smoke agents at a predetermined rate that represent smoke in the atmosphere. Smoke agents move using two methods, the first being carried by the wind, and the second through diffusion to increase dispersion. The local terrain acts as a lower limit for the elevation of smoke agents as they travel over, around, and through mountains and valleys. The simulation outcomes are voxelized, creating voxels with values representing the number of agents they contain, thus the concentration of smoke.

3. IMPLEMENTATION

The study area is located in British Columbia (BC), Canada, where fire and smoke are represented as agents as they move through the mountains using wind patterns and terrain. The 3D spatial extent of the study area is 96km x 45km x 10km with a southwest corner located at 70 Mile House, BC and a northeast corner of approximately 5km northeast of Mahood Lake, where the Murtle River joins the Clearwater River. The study area is approximately 100km north of Kamloops, BC. The simulation uses a hypothetical forest fire started on August 10th, 2017 on the west coast of BC and burns for a total of 15 days. The model simulation outputs are obtained in voxel data format with the locations of smoke agents and at a voxel of resolution of 1km. The ABM simulates the concentration of smoke agents that are represented with five voxel classes. Voxel classes were determined by the number of smoke agents in each voxel based on the information from the Government of BC Air Quality Health Index (Government of British Columbia 2020). Class 0 represents no smoke when voxels contain no smoke agents. Class 1 represents low smoke with 1-3 smoke agents, Class 2 represents moderate smoke with 4-6 smoke agents, Class 3 represents high smoke with 7-10 smoke agents, and Class 4 represents very high smoke with more than 10 smoke agents in the voxel.

Three sets of comparisons were made between different output times and settings of the agent-based model. The first set of comparisons is composed of four simulation outcomes to allow for six comparisons (Figure 2). The study area contains smoke on August 23rd at 3pm, 13 days and 15 hours into the simulation. The model is run twice to retrieve two simulation outcomes, Simulation 1 and Simulation 2. The model also has an option for the forest fire to move under the influence of the wind and is used to retrieve Simulation 3. A second output time was selected for Simulation 4 on the 7th day, or 0:00am on August 17th, 2017 when smoke was also found to be present in the study area for Simulation 4. Comparison 1 is between Simulation 1 and Simulation 2 of the static fire scenarios, where both simulations use the same initial conditions but different automatically generated random seeds that define the set of random values used. This comparison is used to assess the differences caused by the randomness in the model. The randomness originates from two sources, one is the random diffusion of smoke and the other is due to the random order of the agents' movement. Comparison 2 is between Simulation 1

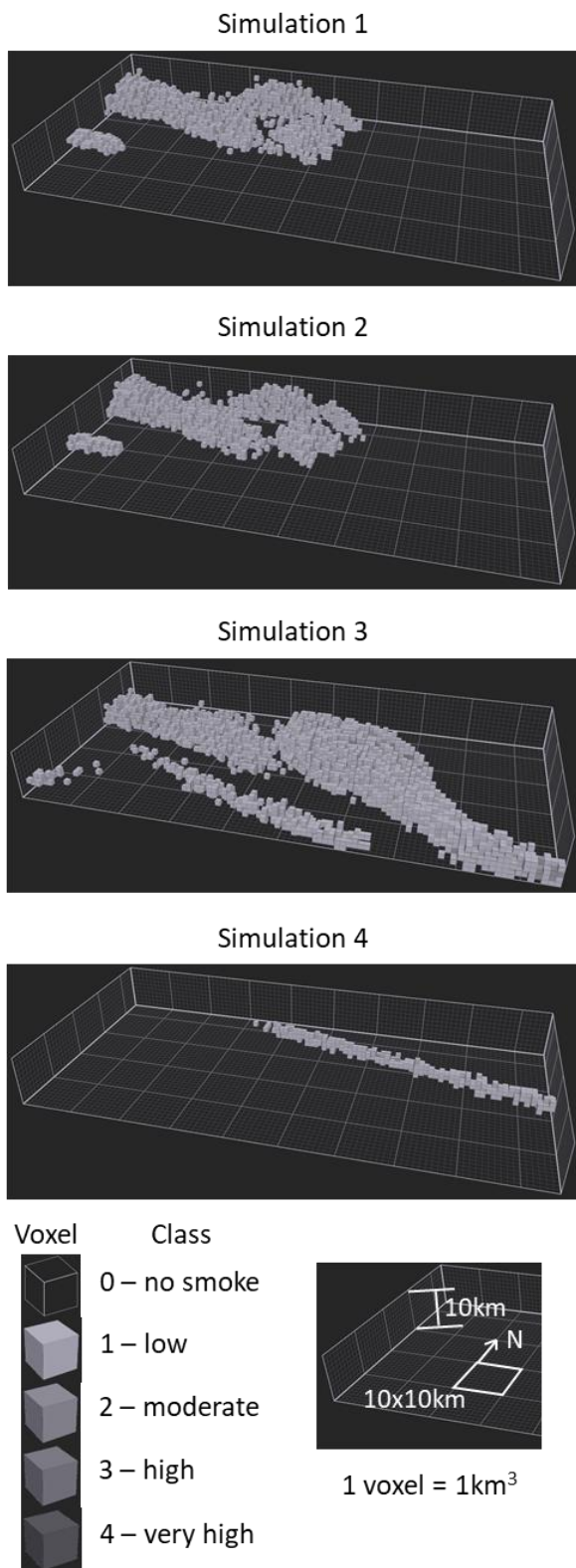


Figure 2. Agent-based model forest-fire smoke propagation simulation outcomes in voxel data format used for six comparisons.

and Simulation 3 using the scenario with a fire that moves over space-time through the landscape of the study area. Comparison 3 is between Simulation 1 and Simulation 4 which is obtained from the alternate time step assuming the 3D terrain is identical, but the smoke will have experienced different wind conditions. Comparisons 4, 5, and 6 are between all other possible combinations of these four simulation outcomes.

Smoke is present in the study area for an additional 9 hours after the output time used for Simulation 1, and a total of eight ABM simulation outputs have been generated from the two runs and presented on Figure 3. The four outputs of one run of the model are compared against the outputs of the second run, where the first outputs are Simulation 1 and Simulation 2 from the first set of comparisons. The 16 comparisons between these two sets of four simulation outputs make up the second set of comparisons. These comparisons are done to investigate a use case of 3D Kappa for comparing simulations to temporally match patterns from two simulations.

The third and final set of simulation outcome comparisons is performed to explore the sensitivity analysis of the diffusion rate of the smoke of the agent-based model. The distance used by the diffusion process is multiplied by the diffusion factor (DF), where $DF = 2$ is the most similar to the calibrated DF of the model. Six levels of diffusion are used and results in $DF1$, $DF4$, $DF8$, $DF12$, $DF16$, and $DF20$. This analysis will help assess the impact of the diffusion factor, for the selected simulation runs. Due to the required computer processing time of the model only one run for each diffusion factor was completed.

4. RESULTS AND DISCUSSION

The obtained results in Table 1 show the 3D Accuracy and 3D Kappa values for the first set of comparisons. Due to the large number of voxels with no smoke in the outputs, the accuracy typically receives high values. The only comparison to receive a moderate or better 3D Kappa value is Comparison 1, between the two runs with the same initial conditions. This comparison was expected to receive much higher 3D Kappa values due to the very similar appearance of the smoke patterns. However, the 3D Accuracy of this comparison is only slightly higher than Comparisons 3 and 5 which obtained the lowest 3D Kappa values. These two Comparisons involve Simulation 4 which along with Comparison 6 achieve the lowest three 3D Kappa values and with exception of Comparison 1, the highest 3D Accuracy values. This is due to the large number of voxels with no smoke in Simulation 4, more than the other simulations.

Comparison	Simulations		3D Accuracy	3D Kappa
C1	S1	S2	0.96	0.53
C2	S1	S3	0.87	0.15
C3	S1	S4	0.95	0.01
C4	S2	S3	0.89	0.15
C5	S2	S4	0.95	0.00
C6	S3	S4	0.90	0.03

Table 1. Values obtained for 3D Accuracy and 3D Kappa for the six comparisons of the forest fire smoke propagation ABM simulation outputs.

The randomness incorporated in the ABM creates the differences between Simulation 1 and Simulation 2 presenting itself partially in the form of randomly distributed voxels within

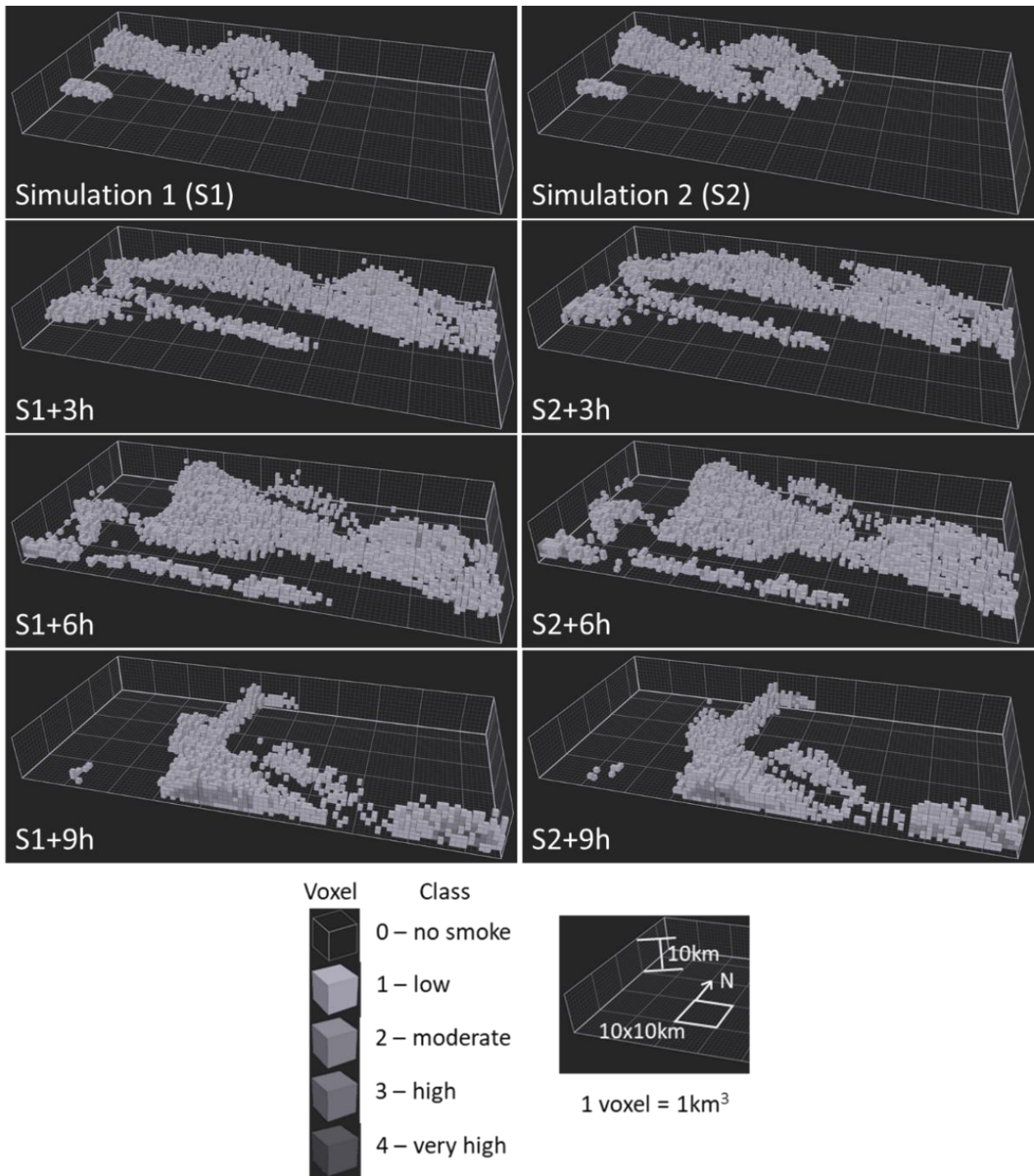


Figure 3. Simulation outputs S1 and S2 at initial stage and after three hours, six hours, and nine hours of smoke propagation.

the 3D space (Figure 4). While this can make two 3D maps appear visually similar, few of these voxels with smoke in one map agree with their counterpart in the other. This leads to noticeably lower 3D Kappa values than expected and slightly lower 3D Accuracy values. In 2D spatial analysis techniques, a moving window is often used to incorporate nearby cells into the analysis of the central cell. Situations like this could be more accurately assessed if this type of moving window analysis could be developed into a more advanced 3D map comparison method. Using a moving 3D window, nearby voxels could be used to partially substitute the agreement between two disagreeing voxels. A similar technique with the moving window was adopted for the calculation of Fuzzy Kappa (Hagen 2003) in 2D and can allow for further expansion into 3D Fuzzy Kappa to accommodate the benefits of the use of 3D window space.

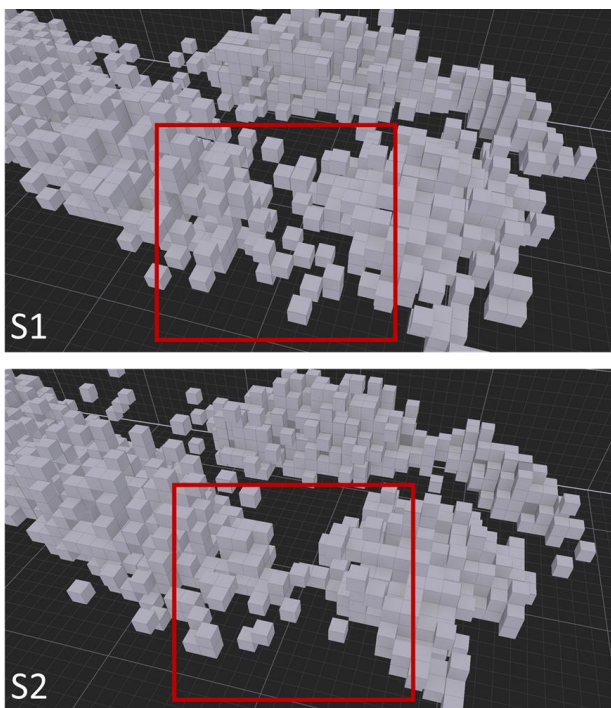


Figure 4. Randomly distributed smoke voxels highlighted in red from ABM outputs for simulation S1 and S2

The values for 3D Accuracy and 3D Kappa for the second set of comparisons are presented in Table 2. Similar to the first set of comparisons, accuracy is very high, with the lowest value of 0.86. The accuracy values of the comparisons of outputs of the same time step are typically the highest in their row and column, although by small differences. However, 3D Kappa produces greater differences between the comparisons of outputs of the same time step and all other comparisons. Neither 3D Accuracy nor 3D Kappa include any time component that can incorporate differences in time to the comparison of data.

Table 3 presents the values obtained for 3D Kappa for the third set of comparisons. These last comparisons are accomplished to investigate the effect of the rate of the smoke diffusion on the simulation outputs generated by the ABM. Similar to the other results, the 3D Accuracy values are very high, but were also very consistent with all values being 0.95-0.96. While the 3D Kappa values are also somewhat similar, they show a clearer trend. The larger the difference in *DF*, the lower the 3D Kappa

values. Additionally, comparing simulation outputs with higher *DF* values is also associated with a lower 3D Kappa. This is most likely caused by the increased diffusion rate increasing the amount of randomness in smoke location in the model. Because 3D Kappa has no spatial fuzziness unlike Fuzzy Kappa, it cannot account for the increasing amounts of randomness in smoke locations with higher diffusion factors. This indicates that more advanced 3D Kappa metrics have to be developed such as 3D and 4D Fuzzy Kappa to incorporate a variable 3D window size and 3D spatial autocorrelation to accommodate the variation of simulation outcomes in 3D and 4D space-time.

3D Accuracy				
Time Steps	S2	S2+3h	S2+6h	S2+9h
S1	0.96	0.91	0.86	0.91
S1+3h	0.92	0.94	0.87	0.90
S1+6h	0.88	0.87	0.92	0.88
S1+9h	0.91	0.88	0.86	0.96
3D Kappa				
Time Steps	S2	S2+3h	S2+6h	S2+9h
S1	0.53	0.17	0.08	0.03
S1+3h	0.17	0.49	0.10	0.01
S1+6h	0.09	0.11	0.54	0.08
S1+9h	0.02	0.00	0.08	0.60

Table 2. Values obtained for 3D Accuracy and 3D Kappa for the 3D comparisons of four forest-fire smoke model simulation outputs for two runs S1 and S2.

Simulation Outcomes	DF1	DF4	DF8	DF12	DF16	DF20
DF1	1.00	0.49	0.40	0.37	0.33	0.28
DF4		1.00	0.46	0.42	0.39	0.35
DF8			1.00	0.43	0.40	0.37
DF12				1.00	0.39	0.36
DF16					1.00	0.34
DF20						1.00

Table 3. Values obtained for 3D Kappa for the comparisons of the forest-fire smoke propagation model simulation outputs with varying diffusion rates (*DF*)

5. CONCLUSIONS

This research study extends two existing methods of Accuracy and Kappa measures into 3D by proposing 3D Accuracy and 3D Kappa for comparing model simulation outcomes in 3D space. Critics of Kappa typically have concerns with how chance is accounted for (Foody 2020; Pontius and Millones 2011), however for use in assessing models such as the one used in this study that have built-in randomness, this concern may be reduced. Kappa is also an often-required metric in many fields, such as remote sensing and GIS, and is applied despite the criticisms. While in this research study 3D Accuracy and 3D Kappa metrics show potential for use in comparison of model outputs dominated by a single voxel class, it also revealed a greater potential to expand these methods to account for spatial neighborhoods or spatial autocorrelation in 3D domain. The existing 2D Fuzzy Kappa method can be extended into 3D and even 4D Fuzzy Kappa counterparts to advance map and simulation outcome comparison into 3D and 4D while incorporating fuzzy logic (Smith and Dragičević in press). The proposed 3D Accuracy and 3D Kappa measures may be the first step towards enabling more comprehensive comparison methods of simulation outcomes and generated patterns in 3D.

While designed for 3D simulation outcomes and 3D map comparisons, such measures may also be useful to other fields, including 3D medical imaging, and in the training and comparison of predicted 3D AI-system behaviour with the real-world data. Based on findings from this research study, further advance methods of 3D simulation outcomes and 3D pattern comparisons should be developed that facilitate the calibration and validation of simulation models such as 4D ABM or voxel automata operating in the 4D space-time domain.

ACKNOWLEDGEMENTS

This study was fully funded by a Natural Sciences and Engineering Research Council (NSERC) Discovery Grant awarded to the second author. The authors are thankful for the feedback provided by two anonymous reviewers.

REFERENCES

- Cohen, J. (1960). "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement*, 20(1), 37–46.
- Cohen, J. (1968). "Weighted Kappa: Nominal scale agreement provision for scaled disagreement or partial credit." *Psychological Bulletin*, 70(4), 213–220.
- Congalton, R. G., and Mead, R. A. (1986). "A Review of Three Discrete Multivariate Analysis Techniques Used in Assessing the Accuracy of Remotely Sensed Data from Error Matrices." *IEEE Transactions on Geoscience and Remote Sensing*, GE-24(1), 169–174.
- Eshraghi, M., Felicelli, S. D., and Jelinek, B. (2012). "Three dimensional simulation of solutal dendrite growth using lattice Boltzmann and cellular automaton methods." *Journal of Crystal Growth*, 354(1), 129–134.
- Fleiss, J. L. (1971). "Measuring Nominal Scale Agreement Among Many Raters." *Psychological Bulletin*, 76(5), 378–382.
- Foody, G. M. (2020). "Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification." *Remote Sensing of Environment*, 239, 111630.
- Van Genderen, J. L., Lock, B. F., and Vass, P. A. (1978). "Remote Sensing: Statistical Testing of Thematic Map Accuracy." *Remote Sensing of Environment*, 7(1), 3–14.
- Gobron, S., Çöltekin, A., Bonafos, H., and Thalmann, D. (2011). "GPGPU computation and visualization of three-dimensional cellular automata." *Visual Computer*, 27(1), 67–81.
- Government of British Columbia. (2020). "BC Air Quality." <<http://www.env.gov.bc.ca/epd/bcairquality/data/aqhi-table.html>> (Jan. 31, 2020).
- Greene, N. (1989). "Voxel space automata: Modeling with stochastic growth processes in voxel space." *Computer Graphics*, 23(3), 175–184.
- Hagen, A. (2003). "Fuzzy set approach to assessing similarity of categorical maps." *International Journal of Geographical Information Science*, 17(3), 235–249.
- Heidke, P. (1926). "Berechnung des Erfolges und der Gute der Windstarkevorhersagen im Sturmwarnungsdienst." *Geografiska Annaler*, 8, 301–349.
- Hudson, W. D. (1987). "Evaluation of several classification schemes for mapping forest cover types in Michigan." *International Journal of Remote Sensing*, 8(12), 1785–1796.
- Jjumba, A., and Dragičević, S. (2015). "Integrating GIS-Based Geo-Atom Theory and Voxel Automata to Simulate the Dispersal of Airborne Pollutants." *Transactions in GIS*, 19(4), 582–603.
- Jjumba, A., and Dragičević, S. (2016). "Spatial indices for measuring three-dimensional patterns in a voxel-based space." *Journal of Geographical Systems*, 18(3), 183–204.
- Narteau, C., Zhang, D., Rozier, O., and Claudin, P. (2009). "Setting the length and time scales of a cellular automaton dune model from the analysis of superimposed bed forms." *Journal of Geophysical Research: Earth Surface*, 114(3), 1–18.
- Pontius, R. G., and Millones, M. (2011). "Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment." *International Journal of Remote Sensing*, 32(15), 4407–4429.
- Roberts, K. C., Lindsay, J. B., and Berg, A. A. (2019). "An analysis of ground-point classifiers for terrestrial LiDAR." *Remote Sensing*, 11(16).
- Rogers, D. J., Suk, J. E., and Semenza, J. C. (2014). "Using global maps to predict the risk of dengue in Europe." *Acta Tropica*, 129(1), 1–14.
- Smith, A. K., and Dragičević, S. (in press). "Map Comparison Methods for Three-Dimensional Space and Time Voxel Data." *Geographical Analysis*.
- Smith, A. K., and Dragičević, S. (2019). "A four-dimensional agent-based model: A case study of forest-fire smoke propagation." *Transactions in GIS*, 23(3), 417–434.
- Tobler, W. (1995). "The resel-based GIS." *International Journal of Geographical Information Systems*, 9(1), 95–100.
- van Vliet, J., Bregt, A. K., and Hagen-Zanker, A. (2011). "Revisiting Kappa to account for change in the accuracy assessment of land-use change models." *Ecological Modelling*, 222(8), 1367–1375.
- Wang, J., Lindenbergh, R., and Menenti, M. (2018). "Scalable individual tree delineation in 3D point clouds." *Photogrammetric Record*, 33(163), 315–340.