

Interactive comment on “TopoSUB: a tool for efficient large area numerical modelling in complex topography at sub-grid scales” by J. Fiddes and S. Gruber

J. Fiddes and S. Gruber

joel.fiddes@geo.uzh.ch

Received and published: 17 August 2012

AUTHORS RESPONSE TO REVIEW 1 BY ETHAN GUTMANN

we would like to thank Ethan Gutmann for his constructive and encouraging comments. We have split reviewer comments (RC) where we have found this useful to do so. Authors comments are given by AC. Original, added or removed manuscript text is given by MT.

RC1: Model description: There is only a limited description of the LSM they are using (GEOtop), while they do supply a reference, a paragraph or two describing the model would be appropriate. e.g. Does it use a Temperature index snow model or a more
C570

physics based model? Does it have an explicit vegetation canopy? How thick is the soil and how many layers in the model, do the two covary?

AC1: Model description has been expanded and given its own subsection 4.2, which now reads as follows:

MT: 4.2 Land surface model. We employ the open-source LSM GEOtop \citep{Endrizzi2010,Dall'Amico2011, rignon06} which is a physically based model that simulates the coupled energy and water balance with phase change in soil, a multi-layer physically-based snow-pack model and surface energy fluxes in 1D and distributed 2D modes. It has been designed specifically for application in mountain regions. The model domain consists of a soil column of user-specified depth (typically of a few meters) from the ground surface, which is, in turn, defined by a Digital Elevation Model (DEM). The heat and subsurface water flow equations are then solved with finite differences schemes.

The multi-layer snow pack scheme accommodates compaction as well as water percolation and refreezing. The influence of topography on micro-climatology is parameterized, allowing for the solution of the surface energy balance for differing topographic situations based on one driving climate time series (Endrizzi and Marsh, 2010, Liston and Elder, 2006). A vegetation canopy was not considered in these experiments. The soil is uniform over the entire simulation domain, parameterised using the Van-Genuchten model (1981), and has 5 layers and total depth of 3.1m. The model is run on an hourly timestep. We apply two years of spin up and then generate 1 year of data.

RC2: A little more discussion of how the "informed clustering" is performed would be helpful. Do they run the training routine for multiple realizations of the K-means clustering algorithm or just run it once for training purposes?

AC2: Informed sampling and its training routine is simply an initial simulation where all input predictors are scaled equally in the clustering algorithm, i.e. the assumption under equal scaling is that all predictors are of equal importance to the simulated target

variable(s). This we call simple scaling (of predictors). In a second step, we perform a regression on each target variable of interest (single target variable is also possible). The resulting regression coefficients are then used to scale each predictor in an 'informed' fashion. This procedure is performed once and the scaling functions can then be used for all subsequent simulations, where predictors used and target variables required, are the same as the training simulation. We have added an equation describing simple scaling (Eq. 3) to accompany existing regression (now Eq. 4), informed scaling (now Eq. 5) and weighted scaling (now Eq. 6) formulae. We have extended the text with an example to further explain the process:

MT: The resulting regression coefficients, β_i provide an informed scaling of the $PRED_i$ for the clustering algorithm by transforming them into equivalents of the TVn dimension and unit (Eq. 5). As an example, elevation (m), slope angle ($^\circ$) would be scaled into equivalents of [$^\circ\text{C}$] if ground temperature was the target variable.

We have also made the distinction between the K-means clustering algorithm and informed clustering clearer (see second paragraph in AC5(b)).

RC3: Forcing data: They state the forcing data come from one met. station. While the quality of that station is not terribly relevant to this paper, it would be useful to know what the climate for this station and surrounding area is like. Mean annual precip, air temperature, peak snow depth/SWE, etc.

AC3: We have added the following text to section 4.1:

MT: Based on 28 years of climate data (1982-2009) from the driving meteorological station, the mean annual temperature is -5.15°C and the mean total annual precipitation is 876mm. The mean annual 0°C isotherm is at approximately 2200m asl. Station data (1985-2010) in the simulation domain from Samedan (1709m asl) and Passo del Bernina (2307m asl) give mean winter (DJFM) snow depths of 43 and 209cm and mean March depths of 36 and 226cm, respectively .

C572

RC4: They allude to the fact that they adjust the forcing data spatially (e.g. temperature follows a mean lapse rate, and SWin varies, presumably purely as a function of the cosine of the angle of incoming solar radiation on topography?). How are these adjustments performed? Is this part of GEOtop or TopoSUB or both? Do they adjust precipitation as a function of elevation? Do they adjust relative humidity to match the air temperature adjustment or does their model use a mixing ratio or specific humidity input?

AC4: Further details have been added to the new Section 4.2 (model description). This is as follows:

MT: The meteorological data, input as point time-series are spatially distributed by GEOtop to each simulation point using principles of the Micromet model (Liston and Elder, 2006) . Specifically: (1) Air temperature follows a mean lapse rate ($6.5^\circ\text{C}/\text{km}$). (2) Since relative humidity is a non-linear function of elevation, the dewpoint temperature is used for vertical extrapolation. (3) Model time step is used to calculate the solar radiation for that specific time. In addition, the influence of cloud cover, direct and diffuse solar radiation, and topographic slope and aspect on incoming solar radiation is accounted for. The distributed version has self and cast shadowing based on DEM, point has self and a uniform horizon elevation. (4) Precipitation is not adjusted.

RC5(a): This is important because it is useful to understand what is driving the improvement in their model. Is improvement caused by a more realistic precipitation field? SW field? temperature field? All of the above? Such discussion of the importance of different variables would be useful as it would help guide future research. In addition, assuming that the fully distributed version of their model is using the same adjustments, then any improvement in their model is dependent on how these adjustments are treated. If, precipitation is not adjusted for elevation in this study, then another study (that did adjust precipitation) might get a different answer (i.e. the number of required iterations/samples/starts might change.).

C573

AC5(a): We think its important to clearly distinguish between (1) the physical attributes (e.g. topography, surface and subsurface properties) of the unit of computation, or sample as we call it in this study, and (2) the methods by which driving meteorology is extrapolated (or adjusted) to a given point or virtual point at the earth's surface where a numerical model may be simulated.

While extrapolation of driving meteorology allow different results in different locations to be simulated and therefore, in one sense the statement: 'the entire reason their model exhibits spatial variability' is valid. However, it is important to remember that it is the physical/ topographical attributes of that location which determine what degree of 'adjustment' is made, e.g. how many vertical metres the lapse rate must be applied. It is perhaps worth emphasising the central argument of this paper, that topography largely drives the variation in climate we see at the earth/atmosphere boundary as well as subsurface (within reasonable latitudinal bounds). Therefore, this manuscript focuses on (1) – that is, how do we make sure that physical attributes of samples best represent (or 'sample' in a true sense) our simulation domain, and therefore enable a result, approximating the quality of a distributed simulations, to be calculated. The response to RC4 gives additional text that has been added to further clarify the partitioning of (1) and (2).

The methods by which the driving meteorology is extrapolated to each sample or grid element (2) is exactly the same in TopoSUB and the baseline simulation and is not the focus of this study, as well established methods exist. While there is little doubt that a combination in the sophistication of the parameterisations used to extrapolate driving meteorology, input predictors used and expected quality of outcome will affect the required number of samples to sufficiently describe (also subjective, how much information is required?) the target variable(s) of interest.

RC5(b): These adjustments to the forcing data are part of if not the entire reason their model exhibits spatial variability, as such understanding how they are adjusted is a critical component of this paper.

C574

AC5(b): We therefore argue that it is not, as stated, crucial to understand the methods by which meteorology is extrapolated from the driving station – as long as the same methods are used in both TopoSUB and the baseline. However, it is crucial to understand on what basis samples are formed, the physical attributes of which, form the basis of meteorology extrapolations.

Finally, we feel it is important to not get distracted by iterations of K-means/ number random starts used (Section 2.1). Section 2.1 serves to acknowledge that these parameters are important and that the K-means method is sensitive to them, but, hopefully this is not confused with the TopoSUB method itself, but seen as a subset of the K-means algorithm which we feel important to address, yet not central to the TopoSUB method or its concept. We have seen that there exists room for confusion in the original text (i.e. 'training' in the K-means algorithm v 'training' in the informed sampling and 'model' referring to TopoSUB method or just the K-means algorithm) and have made the following changes:

MT: For this reason it is highly recommended to run the K-means algorithm with several random starts and average the results. A first run of the K-means algorithm is performed on a sample of input data (105 pixels) with 10 random starts and maximum iterations set to 20 (as previously defined). The cluster centres defined by this clustering of a subset of the dataset are used to initialise K-means for the entire dataset (106 or more pixels). This allows for significant speed up of the algorithm (factor of 10, Table 1) while not compromising on quality of results.

Additionally, it is recognised that a different landscape with, perhaps, less spatially heterogeneity may require less random starts or number of iterations of the K-means algorithm to achieve a result of a given quality level (method of spatialising meteo, as mentioned previously would not affect this). However, this is overprinted by the main message of this paper which is that samples formed by a clustering algorithm (whatever the parameters used in that algorithm) may reasonable represent the heterogeneity of complex topography in an efficient way and additionally provide a means (based on

C575

memberships) to spatialise results to achieve 'distributed' maps of target variables.

RC7: Test simulations: Were the two years of spin-up run by looping over the single year of forcing data (2009-2010). Do they start on January 1? If so, how do they initialize snow?

AC7: We looped over one year of forcing data. We start on July 1 to allow for snow-pack accumulation. Snow was not initialised above the perennial snow line. As land surface above perennial snow line (assumed to be around 3500m) only represents approximately 2% of the total simulation domain, therefore this is assumed to be a reasonable insignificant simplification. Additionally, glaciated areas were not defined as such (simulations can be initialised as ice in GEOtop). This would simply add a further dimension of variability in terms of surface cover, and can be readily accounted for by TopoSUB.

RC8: Results (and Discussion?) The discussion describing what their results mean is very sparse, they should expand on every sub-section.

AC8: We were perhaps a little hasty in our streamlining efforts and agree with the reviewer. We have expanded each section (diverse and numerous changes in revised manuscript). Specifically, we have expanded the discussion of Figure 10 (see RC9(c)) and have added a discussion of temporal error characteristics, however due to arguments outlined below (RC9(b)) we have kept this brief.

RC9(a): Ideally, I would also like to see some discussion of the temporal error characteristics. I think most of their discussion revolves around the mean errors (presumably over both space and time.)

AC9(a): Errors are calculated as mean annual values and as mean values in space (by definition) in aggregated results. However, spatialised results are evaluated on a pixel-by-pixel basis. We have clarified this as follows in Section 4.4:

MT: TVs are analysed as mean annual values in all cases. Distributed results are eval-

C576

uated on a pixel by pixel basis whereas aggregated results are by definition evaluated as a mean value of the simulation domain.

RC9(b): However, there is no discussion of the temporal errors. I realize their method is primarily focused on spatial characteristics, but it is likely to have implications for the temporal evolution of the model as well. For example, how does their model influence the timing of spring snow melt?

AC9(b): We see the temporal component as an extension of the problem to accommodate more than one target variable. Meaning, that in order to gain results well resolved in time, more clusters would likely be needed. We do however include one example of the temporal characteristics of the mean domain values plotted at 5 day intervals for each target variable (Figure 1). This figure shows that there is no obvious temporal signature to the error of TopoSUB, at least at this temporal resolution. Furthermore, we have included the following section on temporal errors in the results (together with Figure 1):

MT: 5.4 Temporal errors Figure() shows the temporal error signature of the 4 tested target variables as 5-day mean values for the whole simulation domain, as absolute values for BASE, TopoSUB and difference between BASE simulation and TopoSUB. This figure shows that the temporal signature of the error of TopoSUB (at this temporal resolution), is relatively small. Swin and SWE show the most obvious temporal trend with a stronger negative bias in SWin during winter months (approx. 2W/m²) and an increasing positive bias in SWE during the main snow melt months (April-June) up to 7mm. The temporal dimension is an extension of the multivariate problem and clustering may need to be tuned to fit certain seasons in much the same way as simulating different target variables (through informed sampling).

RC9(c): They touch briefly on the spatial error characteristics with figure 10, but here too, more discussion of this figure would be helpful instead of just presenting it. Where/when does the TopoSUB model do better or worse? Why?

C577

AC9(c): We have added the following text to section 5.3:

MT: Figure \ref{} provides a visual comparison of the simulation results for GST presented as deviation from BASE simulation (BASE – TopoSUB) as well as a histogram of error distribution. The spatial forcing of this error was investigated through a regression analysis of difference against PREDs. By restricting the dataset to values > 1 and < -1 deg C we could ensure the signal was not masked by the (vast majority of) low error values. The model explained 47% of variance (increased to 62% by including interacting effects). A relative importance metric derived from decomposed r^2 value (Genizi 1993) gives the percentage of variance explained by model attributable to each PRED, as follows: $\sin(\text{aspect})=38\%$, $\cos(\text{aspect})=24\%$, $\text{slope}=20\%$, $\text{elevation}=14\%$ and $\text{sky view factor}=4\%$. This shows that the spatial component of the error is a reasonably complex result of interactions among the PREDs, with only sky view factor being insignificant. The PRED $\sin(\text{aspect})$ explained most of the observed error (38% of variance explained by the regression model).

RC9(d): Given that the characteristics of e.g. the GST maps are so similar, wouldn't a map of the differences be helpful?

AC9(d): We have exchanged this for a differenced map together with histogram of error distribution (Figure 2).

RC10: section 5.4 Model Stability: The authors state that a significant increase is seen in model stability between 25-100 samples; however, the authors should put these errors in the context. While the model is indeed "more stable" I would suggest that the stability is within reasonable bounds already at 25 samples. At this point the errors are negligible ($\text{GST} < 0.15 \text{ deg C}$, $\text{SWin} < 2 \text{ W/m}^2$, $\text{SWE} < 4 \text{ cm}$, $\text{airT} < 0.05 \text{ deg C}$). One would see an order of magnitude more variation than those numbers using different land surface models, different parameters, or different forcing variables (measured or modelled). This is important because it says that their model only requires 25 samples (and perhaps fewer) to achieve model stability. Although, figures 6 and 8 show that 200

C578

samples are required to achieve minimal errors relative to the BASE case, as a result, with reasonable sample numbers to minimize errors, stability should almost never be an issue, so any work to improve the K-means clustering algorithm will have minimal payback at this point.

AC10: We mean that while a reasonable absolute stability is seen at all resolutions, a relative significant increase in stability is seen at resolutions of 25-100 samples. We have decided that the following sentence is perhaps misleading and have decided to remove for clarity:

MT: A significant increase in stability is seen between 25–100 samples in all variables tested.

The key point of this figure is to demonstrate that as the clustering algorithm is an unsupervised technique and will provide a different result each time it is run, that the results are not significantly sensitive to this variation in cluster formation. We have added the following sentence to Section 5.4:

MT: Results of deviation of each simulation from mean values of mean and quantiles 25/75 of all 40 simulations indicate reasonable stability even at low resolutions as indicated by a low absolute deviation. This demonstrates that the result is not significantly sensitive to variations in the K-means clustering algorithm (which are generally small and diminishing with increasing sample number).

MINOR COMMENTS

RC: p1045 I4: PREDs: The authors might want to add a few examples of their predictor variables here, instead of making the reader look in the table for them. I7: TVs: Same thing here, it would be useful to see what they are predicting, along with a description of how it is predicted, e.g. via the LSM or via a direct topographic interpolation (e.g. air temperature lapse rate adjustment and solar radiation cosine(?) adjustment.)

AC: Done this.

C579

RC: Throughout the paper they should select one naming system and stick with it, e.g. is it the "BASE" model, or the "Distributed" model? Is it the "Lumped" model or the "SUB" model? Or is the lumped model the distributed model run at very low resolution?

AC: We have changed the naming convention to BASE and TopoSUB throughout the paper. BASE being the baseline distributed 25m simulation and TopoSUB being the lumped model developed and tested in this paper.

RC: Figure 9: This figure would be easier to interpret if the x and y axis had the same ranges, and/or if a 1:1 line was presented... I'm assuming the line on the graph is a regression line since they list r values (also, I would prefer to see r^2 values as r^2 has a more readily interpretable meaning as the percent of variance explained.) However, because the line is very close to 1:1, I can't tell if there is any bias. Do these data points come from a single time slice, or all are they data points in space and time? (also same comment as above, Is "DIST" the same as "BASE" and "LUMP" the same as "SUB"?).

AC: X and y axis now have the same ranges. It is currently 1:1 line and Pearson product-moment correlation coefficient, r, is the statistic presented (see section 2.3). We have expanded the caption of Figure 9 to make this clear. BASE/TopoSUB changes made throughout paper. The figure presents all datapoints in space and time – by plotting the mean annual values for each pixel (10^6), again caption has been amended to make clear:

MT: Density scatter plot of 1D/2D after informed scaling and fuzzy spatialisation at 258 samples. Data presented is mean annual value for each pixel in the simulation domain. All TVs are reproduced with low error as reported by the correlation coefficient (r) and RMSE (computed over 10^6 pixels). The diagonal line represents $y=x$.

RC: Figures 9,10 need units on their axis / colors.

AC: Units added to axis and legend in Fig. 10.

C580

RC: An additional figure showing a map of some of the clusters would be useful though not necessary. I realize it would be difficult in the fuzzy case. Perhaps maps could be presented for "crisp" membership for some of the key points mentioned in e.g. figure 8 with $n_{\text{samples}}=16, 64, 258$ (and should that be 256?)

AC: We have included a Figure 3 showing a mapping of 128 samples together with a polar plot visualising the sample distribution in terms of predictors. Yes should be 256 – corrected.

FIGURE CAPTIONS

Figure 1: Temporal characteristics of the TopoSUB mean domain values plotted at 5-day intervals for each target variable as deviation from BASE simulation (BASE – TopoSUB). Errors are relatively small in all cases. A stronger negative bias in SWin during winter months (approx. $2W/m^2$) and an increasing positive bias in SWE during the main snow melt months (April-June) up to 7mm.

Figure 2: A visual comparison of the simulation results for GST (128 samples) presented as deviation from BASE simulation (BASE – TopoSUB) together with a histogram of error distribution. Errors statistics show the error to be reasonable: RMSE=0.6, bias=-0.15, standard deviation=0.58. Regression analysis showed that $\sin(\text{aspect})$ explained most of the spatial distribution of error (38% of variance explained by the regression model), although the spatial pattern of error is likely attributable to a complex interaction of PREDS.

Figure 3: Visualisation of 128 samples generated by TopoSUB. (b) Polar plot showing the distribution of samples in terms of predictors used: elevation, aspect, slope and sky view factor.

Interactive comment on Geosci. Model Dev. Discuss., 5, 1041, 2012.

C581

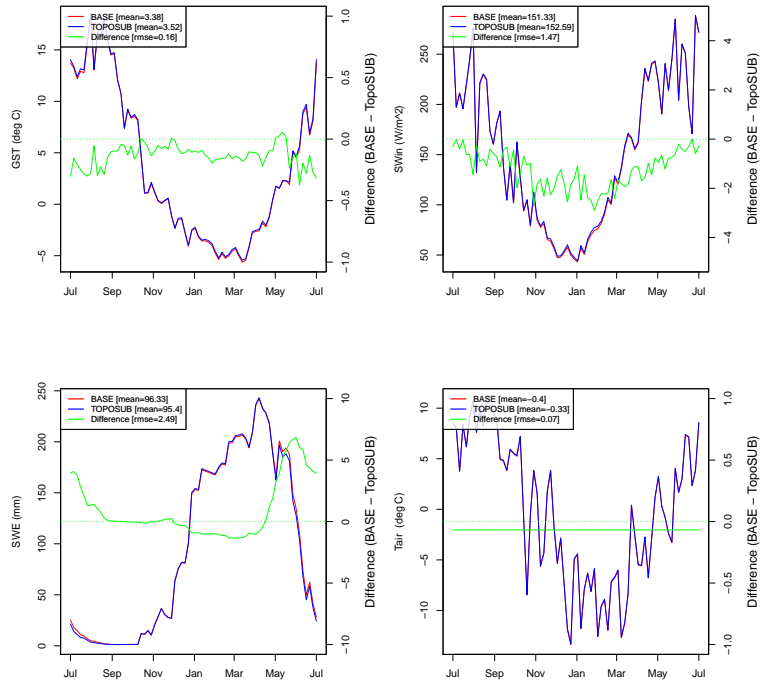


Fig. 1.

C582

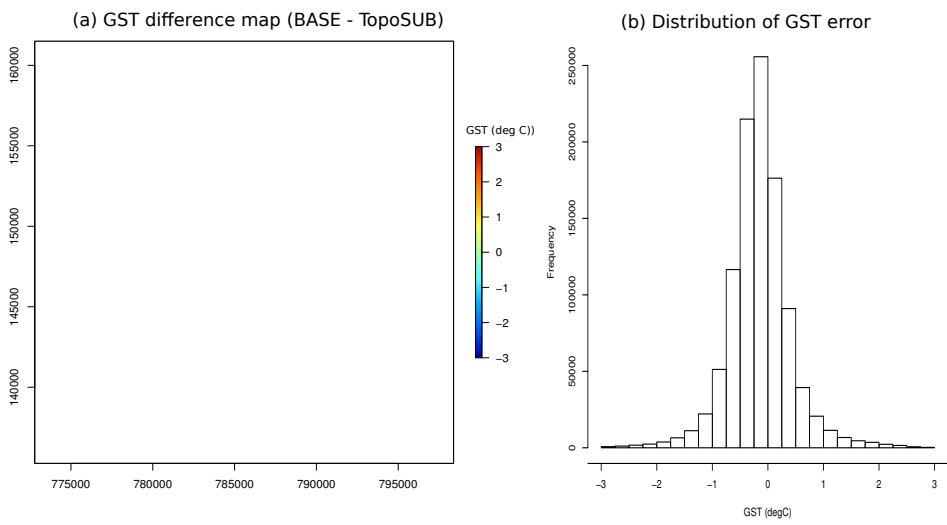


Fig. 2.

C583

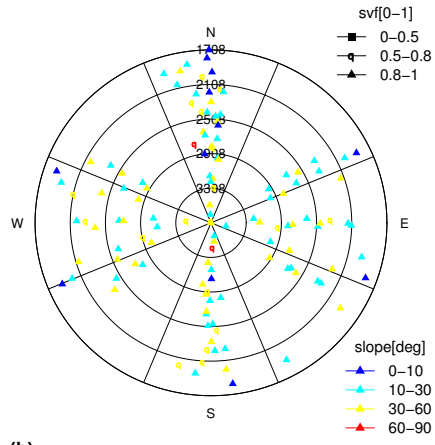
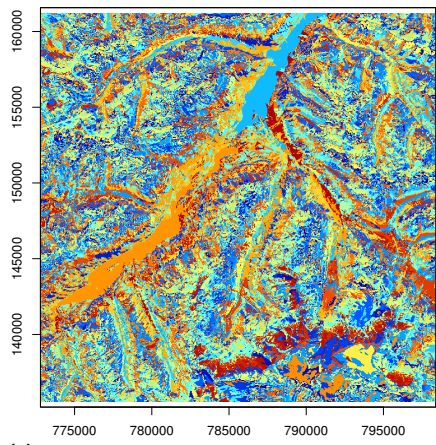


Fig. 3.