Hydrol. Earth Syst. Sci. Discuss., 6, C1262–C1279, 2009
www.hydrol-earth-syst-sci-discuss.net/6/C1262/2009/

Hydrology and Earth System Sciences Discussions

Interactive comment on "A modeling study of heterogeneity and surface water-groundwater interactions in the Thomas Brook catchment, Annapolis Valley (Nova Scotia, Canada)" *by* M. J. Gauthier et al.

C. Rivard

crivard@nrcan.gc.ca

Received and published: 19 June 2009

Reviewer #1

2a) The article lacks of an explicit "material and method" section, with sound descriptions of the model, the modeling strategy and the catchment, the later being treated as a "study case". By beginning the introduction with the description of "The Thomas Brook catchment" the authors induce the reader to treat the paper as a "site specific"

study. In the introduction the study should be replaced in a more general context, and include a review of other works dealing with the effect on soil heterogeneity modeling on water flow at various scale or the way soil heterogeneity is mainly treated in water flow catchment scale modelling studies.

Response 2a: The Introduction has been modified to present a more general context. The first paragraph of the introduction becomes: "The hydrologic response of a small catchment was modelled using a 3D coupled model. The main objective of this study was to investigate the level of complexity required to simulate hydraulic connections and interactions between surface water (springs, overland flow, and streamflow) and groundwater (within unconsolidated sediments and the bedrock aquifer)." The Thomas Brook study area is only presented at the end of the Introduction. The Abstract has been similarly modified. A review of other studies on heterogeneity is already present in the Introduction. Specifically, the paragraph beginning by "The impact of heterogeneity on surface and subsurface processes. . ." is devoted to this topic.

2b) The "Results and discussion" section should be reorganized in order to better illustrate the effect of soil heterogeneity modeling.

Response 2b: This section has been modified. Subsection 5.1 (Model calibration) is now sub-section 4.4 (Model calibration and initial conditions); the section "Results and discussion" now contains: 5.1 Model response 5.2 Effects of heterogeneity ("and other factors" has been removed since the focus is on heterogeneity) 5.3 Catchment behavior for different response variables.

3a) It is unclear in what matter the scenario including snow cover modeling (scenario 9) represents an added value to the issue of heterogeneity modelling and so, unless the authors could discuss it, it should be deleted. Nevertheless, if snow accumulation and melting modeling is a critical issue in yearly water flow modeling, all the scenarios should include these processes.

Response 3a: The addition of snow cover does not represent an added value to the issue of heterogeneity. It was added simply to study the additional effect of the snow accumulation once the model was sufficiently well-represented geologically. The annual budget is of course the same. We especially wanted to see the effect on monthly recharge. This is why we did not introduce it for each scenario.

3b) Finally, the conclusion of the study would be clearer and stronger if the authors focus their attention on the effect of soil heterogeneity on water flow modeling.

Response 3b: A paragraph has been added in the Conclusions that makes a recommendation on the minimum heterogeneity to be included: "For instance, they [the seven scenarios] showed that the description of subsurface heterogeneity must include, as a minimum, hydraulic conductivity values for the three bedrock formations, because of their strong influence on water table position, and for the various surficial units given the dominant role of the surface cover in rainfall-runoff-recharge partitioning." 4) Scenarios 4 to 6 are not enough discussed. Some variables seem to be better simulated for high level of soil heterogeneity scenarios but the improvement is less clear for other variables. Could the authors discuss this point?

Response 4: The text of the first paragraph of the new sub-section 5.1 (Model response) has been significantly modified to better describe the various scenarios. This paragraph has been changed to: "The various scenarios show different performance according to the observed (streamflow) or estimated (recharge) variables. In general, model performance based on streamflow improved from scenario 1 to scenario 8, but for recharge, results are variable (see Table 3). Scenarios 4 to 6 produced similar values for hydraulic components, with the best results for mean streamflows compared to observations; nonetheless, recharge was much too high. Scenario 7, with the most realistic representation of the system thus far, nevertheless yielded water levels that were too low in the North Mountain formation and an annual recharge that was too high, at 675 mm (Table 3), likely due to the available local data not being very representative. This was corrected in scenario 8 when regional values were incorporated. When considering recharge, model performance also improved significantly from scenario 8 to

scenario 9, after snow accumulation was taken into account (see below)."

5) Recommendation on the "minimal description" of soil heterogeneity sufficient to assess water flow at the catchment scale would be of great significance and this point should be discussed in the conclusion.

Response 5: See "Response 3b".

For the specific comments: - part 2 and part 3 should be put together : the "description of the study area" includes the geological context.

We prefer to keep these sections separate.

- p 2755, lines 13-16 : It could be informative to indicate the number of year of surface flow data monitoring.

In section 4.1 (Model implementation; now section 4.2), the following sentence was modified to make it clearer that only one year was used in the model. "Only one year, 2005, was retained because it was the only one containing a nearly complete streamflow time series on record."

- p 2755, lines 26-27. For how many year is this average? Could the authors indicate whether the year 2005 is representative of the mean climatic and/or hydrological behavior of the study zone?

Climate statistics from Environment Canada are currently provided for 1971-2000 (30-year period). 2005 is quite typical, with total precipitation for 2005 being 1201 mm/y, while the average over 1971-2000 is 1211 mm/y.

- p 2758 Please indicate the resolution method of the Richards equation : finite difference? Finite element ?

This specification is now provided in section 4 (Hydrological model of the Thomas Brook catchment): "The three-dimensional Richards equation, solved by the finite element method, represents variably saturated flow in porous media, . . ."

- p 2759: lines 1-3 : Could the authors justify your choice?

The 20 m DEM was coarsened to a resolution of 60 m in order to keep the simulation turnaround times reasonably short. .

- p 2759 lines 5-9: figure 5 should be introduced and commented in part 3 (geological context)

Done. Figure 5 is now called Figure 4 and the sentence introducing it has been moved accordingly.

- p 2759 lines 18 – 28. I believe the authors chose a flat base for the bottom of the flow domain to limit mesh complexity. How was assigned the thickness of 50m at the outlet of the catchment? Could the authors discuss these choices and indicate the possible consequences? Wasn't it possible to decrease the total thickness and decrease the mean thickness of the vertical layers? The bottom layer is the only one to have a nonunique thickness. Therefore the maximal thickness of this layer may reach 200m. Is this thickness not too high to support the assumptions of classical Richards equation resolution scheme? In particular it seems from fig 15 that this layer is not completely saturated near the North Mountain cuesta (between latitude 4.9955 and 4.995 106 m). More generally, the quality of the mesh should be evaluated with for example an aspect ratio (delta(x)/delta(z), Calver and Wood (1989), Paniconi and Wood (1993)).

A very thick bottom layer is not normally a problem for simulations that are driven by surface inputs (rainfall/evaporation), provided the total profile being simulated (soil + aquifer) is deep enough relative to the horizontal scale of the catchment being simulated. As for the quality of the mesh, mesh aspect ratios could indeed be important to examine in more detail. However, once again we should expect that the bottom layer should not be too affected by the aspect ratio factor. This factor will be more influential in the near-surface layers, where very fine vertical layers (relative to horizontal mesh dimensions) are typically needed to resolve surface/subsurface partitioning. This issue has not been examined in detail in this study.

- p 2760, lines 11-13: It means that surface catchment and ground catchment have the same limit and the same outlet. Is this hypothesis supported by experimental data or observations?

This is a very relevant comment. Indeed, surface and subsurface flows probably very rarely coincide in their flow boundaries, but integrated groundwater/surface water studies must by necessity impose the same "domain" for the entire flow system. Even within the same flow domain, however, this limitation can be overcome, for example by imposing source/sink boundary conditions along selected points or regions of the lateral catchment boundaries. The problem is that it is exceedingly difficult to obtain reliable estimates of water fluxes across these boundaries, as was the case also with our study.

- p 2760, lines 13-22 : To my opinion, Fig 6 and its comments should be include in the description of the study site, as an illustration of the hydrological behavior of the catchment.

Although this would be a good idea for two of the curves, we prefer to keep this figure in Section 4 since the graph of Figure 6 also includes some results (simulated streamflows).

- p 2760, line 24 to p 2761 line 29: These are mostly general information concerning the model CATHY and not specific information relative to your study. This general information should be included in a 4.1- model description and only specific information (flow domain geometry and discretization, boundary and conditions, material properties...) should be kept in a 4.2- model implementation.

Done. A section 4.1 (Description of the coupled model) was created.

- p 2762, lines 2-6: Is it "response parameters" or "response variables". To my mind, soil properties such as porosity or conductivity are parameters whereas discharge , saturation dynamics, . . . are variables since they are calculated from state variables such as pressure head or soil moisture content.

“Response variables” is now used in the text.

- p 2762, lines 13-17 and Figure 7: It is not necessary to include conductivity and porosity value in the text. It would be clearer to report those values in a table than to indicate it on figure 7. In Figure 7, soil number should be reported and the boundaries between soils better marked (by black line?).

We prefer keeping the values in Figure 7 so that the figure stands by itself.

- p 2764 §1 and 2: These 2 paragraphs (and part of the following paragraph p2764-2765) mainly present the method of model initialization and calibration and should therefore be included in the 4th part of the article (4.3. Model calibration).

Done. These two paragraphs now belong to a distinct section (4.4 Model calibration and initial conditions.

-part result and discussion : The last paragraphs of 5.1 and section 5.2 should be re-arranged in order to better illustrate the effect of heterogeneity on surface and ground-water flow (the title “effects of heterogeneity and other factors” is not appropriate). The authors should better describe the results of all the scenarios. In particular, scenarios 4 – 5 -6 simulate better the mean outlet streamflow than scenario 8. It could be interesting to plot in a graph the stream-flow simulated by all the scenarios. The section “Catchment behavior for different response variables” is interesting but does not lead to new informations/ideas about the effect of heterogeneity on flow modeling. This section should be removed OR you should compare the difference in “saturation zone” location for the 8 scenarios.

The last paragraphs of the old section 5.1 have been put into the new section 6.1 (Model response for scenario 9). The title of section 6.2 has been changed to “Effects of heterogeneity (“and other factors” has been removed since the focus is on heterogeneity). Scenarios 4 to 6 are now better described (see Response 4). We believe that

the section “Catchment behavior for different response variables” (now section 6.3) is important to the paper as it provides examples of results obtained for two important variables related to SW/GW interactions: saturation and the water table position.

- p 2765 lines 14-16 / 23-25 : How many well measurements were used? In particular, were all the data for well 1 and well 2 used ? If not, why?

30 wells were visited for GW level measurements over the study area (8 km2). The complete records of wells #1 and 2 were only used for comparison with simulated daily data at the same location in Figure 9. Figure 8 only uses mean water level measurements (observed and simulated). Therefore, Wells # 1 and 2 are contained in only one point each in Figure 10.

- p 2766 lines 6-8: How fluctuate the groundwater level for scenarios 1 to 7? It would be interesting to examine “groundwater variables” especially for scenarios 4 to 6 for which the mean outlet streamflow was well reproduce.

GW levels were much too low in the North Mountain Formation for scenarios 1 to 7. Decreasing K to the regional value helped raising them to an acceptable value. Differences with observations usually decreased towards the southern part of the catchment.

- p 2766 lines 22 – 26: the sentence is difficult to understand and need to be clarified.

This sentence was perhaps too long and has been split in two (and a comma added) to make it clearer.

- p 2767 lines 25-27: Did the authors include recharge value for January in the annual value reported in table 3. If simulated value January recharge are too high because of initial conditions, January should be included in the “initialization period” and the comparison between simulated and observed data should be made from February to December.

Recharge for January is indeed included in the annual value reported in Table 3. We prefer to present annual (12-month) estimates, even if we have to explain that the first month may be affected by initial conditions.

- p 2768 lines 1 – 10: This part of your paper should be enhanced (since it was mentioned in the introduction results in terms of numerical performance.)

Initially, a section on numerical performance was included in the paper. However, the paper was too long and we decided to remove it. We have now removed this last part of the sentence ("...as well as on the numerical performance of the model.") from the Introduction, as it was a relic from the initial version of the paper.

- Table 3 : correct the number of "*" for the first note and in the formula.

Done.

- figures 2 to 4: roads should be removed. Figure 2 is not clear: what was drawn :altitudes of soil surface as indicated in the text or potentiometric map as indicated in fig.2 caption?

Figure 2 presents GW elevations (a potentiometric map). The reference to Figure 2 was removed from the sentence discussing the topography. We do not think that roads represent a problem in these figures, at least not if the figures are reproduced in colour. They show that the wells are mostly located along roads in this area, and they convey that data was not available in less accessible areas.

- figure 9 : it would be informative to plot daily net atmospheric forcing and/or outlet streamflow above this graph.

This information is already provided in Figure 6.

Reviewer #2

1. I wonder what would happen if the order of the scenarios would change i.e. first adding the top layers, then refining the bedrock geology. Please discuss this issue.

Response 1: We did not try changing the order of the scenarios, but in the end the final result would be the same. We did not give any "priority" to bedrock as opposed to surficial geology (indeed we started with surface geology in scenario 2; in scenarios 3 and 4 we incorporated the main features of the bedrock geology; in scenarios 5 and 6 we refined the surface geology; and in scenario 7 we refined the bedrock geology). Our strategy was to first include characteristics that we thought would influence most the model response (e.g., adding a surficial cover, defining three bedrock units). Then, known parameters (e.g., porosity) were added.

2. I also wonder how well the simplified scenarios would work if their parameters were all subjected to calibration. And the opposite question: if no calibration is allowed at all.

Response 2: No calibration was performed on hydrogeological parameters (K, n). Scenarios 1 to 7 use local existing and collected information; scenarios 8 and 9 use regional values. A trial and error approach was used for the model parameters common to all scenarios (described in Table 2). Perfect calibration was not the focus of this paper, as we discuss in the Introduction and Model calibration sections.

Specific comments:

page 2752 / line 14 "to a sufficient degree" : define "sufficient".

This phrase in the Abstract is expanded in the Conclusions: "The simulated heads, aquifer recharge, and streamflow at the outlet for scenario 9 were comparable to observed or previously estimated values."

/ line 16-17 "North Mountain basalts": is it really necessary to refer to the formation name in the abstract? just "basalt" should do, here.

Done.

page 2758 - line 9 "and other criteria" : like?

This sentence has been modified: "The distinction between grid cells belonging to the hillslope and stream network systems can be made according to three different threshold-based options, based on criteria such as upstream drainage area, local terrain slope, and land surface curvature."

/ line 12-13 "D8 scheme ... or by ... methods" : which algorithm is actually used?

An addition has been made to this sentence: "Cell drainage directions can be identified by the simple D8 scheme (the method used for this study, whereby one of eight inflow/outflow directions is taken on each cell) or by more recent nondispersive and dispersive methods (Orlandini et al., 2003)."

/ line 20-22 "two different ... schemes allow ..." : they allow updating. But have they been used? which one? or both?

We have added at the end of this paragraph: "The data assimilation feature was not used for this study."

page 2759 / line 16 "gOcad" : what kind of software is that?

gOcad is a software for the construction of geological models. For more details, see their website at: http://www.gocad.org/www/

page 2760 / line 9-10 "a high end laptop computer" : in 5 years time, high-end computers will have completely different specs. Please add more info on the processor speed, like as defined in flops.

True. We have replaced this sentence with: "A grid of this size required calculation times of several hours for 1-year simulations run on a laptop computer with a 1.90 GHz processor."

/ line 15-17 : "the input fluxes are [P-Epot]" : How can P-Epot be used as input fluxes? P-Eactual can, but not Epot, which is not a flux but only the maximum possible flux.

The CATHY model treats both P and Epot as "potential" fluxes, and the boundary condition switching procedure determines whether this flux becomes the actual amount

of water that gets infiltrated/evaporated (Neumann boundary condition) or whether the actual flux is less than this amount (Dirichlet boundary condition). We have added the following sentence to clarify this issue: "Switching between specified head and specified flux boundary conditions occurs at surface saturation (zero pressure head) in the case of rainfall and at the "air-dry" pressure head value in the case of evaporation."

page 2764 / line 3 : "model calibration" : which calibration technique was used? Which goodness-of-fit criterion was used?

See Response 2.

/ line 5 "K and n were assigned" : a prior estimation on this scale requires some averaging procedure. Has that been done? Also. why haven't these parameters been calibrated?

Mean values were obtained for K using medians while mean porosities for bedrock were provided by a parallel study using thin sections and image manipulation and analysis techniques (penetration of the dyed mounting medium) and from the literature for surficial units. Their calibration was not the purpose of this paper. We only wanted to see the effect of increasing the geological complexity and of the assignment of local versus regional values.

/ line 7 : "an adequate agreement" : define adequate.

We were interested in obtaining a good or satisfactory visual agreement; the quest for an optimal calibration was not the purpose of the study.

/ line 9-10 : "once obtained, the parameter values ... where kept fixed for all subsequent scenarios" : is this allowed? The parameters are effective parameters, and as such coupled to the model schematisation. When this latter changes, so should the effective parameters.

This would be the case if we sought, for each scenario, an optimal calibration, but this was not our procedure (see also Response 2).

/ line 20-21 : "the match was obtained after 1-2 months" : figure 6 suggests a response time of 3-4 months (recession during july-august-september)

After 1-2 months of summer recession the flow is around 0.1 m3/s, while after 3 or 4 months the flow is about 0.05 m3/s. This provides a rough estimate of the baseflow range (an estimation using hydrograph separation provided a value of 0.08 m3/s).

/ line 27 : "model performance" : how is this quantified?

We compared mean observed and simulated values.

/ line 27 " improves slightly from scenarios 1 to 8" : according to table 3, the streamflow error is smallest for scenario's 4, 5, and 6.

The intermediate scenarios are now discussed in more detail (see Response 4 to Reviewer #1).

page 2766 / line 16 "mismatches could be reduced" : is seems to me that the peaks in the simulated hydrographs are both too large and of the wrong shape. Perhaps part of the catchments is reacting flashy (small but sharp peaks in the data) and other parts are buffering water, reacting much slower (higher baseflow).

We do not think that the peaks have the wrong shape, although the variations within a peak are not always captured. As already mentioned, we did not pursue intensive calibration to improve individual peaks (or other features) for any scenario. An "adequate" match was obtained for scenario 1 and then the parameters described in Table 2 were kept fixed in order to see the evolution on mean annual streamflow, annual recharge, and position of the water table (the three variables for which observed or estimated values were available) as complexity was added to the geology.

page 2767 / line 23 : "adjusting to initial conditions" : why then is the model initialised for january with conditions that resemble summer baseflow?

It may indeed have been better to start with a higher streamflow, but winter is also

usually a period with low streamflow.

/ line 29 "suggesting a much greater sensitivity to k" : this is circumstantial evidence, and should be tested directly.

We do not think that this is circumstantial evidence; scenarios 7 and 8 for example provide a good illustration of this.

page 2768 / line 26 "what is labeled as Horton ... may arise from shallow subsurface flow ... that saturate" : Sounds like Dunne OF.

What we are describing here is return flow that can saturate near-surface cells while underlying cells remain unsaturated. This is not Dunnian runoff, where the entire vertical profile becomes saturated. It is not exactly Hortonian runoff either (the "classical infiltration excess" mentioned in the paper), but we lump this contribution with Horton runoff because its manifestation is the same (i.e., both result in a saturated surface).

Table 3 : check the use of * ** marks. there are some mistakes. Also in the **100.

Done.

Figures 6, 10, 11: the x-axis tick labels are in French. Why are they on the 5th day of the month?

"05" refers to the year (2005). This has been removed (the year is given in the figure caption) and the labels now give the months in English.

Figure 8: The good fit is partly a result of the topographic gradient. Please provide a similar plot based on groundwater depth (with respect to the surface). (m, ASL) should be (m ASL).

It's true, the fit would not be as good using GW depths, but an error of 2 m over 5 m is much worse than over an elevation of 200 m. An RMSE threshold in hydraulic head equal to 10% of the difference between highest and lowest heads is often deemed to be quite acceptable.

Figure 9: "Time (days)" should be "Time (day)"

Done.

Figures 12-14: Longitude and latitude are measured in degrees, not in meters. If you want to use meters, then the labels should be "easting (m)" and "northing (m)"

Done.

Reviewer #3

1. The introduction gives an overview of previous and recent work in this field but doesn't really explain the necessity for the presented work. What can the results of modeling experiment contribute to the real society, not only from the scientific view?

A sentence has been added to the first paragraph of the introduction: "This work can be viewed as part of an integrated management plan for a rural region, for which spatial distribution of recharge and knowledge of the water budget are crucial."

2. The scenarios themselves are well explained, however, it is not explained why these scenarios have been chosen. Could you please provide a bit more background information to explain about these scenarios? Or, these scenarios are decided randomly.

In the design of the scenarios, our strategy was to incorporate in sequence those features deemed most important and for which data were available. See Response 1 to Reviewer #2.

3. The numerical model is a physically based model with a number of calibration parameters. Uniqueness and robustness of the calibration is not sufficiently demonstrated. The physical meaning of various model parameters (applied in this case study) is not well discussed in detail.

Hydrogeological parameters (K, n) were not calibrated. They were assigned according

to local or regional values. See previous responses for more on calibration issues. The model parameters common to all scenarios are described in Table 2, and references for further information on these parameters are provided as well in this table.

Specific comments

1. P8 Line 9, What was the 'other criteria'?

The sentence has been changed: "The distinction between grid cells belonging to the hillslope and stream network systems can be made according to three different threshold-based options, based on criteria such as upstream drainage area, local terrain slope, and land surface curvature."

2. P9 Line16-17, it wasn't clear to me which version of gOcad and ArcGIS.

ArcGIS 9.0 and gOcad 2.0.8.

3. In section 2 'Description of the study area', why not mention land cover distribution of the catchment which will have great impact on its hydrological cycle? Some details about the horizontal spatial variability of land use and cover could be introduced.

Land use and land cover information is not directly utilized by the CATHY model (unlike more "land surface"-oriented models such as SWAT). This information can be indirectly used when linked to soil parameters (for instance setting a low surface K for urban areas). Land cover maps were not used in our study, which relied exclusively on geological data for model parameterization.

4. The instrumentation (type and location) for the measurement of rainfall, stream flow, groundwater should be described.

Data on total precipitation came from Environment Canada (http://climate.weatheroffice.ec.gc.ca/Welcome_f.html) Streamflow data came from a local study in Nova Scotia (one of our partners). These sources are now mentioned in the Acknowledgments. Groundwater data came either from dataloggers (installed by us) in two residential wells or from instantaneous measurements in visited residential wells (using a probe). In the paper we refer several times to these “measured groundwater levels”.

5. In Section 4.1, because the study area is a major agricultural region of Canada, do people in this valley pump any groundwater for agricultural water use? How to consider the impact of human activities on groundwater recharge? Or, does this model only consider the most ideal and natural condition without any human interruption?

For the moment, the majority of water dedicated to irrigation comes from surface water (approximately 80%). Pumping (withdrawals) probably represents a small percentage of total precipitation (and even recharge).

6. P12 Line 11-12, which hydrogeological parameters were assigned based on fieldwork results? Which parameters are based on database? Where are the databases?

K and n in both cases. For scenarios 1 to 7 (local values), most values came from acquired data (fieldwork). For scenarios 8 and 9 (regional values), data came from the provincial database.

7. Results and discussion: The different stream discharge time series are only shown in figures. It would be useful for comparison if you can quantify their distribution by, for example, the variance.

We do not think that, for streamflow, variance is a relevant characteristic. We present daily data graphically so that the appropriateness (location and height) of the peaks can be discerned.

8. The model also produces overland flow, interflow, and groundwater. Why then not compare them? The title is “A modeling study of heterogeneity and surface water–groundwater interactions”, how about their interaction through the modeling study?

Overland flow and return flow are presented in Table 3 for each scenario. However, these variables were not measured in the field. Recharge is also presented in Table

3 and the mean annual value estimated with another method is provided in the last line. Finally, Figure 8 presents a comparison between measured and simulated GW elevations.

9. P34 (Figure 6), 38 (Figure 10), and 39(Figure 11), is the date in the x-axis of data French? Can you change them to English?

Done.

10. P35 Figure 7, in the block 2, "n=0,20" should be " n=0.20”.

Done.

11. P41 and 42, the numbers and labels in the x and y axis are too small to be read clearly. Can you modify them?

Done.

12. P45, in figure 15, I could not see clearly the dotted line. Which line is for which scenario? Can you improve the quality of this figure?

In colour the dotted lines are quite clear.

13. Figure 8, can you tell me the meaning of ASL?

“Above sea level”.

14. How the model transfers from Potential evapotranspiration to Actual evapotranspiration? Can you briefly describe it in your article since it is very important at a catchment scale study?

See the response to “line 15-17” of Reviewer 2.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 6, 2751, 2009.