Answer to Anonymous Referee #2

We would like to thank Referee #2 for the valuable feedback that highlights some additional points for improving the manuscript. We agree with most of the points brought up and will answer first the general comments, and afterwards the specific comments.

We agree with the major point of criticism concerning the lack of calibration/validation of our model. The Editor also brought up this point. Upon the advice of Referee #2 we suggest to extend the last paragraph of the “further research” section (L750) as follows:

“Although topography-based model estimations can give insight of pesticide transport via hydraulic shortcuts on a large scale, they have not been tested and validated in the field with measurements on flow and transport. Targeted measurements on surface runoff and on pesticide transport through shortcuts are needed to provide evidence on the quantitative relevance of this flow path. A field study in one catchment in the Swiss plateau (Schönenberger, in prep.) demonstrates that pesticide concentrations in shortcuts can be very high. However, more systematic research is needed to quantify the relevance of shortcuts. Ideally, catchment-scale experiments – e.g., with controlled pesticide applications (see Leu et al. (2004);Doppler et al. (2012)) – would be carried out to quantify loss rates from directly and indirectly connected fields. Apart from the practical problems of implementing such experiments in the context of farmers managing their land, this approach will often face the problem that many fields are also tile drained. Consequently, any signal in the stream is a superposition of different potential flow pathways. Given that the transport through shortcuts has no unique characteristic, it is difficult to disentangle and quantify these pathways. This implies that one has to observe simultaneously flow and transport within a catchment at locations where one can differentiate between the flow paths. Such a setup would allow to determine the proportion of total catchment runoff and pesticide load that is transported via hydraulic shortcuts. In addition, isotopic tracers and runoff separation techniques could be used to determine the total amount of surface runoff contributing to catchment runoff. Knowing both, total and indirect surface runoff, the amount of direct surface runoff could be calculated. These measurements of direct and indirect surface runoff could then be compared to the data provided from our connectivity maps. Since shortcuts are only active during rain-events, we suggest an event-based monitoring approach.”

Another point of criticism states that “value of a national map of directly / indirectly connected areas remains unclear without knowing, whether elevated connected proportions correspond to higher or faster hydrological response and respectively increased pesticide wash-off” and that therefore the upscaling to the national level seems premature. Also Reviewer #1 stated this part to be scientifically less interesting.

We agree that we did not make it very clear what the added scientific value of this part remains is, given that there is already a similar model on the national scale. We noticed for example that this part was not introduced as one of the objectives of the paper (L102 ff.). The comments by both reviewers clearly demonstrate that we need to sharpen our arguments and describe concisely what the purpose of that part of the paper is and to eliminate those parts in the manuscript that do not fit the specific objective.

The rationale for the development of our map at the national scale is that there is first a need for connectivity data at that scale (e.g. for people evaluating the Swiss Action Plan on pesticides) and that they rely on the existing connectivity map. Given that we substantially improved on the empirical basis at the catchment scale regarding connectivity, we consider it scientifically justified to evaluate how our findings affect predictions at the catchment and the national scale. Actually, we consider this a necessity taking the societal role of science serious: not providing such a comparison would imply that we demonstrate with our empirical observations that connectivity indeed matters according the best knowledge we have, but leaving it to others (who have no better means than we have) to figure out what our findings mean in the larger context. Hence, what we do with the upscaling step is to bring together existing scientific knowledge in a transparent manner.

Reviewer 2 asked the question whether or not the proposition of a connectivity map at the national scale wasn't premature given the lack of empirical validation of the model. This is a very well founded question that has to be considered seriously. However, this question does not only address our map at the large scale but also the use of the already published map, which is intensively used in practice. What does a statement of prematureness in this context mean? If one accepts that one should use the best available knowledge for rationale decision making, the question is whether one should make use of such maps at the current stage or whether their use does more harm than providing benefits. Causing more harm than benefits could be called a premature use of a model from a decision-making perspective.

Based on the comparison of our empirical data, the existing model and our "new" model, we can conclude that both models can reasonably well represent the field observations. We don't have any indications that would suggest that it is preferable not to use any of the large-scale models for assessing the connectivity of fields to surface water bodies. In this sense, we conclude that is not premature to use either of these models. Therefore, it also makes sense to evaluate how our observations affect the existing model and which differences are induced.

This argument does of course not invalidate the correct critique that neither of the two models have been tested and validated empirically regarding their actual capacity to quantify the connectivity effects on water flow and transport of agrochemicals beyond the few observational studies that triggered this kind of connectivity assessment in the first place. This aspect needs to be clearly communicated.

In summary, we propose the following modifications of the paper to address the concerns of Reviewer 2 (these changes partially also address the issues raised by the first reviewer):

- The objective of evaluating the consequences are clearly stated in the Introduction and an explicit rationale for doing so is provided (How do the additional empirical observation influence the large-scale predictions?).
- The limitations of existing large-scale models are clearly stated, the research need is emphasised, and possible approaches explicitly discussed.
- The focus of the large-scale aspect is on the comparison with the existing model. To be consistent with this objective, we replace the current Fig. 8 with a map depicting the differences between the two models. The focus will be clearly on how the additional empirical observations influence the model predictions.
- Along the same lines, we will skip those parts of the text that discuss the findings specifically for the pesticide issues (L552-561).

Specific comments

(1) Lines 106 ff: Shortcut definition should be moved to section 2.2 Assessment of hydraulic shortcuts:

We will move this.

(2) Lines 125 ff: The probability of selection was proportional to the total area of arable land...How was this represented in the random selection?

We are not sure if we understand the question correctly, but try to answer it as good as possible. We performed a weighted random selection from a list of all catchments. The weights of each catchment equalled the total area of arable land in the catchment. Specifically, we were using the python function “numpy.random.choice”:

```
numpy.random.choice(a = catchment_id_list, size = 20, p = catchment_area_list)
```

(3) Lines 156 ff: How did you prevent selection bias due to drainage plans? I.e. how did you rule out, that no available drainage plan did not correspond to no existing drainage system

We tried to reduce the impact of selection bias as much as possible by using three different acquisition methods. If drainage plans are not available in a certain catchment, the other two methods are to some degree filling this gap and accordingly reducing the selection bias. As shown in Table 5, the drainage plan mapping method had a lower recall than the aerial image mapping methods. Additionally, also the number of shortcuts identified by aerial images is much higher than by drainage plans. Therefore, the aerial image mapping method is more important for the overall result and we expect the selection bias due to drainage plans to be small. However, we are sure that we still missed some of the shortcuts and addressed this in L424-226 by writing that the numbers reported are a lower boundary estimate.

(4) Line 243: Elaborate under which circumstances hedge infiltration may be active or inactive

We could imagine various factors affecting the runoff capturing efficiency of a hedge. For example, the width of the hedge, shrub species, or the degree of runoff concentration. We did choose this parameter to have a binary distribution since no further information on the hedges (such as hedge width) were available. As shown in the sensitivity analysis (Figure S21 and S22), the hedge width parameter only has a minor influence on the overall results. Therefore, we also did not spend time in refining this parameter further.

(5) Line 271 ff: Conceptually, the parameter “maximal flow distance” should depend on soil properties and cultivation phase, was this considered?

This was not considered directly, but indirectly. Since no data on soil properties is available on national scale in Switzerland, we tried to identify potential differences in soil properties by calculating the topographic wetness index (TWI) (see L295 ff.). We did not find any systematic differences between the TWI distributions of directly and indirectly connected areas (see L508 ff.). We therefore also do not expect systematic differences in soil properties between directly and indirectly connected areas. We did not address the cultivation phase specifically. This influence factor is expected to cause large differences in maximal flow distances in time and space. However, we again do not expect systematic differences of this influence factor between directly and indirectly connected areas.

Consequently, we expect that the maximal flow distances found on directly and indirectly connected areas are not systematically different from each other.

(6) Line 283: It is unclear how "all possible flow distances were evaluated" with 100 model realizations

We agree that this is not clear. We will change this to:

"For the parameter maximal flow distance, all possible flow distances were evaluated for each Monte Carlo simulation."

(7) Line 295: Suggested section title change to 'hydrological boundary conditions'

We think the more specific term "hydrological activity" fits better here, since this term is usually used in the context of critical source areas, for example see: Pionke et al. (2000)

(8) Line 312 ff.: Although crop type and rain intensity probably don't differ systematically within one catchment, they do impact runoff generation. Should this not be systematically evaluated e.g. across catchments?

Yes, these factors affect runoff generation and are expected to differ systematically between catchments. As also mentioned in L299-301, this manuscript focuses on comparing indirect surface runoff to direct surface runoff. We therefore were looking for systematic differences between indirectly and directly connected areas. Comparing surface runoff generation across catchments was not within the scope of this manuscript. Currently, except from rainfall data, the data availability for such a comparison is not given. For example, crop data are currently not available in sufficient resolution (see also L349-350), soil maps are not available on a national scale. It is also unknown how farming practices (e.g. pesticide application or soil management) differ between catchments.

We however agree that this could be an interesting direction to go and suggest this in the "further research" section.

(9) Line 384 ff / Tab. 3: why is the destination of such a large number of drainage structures unknown? The maps (Figs. 5, S 2.2.1) suggest line / network structure for most inlets, was the outlet of these unclear? How were unknown drainage locations treated in the connectedness classification?

Three reasons were mainly responsible for this problem:

1) There was no drainage plan available in the whole catchment.
2) Drainage plans were available in the catchment, but did not cover the specific region where the potential shortcut was located.
3) Drainage plans were available in the catchment and did cover the specific region where the potential shortcut was located, but the potential shortcut and its drainage

structure were not shown on the plans. (They were however identified during the field survey or on the aerial images.)

For the inlets with known drainage locations 99 % were connected to the surface waters (87 %) or via WWTP/CSO (12 %). Therefore, we assumed in the connectivity model that all shortcuts with unknown drainage locations drain to surface waters (see L252-255).

(10) Lines 419 ff: In Lines 147 ff the field survey was described as "we walked along roads and paths and mapped all the potential shortcut structures." How was mapping accuracy of inlets (5%) and manholes (25.5%) on fields and other areas validated? Lower accuracy esp. in fields with dense vegetation seems likely.

The accuracy of field mapping could not be validated since we would need a "ground truth" for this, which was not available. We agree that lower accuracy in fields with dense vegetation is likely and discussed this issue in L608-612.

(10 cont.) How are false positives from mapping ruled out to be false negatives (overlooked structures) from the field survey? How are false negatives from the aerial images quantified altogether?

From your question, we noticed that Table 5 is not clear enough. As you state correctly, we cannot differentiate between false positives from the aerial image/drainage plan method and false negatives (=overlooked structures) from the field survey when looking at the *identification* of a shortcut structure. For the *identification,* we therefore only report the recall. However, given that a shortcut structure is identified, we can analyse if it is *classified* correctly by the aerial image/drainage plan method (see also L414-418). For the *classification,* we can also report false negatives. For example, a shortcut structure was identified by the aerial image method and was classified as a maintenance manhole. In fact, the field survey showed that the structure showed that the structure is an inlet. This would correspond to a false negative classification.

We will adapt Table 5 as follows, to make the difference between identification and classification accuracies more clear.

Table 1: Recall and classification accuracies of the mapping methods aerial images and drainage plans. The recall corresponds to the probability that a potential shortcut is found by the mapping method. Percentages indicate the recall of each individual mapping method. In brackets, the recall of the combination of both methods is given. The accuracy corresponds to the sum of true positive fraction and true negative fraction.

Mapping method	Manhole type	Identification	Classification				
		Recall	True positives	False positives	True negatives	False negatives	Accuracy
Aerial images	Inlets	53 % (60 %)	61 %	1.3 %	33 %	4.9 %	94 %
	Maintenance manholes	62 % (69 %)	32 %	5.3 %	61 %	1.3 %	93 %
Drainage plans	Inlets	32 % (60 %)	67 %	4.5 %	22 %	6.6 %	89 %
	Maintenance manholes	21 % (69 %)	20 %	7.1 %	68 %	5.3 %	88 %

(11) Line 463: Are the 21 % in Müswangen and 97 % in Boncourt medians / means of the MC ensemble results?

Yes, those are the medians. We will adapt the sentence as follows:

“However, this fraction varies strongly between the study areas, with median fractions ranging from 21 % in Müswangen to 97 % in Boncourt.”

(12) Lines 504 ff: Apart from distribution similarity, how do wetness index and slope affect connected area proportions? I.e. do catchments with higher “hydrological activity” exhibit higher connected proportions?

Our connectivity model produces something like a “theoretical connectivity map”, i.e. the areas reported are connected under the assumption that surface runoff is produced and that this surface runoff is not infiltrating before reaching the recipient area. In contrast, the “effective connectivity” depends on the amount of surface runoff produced and the amount of surface runoff infiltrating before reaching the recipient area.

Wetness index and slope are positively correlated to the probability of an area to be hydrologically active, i.e. more surface runoff is produced. Additionally, they are negatively correlated to the probability that surface runoff infiltrating before reaching the recipient area.

Accordingly, areas with higher wetness index and slope are expected to exhibit a higher “effective connectivity”. In fact, those considerations were the reason for performing the analysis described in L504 ff.

(13) Lines 515 ff: I suggest to quantify the deviation between NSCM and LSCM with RSME or another goodness of fit measure.

We calculated the RSME between the NECM and the LSCM, and between the NSCM and the LSCM. We will adapt L520-523 as follows:

“The differences to the LSCM were strongly reduced by this transformation. The root-mean-square error (RSME) reduced from 17 % to 9.5 % for directly connected fractions, from 12 % to 7.6 % for indirectly connected fractions, and from 18 % to 7.6 % for not connected fractions.”

(14) In Fig 7 the mean fraction not connected of LSCM appears to be roughly 15% higher than of NSCM but the text states it is 3% larger (l. 545), why do text and figure differ?

There is an error in the legend of Figure 7. While the colours were described correctly in the figure description, in the legend (on the top right of the figure) the colours “red” and “blue” were interchanged. We corrected Figure 7 (see below) and will adapt it in the manuscript.

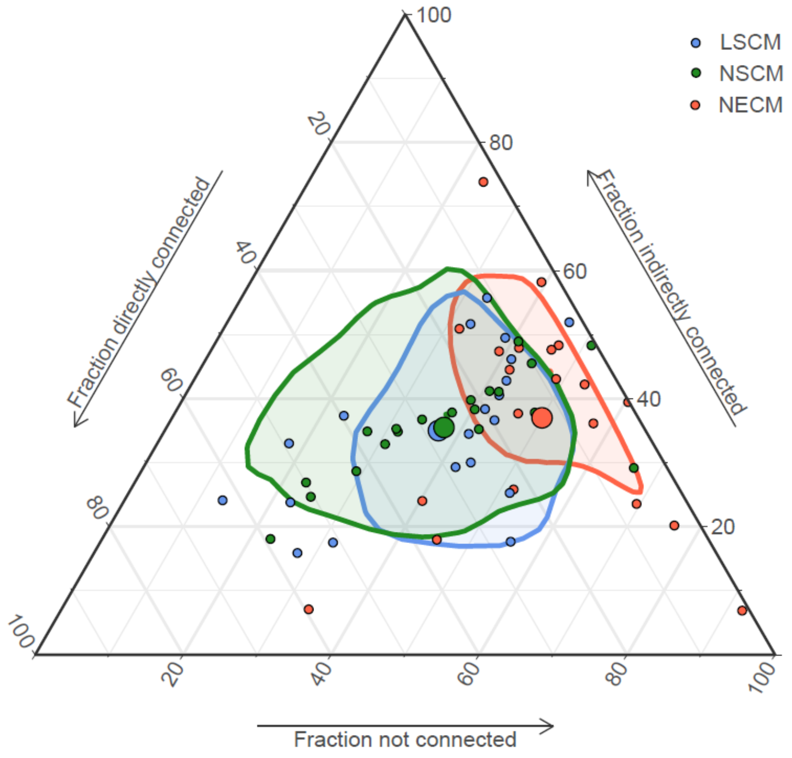

Figure 7: Fractions of directly connected (f_{dir}), indirectly connected (f_{indir}), and not connected areas (f_{nc}) per total agricultural area for the local surface runoff connectivity model (LSCM, blue), national erosion connectivity model (NECM, red), and national surface runoff connectivity model (NSCM, green) in the 20 study areas. Small blue circles represent the catchment medians of all Monte Carlo simulations of the LSCM, small red circles represent the data reported by the NECM, and small green circles represent the catchment medians of the NSCM. Large circles represent the means of the LSCM (blue), NECM (red), and NSCM data (green). Shaded areas represent normal Kernel density estimates of the LSCM, NECM, and NSCM data.

(15) Lines 554 ff: I disagree, that "map corresponds to a risk map of pesticide transport via hydraulic shortcuts". Although hydraulic shortcuts contribute to the risk, surface runoff proportion and volume, pesticide application intensity and retention in treatment facilities contribute to this risk as well. This is acknowledged in the discussion of NSCM (Lines 700 ff). (See also general comment above)

We agree with this point. With revising and shortening the second part of the paper (as written in our answer to the general comments), we will replace Figure 8 and this sentence.

(16) Line 599: phrase starting with "In Buchs,..." is unclear.

We will rephrase this to:

"In Buchs, around 60 % of the channel drain and ditch length consists of ditches that cannot be clearly distinguished from small streams."

(17) Line 633 – 687 the parameter discussion is somewhat self-referential without calibration data: Road carving depth, sink depth, and shortcut definition should be evaluated / calibrated by observed events. Assumptions such as 'higher DEM resolution is better' or 'manhole sinks should not be filled completely' seem plausible

but can’t be substantiated by the results of this research. The impact of hydrological activity parameters may not differ between directly and indirectly connected areas of the same catchment, but among catchments and should be evaluated accordingly.

We agree that our quantitative results cannot substantiate statements that higher DEM resolution was better. However, the field observations – e.g. based on visual inspections of sediment deposition along roads – clearly revealed that the microtopography can play a major role in controlling the flux of water, solutes and sediments into shortcuts. Furthermore, one has to be aware that these parameters are global parameters in the model despite the fact that there might be regional differences. The optimal road carving depth for example may differ according to topography, regional construction standards, etc. Given this situation, we also think that the optimal way to go would be to calibrate these parameters by observed events. However, on a scale of 20 catchments this would be an extremely laborious task. Additionally, field observations of these parameters also underlie high uncertainties. For example, sink depths are strongly variable in time and space. We therefore aimed on discussing other options that could be used to improve our results. To clarify that these statements are not findings that can be substantiated by our results, but a discussion of improvement options, we will modify the manuscript as follows:

- Higher DEM resolution: We will replace the word “would” in the sentence in L637-638 by the word “could”, since we can’t tell from our results that this would really improve the model.
- Sink filling: Similarily, we will replace the word “can” by the word “could” in L650-651.

(18) Line 733: Suggested extension: End of pipe measures at shortcut / pipe outlets(treatment, sedimentation, filtration)

We agree on this and will adapt this sentence to:

“Other measures could aim on the shortcut structures themselves (e.g. construction of shortcuts as small infiltration basins, removal of shortcuts, or treatment of water in shortcuts) or on the pipe outlets (e.g. drainage of shortcuts to infiltration basins, treatment of water at the pipe outlet).”

(19) In my view, the 2nd research question (line 104) can’t be fully answered with the present approach. It should be rephrased and / or referred in the conclusions section.

In our view, we can actually answer the second research question with our approach. We did not find any evidence on systematic differences in hydrological activity. In addition, we do not expect systematic differences in farming practices, precipitation, or crop types between directly and indirectly connected areas. Given the current knowledge, we therefore expect that the proportions of direct and indirect surface runoff related pesticide transport are proportional to the directly and indirectly connected area. However, we think that this is not formulated clear enough and we will revise the conclusions (L758) and results (L512) accordingly.

(20) Line 1116: the table title is confusing. I assume directly connected local surface connectivity model fraction was correlated to directly connected national erosion connectivity model fraction and so on...

We agree and will adapt the table accordingly (see below).

Table S 1: Correlation of catchment statistics with fractions of connected area connectivity. NECM: National erosion connectivity model, LSCM: Local surface runoff connectivity model.

Variable	Fraction directly connected $f_{LSCM,dir}$ (-)			Fraction indirectly connected $f_{LSCM,indir}$ (-)			Fraction not connected $f_{LSCM,nc}$ (-)		
	R^2	Slope	P	R^2	Slope	P	R^2	Slope	P
NECM: Directly connected agricultural area per total agricultural area $f_{NECM,dir}$ (-)	0.71	1.0E+00	**< 0.001** ***						
NECM: Indirectly connected agricultural area per total agricultural area $f_{NECM,indir}$ (-)				0.52	6.0E-01	**< 0.001** ***			
NECM: Not connected agricultural area per total agricultural area $f_{NECM,nc}$ (-)							0.26	4.0E-01	**0.022** *
Surface water body density (m^{-1})	0.51	2.2E+02	**< 0.001** ***	0.35	-1.4E+02	**0.006** **	0.14	-7.6E+01	**0.10** *
Paved road density (m^{-1})	0.20	-2.2E+01	**0.049** *	0.19	1.7E+01	0.053 -	0.04	6.5E+00	0.41 -
Inlet density (ha^{-1})	0.07	-1.3E-01	0.28 -	0.10	1.2E-01	0.17 -	0.00	1.0E-02	0.90 -
Manhole density (ha^{-1})	0.15	4.0E+02	0.09 -	0.07	-2.0E+02	0.27 -	0.07	-1.8E+02	0.27 -
Yearly rainfall (mm/year)	0.10	-5.2E-02	0.17 -	0.06	3.2E-02	0.28 -	0.04	2.0E-02	0.43 -
Total road density (m^{-1})	0.05	2.6E-01	0.35 -	0.05	-2.0E-01	0.33 -	0.00	-4.5E-02	0.80 -
Subsurface waterbody density (m^{-1})	0.11	-7.5E+00	0.14 -	0.04	3.3E+00	0.40 -	0.10	4.5E+00	0.18 -
Fraction of agricultural area (-)	0.00	2.6E+01	0.94 -	0.03	-1.7E+02	0.48 -	0.03	1.7E+02	0.43 -
Unpaved road density (m^{-1})	0.15	4.4E-04	0.09 -	0.02	-1.2E-04	0.55 -	0.18	-3.2E-04	0.063 -
Lake shore density (m^{-1})	0.03	1.3E-02	0.49 -	0.02	7.7E-03	0.60 -	0.13	-1.9E-02	0.13 -
Slope on agricultural areas (°)	0.04	-5.8E+00	0.41 -	0.00	2.2E-01	0.97 -	0.09	6.0E+00	0.19 -

References

Doppler, T., Camenzuli, L., Hirzel, G., Krauss, M., Lück, A., and Stamm, C.: Spatial variability of herbicide mobilisation and transport at catchment scale: insights from a field experiment, Hydrol Earth Syst Sc, 16, 1947-1967, https://doi.org/10.5194/hess-16-1947-2012, 2012.

Leu, C., Singer, H., Stamm, C., Müller, S. R., and Schwarzenbach, R. P.: Variability of Herbicide Losses from 13 Fields to Surface Water within a Small Catchment after a Controlled Herbicide Application, Environ Sci Technol, 38, 3835-3841, https://doi.org/10.1021/es0499593, 2004.

Pionke, H. B., Gburek, W. J., and Sharpley, A. N.: Critical source area controls on water quality in an agricultural watershed located in the Chesapeake Basin, Ecol Eng, 14, 325-335, Doi 10.1016/S0925-8574(99)00059-2, 2000.