Answer to Anonymous Referee #1

We would like to thank Referee #1 for the constructive and valuable feedback. We mostly agree with the referee's comments and will address them individually in the following. We still think that the manuscript part about the up scaling to the national level is important and scientifically interesting and will explain our arguments in more detail below. Nevertheless, we agree on the criticism brought up concerning to this part and proposed some adaptations to address it.

L42: Please consider rephrasing to: "relevant process have to be understood".

We agree on this. Will be rephrased accordingly.

L116-117: I am curious to why shortcuts that drain into surface waters or treatment plants are treated same by the model. If you are looking at pollutant transport, shouldn't there be a difference?

We agree that there is a difference between these two processes. In waste water treatment plants pesticides are removed to a certain degree. However, during heavy rain events (i.e. the point in time when the largest pesticide loads are expected in shortcuts) rain water is often not reaching WWTPs but directed to surface waters through CSO. Pesticides have been reported to transport pesticides (Mutzner et al., 2020) and we expect the transport via CSO to be very similar to "normal" shortcuts. In addition, from our field studies we know transport via WWTPs/CSO is less important. Only 12 % of the inlets mapped drain to WWTPs/CSO while 87 % drain to surface waters (see L382-384). We therefore do not expect these differences to have a major influence on our results and decided to neglect them.

L182: Please consider rephrasing to: "In order to better understand: : :"

We agree on this. Will be rephrased accordingly.

Figure 2: Please define WWTP/CSO in the legend.

We will add a definition for these abbreviations in the legend.

L236: What are internal sinks?

Internal sinks are depressions from where water cannot flow on the soil surface to a receiving water, but infiltrates water locally. This term is for example also used in Frey et al. (2009).

L258: Regarding the connectivity model: Maybe I missed something, but how do you go from the upslope dependence output raster to defining if a cell is directly or indirectly connected?

For every Monte Carlo parameter combination, we ran the tool "D-Infinity upslope dependence" three times. Each time we used a different type of recipient area (surface waters, shortcuts, infiltration areas) as an input. This resulted in three upslope dependence output rasters – one each for each type of recipient areas. These rasters then defined if a cell is directly, indirectly, or not connected.

L265: How was this "carving" performed? Did you use a stream burning algorithm?

We carved the recipient areas (as defined in L243-257) into the DEM by rasterizing topographic data and lowering the corresponding cells. For the recipient area type "surface waters" this corresponds to a (very simple) stream burning algorithm. We did not impose an additional gradient towards the stream as it is done by some stream burning algorithms. We think that using such a gradient would not change our results, since the river course of the raw DEM and the burnt-in river course align very well.

L274-275 & Table 2: The 2m limit for the maximal flow distance seems unrealistically low. Is there a reason why you chose this value?

Yes, we agree that this is unrealistically low. However, for the maximal flow distance we simply calculated all possible values. Since our DEM has a resolution of 2m, this was simply the shortest flow distance that could be calculated. For analysing the effect of flow distance on our results, we used 100m, 200m, 500m and infinity as boundaries (see L503 and Figure S24). Accordingly, this lower limit has no relevance for the final results.

L303: I think a reference to Beven and Kirkby (1979) should be given where you explain the TWI, not Tarboton 1997.

We will adapt this.

L329: What model output data did you use for the regression? The median of the MC simulations per catchment? Results from all simulations? […]

For finding the best explanatory variables, we used the median of the MC simulations per catchment. For the extrapolation to the national scale we used the results from all MC simulations.

We will adapt the manuscript as follows:

[329-331]: "We created a linear regression between each of those catchment statistics to the median fractions of agricultural areas directly, indirectly, and not connected to surface waters, as reported by the LSCM ($f_{LSCM,dir}$, $f_{LSCM,indir}$, $f_{LSCM,nc}$)."

L329 (cont.): Here I must say that I found using the NECM as an explanatory variable somewhat strange. If I understand this right, you are fitting a linear model to predict the outcomes of your model (LSCM), based on the results of another model (NECM). But if the latter is such a good predictor of runoff connectivity, couldn't you just

recommend using it at national scale? At least until you have enough data to parameterize the LSCM for all of Switzerland?

We indeed think that the NECM is a good (but not a perfect) predictor of runoff connectivity. However, how the NECM performs in describing our data could only be evaluated in retrospect. A priori we did not know that nor did we know whether including other variables could improve on the predictions. Furthermore, our results demonstrate that we have to transform the NECM data to better represent our observations. In summary, the value of the NECM and its limitations for predicting connectivity based on our observations could only be evaluated by building an independent model for comparison.

The NECM model is currently used in practice (e.g. by farmers and local authorities) and we still recommend to continue using it, since it can a) be used for pinpointing critical source areas within a catchment, and b) is a good predictor for connectivity risk in relative terms (e.g. which catchments have a high risk for indirect connectivity). However, looking at connectivity in absolute terms (e.g. which fraction of a catchment is connected directly) our model improves the predictions of the NECM by using additional information from the data we gathered in our 20 study catchments. (Note: We think that the NECM and the NSCM are in fact different from each other and show this in our comment to L547-551.)

In order to parameterize the LSCM for all of Switzerland we would need a map with shortcut locations for the whole country. We don't expect that such a map will be available in near future. Therefore, we had to rely on other nationally available data.

L355: Is meadow the correct term here?

After checking again with a native speaker, this includes both, meadows and pasture. We will adapt this accordingly, and use the term "meadows/pasture".

L436-437: How?

The directly, indirectly and not connected areas were simply a result of the MC analysis using the surface runoff connectivity model. We suggest to replace this sentence by:

[436-437]: "From the Monte Carlo analysis of the surface runoff connectivity model, we obtained an estimate for the fractions of agricultural areas that are connected directly, indirectly, or not at all to surface waters."

L443: Here I had the impression you are changing the language with which you describe negative ("While certain areas change their classification: : :") or positive ("for other parts results are very consistent)" results. Please consider rephrasing.

We are not 100 % sure if we understood this comment correctly. But we suggest to rephrase this to:

[L443:] “The classification of certain catchment parts is changing depending on the model parametrisation (e.g. letters A to C). However, for other parts, the results are consistent across the different MC simulations (e.g. letters D to F).”

L443 (cont.): Anyway, I found these results quite interesting. Would it be a good idea to look at where the model is consistent and where it is not? I mean, considering model uncertainty, which fields are consistently identified as highly connected? Could these areas be regarded at higher risk of pollutant transport than others? Moreover, if you find out where the models are inconsistent, you can try to figure out why?

Areas that in reality pose a high risk for surface runoff are not necessarily consistently classified as connected by our model. Whether a high-risk area is consistently classified as connected by the model, depends strongly on how well the DEM is able to represent the responsible flow path, given the DEM uncertainty (resolution and elevation errors) and the uncertainty of infiltration processes (e.g. which flow length leads to infiltration in a forest). Flow paths that depend on the coarse terrain structure and on large landscape features, are represented well (and the connectivity classification is very consistent). However, flow paths that depend on microtopography or small scale landscape features, underlie larger uncertainties (and the connectivity classification is less consistent).

As an example, you could imagine a field close to a river with a steep slope towards the river and without any protection by forests or hedges. The DEM is easily able to reproduce the flow path in this case and the field will be consistently classified as a high-risk area.

Another example could be the case of a field with a steep slope towards a narrow drained farm track. A small ridge along the center of the road stops the water from crossing the road and drains it into the next inlet. This makes the close-by field a high-risk area. However, depending on the model parametrisation, the model will not be able to reproduce the draining effect of the road, since the DEM resolution is too small. Accordingly, this field will not consistently be classified as a high-risk area in this case.

As a consequence, the model can be used to identify some higher-risk areas (which are the areas which are consistently classified as connected). However, many areas that are not consistently classified as connected may pose a similar or higher risk. In this context, it is also important to say that also other factors besides connectivity have an influence of the risk posed by a certain area.

Figure 5: I didn’t understand the colour-ramp bars in the figure legend. Are they necessary?

When using a D-infinity flow algorithm, the upslope dependence of a raster cell is not classified as “dependent” or “independent”, but receives a certain probability to be drained into the receiving area (see also L269-270). E.g. the model MC28 shows some orange areas west-southwest of the letter B. For this model realisation, these areas are indirectly connected with a certain probability ($0 < p_{direct} < 1$) at the same time not connected with a certain probability ($0 < p_{notconnected} < 1$). The sum of the probabilities (p_{direct}, $p_{indirect}$ and $p_{notconnected}$) equals 1.

L547-551: If you propagate the uncertainty in your linear model (e.g. by simulating posteriors of the slope and the intercept and then bootstrapping model predictions), it is likely that these differences will be within your error bands. I guess my question here is: is your extrapolated national model sufficiently different from the NECM to justify its usefulness?

We agree on this comment and think that the bootstrapping approach is a good approach for addressing the uncertainty. We therefore bootstrapped our linear model and calculated the distribution of the mean of the fractions reported by the NSCM. Since the approach uses a lot of computer capacity we could only run our model for 50 random samples so far, but we will increase the number of random samples to 100. The figure below shows distribution of the mean area fractions resulting from the bootstrapping approach. (Same plot as in Figure S27, but showing the distribution of bootstrapped means.) These preliminary results show that the fractions of directly and not cot connected crop areas reported by the NSCM are significantly different from the results reported by the NECM.

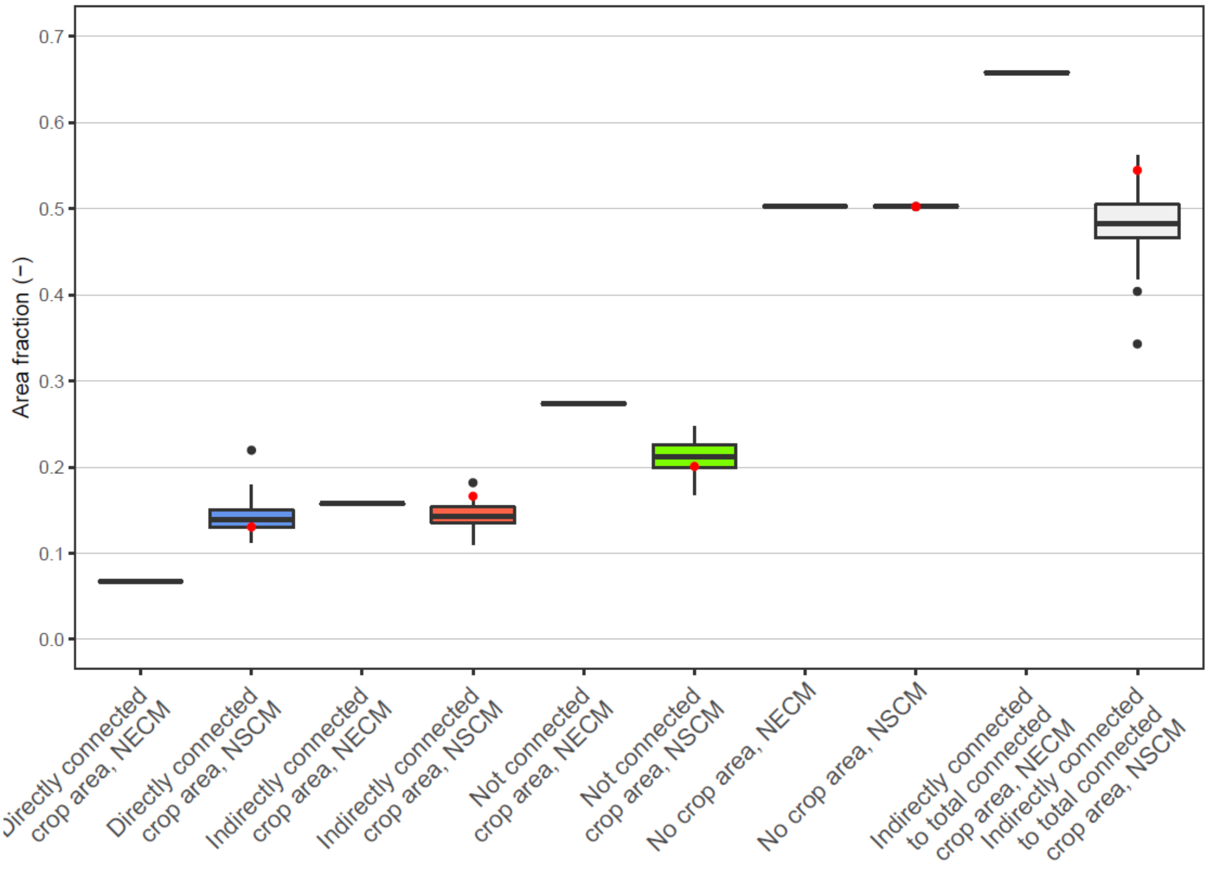

Figure S28: Area fractions reported by the NECM and distribution of the bootstrapped mean area fractions reported by the NSCM. Directly, indirectly, and not connected crop areas per total agricultural area, non-cropping area per total agricultural area, and indirectly connected crop area per total connected crop area for all catchments in Switzerland. The red dots report the means reported by the NSCM without using a bootstrapping approach.

We suggest to update the figure above, as soon as the bootstrapping results are available, and to include it as Figure S28 in the manuscript. Additionally, we suggest to modify the manuscript as follows:

[L338-339, additional sentence]: In order to address the uncertainty introduced by the selection of our study catchments, we combined this model fit with a bootstrapping approach.

[L1050 (SI), additional sentence]: To address the uncertainty introduced by the selection of our study catchments, we bootstrapped our model 100 times. For each of the bootstrapping iterations 20 of our study catchments were resampled randomly.

In addition, the numbers reported in L547-551 will be updated based on the results of the bootstrapping approach and a reference to Figure S28 will be added.

Figure 3 will be updated as follows:

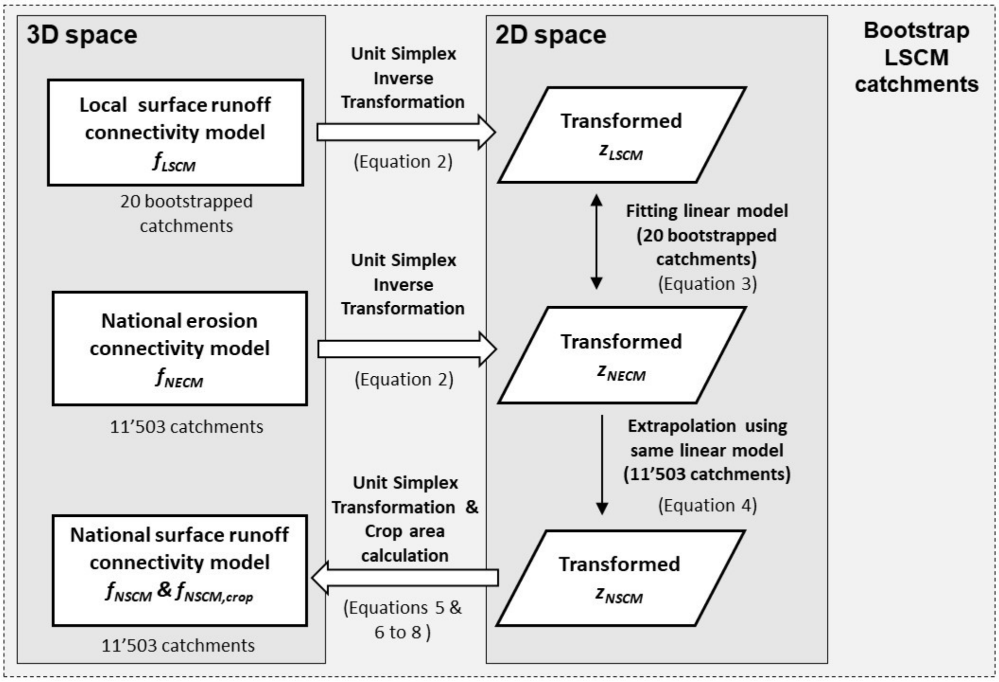

Figure 3: Extrapolation of the local surface runoff connectivity model (LSCM) to the national scale (NSCM) using a 341 **unit simplex transformation approach.**

L652: I am curious: which kind of sink filling algorithm would you recommend in this case?

We don't think that there is a single sink filling algorithm that can deal with this problem in general. In our model, we incorporated our process understanding and our knowledge from field observations to come up with a sink filling algorithm that seems realistic to us (see also L280-282). Additionally, the sensitivity analysis provided insight into how the choice of the sink depth parameter influences our results. For other projects, we would recommend a similar procedure adapted to the question to be answered and meaningful for the topographic characteristics of the landscape.

L693-697: Here you explain the improvements of the NSCM over the NECM regarding the representation of runoff connectivity, which was helpful. While I agree that the information on crop statistics might help your model, I do not see how the national map incorporates the advantages of the LSCM (i.e. all the impressive field data you collected). In the end I have the impression that this upscaling doesn't do justice to all the work you went through in the small catchments. Moreover, while you appropriately

represent the uncertainty of the LSCM, this is somewhat neglected in the extrapolation to the national scale. Would you not expect greater errors in the NSCM than in the LSCM?

We agree that the NSCM is not able to incorporate the main advantage of the LSCM, which is the availability of field data on shortcut locations. Obviously, we expect larger errors for the NSCM than the LSCM, since the NSCM is simply an extrapolation lacking additional field data besides the data from the 20 catchments. However, since the empirical data used for the NECM were extremely sparse, we wanted to use the results of our field study to come up with an adapted version of this national model based on our field data. We think that this modelling step is a scientifically important step independent of its result. If the resulting NSCM were very similar to the NECM, this would give additional validity to the existing NECM. If the outcome were different, this would give insight whether it over- or underestimates the actual connectivity. This cannot be known a priori but needs the actual comparison.

In our case, we concluded that the models are similar with respect to the areas reported to be indirectly connected. However, the differences reported for the directly connected areas (and also the not connected areas) are quite substantial. For the fractions of directly connected areas, the median of the NSCM corresponds to approximately the 75 % quantile reported in the NECM (see Figure S27).

We agree that the uncertainty was not included sufficiently in the extrapolation to the national scale. We addressed this partially by running our extrapolation model for each of the 100 Monte Carlo runs of the LSCM (see L1049-1050 of the SI). However, we agree that we did not address the uncertainty introduced by the catchment selection. We would like to address this issue by including the results of the bootstrapping approach that you suggested in L547-551.

References

Frey, M. P., Schneider, M. K., Dietzel, A., Reichert, P., and Stamm, C.: Predicting critical source areas for diffuse herbicide losses to surface waters: Role of connectivity and boundary conditions, J Hydrol, 365, 23-36, 2009.
Mutzner, L., Bohren, C., Mangold, S., Bloem, S., and Ort, C.: Spatial Differences among Micropollutants in Sewer Overflows: A Multisite Analysis Using Passive Samplers, Environ Sci Technol, 54, 6584-6593, 10.1021/acs.est.9b05148, 2020.