We appreciate the anonymous referee for reviewing our manuscript. The comments encouraged us to advance the content and clarity of our study. Please find below a point-by-point response to the referees' comments (comments of the reviewer in black and our response in red)

On behalf of all authors,

Yours sincerely,

Maximilian Oskar Kottwitz

I have read and reviewed the manuscript "Equivalent continuum-based upscaling of flow in discrete fracture networks: The fracture-and-pipe model" by Kottwitz et al. The purpose of the study is to add the effects of "intersection flow localization" into equivalent continuum models. The authors argue that these localized effects are important to accurately capture the effective permeability and present several test cases. Overall, I found the presentation of the paper to be slightly confusing and am not left with a motivation to integrate such a model into my own continuum model as it seems like a solution to a problem that wouldn't exist in practical applications.

A few high-level comments:
This could be my own ignorance but I'm not clear on why you would use an ECM model over a DFN model with an impermeable matrix in the first place. Modern computing power allows us to simulate DFN models with thousands (and more) fractures. At a certain number of fractures, the ECM will actually be harder to accurately simulate because the resolution needed to preserve the topology of the DFN will quickly approach very low mesh sizes as the fracture network density increases. While some metrics might actually not be affected by high network density and artificial connections in the ECM, transport metrics like breakthrough of solute certainly would. Furthermore, there is no comparison between the ECM models presented here and DFN models in terms of effective permeability. Could that be used as an additional point of comparison rather than to the authors' own model? Previous ECM models (many cited within this paper) have been compared favorably (or unfavorably) to DFNs.

> From our point of view, the fact that ECM's produce similar effective permeabilities as DFN's has sufficiently been demonstrated by the studies of Leung et al., 2012, Hadgu et al., 2017, and Sweeny et al., 2020 (all cited in this paper). We thus haven't seen the necessity to repeat this work in our studies. Our main intention rather was to promote existing ECM methods by incorporating the effects of IFL and local permeability anisotropy in a massively parallelized manner.
> Due to the demonstrated comparability, we think that choosing an ECM or a DFN method for flow-based upscaling in fracture networks is mainly dependent on the preference/experience of the user, the available computational resources (e.g., single workstations or HPC clusters), and the natural application (deterministic DFN structure or stochastic DFN's). In our opinion, the big advantage of massive parallelization (up to 10'000s of cores) with efficient multigrid solvers make continuum methods highly suitable for situations, where very large numbers of simulations must be conducted (i.e., in Monte-Carlo-like sensitivity studies or potentially in adjoint based inversion approaches). For example, with our parallelized approach, ECMs with a resolution of 256 cubic voxels can routinely be solved on a decent workstation (here, 16 cores) in about a minute. This resolution was enough to adequately capture the structure of our test DFNs from section 6 (10000 fractures). If higher fracture densities are considered, we can routinely simulate ECMs with a resolution of 1024 cubic voxels on HPC systems (with 512

cores), which is enough to capture 3 orders of magnitude size scaling of fractures and very high fracture densities (e.g, with an average fracture spacing to system size ratio of 1e-4).

Yet, we agree that a direct comparison of effective permeabilities obtained by our and the DFN method would indeed be interesting, but not feasible given the brief time period (4 weeks) we received from Solid Earth to make corrections to our manuscript. However, if you know about existing DFN benchmark studies, we are willing to compare (1) the time spent on meshing of and simulating flow in a DFN structure as well as (2) the effective permeabilities resulting from DFN-software packages to our presented approach, favorably for DFNs with very high fracture densities (~ 100'000 fractures).

My biggest issue with this paper is the actual verification/validation of the IFL scheme. I encourage the authors to think about a way to clearly articulate: (1) when the IFL matters, (2) why it matters, and (3) and how much it matters (in a quantitative way). As it stands, there's some discussion points related to these, but they're not sufficient. In general, the authors have shown that a difference exists (e.g., in Figure 4), but not shown that this difference is indeed physically correct. I found section 5 to be very convoluted and not convincing with the self-comparison.

This point was also raised by the first reviewer, and we fully agree on the necessity to clarify the scale effects of IFL in a quantitative manner. In section 5, we have demonstrated that IFL increases the permeability of systems at the local ECM scale (system size of ~10times the fracture aperture, i.e., at fracture scales) if the intersection aligns with the applied pressure gradient. We did this by comparing the z-component of an analytically derived permeability tensor to the numerically computed permeability from high-resolution Stokes simulation (the benchmark). To explore the effects of IFL on larger scales (scales that cannot be resolved with current imaging techniques, i.e. larger than a few decimeters), we conducted effective permeability computations with the method described in the appendix on ECM's of several test-DFN's discretized with (1) the ECM method presented in section 4 (fracture-and-pipe method) and a fracture only ECM method (i.e., the method from section 4 without the intersection parametrization), similar as the ECM methods presented by Hadgu et al., 2017 or Sweeny et al., 2020. We added a new section in the revised manuscript that describes the details of this parameter study (section 6):

With this section, we hope to properly address the points indicated by both reviewers. We identified orthogonal fracture formations (so-called cross joints) as a natural candidate where IFL potentially matters most. With the conducted parameter study, we quantified the effects of IFL at these systems with various scales, apertures, fracture densities and matrix permeabilities. The general conclusion of this is that only for systems that have fracture porosities in the range of 30% (P33 is a non-dimensional quantity that combines all of the above-mentioned measures), IFL effects on the effective permeability become noticeable.

This lead to following changes in the abstract, starting from Line 13:

"Our approach is compared against an ECM method that doesn't take IFL into account by performing ECM-based upscaling with a massively parallelized Darcy-flow solver capable of representing permeability anisotropy for individual grid cells. While IFL results in an increase of permeability at the local scale of the ECM cell (fracture scale), its effects on network scale flow are minor. We investigated the effects of IFL for test cases with orthogonal fracture formations for various scales, fracture lengths, hydraulic apertures, and fracture densities. Only for fracture porosities above 30 %, IFL starts to increase the system permeability. For lower fracture densities, the effects of IFL are smeared out in the upscaling process. However, we noticed a strong dependency of ECM-based upscaling on its grid resolution. Resolution tests

suggest that, as long as the cell size is smaller than the minimal fracture length and larger than the maximal hydraulic aperture of the considered fracture network, the resulting effective permeabilities and anisotropies are resolution-independent. Within that range, ECMs are applicable to upscale flow in fracture networks."

This also lead to the following changes in the introduction at Line 78:

"Using these results, we extend the current state-of-the-art methodology for equivalent continuum representations of DFN's to account for IFL in a quantitative manner and analyze its impact on effective permeability computations at fracture and network scales."

We changed Lines 285 - 295 of the discussion to:

"Including a pipe-flow model into the ECM generation process improves the representation of permeability anisotropy therein and can have impacts on overall permeabilities as well. For example, at the scale of the intersection itself, it significantly modifies the shape and absolute values of the permeability tensor (figure 4). However, the errors of the intersection benchmark (2.7 % and 7.9 % for the fracture-and-pipe and fracture-only model, respectively) indicate that, from a statistical perspective, the effect of IFL on overall permeability is usually a second-order effect. This is because the fracture-only ECM discretization approach by default accounts for the increased permeability of intersection cells in an isotropic manner, simply by the summation of the individual fracture contributions per cell. The parameter study from section 6 shows that incorporating the effects of IFL into an ECM-based upscaling approach is only necessary for fracture systems that (1) produce a significant amount of fracture intersection (i.e., for fracture systems with two fracture sets of strongly different directions) and (2) have sufficiently high fracture porosities (above 30 %), regardless of scale. Achieving the latter in our test-scenarios is strongly coupled to large fracture apertures (in the order of 1 cm, see figure 8) and a ratio between average intersection and system size close to or above unity (i.e., intersections fully penetrating the system). So, if two dominant orthogonal fractures with large apertures form an intersection that is penetrating the whole system along the direction of flow, the effect of IFL influences the systems effective permeability. Due to the non-linear radius-permeability relation, this may become more important for fractures with apertures above 1 cm. However, for apertures above 1 cm, Reynolds numbers can easily exceed the critical threshold of unity (e.g., Zimmermann and Bodvarsson, 1996) which would require non-linear concepts to relate fracture and intersection permeability (e.g., Forchheimers law), as well as Navier-Stokes rather than Stokes simulations."

This lead to following changes in the conclusion at lines 348 - 351:

"By assessing the permeabilities of fracture intersections numerically, we could demonstrate that for system sizes close to the approximated pipe radius (mm to cm), the effect of IFL on permeability can be almost one order of magnitude. At network scales (m to km), the impact of IFL on the systems effective permeability is generally minor. Only for fracture systems with high fracture porosities (above 30 %) IFL effects become noticeable.
Furthermore, we investigated the resolution dependency of current ECM-based upscaling approaches, as the cell size with which the ECM is discretized represents the most crucial aspect for the accuracy of ECM-based effective permeability predictions."

Because of this, we also decided to change the title of the paper to:

"Investigating the effects of intersection flow localization in equivalent continuum-based upscaling of flow in discrete fracture networks"

The relationship between section 6 and the previous parts of the paper is not clear. My interpretation of section 6 is that it is showing the needed resolution of the ECM mesh size to have a solution converge. I think this is interesting, but not clear how this is related to the effects of the IFL. Seems like two different papers here.

In our opinion, checking the consistency of an employed method is inevitable in scientific works, especially if problems like resolution dependency are well-known to exits. To clarify this, we changed the beginning of section 6 (Lines 240 -244) to:

"The resolution dependency of ECM methods to upscale the permeabilities of fracture networks is a crucial aspect that has to be considered to provide accurate upscaling results. Artificial connectivity is one of the main issues that arises, if the resolution of the ECM is insufficiently low as demonstrated by Jackson et al., 2000 and Svensson et al., 2001. Their results suggest that we can expect accurate upscaling results, only when the resolution of the ECM is sufficiently large to resolve the structure of the DFN (i.e., a maximum of two fracture segments and one intersection per cell)."

We also changed the title of former section 6, now section 7 to:

"Resolution dependency of ECM methods"

Some more specific comments:
L42: With modern computing power, DFNs are commonly used at reservoir scale. I'd argue the difficulty is including matrix properties in DFNs/DFMs and meshing multidimensional DFMs.

We changed the sentence at Line 42 to:
"Improved discretization techniques with individual fracture treatment (DFN method), inclusion of matrix properties in multi-dimensional meshes (discrete fracture and matrix - DFM - method) or multi-continuum methods come at the cost of high computational expenses."

L74: Does a local increase in effective permeability really matter? Isn't the global effective permeability usually the quantity of interest? Which is being reported here?

You're right, the effective permeability of the system is the quantity of interest. We changed the sentence to:
"As a consequence, the systems effective permeability should increase by a certain amount due to a local permeability increase within the intersection."

I think Figure 5 is mentioned in text before Figures 2—4.

We changed the wording at Line 110 to:
"… (as will be shown later by a comparison of errors to the result at largest resolution in plot b, figure 5)"
If this still represents an issue, we can simply remove the line at this stage.

L155: Not all ECM models use regular grid spacing. Ref L575 used octree grids.

Indeed, Sweeney et al., 2020 use octree refinements for their grids. We changed Line 155 to:

"The goal of the ECM method is to generate a 3D of computational cells that contains a symmetric, positive definite permeability tensor that is based on the fractures and their intersections. For simplicity, we prescribe a regular grid with constant x-y-z spacing dx instead of octree refined grids as utilized by Sweeney et al., 2020."

L175-185: I'm not following how the calculation of V_f is completed regarding this point counting. Why not just directly compute the intersection areas of the fractures with the cells? It would be more accurate. Also why is the z entry in the permeability tensor (13) 0 and not 1? Is the matrix permeability within an intersected cell taken into account?

Calculating the exact intersection areas of each fracture with a cell would of course be more accurate. However, it is very expensive from a computational perspective to calculate the intersection polygon of a disc with a 3D box, especially if cell size is very small compared to the fracture size. The point counting approach we use is much faster from a computational point, because we simply represent each fracture by a grid of equally spaced points and find the respective fracture points that belong to each computational cell. By multiplying the area of the grid spacing (dp^2) with the hydraulic aperture we get the volume fraction for the specific subregion of the fracture that is represented by the point. Multiplying this with the number of points per cell gives the total fracture volume per cell. Finer point spacing (dp) of course results in a better fit, which we demonstrate in figure 4, plot c.

The z-entry in the permeability tensor is zero, because we assume an anisotropic permeability tensor for a fracture that is aligned with the x-y plane. In that case there is no connectivity in the z-direction and hence no permeability. This approach results in the same permeability tensor as produced by the approach of Chen (1999). In eq. 14 we assume the same for an intersection that is aligned with the x-direction, hence no connectivity and permeability in the y- and z-direction.

Figure 4: I don't understand the point of this plot (c). What is the purpose of showing the norm of the tensor? It's showing a difference between fracture & pipe and just fracture, but what should we glean from this? What is the ground truth for comparison? How do we know which one is correct? Too much is being assumed of the reader here and elsewhere. Take us through the key points. There's interesting work here, but it's hidden in lots of verbosity.

The purpose of plot c in figure 4 is to demonstrate the quantitative difference in permeability magnitude of the resulting permeability tensors for a single cell ECM disretized with both methods, the fracture-and-pipe and the fracture-only approach. We've chosen to compare the norm of both tensors, as they also incorporate the difference of the off-diagonal components of the tensor.
Hence, the main point we want to make here is that we can actually have a large difference in permeability at the local fracture scale between both discretization methods, which is also evident by looking at the visualized tensors in plot b.
Admittedly, we haven't described the fracture-and-pipe and the fracture-only in the caption which is why this part can be confusing. We therefore expanded the caption of figure 4:

"c) presents the norm of the permeability tensor K_ijk as a function of the ratio between the ECM grid spacing dx and the initial point spacing dp for the discretization approach described in section 4 (the fracture-and-pipe model) and an approach, where we didn't take the IFL parametrization into account (i.e., leaving out eq. 14 in the discretization procedure, hence the name fracture-only)."

The presentation of section 5 is not very clear. What exactly is the benchmark solution being used for comparing the two different ECM approaches? How was this error metric (eq. 15) chosen – isn't this just a measure of convergence – why is it being compared between the two ECM approaches? The way I read this is that the method in section 2 is being used as the benchmark. So my two questions are then: (1) has the solver in section 2 itself been verified? And (2) Doesn't the method in section 2 include the "fracture & pipe" model (or is it DNS?)? Meaning that comparing the ECM without the intersection permeability will of course yield a worse fit.

The one thing I noticed throughout this paper is the fracture apertures are quite large (mm to cm). Most practical applications the authors listed in the beginning of the paper can include fractures with much smaller apertures. How will this affect the influence of the IFL? This might be interesting to investigate here.

Just a note that it's not "DFN's" and "ECM's", but "DFNs" and "ECMs"

We changed the spelling accordingly throughout the paper.

L262: I think this is the first mention of matrix permeability in the paper. How is the matrix permeability included in the fracture intersection cells if their dimensions are larger than the fracture apertures? Also, how do the effects change with different matrix permeabilities?

Matrix permeabilities are prescribed as isotropic (constant k_xx, k_yy and k_zz values) in the computational grid prior to adding fracture and intersection permeabilities to respective cells. We assume that the prescribed value is already properly rescaled to the cell volume. We incorporated different matrix permeabilities (1e-17, 1e-15, and 1e-13) in the parameter study of the new section 6. There was no noticeable dependence of the IFL effect on matrix permeability.

What are the readers supposed to learn from Figure 9? That changing the cell size changes the permeability? That's obvious based on the upscaling equations. Please explain further. I think it's showing the effects of artificial connections in the ECM, but again, this doesn't mean it's physically correct. It just means it's converging. Please expand on the discussion here.

With this figure and section, we wanted to demonstrate the resolution dependency of ECM methods. We'd argue that if one wants to upscale a structure properly based on an ECM, the permeability should not change as a function of cell size but rather be consistent throughout scales.
The changes for large cell sizes are indeed caused by artificial connectivity. However, they vanish at sufficiently high resolutions. So if the resolution is sufficiently high and artificial connectivity is avoided, we have demonstrated that we actually produce constant/converged permeability tensors with our ECM method. I.e., the permeability does not change anymore with cell-size changes, and which we argue to be the most accurate ECM representation of a DFN.
Determining whether the resulting permeability tensor is physically correct is indeed a challenging issue. We guess that this can only be properly addressed in an extensive benchmark study with a comparison to natural field measurements, which we are actually working on at the moment. However, due to the seminal results of Leung et al., 2012, Hadgu et al., 2017 and Sweeny et al., 2020, we at least assume that the result is at least similar to the results of the DFN method.

In the same light, what is the point of section 6? The topic sentences in the beginning of the section promise some presentation of the accuracy of the method for large DFNs. I see no quantitative measures of error in this section. Just numerical convergence.

Please see previous answers. Using the term "accuracy" was indeed misleading and could cause confusion. We changed the title and the first sentences of the topic to highlight that we address the resolution dependency of ECM methods.