Thanks to the reviewers, we could greatly improve our manuscript and are now hopeful that all stated issues are solved to satisfaction. Most of the manuscript was completely rewritten and we therefore kindly ask the reviewers to have a look at the revised manuscript where all changes have been marked. Citing all changes in this letter would be too messy. Nevertheless, our answers are marked in red and if the changes were just slightly, we still cited them here in green.

To give a broad overview over the changes, here a small summary:

We changed the model setup in a way that the coordinate system follows the maximum melt fraction, which allowed us to zoom into the initial perturbation, which in turn helped us reaching the required resolutions for the compaction length. With this improved resolution we no longer observe channels, but solitary waves. These will build up independently of how slow the segregation velocity is, if the ascend time is high enough. For very high radii a diapir will split up into numerous solitary waves but ascend as a whole, mostly affected by the surrounding matrix.

We got rid of the retention number in our mathematical description and replaced it by the squared ratio of compaction length to model length scale r. Following this the segregation to Stokes velocity ration analysis was expanded by a figure, which shows the results of it for a few initial perturbation maxima. They fit nicely to the observed results of our models.

Reviewer 2

Suggestions for revision or reasons for rejection (will be published if the paper is accepted for final publication)

The submitted revised manuscript by Dohmen and Schmeling systematically investigates magma ascent dynamics in order to capture the transition from the solitary wave regime to diapirism. The authors explore this transition by varying the relative compaction length of the system - here by changing the model extend while keeping the compaction length constant.

The authors addressed majority of the concerns raised during the first round of revisions. However, the new version of the submitted manuscript still suffers from major design issues, both in the content and form. Rather than "time investment and good enough-ness", focus should be on scientific approach and accuracy. To the point, the main story of the manuscript -channelling- is not receivable as such. The authors motivate their revised study by unveiling apparent channelling mechanism occurring while transitioning from the solitary-wave to the Stokes regime. Their argumentation would only be receivable if following a scientific approach, i.e., including more than wishful thinking.

Simple words for simple things; Assuming there is a not yet discovered channelling mechanisms, natural steps would be following: (1) provide a parameter accounting for it; (2) test and report the influence of this parameter in a systematic study; (3) prove the robustness of the suggested results by providing (numerical) convergence tests (physical results should no longer vary with further increase in numerical resolution - proof of a robust numerical implementation) targeting the configuration of interest. To date none of these steps are successfully implemented.

Now, and unless proven otherwise, the underlying equation do not contain any channelling mechanism as such. Thus, the reported channels may rather be the expression of a lack in numerical resolution. This conclusion still confirms the outcome of previous reviews.

Channelling ultimately requires an asymmetry in compaction versus decompaction regimes, obtained upon nonlinear bulk rheology by mechanisms such as e.g., decompaction weakening (Connolly and Podladchikov, 1998; Räss, 2018; Räss, 2019) or brittle failure (Keller, 2013; Yarushina, 2015). Moreover, including the full shear stress tensor for the mixture velocities and total pressure won't produce extra focussing and asymmetry; neither would porosity dependent and even strain-dependent shear rheology. Both may impact the compaction length which may further influence the relative inclusion size, at most.

Now, after following, the remarks about numerical solution wo do not, as supposed above, observe channeling anymore. The finger-like features turned out to be not-sufficiently resolved solitary waves. We therefore got rid of all text passages concerning channeling.

Finally, the effort spent in providing further insights into the underlying physics and mathematical model (Section 2.1) is very much appreciated. However, this new section reports inconsistent derivations. Equation (3) reports de analogy of the fluid momentum equations as a generalised Darcy law that contains de gradient of the fluid pressure P minus the buoyant fluid force ρfg. Equation (4) reports the total force or momentum balance, where the viscous stresses and total (mixture) pressure P equilibrate the total buoyancy force $\bar{\rho}$g. The pressure term in equation (3) represents as such the fluid pressure Pf, while the pressure in equation (4) stands for the total pressure Ptot. Equation (10) and line 96 is thus wrong. Fluid pressure ≠ total pressure (Pf≠Ptot) and CANNOT be eliminated.

While the original study needed some revision, the here submitted revised version addresses none of the early design issues. Instead of providing scientifically robust proofs about potential new transient regimes, it further motivates wishful thinking instead of results.

To accept the claims made by the authors about the existence of an intermediate regime leading to flow channelling while transitioning from solitary waves to diapirism, following steps should be included:
1) identification of a physical and testable parameter accounting for focusing
2) systematically testing and reporting of the influence of this physical parameter
3) numerical convergence test to support the robustness of the numerical results (independent of the chosen numerical implementation)

Due to a new procedure in in modeling our solitary waves, we are now able to zoom into the wave, which helps a lot to reach sufficient resolution for solitary waves. We now restrict our models in the paper to models where the compaction length is at least resolved by three grid lengths.

-- Further detailed comments (line numbers refer to the manuscript version 4):
l.9: Not only size but related to compaction length. Size could be kept constant but change in compaction length may lead to similar results

This passage is no longer part of the paper.

l.19-22: No channels will form. Results seem to report a lack of resolution here. To form channels, one needs an asymmetry in compaction versus decompaction rheology. This asymmetry one does not get with the shear rheology. Including porosity dependence in bulk and shear rheology may induce a change in compaction length but no asymmetry.

s.a.

l.45-47: It's the same, as compaction length and radius are interconnected. Changing compaction length using Rt may be the same than changing the bulk to shear viscosity ratio, which will ultimately also impact the compaction length.

Yes, ultimately, we both are changing the compaction length. But in contrast to Scott (1988), we use a porosity dependent viscosity, while he uses a constant viscosity ratio. We changed the sentence to:

Scott (1988) already had a look at a similar scenario. He calculated porosity waves changing the compaction length by altering the constant shear to bulk viscosity ratio, while we want to change the radius of a partially molten perturbation in terms of compaction lengths but keeping the porosity dependent viscosity laws the same

l.68-72: Boussinesq approximation. There is no need for abbreviation since you only use "BA" twice. Also, the wording could be improved here as it is not very clear in the current form.

We got rid of the abbreviation, but we don't know how to improve the wording, as the sentence seams not to be too long and difficult.

---- Section 2.1
Equations (3) and (4) have a pressure issue. How can the same P both be used in the Darcy flow and in the total momentum balance, once relating to fluid density, once relating to total density? Needs revision, modification and clarification.

Equ 10 contains the fluid pressure, indeed, and not the total pressure as claimed by the reviewer. A rigorous derivation of this equation from basic principles can be found in McKenzie (1984, J. Petr. 25, 713 – 765) in Appendix A. In that Appendix equ A9 gives the interphase force and contains the fluid pressure. This interphase force is inserted into equ A7, the momentum equation of the matrix. Inserting also other terms into that equation McKenzie arrives at A16 and furthermore A21 which then is the momentum equation of the mixture (see the mixture density) in the limit of low viscosity fluid (see the deviatoric stress tensor which neglects fluid shear stresses). The pressure in that equation is still the original fluid pressure. McKenzie eliminates the fluid pressure when arriving at A23, and that is what we are doing. In the late 80's many papers used that way eliminating the pressure…

Here from my personal notes on that issue:

It should be noted that the fluid pressure P also occurs in the momentum equation for the mixture, and the intrinsic (averaged) matrix pressure does not explicitly occur. Usually it is different from the fluid pressure (also for the case of neglecting surface tension). If $\nabla \cdot \vec{v}_s = 0$ the intrinsic matrix pressure, the fluid pressure and the effective mixture pressure become equal. In the limit of zero melt porosity the effective mixture pressure and the intrinsic matrix pressure become equal, and the P in equ 4 approaches the intrinsic matrix pressure. This happens smoothly as long as in the limit of vanishing porosity $\nabla \cdot \vec{v}_s$ approaches faster to zero than η_b approaches infinity.

Minor notation issue: this section could be enhanced with notation homogenisation. Either adopting the ∇ or ∂/∂x notation. Also, some i,j,k may be missing if including those.

The ∇ notation is used everywhere but in the momentum equation of the mixture and for the viscous stress tensor. Even though these equations look quite messy in the other notation we think it might be more clear what is done this way.

---- Section 2.3
What linear and nonlinear absolute and or relative tolerances are used (criterion to stop iteration and accept the current solution before starting the next physical time step)?

We think this paragraph is already very detailed and does not need more technical insights. Anyways, we now cite Schmeling et al. (2019) here where the used code is described in more detail.

---- Results Section:
l.233-234: What is observed in Fig.1a-d is simply the evidence of the problem's internal length scale, the compaction length. Although the initial melt anomaly becomes larger, flow still re-organises within a blob of characteristic size given by the compaction length.

Yes, that is true, and we do not state that this is a problem. The shrinking of the waves in comparison to the initial perturbation is totally expected and we just describe this observation. To make it clearer, we added:

This shrinking of the wave in the model is a consequence of reducing the compaction length. The resulting solitary waves have always the same size in terms of δ_c and become smaller compared to the initial perturbation.

l.244-245: There is no channelling. You may see focusing of melt from an original distribution into a new circular one, but the channels you refer to are numerical artefacts. A model including at least >10 grid points to resolve the channel width will be needed to validate your statement.

s.a.

l.245-267: If you can both model Stokes and porosity waves, why do you need analytical velocity formulation for your lines in Fig. 2? Basically, Fig. 2 should be obtained by tracking your results, or at least some of those. How does one know whether your results match the curves on Fig. 2? One does not need to run any numerical model to reach your conclusions if they're based upon evaluating semi-analytical solutions from literature.

We use these semi analytical solutions from literature, because this solution is for a perfect solitary wave and benchmarked, whose results fit quite nicely to our observations made in this paper. Nevertheless, your point is valid and the reader does not know, whether our waves match these results, but we already compared our waves with the results of Simpson & Spiegelman (2011) in Dohmen et al. (2019). We now state this in the manuscript:

These semi analytical solutions fit quite nicely to our solitary wave models, as already shown in Dohmen et al. (2019).

l.319-325: Be careful with statement. You still do not resolve the small instabilities and those may just be small blobs if properly resolved.

s.a.

l.339-341: Receivable statement upon successful proof that those are not numerical artefacts.

s.a.

---- Numerical issues:
As long as no convergence test neither benchmark (available in e.g., the Appendix of Räss (2019) and Keller (2013)) is provided, the reported results, especially channels, could well be under-resolved and thus numerical artefacts.

We now added the small resolution test to section 2.3., where we can show that the features we observe can still be seen for not sufficiently resolved solitary waves. Anyways, a solitary wave benchmarking is not carried out in this publication, but we carried out a bigger resolution test, and compared our wave solutions to the semi-analytical solutions of Simpson & Spiegelman (2011), which we now cite within the numerical tests in this publication:

The model resolution is a critical parameter in this kind of numerical calculations and should always be kept in mind. With increasing length scale ratio, the compaction length in the model gets smaller and the resolution needs to be increased to keep it equally resolved.

According to several authors (e.g. Räss et al., 2019; Keller et al., 2013) the compaction length should be at least resolved by 4-8 grid points to accurately solve solitary waves. For small length scale ratios this is no problem, where, with a model resolution of 201×201, up to nearly 30 grid points per compaction length can be achieved. The highest resolution our code can run is 601×601, which is enough to resolve the compaction length by three grid points for the model with a length scale ratio of 40. Everything above that cannot be sufficiently resolved.

Fig. 1 shows the resulting models for a length scale ratio of 10 for three different resolutions. The pictures were taken after φ_{max} has risen approximately 0.25 times the initial Stokes radius ($t' = 0.25$). With increasing resolution, the maximum melt fraction increases strongly from 101×101 to 401×401 by approximately 20% but the velocity of φ_{max} decreases by 7% (not shown in the figure). Both values converge. Even though the compaction length is not sufficiently resolved in Fig. 1a), one can still observe the main features of the model: A main solitary wave has emerged from the original gaussian perturbation and secondary porosity waves are beginning to emerge within its remains.

The solitary waves modeled with our code have been compared to the semi-analytical solution of Simpson & Spiegelman (2011), and more benchmarking was carried out in Dohmen et al. (2019).

4.1 channelling: As long as bulk rheology is linear; no literature reports any growth of instabilities besides splitting of original wave into new size owing to dynamical change in compaction length. Richardson (1998) shows minor impact of external stresses on blob's shape, but no channelling as such is presented.

s.a.

l.404-405: No channelling here, changes in compaction length will change the characteristic diameter of the spherical wave which needs to be resolved.

s.a.

l.411-414: Unfounded claims. Connolly & Podladchikov (1998) do not suggest following "this upward weakening might not be strong enough to lead to the focusing needed for the nucleation of dykes".

This sentence was deleted, due to the complains above and is not really of importance.

l.422-428: No channelling. As long as there is no asymmetry in viscous compaction versus decompaction, you won't get channels out of blobs. Taking full stress tensor into account and having porosity dependent viscosity will just impact compaction length, nothing else (Räss, 2019).

s.a.

l.439: "the velocities fit quite nicely to the observed model velocities" where does one sees this? You report analytical solution from other authors and your analytical solutions, but nowhere your modelled results. Since your model includes the stress tensor and velocities, it would be very interesting to report those to support your statement and make them receivable.

We now gave the comparison of our results a small paragraph in section 2.3:

The solitary waves modeled with our code have been compared to the semi-analytical solution of Simpson & Spiegelman (2011), and more benchmarking was carried out in Dohmen et al. (2019).

In a single-phase flow case, where the melt is not allowed to move relatively to the solid, the initial perturbation ascends, shortly after beginning, with a velocity of 0.95 times the calculated Stokes velocity, and then slowly decreases as the original Gauss-shaped wave deforms and loses in amplitude.

l. 447-448: A mechanism needs a testable physical parameter, and a verification that this parameter delivers robust and resolution independent results.

s.a.

l. 451: (2) see previous comment.

s.a.

l. 459-463: Good point. Apply it; re-run the suggested simulations with 10 times higher resolutions and longer travel path to convince the reader that you won't get blobs but some real channels being resolved at least with more than 10 grid points.

s.a.

As of the current state, major revisions are warmly suggested.

-- References

Connolly, J. A. D., & Podladchikov, Y. Y. (1998). Compaction-driven fluid flow in viscoelastic rock. Geodinamica Acta, 11(2-3), 55-84. https://www.tandfonline.com/doi/abs/10.1080/09853111.1998.11105311

Räss, L., Simon, N. S., & Podladchikov, Y. Y. (2018). Spontaneous formation of fluid escape pipes from subsurface reservoirs. Scientific reports, 8(1), 1-11. https://www.nature.com/articles/s41598-018-29485-5

Räss, L., Duretz, T., & Podladchikov, Y. Y. (2019). Resolving hydromechanical coupling in two and three dimensions: spontaneous channelling of porous fluids owing to decompaction weakening. Geophysical Journal International, 218(3), 1591-1616. https://academic.oup.com/gji/article/216/1/365/5140152?casa_token=ffbIm7VK8EsAAAAA:L7LuROOcMXTBgosYEdylrae-1rhNCS2E_kVvfn9aOpM3-LnRn5RFtmHEvvFOLvpPlCRssxARrVVa4Zg

Keller, T., May, D. A., & Kaus, B. J. (2013). Numerical modelling of magma dynamics coupled to tectonic deformation of lithosphere and crust. Geophysical Journal International, 195(3), 1406-1442. https://academic.oup.com/gji/article-abstract/195/3/1406/2874184

Yarushina, V. M., & Podladchikov, Y. Y. (2015). (De) compaction of porous viscoelastoplastic media: Model formulation. Journal of Geophysical Research: Solid Earth, 120(6), 4146-4170. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014JB011258

Reviewer 3

Suggestions for revision or reasons for rejection (will be published if the paper is accepted for final publication)

This study focusses on the interesting and relevant question of the trade-off between compaction waves and melt-rich diapirs in partially molten systems of the upper mantle and crust. The topic has been covered in the literature in the past but perhaps there would still be room for a study to systematically investigate the transition between the two well-accepted end-member regimes. However, in my opinion the manuscript in its revised form still suffers from a number of critical flaws that must be addressed fully before publication of the study is advised. The most important issues to be addressed in my view are the following:

- Ratio of length scales as governing model parameter. Previous literature shows clearly that the ratio between the emergent physical length scale, the compaction length (here, "delta"), to the set system length scale (here, "r") is the crucial control on flow regimes between compaction waves and diapirism. If the compaction length is similar or larger than the system length scale, pore fluid segregation is more or similarly rapid than collective flow of both phases as a mixture. Conversely, if the system is much larger than the compaction wave, segregation becomes less relevant and collective flow becomes dominant. Dimensional analysis shows that the square ratio of length scales R = delta^2/r^2 is a dimensionless parameter arising in the governing equations if the system length scale is used to non-dimensionalise length. The parameter in fact arises from taking the ratio of characteristic Darcy and Stokes speeds driven by the buoyancy-contrast between phases (see the recent discussion in Keller & Suckale, GJI, 2019). As this ratio of length scales is at the heart of this study, it is surprising that the authors choose a different, rather more circuitous route in their dimensional analysis of the governing equations. They first introduce the retention number, Rt, not recognising that it is in fact both the ratio of segregation to diapirism speed as well as the square ratio of length scales. Later, apparently as a mere afterthought, the authors bring up the compaction length without putting it into context with their dimensional analysis. They then seek to explain in rather convoluted language how they are varying Rt to obtain an increase in system length scale compared to compaction length. I would regard it as critical to the clarity of their model description and the entire study to instead use a form of dimensional analysis that introduces the governing ratio of length scales clearly from the start and sidesteps the confusing and unnecessary reference to the retention number.

We just never thought of the retention number as a ratio of length scales, but now, as the reviewer stated, we use this ratio as it is much more intuitive than the retention number. We use it in the description of our equations and in the small analysis.

- Numerical benchmarking and resolution testing. One of the main results of this study is the apparent third regime at the transition between compaction waves and diapirism, which the authors characterise as "channelling instability" and discuss in context with rheological melt-shear localisation first introduced by Stevenson (1989). However, as matters stand, the reader can have no confidence that the feature in question is in fact a robust model result rather than a numerical artefact. The numerical setup used in this study is critically flawed. To avoid interactions with boundaries, the domain is 20x larger than the radius of initial perturbations, r. The standard resolution is 200 cells in each direction, meaning r is resolved by 10 grid cells. Unfortunately, the authors then proceed to test parameters where r is roughly equal to or much larger than the compaction length, meaning that in most parameter tests, and notably the ones showing the apparent "channeling instability", the compaction length is not resolved even by one grid step. It would be best practice to present benchmarks of the numerical method against known solutions (or reference published literature if the numerical method has been benchmarked elsewhere). As an absolute minimum requirement, the authors would need to provide a resolution test where it is clear that the solution converges towards a well-resolved geometry, where the compaction length is resolved by at least 8 grid cells. Unfortunately, the authors do not provide either. Nor is it, to my understanding, possible to push the resolution to the necessary levels given the present implementation and model setup.

The reviewer is correct in the statement that the resolution was much too slow to get trustable results. As in the original model setup the required resolution was not reachable, we now changed the model setup a bit. We now let the coordinate system follow the maximum melt fraction in the model, which allows us to zoom into the initial perturbation, as we no longer need model space for the ascend. Due to that we are now able to reach higher compaction length resolutions. Additionally, a new resolution test shows that the same geometries can be expected, even for lower resolutions.

- Discussion of the apparent "channeling instability". Even if the feature in question be confirmed in well-resolved simulations, its discussion in context of rheological melt-shear channeling remains questionable at best. Previous discussions of channelling instabilities demonstrate that it can arise from strong melt-weakening or non-Newtonian stress-weakening of matrix shear viscosity, decompaction weakening or tensile plastic failure. None of these effects are included in the present model. Therefore it is most likely that the feature in question, if sufficiently well resolved, will turn out to be simply a small compaction wave temporarily escaping ahead of the diapir. However, that would not necessarily be the expected outcome, since the diapir rise speed under the relevant conditions should exceed compaction wave speed. Either way, referring to the feature, if it should persist in revised models, as "channelling instability" is highly misleading.

After changing our model setup and being able to increase the resolution of compaction length we no longer observe channels. Because of that we revised our whole results and discussion chapter.

- Quality of language. The use of language and style in this manuscript falls short of expected standards in international journals. Clear and concise language is important to foster unambiguous udnerstanding. It is highly recommended to have the manuscript edited by a native speaker or professional editing service before resubmission.

Some more detailed comments are given as annotations in the attached PDF file.

L20
The abstract has been rewritten.

L79
The pressure is no declared right after the equation and not later.

L92

L95
Yes, the intrinsic shear viscosity is constant in our models and it is now stated within its declaration.

L102
Thank you for pointing out this mistake. The following equations are based on Sramek et al. (2010) and not Sramek et al. (2007) as falsely stated in our manuscript.

L120
Yes, the pressure was already eliminated, but in another equation. Earlier we use equ. 4 to eliminate the pressure in equ. 3. But now we eliminate the pressure of equ. 4 by taking the curl. This way we do not have the segregation velocities in our equation.

L139
We now use, as stated already above, the ratio of compaction length to model length scale r for the equations and the following analysis.

L154
The ratio was flipped back again.

L167
s.a.

L172
We now use the compaction length earlier in the description of our equations and do not need to attach it at the as an afterthought as stated by the reviewer. Therefore, it is now introduced above at the non-dimensionalization, when it first appears.

L177
Thanks to the reviewers point we do it now as he stated. See above.

L179
Yes, we had concerns about the Stokes flow being affected by the boundaries and we needed some space for the wave to travel to be able to observe the evolution. Periodic boundaries would be a way to tackle this problem but would have been quite hard to implement in our already existing code. We now use a moving coordinate system which basically should lead to similar results and was not too hard to implement. The boundary effects to the Stokes flow get tackled by choosing the pre-factor of the calculated Stokes velocity according to a numerical solution of a cylinder rising within another cylinder and applying it to our square model box. Even though in nature "boundary effects" are never far away, we think it would be best our model as theoretical as possible.

L180
We changed A to phi_max as stated by the reviewer, but we still think the description of the initial perturbation fits best in the model setup. Nevertheless, we now give more information on the perturbation, earlier at the non-dimensionalization.

For r the half width of the prescribed initial perturbation, consisting of a 2D Gaussian bell, is chosen. This is reasonable as the rising velocity in our code is best described by the Stokes velocity, using this radius. The exact shape of the perturbation is given later in the model setup.

L183
s.a.

L189
s.a.

L195
Yes, this is correct, but as stated above periodic boundaries would have been not easy to implement. Anyways, with our prescribed boundary velocities we tackle the problem of melt accumulations at the top satisfactory and do not have any problems. We still need to prescribe them with our new moving coordinate system.

L199
We already thought about mirroring the solitary waves on the boundary, but this would probably lead to more problems. With our old model setup, the observed geometries were about the order of just one grid length, which would have probably led to even more erroneous conclusions. Additionally, even calculating just half of the domain would have been not enough to reach the required resolutions.

L204
The model setup was completely rewritten

L210
s.a.

L212
All used equations were referred to during the description of the code. We think that should be enough.

L213
We now cite Schmeling et al. (2019) where the code was described in detail.

L214
The sentence was removed.

L218
We exchanged the word “damping” by “underrelaxation”. We hope that that makes it clearer.

L225
s.a.

L248
s.a.

L251
s.a

L254
s.a.

L293
Sure, we agree. In our code, by default, we solve for advection of composition within the separate phases, and get local evolution of the bulk compositional field. However, in this study we do not distinguish between the chemical composition of melt and solid, melting/solidification is switched off, so we do not have compositional gradients. Pointing this out here may distract the reader.

L318
We now remember the reader that we use a viscosity law that evolves with the melt fraction, while Scott (1988) uses a constant viscosity ratio as model parameter:

This switch from negative to positive mass flux was already observed by Scott (1988), but while he changed the viscosity ratio as an independent constant model parameter, we change the radius and keep the viscosity law the same, still evolving with φ.

L340
s.a.

L345
The benchmark is now presented in 2.3 Numerical approach

L348
Yes, the resolution test could not reach the resolutions required to optimally resolve our models. With our new model setup we can now reach the required resolutions.

L356
s.a.

L364
s.a.