Comments of Reviewer 1 (Cooke)
Replies of Authors
Changes in manuscript

A. The third classification of experiments as 'scale' models might more accurately called scaled crustal models. Some fault block models are scaled and so the term scale models doesn't fully describe the models of this classification. I recommend replacing with the term scaled crustal models.

We used "scale model" as a standing term (e.g. in title of Hubbert's seminal paper), here applied to earthquakes and tectonics.
Fault block models are labscale but not "scaled" in the sense of Hubbert's scaling theory, i.e. showing geometric, kinematic and dynamic similarity. Fault block models do rarely show geometric similarity (instead of representing a certain structural setting at a given geometric scale they represent an arbitrarily oriented and scaled fault interface), typically there is no defined length scale and the deformation is highly exaggerated. Fault block models show no kinematic similarity (which would mean loading rate and rupture speed etc. scale down consistently) and consequently cannot, by definition, be considered dynamically similar. Observations from those models cannot be directly scaled up to nature but only extrapolated qualitatively.
This is why we suggest to "seismotectonic scale model" to better describe these models. This also better describes the multiscale nature of the approaches.

We changed the name of the third category to "seismotectonic scale models" in order to emphasize the multiscale nature and more quantitative applicability of these models.

We changed the text in Sect. 2 to make clear the difference also in terms of scaling:

*"Here, we categorize analogue earthquake models into three groups with decreasing level of abstraction and applicability (Fig. 1, Table 1): (1) "Spring-slider models" in which elastic and frictional elements are physically discrete components of the setup (Sect. 2.1). **These models can only be applied conceptually to nature**. (2) "Fault block models" in which two elastic blocks, with similar or different elastic properties, are in frictional contacts (Sect. 2.2). **Observations from these models can be qualitatively extrapolated to nature**. (3) "Seismotectonic scale models" in which a distinct tectonic setting is realistically simulated at small scale and with boundary conditions mimicking as closely as possible the natural prototype (Sect. 2.3). **These models can be directly and often quantitatively upscaled to nature**."*

B. Page 2, line 3; Why not consider all earthquakes? While the scaled crustal models primarily pertain to tectonic earthquakes, the spring slider and fault block model certainly can be applied to any earthquake, regardless of setting. The mechanisms presented in the paper are broad enough to consider all earthquakes.

Earthquakes are the result of a variety of source mechanism, only some of which involve the here considered Mode II crack-like displacement in rocks (i.e. tectonic earthquakes). Many non-tectonic earthquakes involve very different kinematics (e.g. hydrofractures = Mode I crack, landslides = non-double couple, explosions = Mogi source) or mechanisms (e.g. resonance in volcanic conduits). We therefore don't think that we can consider all earthquakes.

No changes in manuscript

C. The paper misses an opportunity to promote the benefit of experimental results over numerical simulations. The discussion mentions that numerical simulations can be used to understand the experimental results and a reader could be left with the impression that one could dispense with experiments entirely and go right to numerical models. Should we not bother with the challenge of scaling the crustal experiments to both short and long term processes and just develop numerical models?

We hope this impression is wrong! We agree that there are pros and cons but also benefits on both sides.
Numerical models are an approach to describe nature mathematically and since the physical laws are not known (at best described by simplistic models or empirical laws) any simulation is a somehow arbitrary and possibly biased solution of a highly non-unique problem. The fact that numerical models replicate natural observations does not mean they are a proper physical description. Given the many parameters to be tuned, any result can be obtained by a numerical model. Experiments on the other side are physically correct in itself; however, extrapolation to nature might not always be valid. Additionally, not all quantities necessary to infer the physical laws from experiments might be observable directly but they might be inferred from the experiment using numerical models. Finally, numerical models can implement more complex scenarios, be it in terms of geometry or rheology, and a wider parameter spaces. However, computational limitations often limit the spatial and temporal resolution or reduce the model to 2D. We think it is here were analogue and numerical models should complement, rather than compete, with each other.
Experiments suffer from variability while numerical models always give the "right" answers; however, precision should not be confused with accuracy. In fact, variability is present in nature as well and we might learn a lot from analyzing it in the models. Experiments are carried out in a space-time continuum, while numerical models are discretized. Especially in terms of cross-scale modelling, this may be a source of bias.

However, since the paper does not explicitly deal with numerical models we feel uncomfortable with adding this general debate to the manuscript. Also we do not know if this paper is right place to discuss this at length. We hope the current version advertises experiments enough and leaves it to the reader to judge whether experiments are needed or not.

We added however some discussion on this in the intro & conclusions and more details on the comparisons mentioned to illustrate the interplay between experiment and simulation:

Intro:
"Deformation of the earth involves timescales ranging from nanoseconds (processes at the rupture tip) to hundreds of Millions of years (Wilson cycle) and all spatial scales from atoms to the earth itself (e.g. Ben-Zion, 2008). Such a multiscale process pose major challenges to observation in nature as the instrumental and historical records are too short to capture a significant amount of the evolution. Simulation is a way to overcome such limitations. However, our knowledge of the physics of earthquakes and earth deformation in general is incomplete and does not allow to setup realistic numerical scenarios. Non-uniqueness of numerical solutions for typical problems of earth deformation at various time scales (e.g. inversion of rupture kinematics or mantle rheology from co- and postseismic observations, respectively) is another limitation of computer models. Experimental

approaches have been traditionally used to address physical problems like earthquakes and seismic cycles and, more recently, bridging the gap between short-term instrumental/historical observations and long-term paleoseismological/geological observations."

Ch. 2.3:

"The analogue models of Rosenau et al. (2009, 2010) have been cross-validated using finite element models (Pipping et al., 2016;).The numerical model of Pipping et al. (2016) replicated the laboratory results by means of the general deformation pattern of the wedge through seismic cycles, the recurrence behaviour (recurrence times and periodicity) as well as principal source parameters (slip distribution). Using numerical simulation, the frictional properties of the analogue model material were validated and augmented if not measured physically. Moreover the numerical model provided a highly-resolved image of the rupture dynamics beyond what was observed in the analogue model (see e.g. animation of analogue vs. numerical subduction earthquake cycles in Rosenau et al., 2016). Vice versa, the laboratory example served as an object for testing the general performance of the numerical modelling scheme and for the verification of the effectiveness and reliability of the algorithm for frictional contact modelling introduced there."

"The analogue models of Corbi et al. (2013) have been tested against numerical models using the finite element technique (van Dinther et al., (2013a, b). An extensive benchmark has been carried out analysing a set of parameters that are characteristic of both the interseismic and coseismic stages (i.e., recurrence time, coseismic displacement, rupture duration, rupture velocity, slip rate, rupture width and hypocentre location). A robust fit has been obtained for the majority of the investigated parameters of the reference model (van Dinther et al., 2013a). In particular, the mean of individual source parameters of the numerical model fall systematically within the standard deviation of the same parameter of the analogue models. The largest discrepancies between the two modelling techniques are observed for rupture width and hypocentre location and are attributed to difference in boundary conditions (i.e., the aseismic part of the megathrust doesn't create stress build up in the analogue models) and sampling rate, respectively."

Conclusions:

"We presented an overview of experimental approaches to model earthquakes,seismic cycles and seismotectonic deformation. The processes involved are multiscale posing the challenge to cross time scales from seconds (seismic deformation) to Millions of years (tectonic deformations) both in natural observation as in simulation and experiment. Since natural observations are intrinsically limited in resolution and period of observation, simulation by means of analogue and numerical modelling are key to understanding multiscale processes. An experimental approach to multiscale problems seems most natural because experiments happen in a time and space continuum in contrast to numerical models which need to be discretized."

(…)

"3. Coupling of analogue models with numerical models help to overcome the respective limitations and better exploitation of the respective potentials. For example, numerical models can be used to infer quantities from the experiment that are not directly observable (e.g. details of rupture propagation). On the other hand, experiments can help in validating numerical models by means of testing their predictions. Cross-validation and benchmarking in general should be promoted in the respective communities.

4. Properly scaled analogue earthquake models may help to improve seismological and geodetic inversion techniques and overcome non-uniqueness of numerical solutions. They provide a large number of well constrained and self-consistent "case studies" which display both natural complexity and variability. Analogue earthquakes may thus serve to minimize the solution space and more adequately constrain e.g. slip variability."

D. The review is comprehensive but its utility could be improved by additional presentation of particular configurations, rheology etc that would be adept at capturing particular processes. For example, in the scaling discussion, the appropriate scaling to use in the model depends on the questions of interest. If you are more interested in the directivity and details of rupture evolution you will probably scale the experiment differently then if you are interested in the statistics of thousands of rupture events. Examples of this could be very helpful.

We agree that it would be desirable to have more concrete examples and descriptions as in similar reviews (e.g. Dooley and Schreurs, 2012; Graveleau et al, 2012). However, it became clear in an early stage of writing that we cannot provide such granularity in the description of particular setups and results. This is not only because of the large number of papers but also the variety of research questions and approaches. We therefore decided to generalize as much as possible leaving space for individual application of the material presented to the reader.
At the same time we provide a rather long reference list of applications to be exploited by the reader once he/she found his/her focus (see also Table 1).

Examples of adaptation of scaling to the research question (short vs. long term) are given in current version e.g. by comparing the "long term models" of Rosenau et al. (Chapter 6.1 and 6.5) to the "short-term" models of Corbi et al. experiment on rupture dynamics (chapter 6.3). We consider this sufficient.

We added figure 2 and 3 on particular setups; also we describe the ring shear tester used here to generate new data:
We describe the setup now in chapter 2.1 and show it in Figure 2 as an example of spring-slider setup.
We added:
"Several studies which focus in frictional behaviour of granular rock analogue materials (e.g. sand, glassbeads) at low loads (kPa) used a Schulze ring-shear tester (Schulze, 1994, Ritter et., 2016; Klinkmüller et al., 2016; Panien et al., 2006; Lohrmann et al. 2003) which serves here as an example of spring-slider device used to generate analogue earthquakes (Fig. A1). The ring-shear tester consists of a 4 cm high annular shear-cell made of stainless steel holding approximately 0.1 and 1 liter of the sample material. A ring-shaped-lid is placed onto the filled cell. The lid is subjected to a normal force in order to control normal load on the sample. While the cell is rotated, the lid is prevented from rotation by two tie rods connected to a crossbeam. The force necessary to shear the material is measured continuously. To ensure shearing inside the material and prevent slip between the lid and the granular material, the lid has 20 vanes protruding 4 mm into the material.The loading system is compliant enough (~1.3 kN/mm) to generate sticks-slip in a variety of materials at loads below 20 kPa. Results of this setup are presented on several occasions in this paper."

E. The implementation of equation (4) is based on the assumption that the Coulomb failure dominates within the brittle regime. Since it is the creation of new faults and not the sliding along existing faults that is the process of interest, the scaled

parameter should be inherent shear strength rather than cohesion, which describes the strength of existing faults. For dry sand, this distinction is blurry because sand has many surface, grain boundaries, along which to slide and there is no explicit material failure. For this reason, many people have used the term cohesion for scaling of experiments but since this discussion paper goes beyond dry sand, the formulation should be clear that the assumption is failure strength of the material. This also applies to section 4.2.1 on Mohr-Coulomb plasticity. The parameter of interest there should also be inherent shear strength, So, which for dry sand happens to be the same as cohesion.

The long term scaling as shown here is adopted from sandbox studies. This is why the Smoluchowski Number is defined using the Cohesion here. But it is true that all quantities which share the unit of stress can be used here (including all strengths).

Shear strength in the brittle regime is pressure dependent while it is strain rate dependent in the viscous regime. So "inherent shear strength" is not a fixed parameter of our models. In the brittle regime it can only be defined by giving a cohesion (shear strength at zero load) and friction coefficient ((describing the pressure dependency of shear strength) according to the Mohr Coulomb theory.

Strain localization in sand follows a qualitatively and quantitatively very similar stress-strain curve compared to brittle failure of intact rock. And it shows Mohr Coulomb failure envelopes characterized by cohesive and pressure-dependent shear strength. While the micromechanics are different, the bulk behavior mimics that of the prototype rather well. This is why we view cohesion in combination with the friction coefficient is the most useful set of parameters to describe brittle deformation, be it frictional failure or sliding. Both failure and sliding can be adequately described using the graphical model known as Mohr Circles.
We make these points clearer now.

Page 9, Line 18:
"Note that all quantities with the unit of stress, in particular all strengths, share the same scaling and are substitutes for cohesion in Eq. (4)."

Page 14, Line 26:
"While the Mohr-Coulomb criterion originally describes frictional faulting of intact rock, the same graphical method can be applied to describe frictional sliding on pre-existing faults (Byerlee's Law)."

F. Please explain why the characteristic length scale for the quasi-static model should be peak slip. The peak slip may a consequence of dynamic processes in the model. If the dynamics aren't properly scaled then the peak slip might not scale regardless of whether the quasi-static regime is properly scaled. For some models the more appropriate length scale for the quasi-static regime would be thickness of the brittle material, which should scale to locking depth of the crustal system.

We totally agree. We used peak slip here to illustrate the challenge in observing analogue earthquakes. However, we see that this causes confusion in the context of static vs. dynamic scaling and because it might appear as an a-priori set parameter. But as you say, it is a result of the dynamics and thereby rather a verification criteria for proper dynamic scaling. In fact the length scale is set by the stress scale which in turn is related material *properties (cohesion, strength, density). And the length scale is the same in the static and dynamic scaling (which only changes time scale).*

We deleted the sentence referring to peak slip in the text in order to avoid confusion.

G. For the dynamic regime, the scaling of Dc, slip-weakening distance, should be discussed. If the Dc of the material is artificially high, this can change the nature of dynamic rupture. This parameter is challenging to scale within numerical models and should be addressed within the scaling of the dynamic regime. Dc is mentioned in equation 18 but its scaling is not discussed.

We thank the reviewer pointing us to this issue. In fact Dc is one of the key material parameters to be chosen according to the length scaling derived in the scaling chapter. However, few constraints exist for scale models. We added what we know about it in chapter 4.2.2.:

Page 16, Line 25:
"The slip weakening distance Dc is strongly scale dependent: It is in the order of decimetre to meter for natural earthquake slip events and millimeters in rock mechanics experiments (e.g. Hirose and Shimamoto, 2005, and references therein). In studies using rock analogue materials Dc scales down to micrometres consistent with typical length scales derived from applying Eq. (4) (e.g. Mair and Marone, 1999; Rosenau et al., 2009)."

H. The discussion of the incompatibility of scaling the Froude and Reynold's numbers is very interesting and relates to the issue that the most important aspect of scaling is to ensure that you scale the processes important for the questions asked. Some scaled models may aim to capture all of the processes acting within the crust but many very useful models will investigate a subset of processes. For most models it will be critical to scale some but maybe not all of the crustal processes. The scaling section of this paper should make note of the importance of matching your scaling to the processes of investigation.

The Froude vs. Reynold scaling is a prominent example of the conflict that may arise if you want to "scale everything". While we are friend of rigorous scaling we acknowledge the limits which we illustrate with this example. We think our scaling approach is already quite limited in terms of the crustal processes (brittle and viscous deformation, slow vs. fast deformation) and sets a minimum number of constraints by means of dimensionless numbers. It becomes clear from this paragraph that a dynamic scaling in the viscous is not possible. However, because the earth behaves viscoelastic, i.e. elastic at the timescale of aan earthquake, this limitation might not be a problem.

I. The text on rate and state friction within section 4.2.1 overly relies on the textbook of Scholz. As a review paper, this manuscript should cite additional resource. Even Marone, 1998, which is cited in the figure caption, does not appear in the text.

RSF theory is rather well established and consensus arrived to the details of RSF that we report here. This is why we restricted ourselves to a short description of RSF theory showing a minimum set of useful equations. We consider citing textbooks (Scholz) and review papers (Marone 1998, Scholz, 1998) along with the original papers (Dieterich, *Ruina) as sufficient and least biased. A critical review on RSF is beyond the scope of the paper.*

We added missing reference to Marone (1998) in the text.

K. The paper misses an opportunity in the discussion of seismic versus aseismic faulting to highlight the simple block slider experiments that are used in teaching classrooms. These simple brick on sandpaper (or variations) demonstrations very effectively convey the concepts of stick slip to students and the public. Some apparati include accelerometers, acoustic emissions, force data etc for students to analyze various earthquake properties.

We agree that spring slider setups are useful for teaching purposes, but we think they have been exploited scientifically rather to their end. That's why we only refer to them in a historical way.
However, we now dedicated a short paragraph in the conclusion on outreach applications.

In the conclusions:
"4. experimental techniques are a superb method for visualization and teaching complex processes. For example, simple spring-slider experiments equipped with force sensors and accelerometer are easy to realize and provide fascinating hands-on experience on earthquakes."

L. There is a rich literature of stick slip experiments with glass beads that is not included in this review (e.g. Savage and Marone 2007 JGR). Since glass beads are analogs for rock they should probably be considered within this review paper.

We consider the early papers from this group laying the foundation of glassbead-based analogue earthquake studies (Mair and Marone, 1999; Mair et al. 2002). The Savage and Marone papers (2007 and 2008) focus on dynamic triggering, a topic that we omitted for sake of respecting space limits. These are very interesting papers but dynamic triggering is not yet a topic studied by more than individual groups by means of analogue earthquake modelling. That's the reason why we cut those studies out.

No changes in manuscript.

M. Section 6.4 may need a different title or become more broad in scope. For many researchers, site effects include the very local effect of the geotechnical layer on ground shaking. For example, some civil engineers care only about the attenuation and amplification within the geotechnical layer.
We agree.
We changed it to "Ground motion".
.

Page 2 Line 3:
Anthropogenic pumping can also produce earthquakes
We agree.
We added it.

Equation 3.

The Ramberg Number doesn't seem to be utilized in this paper. Please explain its utility.
It is!
Eq. (5) is derived from it and it is used to set the time scale in viscous models. It is applied when discussing the postseismic relaxation in Caniven's model for example.

We clarified it in the text:

"In the viscous deformation regime, the ratio of gravitation and viscous flow strength is used and has been labelled the Ramberg number"

"Note that the Ramberg number has been derived from the Stokes number (Table 2), which characterizes more generally slow (small Reynolds numbers Re << 1, Table 2) flow typical oftectonic applications.To achieve similitude these numbers have to be the same in the model as in the prototype. For a given length scale (usually suitably chosen for handling the model in a lab), Sm and Ra dictate the stress scaling in the brittle and viscous regimes, respectively."

Page 9 Line 25:
The rate and state parameter Dc has dimension and should be scaled.
We agree, see "G" from above.

Page 10 Line 9:
I may have missed something but I didn't see where Mo* was defined.
Shortly before: "…it is defined as the product of rigidity of the earth, mean coseismic slip and rupture area"

End of section 4.1.1.
You could point out that in the case where DL equals Ds the sum of the interseismic deformation and the coseismic deformation produce a set function of the tectonic velocity.
We agree!
New sentence:
"In the case where DL equals DS the sum of the interseismic deformation and the coseismic deformation produce a set function of the tectonic velocity."

Page 17 Line 1:
Please Explain why the Maxwell Model is considered more relevant than the Kelvin-Voigt model.
New sentence:
"This is because the earth shows, in case of an earthquake stress drop (that can be viewed as large scale creep test), the instantaneous coseismic and transient postseismic deformation characteristic of a slowly relaxing Maxwell body. A Kelvin-Voigt body would show no instantaneous elastic response to sudden loading but a transient."

Page 19 Line 23:
Can Tchalenko, 1970 GSA Bull V 81 Pm 162501640. A Very cool paper that like Niewland Et al. (2000) Shows the pressure changes associated with faulting.
True, but this is a "global" monitoring tool while Niewland is a "local" one.

We added Tchalenko and other similar in the Global monitoring techniques section:

"Several studies used force sensors to measure the force exerted by a backwall in sandbox experiments of strike-slip (Tchalenko, 1970) and thrusting (e.g., Cruz et al., 2010; Souloumiac et al., 2012; Herbert et al., 2015)."

Page 19 Line 27:
Paul Young Has also done a lot with AE For precursory failure.
True, he is second author of the three Thompson et al papers cited!

Page 20 Lines 3-6:
What Are the drawbacks and benefits of laser techqniues?
The reported high temporal resolution vs. the spatial limit. We reported this already.

Page 20 Line 10:
Karen Daniels Has some very nice granular experiments using photoelastic materials that should be included.
We agree.
Page 4:
"Shearing of acrylic or polymeric discs (e.g. Daniels and Hayman, 2008; Reber et al., 2014) provided insights into the dynamics of sticks-slip in granular media by means of 2D "see through" experiments."

Page 22:
"Based on the photoelastic effect, Daniels and Hayman (2008) visualized the dynamics of force chains in sheared granular media undergoing stick-slip."

Page 21 Line 13:
PIVLAB Is another used by some groups
We agree and added:

Page 23:
"PIVlab (Thielicke and Stamhuis, 2014)"

All other technical comments:
We agree.
We changed the manuscript accordingly.