I have reviewed the paper "Landscape-scale water balance monitoring with an iGrav superconducting gravimeter in a field enclosure," by A. Güntner et al., which presents a new capability for deploying gravimeters in a field setting. The paper nicely shows how the gravity response to infiltration becomes more linear as a direct result of the sensitivity allowed by the pillar/enclosure. By demonstrating a novel application that could have wide applicability, I find the paper deserving of publication in HESS-D.

I found no major errors or concerns with the paper. A number of minor revisions are included below.

I suggest the authors cite the Wilson, et al. paper that describes an earlier attempt at using the superconducting gravimeter as a field instrument. Although the present effort using an iGrav is an advance over the Wilson field enclosure, it is worthwhile to show that others have also considered ways to improve gravity data collection for field studies. [Wilson, C.R., B.R. Scanlon, J. Sharp, L. Longuevergne and H. Wu. 2012. Field test of the superconducting gravimeter as a hydrologic sensor. *Ground water* 50, no. 3: 442–9.]

For studies of aquifer recharge (i.e., infiltration that has entered the aquifer), near-surface soil moisture is "noise" that must be estimated and removed in order to identify recharge. In that case, the decreased near-surface sensitivity afforded by the umbrella effect beneath a large building/enclosure is an advantage, and a small pillar/field enclosure would decrease the ability to identify recharge. It may be worth noting the iGrav/tall pillar combination isn't useful for every hydrologic investigation.

Lines 16-18: I find the description of the instruments confusing. It might be clearer just to state relative gravimeters measure relative gravity differences (between stations or over time). It could be just as accurately said that absolute meters also measure the impact of the resulting force changes on a test mass. Also consider explaining what you mean by "continuous" (including the ~45 second filter on relative meters).

Line 17: the magnitude of the vertical gravity vector (as written, the implication is the meter measures the 3-d vector).

Line 22: It would be helpful to state explicitly that spring and superconducting meters are relative meters.

Line 26: Usage of "time-lapse" is vague; in the previous paragraph you use it to mean "periodically repeated", but some of the references cited here use SG (continuous) data.

Line 33: Consideration of these effects (tides, loading) are important for all gravity studies, not just those using SGs. Lines 33-36 could be deleted to help the paragraph flow better.

P.3, Line 23: It might be useful to indicate this study is concerned with near-surface soil-moisture changes, which weren't relevant or observed in the Kennedy (2014) paper, which also demonstrated the iGrav in a field enclosure.

Line 28: gneiss should be lower case.

Figure 1: What is the coordinate system? Suggest changing "gauging station" to "streamgauging station".

Page 6, line 3: Usually "on the one hand" and "the other hand" are used to contrast two items. In this case the two enclosures are just two parts of the system, not contrasting parts.

Page 6: It would be helpful to indicate the power requirement of the iGrav (and that it requires AC (line) power, if that is the case).

Page 7, line 20: Suggest including the URL for the Atmacs service (I wasn't previously familiar with it).

Page 8, line 5: meaning of "could not be carried precisely enough" is not clear. What are the implications for determining drift for other field studies, where a lysimeter is not present? If drift cannot be determined using the abundance of gravity meters available at Weitzell, is there any hope for other field sites? Can you make any inferences regarding the linearity of drift (perhaps there are other periods with identical soil moisture/groundwater levels)?

Would it be possible to show a cutaway schematic of the field enclosure? From the description the annular air space and configuration of cooling grills isn't entirely clear.

Page 9, line 10: Can you indicate the temperature at which PCB effects began to be seen? I don't see anywhere the maximum temperatures observed at Wetzell, but I'm guessing it's relatively mild. Would these temperature effects be expected at pretty much every installation? Can you comment as to whether these temperature effects have been seen in other (non-iGFE) iGrav installations? Perhaps it would be useful to include a table showing the range of environmental conditions (temperature, wind speed, humidity, etc.) under which the iGrav operated with satisfactory results.

Page 10, line 16: Cold head = cryocooler, correct? Consistent terminology will help tie in with the discussion on page 6.

P.11, line 10: suggest replacing perpendicular with horizontal.

Figure 6: A useful figure, as it's not intuitive the departure from an infinite slab would be so great for a 1m high pillar. I believe the purple dashed line (infinity symbol in the legend) is for the infinite radius case, not infinite height, as shown? The gravitational attraction of a finite-radius cylinder would approach zero if the gravimeter were at infinite height.

p. 13, line 11: I think the 2.17 and 0.15 numbers are reversed, the iGrav should have a larger gravity effect?

p.13: Can you talk about the difference in sensitivity to storage change at the water table for the 2 sites? Is it equal?

Figure 8: Does the "double-peak" behavior warrant explanation? I assume the increase at distances between 500 and 3000 radius comes from valleys surrounding the site?

Its obvious, but it may be helpful to say explicitly that the gravity signal never decreases as the wetting front moves deeper – it can only decrease if soil water is evaporated, or somehow moves laterally (usually as groundwater, after recharge). That feature makes the method equally useful for thick unsaturated zones.

p 15, line 17: what is the soil porosity? The precip flux and Ks are both very high, although the response in figure 9 would be similar for a broad range of parameters.

p.18, line 5: there are a few references where e.g. is used but not needed

Figure 10: It is worth noting in the caption or on the figure that only the first 10 days are used for optimization. I would move "after optimization" to later in the sentence, so it is obvious the blue line is data. Maybe delete "(as a deviation from…)"

It seems that P-E-R misses some of the high-frequency behavior in the gravity residuals. Can you comment? Is it because s is slightly too high for near-surface soil moisture change? I.e., when converting from dg to ds, there is a slightly non-linear effect that would make s, as shown in figure 10, slightly less dynamic.

p.19, line 19: Worth noting there are many factors (including all those mentioned), with the same 24-hour period as ET.

line 23: State explicitly why day-to-day ET can't be determined directly from the data (i.e., rearrange eq. 3 to solve for (Eref – c) without stacking or a moving window. You mention the main obstacles earlier, but it's not obvious these are the exact (and only) reasons it can't be done. line 30 would indicate the estimate improves as the window gets even shorter than 9 days?

p. 20, line 17: sign not sing

I find the discussion about the E discrepancy very interesting, as it's not obvious to me how a gravimeter at the land surface responds to ET – to decrease gravity seems to require advecting the evaporated water away from the region of sensitivity.

p.22, line 9: I'm unclear what a "dedicated sensitivity" is. Is it appropriate to cite a paper in preparation?

line 21: I'm not sure "fundament" is widely known (it's unfamiliar to me). Also, the significant limitation of requiring AC power.