Hydrol. Earth Syst. Sci. Discuss., 10, C2544–C2563, 2013
www.hydrol-earth-syst-sci-discuss.net/10/C2544/2013/

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Three perceptions of the evapotranspiration landscape: comparing spatial patterns from a distributed hydrological model, remotely sensed surface temperatures, and sub-basin water balances" *by* T. Conradt et al.

T. Conradt et al.

conradt@pik-potsdam.de

Received and published: 14 June 2013

First of all, we would like to thank Professor Dr. Pegram and the anonymous referee that they took the time to review this – indeed a bit lengthy – manuscript and provided us with many ideas for improvement.

On the comments of Anonymous Referee #1

Regarding the general critics, we agree that "a general excessive verbosity of the presentation often makes the analysis difficult to be followed." Therefore we deleted or condensed some less relevant passages from the manuscript, namely

- the introductory paragraphs on spatial calibration (starting at P1129-L22),
- the digression into the Monin–Obukhov theory (starting at P1140-L17),
- the ET-integration thereafter (P1141-L10), and
- the second variant of resistance estimation (starting at P1142-L13).
- Finally, the Conclusions (starting at P1151-L20) were abridged.

The general comments further critisize the many assumptions being made for utilising the remote sensing data and recommend their reconsideration. However, any evapotranspiration estimation based on remotely sensed thermal radiation is forced to make lots of assumptions. We tried at least to clarify all our assumptions; especially the temporal integration of the remote sensing data which needed refinements. Even if our approach might be difficult to be "clearly contextualized in the wide range of approaches already available in the literature" we would like to present it as it had been used. Referee #1 points out him/herself: "This topic is still an open issue in the remote sensing scientific community [. . .]"

Using only three names/labels for the three methods consistently through the text is a good idea. We now use the signal words

1. SWIM
2. Remote sensing

3. Water balance

in referring to the three methods, even in figure captions and axis labels.

Specific comments of Referee #1

P1135-L10: It is not clear to me if the rain gauge stations are 501 or 853; please clarify this point. Moreover, since the authors say that heterogeneity in network density is one of the possible justifications for the discrepancies, it would be useful to see the spatial location of both fully instrumented and rain gauges, perhaps in Fig. 2.

We changed the wording a bit, now it should be clear that there are 501 additional rain gauges, thus 853 altogether. And a new Figure 3 shows their spatial distribution.

P1135-Eq.1: To the best of my knowledge, Turc equation is based on solar radiation instead of net radiation. Additionally, I can not recall the dimensionless factor in the original formulation. A better reference for this approach should be provided, as well as a justification on the use of this alternative approach.

It is explicitly stated that this is a modified Turc–Ivanov approach, hence the deviations from the original Turc formula. To justify the use of this alternative approach, the remark that "it had been originally developed for East Germany largely resembling the German part of the Elbe basin" has been added here.

P1136-L1: The reported crop coefficients (land use specific factors) are quite high (> 0.9). Are these including also sparse vegetated areas?

The crop coefficients are a modification of the reference evapotranspiration for idealised vegetation. Reductions for sparse vegetation are made afterwards through LAI and root state (see the following paragraphs).

P1136: The connection between EP and WU is not clearly highlighted. Also, how many soil layers are modeled? Is the temporal variability in root depth accounted?

Yes, the final step of calculating the actual plant transpiration $EP_a = \sum_i WU_i$ was indeed missing. The other suggested informations regarding the number of soil layers and root growth have been inserted, too.

P1136-L21: 30 cm is a rather tick layer for soil evaporation. Do you have any justification for this value?

Of course, soil evaporation reduces soil water content mostly in the uppermost millimetres of the soil profile, but as SWIM runs on daily timestep, also the overnight replenishment from deeper layers has to be considered. The idea behind the 30 cm are to completely capture the vertical profile of evaporative losses from a freshly ploughed soil, the typical condition when soil evaporation really matters. The SWAT (2009) approach, for comparison, uses an exponential decrease of the evaporative demand with depth but has no maximum depth limit at all.

P1138-L1: G is generally negligible only for full covered densely vegetated surface. Several formulations are proposed in the literature for accounting for daily G in case of sparsely vegetated areas. This can cause distortions in ET spatial distribution.

The vegetation in the Elbe River basin is mostly dense; freshly ploughed agricultural fields may be the only relevant exception. But these are local short-time phenomena, so the distortions will be small.

P1138-L24: The SEBAL model aims at reducing the inconsistency between remotely observed land-surface temperature and ground-observed air temperature. The use of a single Ta value or a Ta map is a different issue; this must be clarified. The use of conventional ground-measured temperature data does

not overcome the problem of the inconsistency, which is the main issue in using thermal remote sensing data. The authors should rethink this assumption.

SEBAL has in fact the advantage of working without measured air temperatures, but this was not really made clear. In the revised version, the misunderstandable reference to SEBAL has been removed, and it reads simply and clearly: "We utilised spatially interpolated, ground-measured air temperature and radiation data: [. . .]"

P1139-Eq.11: This equation is valid only for fully vegetated areas (Te $\approx$ Ta). Since the SWIM uses this Eq. in the Turc model, this is a valid assumption; however, the same approximation is not valid in the framework of the surface energy budget (Te $\neq$ Ta). Again, this can be a cause of distortions in ET spatial distribution.

Again, the vegetation in the Elbe River basin is mostly dense; freshly ploughed agricultural fields may be the only relevant exception. However, the equation is of course a simplification. Added "assuming $T_a \approx T_e$" in parentheses to make that clear.

P1140-L9,Eq.15: The temperature gradient in Eq. (15) is wrongly defined. It represents the gradient between aerodynamic temperature (see Norman and Becker, Agr. Forest Meteorol. 77:153-166, 1995) and air temperature above the surface. In your approach there are two strong assumptions: 1) the aerodynamic temperature corresponds to the surface temperature, which is not true for heterogeneous surface as hydrotopes likely are; 2) the ground-measured air temperature at 2-m corresponds to the value above the surface, which is not true for tall vegetation (i.e., forest but also crops taller than 2 m). The latter assumption is even stronger if we consider that Ta and landsurface temperature are not collected at the same time. Again, SEBAL model (as well as other commonly used thermal-based surface energy balance models) is specifically designed to circumvent both these problems rather than ignore them as your approach does. Given that one of the main goals of the paper is to quantify the value of the information provided by remotely observed land-surface temperature, this information must be use

correctly within a physically based framework. In my opinion, considering also another limitation of the available dataset (i.e., absence of wind speed data), the author should consider to adopt a simpler (but widely tested) approach, as for instance the triangle method, rather than drastically simplify a physically based formulation by means of unreliable assumptions. I'm not at all convinced that the ET spatial distribution obtained under these assumptions (as well as the ones on Rn and G) reliably represents what a "standard" thermal-based method would provide.

We are fully aware of the fact that the difference between 2 m-air temperature and surface temperature ΔT does not directly equal the vertical temperature gradient $\frac{\partial T}{\partial z}$. Therefore we wrote: "In practice, this gradient is represented by [. . .]" And $\Delta T = T_s - T_a$ is a common approximation used by numerous remote sensing methods, cf. the review by Courault et al. (2005) in *Irrigation and Drainage Systems* 19: 223–249.

Furthermore, the SWIM hydrotopes are defined by landscape characteristics to be homogeneous units, so the referee's objection "[. . .] not true for heterogeneous surfaces as hydrotopes likely are" does not apply here. That 2 m-air temperatures are not ideal to describe above-surface conditions is clear (Carlson & Buffum 1989 therefore proposed measurements at 50 m height), but the meteorological stations are never placed below forest canopies.

However, we took account of many concerns by respective amendments to the text of Page 1140, finally clarifying the advantage of our method compared to SEBAL: "The explicit consideration of different air temperatures in space is an advantage over the uniformly calibrated SEBAL algorithm that should balance negative side effects of some necessary assumptions."

P1142: Two different calibration methods are introduced here, however, only the first one seems to be used. This has to be clarified. Also, some discussion has to be made on the assumption of unbiased estimates for the two land-use

classes. Does this explain differences in the ET maps (e.g., are the hydrotopes with high/low errors characterized by high/low forest fractions)? I'm wondering if this is another possible source of distortion in ET spatial distribution that it is not commonly present in remote sensing estimates.

The proposal of a second calibration method has already been deleted. Regarding the possible errors, the following paragraph has been added: "The modelled averages of the respective land cover evapotranspiration $ET_{S,f} + ET_{S,n} = ET_{SWIM}$ are very reliable, because they represent large sub-areas of the model domain [...], and statistical analyses of the results showed no relationship between deviations of sub-basin evapotranspiration estimates and sub-basin forest shares."

P1143: This upscaling procedure is rather confused, especially because some terms are not well described: e.g., What ETtot is? Is it the annual average of ETSWIM (previously introduced)?

To clarify what is meant by "total ET" added "of the entire model domain within that period" in parentheses.

P1143: Is the linear relationship between Delta_T and ET supported by any evidence? Additionally, the time invariance of aerodynamic difference is a further strong assumption, given that is well know its dependence from vegetation height and mass (as well as seasonality in wind speed that can not be accounted). This is another assumption not commonly made by thermal-based remote sensing approaches. I would suggest to consider at least the actual roughness parameter for each hydrotope parameters (in Rah), including it temporal variability. The optimization could be performed by calibrating the effect of wind speed at year-scale. This would probably partially reduce the discrepancies with commonly used approaches. In any case, effects of atmospheric stability are ignored (or lumped in the calibration procedure). This must be highlighted.

The linear relationship between ΔT and ET – given R_n and r_{ah} remain constant –

is shown by Eqs. 14 and 15 on Page 1140, so this is not a new assumption. The assumption here is rather the applicability of "effective temperature gradients" for longer time periods.

Regarding the time invariance of the aerodynamic resistance, our method delivers always the average values for the respective time period, confined by the radiation and temperature measurements as well as the average ET from the hydrological model. To consider the temporal dynamics of resistances is a future research topic, but outside the scope of this paper.

Anyway, we re-wrote the paragraph above Eq. (24) to make things more understandable.

P1143: Another inconsistency in Eq. (24) is between the time-scale of Rn and Delta_T . In fact Rn is a daily value while Delta_T is measured at a specific time of the day. How the authors deal with the upscaling of Delta_T from one time-of-day to daily value? If another assumption here is made, this must be clearly highlighted.

Net radiation is also summed up over time, so there is no inconsistency. This is stated now.

P1144-sec.2.3: I agree that the water balance method is rather simpler than the other approaches; however, some more details must be provided. For instance, it is clear successively that it was applied separately for each gauged sub-basin, but this should be highlighted here. Was Eq. (25) applied at year scale (separately for the 3 years) and then the ET map averaged? Is P derived from the same interpolated fields used for SWIM?

Details have been added as suggested: "We calculated average annual balances from the three years of interpolated precipitation data that were used to drive SWIM and from the observations at 133 runoff gauges. For each gauge with a catchment area containing further gauged sub-catchments upstream only the part below these upstream areas was considered by subtracting their shares in precipitation and runoff."

P1144-L15: Given that the aim of the work is to evaluate the impact on interior gauges, the authors should demonstrate that the SWIM model performs accurately at lest globally. If discrepancies between SWIM and water-balance are observed also globally, these should be minimized before to proceed in the analysis.

The behaviour of the SWIM model during the three years has been illustrated: "The eco-hydrological model SWIM [...] was run for the three years 2001–2003. During this period, runoff at Neu Darchau was slightly over-estimated by 12 %. The resulting balance error for evapotranspiration remains below 5 % though, because the runoff coefficient is below 0.3 (cf. Conradt et al. 2012b). The Nash–Suttcliffe efficiency of the daily values was at 0.87. Figure 4 shows the hydrographs."

P1145-L17: Several methodologies are available in the literature (generally based on sinusoidal function), to reconstruct air temperature at a specific time-of-day from daily min, max and mean values. The correspondence between Ta,max and Ta,overpass is another assumption not really required.

One general assumption of our approach is the stability of the spatial relationship between temperature differences, even if there are little time shifts between the observations of the components. But this is an interesting point: how strong such time-shifts between the temperature observations affect the results might be a question for future research.

P1145-L20: Is seven the number of images 99% clear in this study case? Please clarify that this number coming from the analysis of the data.

Inserted "within that period" to make it clear.

P1145-L26–P1146-L3: It is not clear to me what "no time-dependent weighting

scheme... had been applied" means. Please detail more this consideration.

Yes, there arise more questions from this passage than it may answer. It is really difficult to recall the original considerations about a "time-dependent weighting scheme", so we simply deleted the sentence.

P1146-L7: This is another critical point. LST maps provide information on cloudy condition at the satellite overpass time, but nothing is said on the whole day. What about mixed conditions, when the sky is cloudy at the overpass time and clear the other daytime hours or vice versa? This point has to be clarified.

Reformulated the paragraph to consider this: "On the other hand, radiation and accordingly heat gradients and evapotranspiration rates are much lower under cloud cover compared to blue sky conditions, so cloudiness has to be considered. The assessment of longer time periods (full years or our three year period) with hundreds of LST maps minimises the error from specific cloud distributions at satellite overpass times; it allows for utilising average cloudiness maps as shown in Fig. 7."

P1146-L9–15: In my understanding, the attenuation factor represents the relationship between DT during "clear-sky" and "cloudy" conditions. Why do you assume that this is a constant value? Actually, it was observed in the literature (see e.g., Gallo et al., J. Appl. Meteorol. Clim. 50:767-775, 2011) that Delta_T tends to be = 0 under cloudy conditions, but it is a function of vegetation coverage under clear-sky conditions. This suggests that the attenuation factor is a function of vegetation coverage. It would be interesting to analyze the effect of attenuation factor = 0 on the results (which is a rather simpler assumption than your approach, but likely more close to the reality).

We assumed a constant because we consider the average value for the time period of our investigation. Of course, ΔT is a function of vegetation coverage, but why does this suggest the attenuation factor having the same dependency? The reason why the attenuation factor was not set to zero (which is of course true for really cloudy conditions) should be clear from what is said in this paragraph: “White pixels include all conditions from thin cirrus with hardly dimmed radiation to dense stratus.” We tried at least $\eta = 0.25$ for comparison, but this did not visibly change the resulting map (now Fig. 10).

P1146-L21: Is this map the 3-year average? Please clarify.

Inserted the time-frame into the figure caption.

P1146-L21–22: Why mountainous area should have Delta_T close to 0? Delta_T is an indicator of water stress, which can occur both in mountainous and low-land areas. The small values in the upper regions can be related to an incorrect evaluation of the effects of elevation on Ta.

We totally agree with what the reviewer points out here, but this passage simply describes the observation that mountainous areas, amongst other things, can be distinguished on the map from low ΔT values. (However, in Central Europe the mountainous areas regularly receive much more precipitation than the lowlands, and water stress is therefore tendentially reduced there.)

Sec.3.2: This should be the main section of the paper, and it is only 1 pag. (compared to 12 pags. of methodology). In my opinion the author should clearly state if they consider the water balance the “target” reference to evaluate the other two models and discuss the results accordingly.

This main part of the Results section could be substantially extended with the “errors” from anthropogenic groundwater extractions discussed in Sec. 4.2 on Page 1150, now positively presented as example for the practical use of the methods comparison. (It was principally difficult to clearly separate between Results and Discussion.) Given the fact that Figures 9–16 also contribute to Section 3.2, it extends over 4 print pages now.

The term “ground truth” has been removed – water balance results should not by no means be considered as some kind of target reference superior over other methods.

P1147-L14–17: This is true in general, but when the outliers are removed from the analysis (see Fig. 12) the ranges of variability of SWIM and remote sensing are similar. Some more comments on this should be reported.

Changed “main message” into “impression” to avoid the misleading interpretation of the noise not removed at this point of the analysis.

P1148-L4–9: The correlation remote-sensing vs. water-balance and SWIM vs. waterbalance are rather similar (on the German sub-basins). This seems to suggest no value in using remote sensing data. This point has to be deeply discussed, since it is the main topic of the paper.

In order to show the special value amended at least: “– and this clearly highlights the potential of spatial calibration by means of remote sensing in order to reduce the modelling errors especially for smaller sub-basins.”

P1148-L23–24: Is the density of ground measurements a real issue, especially over lowland areas where Ta varies smoothly? A map with the spatial distribution of the stations would be very helpful. Also, the SWIM model is affected by the same issue too, and network density is an even more relevant issue for precipitation. Some discussion on this should be added.

Referenced the added station map again and re-edited the paragraph. Finally, it reads now: “Additionally, the southern exposition of the area at the foot of the Ore Mountains might have be contributed to locally increased air temperatures not captured by the station network.”

P1149-L5–10: This statement on the “remaining noise” must be supported by numerical evidences. What are the average bias and accuracy of the remote sensing estimates in this study case (assuming water balance as a reference)? Are these values comparable with the one obtained in the reported studies?

The “remaining noise” of the remote sensing method meant indeed the correlation with SWIM (upper left panel of Fig. 12) only, because the water balance still exposes extreme outliers. This has been clarified now, but the long list of references should not be blown up further by re-citing all the correlation coefficients or coefficients of determination given there.

P1149-L12–16: This seems out of place in the discussion.

Agreed. Deleted.

P1149-L21–P1150-L4: These differences should be reported in terms of impacts on ET and compared to the differences with modeled (SWIM and remote-sensing) values. Are the differences in these sub-basins explained by the magnitude in water unbalance?

We re-calculated the example with the actual runoff measurements of the years 2001–2003 showing that the large ET deviations from the water balance method – in the way it had been applied – are not at all implausible. The focus here is just on this problem, so we won't introduce comparisons with error ranges of the other methods in these passages.

P1151-L23–P1152-L2: I don't understand the sentence "the water balance approach does not seem to be more exact that the two other methods". In my opinion, the fact that the two methods do not agree with the water balance is not a valid indicator of the accuracy of the water-balance. Only and external observation of ET (or storage) can quantify the accuracy of the water-balance approach. If the authors consider the strong correlation between SWIM and remote sensing approaches an indicator of the good accuracy of these two estimates, what about the likely disagreement between discharges modeled in the interior points with the observed values?

The reviewer is right in that an agreement between the results of two methods does not prove a third method with deviating results to be incorrect. We agree that this does

not really substantiate a 'key finding' as originally advertised here. The paragraph has been rewritten into three short paragraphs, now pointing out the principal value of using and comparing several methods for optimally assessing hydrological reality.

P1152-L10: The reported reasons are in some cases true for all three approaches, and not only for the water-balance.

True. Deleted the reference to the water balance method.

P1153-L13–P1154-L6: This is not a conclusion in my opinion. / This is not related for the result reported in this paper, and it is out of place (it may be part of the introduction).

Yes, it reads more like a part of a general introduction than the conclusions. The idea was to present the common "solution" to differences in multiple data before coming up with our recommendation to (better) actively investigate these differences. We deleted this section and worked its relevant sentences into the Recommendations to make our point clear.

P1154-L13–L17: Again, the value of SEBAL (or other similar methods) is in minimizing the errors associated to the inaccuracy in thermal data calibration, as well as removing the need to separately estimate aerodynamic temperature and air temperature above the surface. The misunderstanding on the value of SEBAL (or other similar methods) must be clarified. The problem of spatially variable air temperature is also faced and addressed by SEBAL (or other similar methods) for applications over complex areas.

How should the value of SEBAL be misunderstood here? We do recommend SEBAL, because we are fully aware of our method's weaknesses that SEBAL avoids. We do not think that this is the place to discuss the advantages and capabilities of SEBAL in detail.

Technical corrections mentioned by Referee #1

P1131-L24: Spell-out the acronyms SWAT, SEBAL, as well as the other acronyms through the text (e.g., SWIM).

Spelled out at first occurrences with the exceptions of SWIM, SWAT, and SEBAL: these are explained in the Methods section together with the literature citations.

P1138-L15: The Stefan-Boltzmann equation reported here is redundant, since it is also reported in Eqs. (7) and (8).

Yes, but the little redundancy makes clear what coefficients are specific for R_{la} and R_{le} compared to the general formula.

P1138-L17: The correct value of the Stefan-Boltzmann constant is 5.67 x 10-8 W m-2 K-4.

Value corrected.

P1139-L20: Is the effect of elevation accounted in Rmax?

No, added "disregarding elevation" in parentheses.

P1140-L8: It is not necessary in my opinion to introduce the concept of Bowen-ratio (with 2 references), since it is barely used successively.

Although the Bowen ratio is explicitely addressed only once in the text, the headline explains what this section is all about. Therefore, references to the publications of Bowen's original work should be allowed.

P1141-L6: This concept must be introduced early in the paper.

Using the SWIM ET averages for estimating the aerodynamic resistances solves the problem explained above, so this is the natural place to come up with that idea. The passage has been reformulated in the revised version anyway.

P1145-L2: In my opinion, it is not necessary to refer to other products not used in this study, as NDVI and LST nighttime maps.

Here we fully agree. Deleted the sentence.

P1145-L26: Blue-sky fraction is not defined.

Explained blue-sky fraction as share of the non-cloud covered pixels in the model domain.

P1146-L9: Please define "white pixels".

Added some words to explain the white pixels.

References: The reference list is too long. Please review all the reported references and keep only the ones that are strictly necessary.

The list is now shorter due to deletions from the original manuscript text containing citations.

Table 1: Are the data reported here derived from SWIM? Please clarify.

Partly SWIM output; extended the table caption with the data sources.

Fig. 1. & 3. I'm not sure that these figures are really useful.

Figure 1: It may seem a bit out of this paper's context, but this figure illustrates not only the motivation for tapping alternative data sources, it also shows the general interior-point problem of distributed hydrological modelling in a way that we have not observed in modelling publications yet. We'd like to keep it.

Figure 3: Although we did not account for the varying overpass times of the satellites in the ET calculations, this figure makes immediately clear how much these times shift from day to day, if the satellite is changed, and how the structure of the changes looks like. Maybe useful for other researchers who consider considering this factor in their calculations. We'd like to keep it.

Fig. 5: Are these maps obtained as 3-year average? It is not clear the definition of "absolute" and "relative" cloud-freedom.

Yes, these maps are the three-year averages, and this is now also written in the figure caption. And the definitions of "absolute" and "relative cloud freedom" have already been given there.

Fig. 6: Again, is this the 3-year average?

Yes, again the three-year average; extended figure caption accordingly.

On the comments of Referee G. Pegram

Some concerns given in Prof. Pegram's general remarks are already addressed by the enhancements made with respect to the comments of the anonymous referee, and the remaining ones have been considered, too. Here are, partly again, his major points:

A huge amount of work appears to have made by the authors [...] but the argument, instead of being crisp and neat in explanation, is swamped by too many words and side-tracks. – The number of words and side-tracks has been reduced. Deletions and abbreviations include the introductory paragraphs on spatial calibration (starting at P1129-L22), the digression into the Monin–Obukhov theory (starting at P1140-L17), the ET-integration thereafter (P1141-L10), and the second variant of resistance estimation (starting at P1142-L13). The Conclusions (starting at P1151-L20) were also condensed.

It is not convincing that the relatively high correlation between the German sub-basins' remote sensing and quasi ground-based estimates of ETa shows that the water balance calculations are not to be trusted. – This unconvincing conclusion about the questionableness of the water balance calculations has been removed.

Why was the observation input not used? – Observation data (meteorological and discharge time series) have been used for calibrating and running the SWIM model and for the water balance calculations. If this question was pointed at ET measurements from lysimeters or evaporation pans, we must regret that these data had not been available.

Why was the excellent work of Tom McMahon et al. ignored? – Because it appeared too recently in HESSD; just a few weeks ago it made its way into HESS (17, 1331-1363, 2013) doi:10.5194/hess-17-1331-2013. Now we have cited it in the Methods section; we can confirm that this paper is really an excellent guide into evapotranspiration calculation.

Aren't 133 basins too many? Would spatial clustering to 30 (say) not provide better averages and get away from the 'leaky aquifer' syndrome? – The idea was to use each and every gauge data available to obtain the highest possible spatial resolution. Both the SWIM model and the remote sensing approach can produce evapotranspiration maps with much higher resolutions (there are approx. 48 000 model hydrotopes and 100 000 LST pixels in the model domain); their output has been resampled to the 133 sub-basins.

Groundwater exchange between the gauged sub-basins contributes definitely to the noise observed although the primary mistake was the error amplification through calculating difference values for subsequent gauges along downstream river branches. We therefore take this point as suggestion for future research; a possible question could be for the minimum size of sub-basins keeping the distortions from lateral subterraneous fluxes below a certain threshold.

Why not make direct ET_a estimates from EUMETSAT's LSA SAF product? – This would of course provide one more perception of the evapotranspiration landscape; maybe we'll consider this data source in future assessments.

Particular remarks of Referee G. Pegram

P1128-L15: I'm not sure what this sentence means; it makes better sense if 'epistemic' is omitted from the abstract. The Webster dictionary definition of epistemic is: 'of or relating to knowledge or knowing' and this usage only clouds the message.

The philosophical excursion into aleatoric and epistemic uncertainties has been removed.

P1135-L11 and P1135-L18: Why not include wind speed? This is a richly instrumented basin. / Aha - no wind data. However, it may be of interest that we found that Weather forecast wind data are very good for calculating ET0 - see: Sinclair S. and G. G. S. Pegram (2010) A comparison of ASCAT and modeled soil moisture over South Africa, using TOPKAPI in land surface mode, Hydrol. Earth Syst. Sci., 14, 613-626

The SWIM model can theoretically be driven with a Penman(–Monteith) formula including wind speed, but we never considered this so far due to missing data and/or uncertainties about aerodynamic and canopy resistances. The Turc–Ivanov approach has some history of development and application in East Germany and was therefore favoured. This is now explained in the text.

P1151-L25: The correlation is 0.613 for the German part of the Elbe basin - see Figure 12. But does high correlation between these estimates really justify the mistrust of the water-balance method? Wouldn't concatenation of the 133 small subcatchments into fewer larger ones with trusted gauging sites remove some of the difficulty of coping with 'leakage' at the small catchment boundaries?

As already stated regarding the similar concern of reviewer #1, it is true that an agreement between the results of two methods does not prove a third method with deviating results to be incorrect. This does not really substantiate a 'key finding' as originally

advertised here. The paragraph has been rewritten into three short paragraphs, now pointing out the principal value of using and comparing several methods for optimally assessing hydrological reality.

Regarding the idea of merging sub-basins to avoid water leakage problems, we suspect the gauging errors being the principal reason for water balance derived ET deviations, but explicitely quantifying the leakage errors (which also play a role) seems to be worth future research.

P1152-L12: This is a pedantic description. Try: These uncertainties might be driven by randomness or by lack of knowledge. Webster says 'aleatoric' is a musical term: aleâËŸA′caâËŸA′ctorâËŸA′cic: adj: characterized by chance or indeterminate elements âËŸA′z musicâËŸAž. Also, 'epistemic' is not negative: epâËŸA′ciâËŸA′csteâËŸA′cmic: adj (1922) : of or relating to knowledge or knowing : cognitive

As already stated above, the aleatoric and epistemic drawers have been completely removed.

1170: Fig 1 has very faint plotted points - difficult to read

Fig. 1 has been enhanced with respect to the line width of the symbols: they do not look faint anymore.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 10, 1127, 2013.