Response to Referee #1

(All referee comments are in Italics.)

1. *The article suggests a very promising title. Upscaling and the sensitivity of the coarse scale mean to small-scale heterogeneity are long standing problems, and are of particular interest in hydrology. However, I do not think it lives up to the promise. Mostly because it ends up with a "null" result: that small scale heterogeneity creates a considerable bias which cannot be fixed. I was very disappointed that the authors didn't push their own derivations and results one step further to actually quantify the effectiveness of a potential empirical scaling method that they propose, but instead, declared that these could not be used.*

We did not say that the scaling method could not be used?

We used a simple example to show exactly how fine scale temporal heterogeneity creates a bias. We did that using a (known) theoretical framework. The same theory also points towards a solution – the use of covariances. Herein lies the problem. It is not currently routine practice to report covariances. So the solution is not to give up – the solution is to begin reporting covariances. In our example, based on evaporation from a free water surface, the covariance that mattered for scaling in this study turned out to be between wind speed and humidity. Whether that is general in other regions remains to be determined. In that respect our aim was not to propose a general parameterisation for scaling evaporation from a free water surface (see point 2 below). Rather, we were interested in the more generic question of how this could be done.

2. *The paper spends a long time on explaining the basics of sub-grid-scale covariance. In this work the "grid" is not physical in space but has a discrete resolution in time, but space and time are exchangeable, and the introduction and derivations are the same general principals used in every eddy-covariance analysis of flux measurements, or in the sub-grid-scale parameterization of every meteorological model. As one classic example, "k" theory relates the covariance of the velocity components in scales smaller than the atmospheric boundary layer to k(dU/dz) (i.e. proportional to the vertical gradient of horizontal wind speed). I expected something very similar here, some theory (empirical or derived) that will relate the covariance terms (further derived here into the correction terms, chi). Some hints of it showed through particularly in figure 5 (that very disappointingly is wrongly referenced in the text and not discussed at all) and figure 7, but instead of using these important relationships to progress (or at least test) a generalized model, they are being quickly dismissed (or ignored, as was the case with figure 5).*

Our eventual aim is to develop a general theory like the k theory the referee mentioned. However, in this initial study our aim was much humbler – to bring the issue to the attention of our scientific community and to propose a way forward. The turbulence and meteorological community made extensive examinations of data before a synthesis using various closure theories was achieved. The same should be true here. The key is to begin reporting data (especially covariances) in a way that will eventually lead to the type of synthesis being requested by the referee. Think for example about the so-called “wind functions” reported for evaporation from a free water surface. There are many and they all have different numerical values. (For a good example of that see the compilation in the appendix of Linacre 1993 Agr. Forest Meteorol., Vol. 64.) No doubt there are measurement problems but it is also possible that there is only one underlying physical equation and that the different numerical values all arise because of different covariances between model variables at the various sites and over different time periods. That simple example highlights the magnitude of the task in front of us in terms of convincing the relevant scientific community to begin reporting covariances.

In relation to Fig. 5 (*that was noted to be wrongly referenced in the text by R1*) we noted that figure in two places in the main text (p. 10, line 8 and p. 12, line 10) but we are unable to locate the error noted by the referee.

3. *As stated in Section 6.4 “if short-term data was available” (I believe it meant highfrequency data). However, if high frequency data would exist, one could directly integrate it at the high frequency to avoid the up-scaling bias. However, in most cases, such data does not exist, and that is the entire reason for this manuscript. Nonetheless, after determining which covariance terms are important and which are negligible, figure 5 and 7 show that is would be possible to parameterize the important covariance terms if only high frequency data existed for short (and hopefully representative) periods or in other (hopefully similar) sites. Using such empirical parameterization, U2 and h can be used to estimate the covariance terms (i.e. the correction factors) with at least some degree of accuracy. I was hoping that now will come the step of trying to estimate this accuracy given the most accurate possible parameterization (using all the high frequency data for the entire period, but reduced through the empirical functions presented in figures 5,7) and trying to establish the relationship between the length of time for which we know the high frequency data and the accuracy of the parameterized prediction over the entire period (e.g. would 10 days of high frequency be enough? How much more accurate would be predictions made on 100 days of high frequency data than 10 days?). Unfortunately, this step did not happen and instead, the rather trivial and disappointing conclusion was reached in P6214L10-14. In my opinion, this missed the point for which this paper was written (at least as far as the claims in the title and introduction go, and as far as I have wished to see).*

But that is partly what we did?

Fig. 5 has been prepared using only the key covariances terms (following Eq. 10).

It would be straightforward to extend that result to ask the question posed by the referee; how many days of high frequency data are needed to correctly estimate the key covariance terms. That is a good idea and would be a neat addition to the paper. We should have done that originally.

Inspired by referee's comment, we have completely re-thought, and substantially extended, this part of the analysis. Our new results are summarised below.

Our previous results showed that there was little difference in the scaling correction at either daily or monthly time steps. This demonstrates that the correction must largely arise from a diurnal phenomenon that we have already identified – the persistent correlation between wind speed and the temperature of the evaporating surface that occurs in all seasons. Those results also suggest that the monthly mean diurnal cycle may well be sufficient to accurately estimate the total correction factor.

The key question (posed by the referee) is how many days do you need to characterise the diurnal cycle with sufficient accuracy to make the total correction factor?

To address that, we first lumped all days-of-the-year over the 2007-2010 period into a single composite year and used that to calculate the mean diurnal cycle in each of the four seasons (Fig. A). The results confirm that the correlation between wind speed and the temperature of the evaporating surface holds in all seasons.

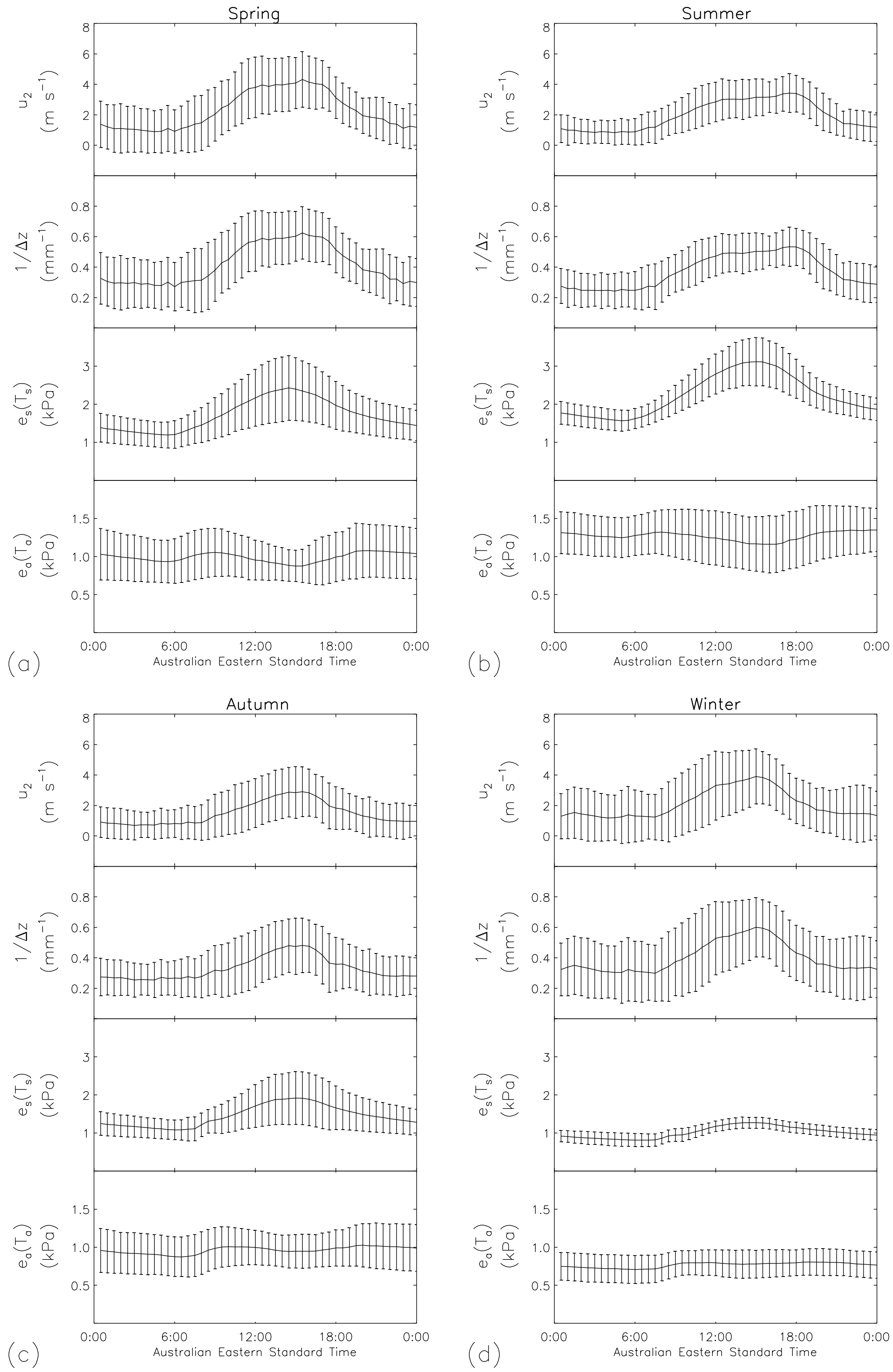

Fig. A. Mean diurnal cycle (based on half-hourly time intervals) during (a) spring (36 days), (b) summer (64 days), (c) autumn (68 days) and (d) winter (69 days). Plots show wind speed at 2 m above ground level (u_2), inverse of boundary layer thickness ($1/\Delta z$), vapour pressure at the evaporating surface ($e_s(T_s)$), air vapour pressure ($e_a(T_a)$). Error bars denote ± 1 standard deviation.

Taking the summer data (64 days available) as a good example, we then estimated the accuracy of the total correction factor as a function of the number of days used to estimate the mean diurnal cycle. In particular, we first calculated each day separately, and found the total correction factor (=1+χ_{All}) to vary between 1.0 and 1.4 (Fig. B, top panel). If the diurnal cycle was then calculated using data from two successive days we have 32 (= 64 days / 2 days) estimates of the total correction factor. In that case, the range reduced slightly (~1.1 to ~1.3). We continued this procedure using various levels of aggregation until the entire 64 days were used to calculate a single mean diurnal cycle for the summer. The resulting total correction factor was around 1.16. The exciting result was that the total correction factor stabilised (at ~ 1.16) when the mean diurnal cycle was calculated using 16 days. There is only a small loss in overall accuracy with the RMSE going from 0.6 mm d^{-1} (using all the data) to 0.9 mm d^{-1} at the threshold of ~ 15 days. This result confirms that relatively few days are required to calculate the mean diurnal cycle with sufficient accuracy to make an accurate estimate of the total correction factor.

We suspect that this would confirm the intuition of the referee.

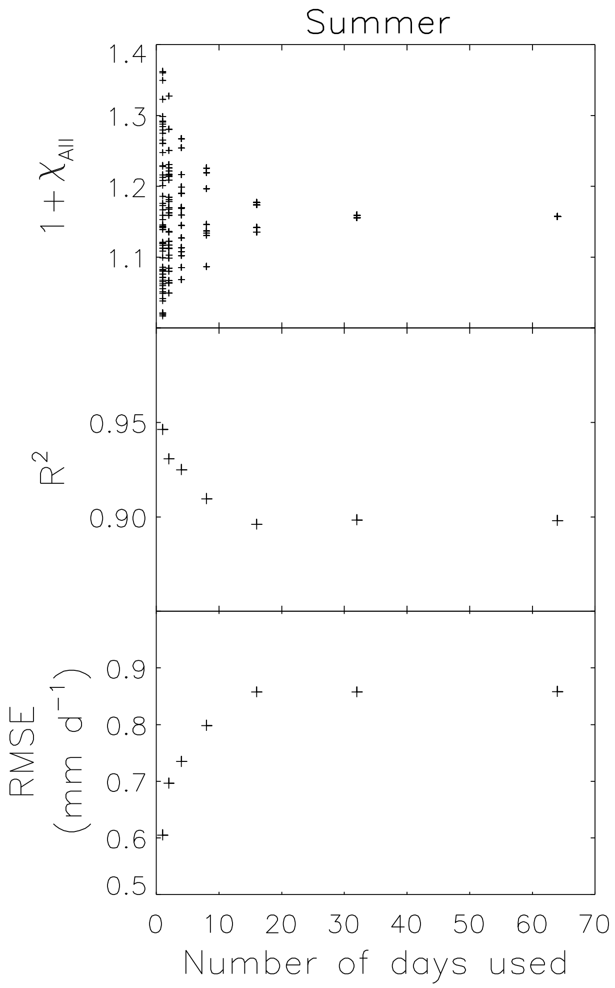

Fig. B. Accuracy of the total correction factor (1+χ_{All}) versus the number of days used to calculate the mean diurnal cycle. (Data are for summer only.)

With this extension, this approach has potential practical application. To test that, we then repeated the above calculations by calculating a mean diurnal cycle for each season using 15 days. With that single mean diurnal cycle we then estimated the total correction factor (spring ~1.18, summer ~ 1.18, autumn ~ 1.12, winter ~ 1.10) and applied that to all days in each of the four seasons. The results (Fig. C) confirm that one can make an accurate correction using a single mean diurnal cycle.

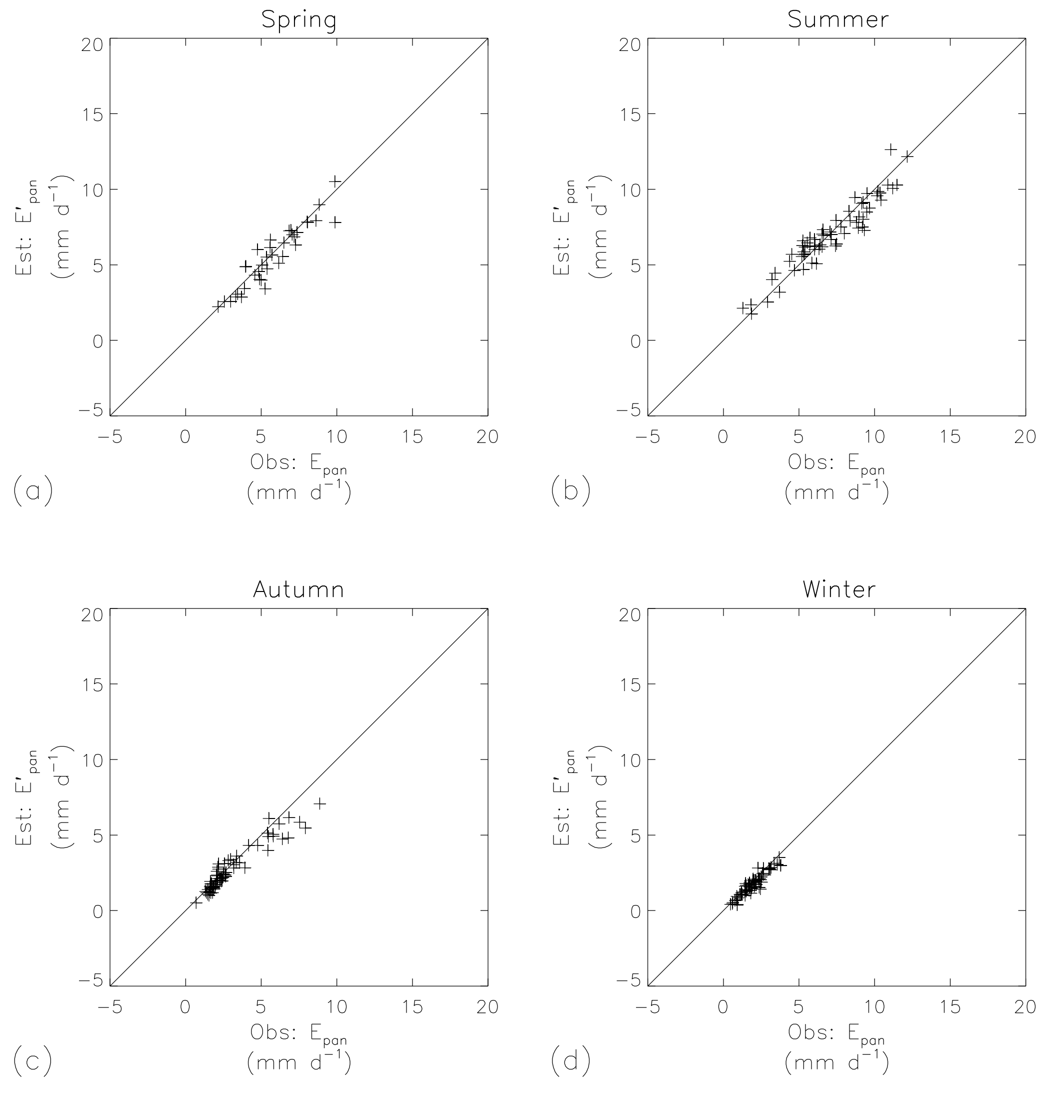

Fig. C. Estimated E'_{pan} versus observed E_{pan} using a constant total correction factor ($1+\chi_{All}$) calculated separately (using 15 days) for each season. (1:1 line shown.) (a) Spring (Regression: 0.95x + 0.09, R^2 = 0.87, n = 36, RMSE = 0.75 mm d^{-1}). (b) Summer (Regression: 0.84x + 1.06, R^2 = 0.90, n = 64, RMSE = 0.82 mm d^{-1}). (c) Autumn (Regression: 0.78x + 0.40, R^2 = 0.91, n = 68, RMSE = 0.68 mm d^{-1}). (d) Winter (Regression: 0.87x + 0.02, R^2 = 0.89, n = 69, RMSE = 0.35 mm d^{-1}).

In summary, the suggestion made by the referee has led to a major improvement in the practical application of the theory developed here.

4. *P6204 - I think the proper term is “up-scaling” and not “scaling up”.*

Noted.

5. *P6206L5-10 “Bias” is a very general term. I think it’ll be better to call this “scale bias” or “up-scaling bias”.*

Good point. Thank you.

6. *Eq.2. fv(tao) (meaning fv as a function of tau) is a bit confusing because later on (eq. 6) you use fv(es(Ta) – ea(Ta)) but there fv is not a function of (es(Ta) – ea(Ta)) but multiplying it. I suggest putting the (tao) in superscript. Same for E(tao).*

Good point. Thank you.

7. *Eq.8. It would be more elegant to simplify the equation by pulling the 3 terms for es(Ts)/(es(Ts)-ea(Ta)) together, and the same for the 3 terms for -ea(Ta)/(es(Ts)-ea(Ta)).*

We originally tried that and we agree that the rearranged equation is more elegant mathematically BUT the three terms can no longer be separately recognised and the physical meaning becomes less clear.

8. *Section 5 – Data – this section does not provide enough information to be able to understand and replicate what was done in the study. What instruments where used to measure exactly what variables? According to what formulation where derived variables such as deltaZ, es(Ts) and Dv calculated? Was there any quality control in the data processing? Gap-handling? These are critically important as they can strongly bias the covariance between observed variables.*

Comprehensive details are available in Lim et al 2012. Briefly, we measured surface temperature directly using a radiometer aimed at the water surface and subsequently calculated $e_s(T_s)$ assuming the air immediately adjacent to the surface was saturated. The boundary layer thickness Δz (deltaZ) was estimated using theory we developed and was mostly a function of the wind speed ($\Delta z \propto u^{-2/3}$). D_v (diffusivity of water vapour in air) was taken from standard handbooks (adjusted for temperature and air pressure). Air temperature and vapour pressure were all estimated using a standard installation with wind speed estimated using a cup anemometer located ~ 2 m above the ground.

Each day used here contained no missing data.

9. *P6210L10-17 - The coefficient of variation of 1/Ta is much larger at smaller time scales. For example, direct measurements of evaporation (latent heat flux) using the eddycovariance method calculate the covariance between Ta and es(Ta) from measure-ments in 10 Hz frequency and average the covariances (fluxes) over 30-minute periods.A resolution of 30 minutes is just not suitable for this measurement (temperaturehumidity covariance). It would have been great (though I suspect impossible from the data that is described here) if the analysis was expanded to the turbulence-eddy time scales (where the physical processes that are responsible for evaporation occur) and not stop at 30 minutes. I am sure that data sets of evaporation pans next to ultra-sonic anemometers and high frequency gas analyzers for water vapor exist (try looking in the ameriflx and fluxnet datasets if you do not have access to such data in your site).*

Our experimental data were originally sampled at shorter intervals (~ 5 mins) but we found that the evaporation measurement, done by measuring the change in height of the water using a very accurate linear transducer in a baffled column, was very noisy at this time resolution. Hence the decision was made to go to 30 min time steps. This largely solved the noise problem with the evaporation measurements.

At any rate, the 30 min time step does not detract from the analysis presented here in that we were able to show that model (derived from measurements made at a 30 min time interval) could be scaled rigorously if the covariances were known.

10. *Fig 5 and 7 – at what resolution where U2 and h calculated? Are these the daily averages for the daily fv(tao) (fig 5) and error (fig 7) and monthly averages for the monthly fv(tao)? Or am I missing something? Please explain.*

All calculations were based on the specified time interval. For example, the daily (monthly) data were based on daily (monthly) means with appropriate covariance corrections (per Eq. 10).

11. *Fig 5. Is fv(tao) in fig 5 calculated using equation 12, or another general empirical function?*

We do not understand this comment? Fig. 5 is calculated using Eq. 10 (at the appropriate time scale as per point 10).

12. *Fig. 5. Typically, scaling such as the one you show in fig 5 are done based on U*, the frictional velocity, which given the roughness length can be related to U at any given height.*

We started with measurements at 2 m height and did not require U*.

13. Appendix A is not needed, the derivation described there is well known.

We agree that the result is well known in mathematics and has been given several times previously in the "HESS-type" literature. However, it is still not well known in the Hydrology and Earth Science fields. For example, we regularly have colleagues tell us that the use of the theory in the appendix requires an assumption about the distribution (e.g., normal distribution, etc.) of the data. That is wrong as shown by the derivation in Appendix A.

While we respect the referee's viewpoint we would still prefer to retain the derivation.