

Reply to Reviewer 2

1 General Comment

This a very interesting paper, addressing an important environmental issue (evaporation from the Dead Sea), presenting very important data and (apparently for the first time) directly measured evaporation rates and thus adding important new information to our knowledge. There is no doubt, therefore, that the paper should eventually be published in one or the other form. There are three main pillars: first, to measure, directly, the lake's evaporation using an EC station on the shore (therewith employing the footprint of the water surface). This provides direct measurements for about 70% of the time. I think this approach is well motivated, well explained and also 'well executed' (all the necessary data treatment, corrections, QC, etc.). To estimate, as a second pillar, evaporation during the remaining 30% of the time, a statistical model is trained using the onshore wind conditions and the available information during these conditions, to estimate lake evaporation during offshore wind direction. This is very appropriate (and possibly novel) as an approach, and so is the statistical approach, its presentation (and results). However, it has the drawback that the estimated evaporation rates are among the largest during the whole year and thus contribute a substantial fraction of the total (yearly, monthly, daily – the shorter the more variable) evaporation. Therefore, the statistical model does not only have to be tested using the 'usual' tests (cross-validation, etc.) based on the available onshore conditions (what is convincingly being done), but attempts should be made to support the hypothesis that the same statistical model applies (yields the claimed 'good' statistics) when the input data stem from offshore situations (high wind speed in combination with low water vapor pressure deficit). Some suggestions are provided in major comments 2 and 3. As a third pillar, empirical estimates (for evaporation) are compared in their performance to the measured evaporation. This is adding a great deal of value to the paper. Unfortunately, for three out of four methods the comparison is not valid (because net radiation is not measured over water – which basically invalidates all the estimates; see major comment 1). Furthermore, the estimation of heat storage (in the water) is also flawed (see again major comment 1), so that all the different 'versions' tested are not really conclusive (with the exception of V1). Indeed, the results show that all the empirical methods using net radiation give results quite different from the observations (if only measured – and not estimated – values would be considered, this finding would probably be even more pronounced). The only 'reliable' empirical method is the aerodynamic approach (not using R_n). The problem with the different fields of view (for radiation and turbulent fluxes) could somehow be overcome (for example by using satellite – or other – observations [even literature values would be better than nothing] to correct for albedo differences between land and water, and by using the [simultaneously measured] land surface temperature [from the two additional EC sites over land] and the estimated lake surface temperature to correct for different longwave outgoing radiation). To actually estimate heat storage in the water from the available data I consider virtually impossible – so that the hysteresis model is possibly the only available source. Overall, the addressed issues call for truly major revisions before this paper can be published in HESS.

We thank the reviewer for the detailed and insightful review. We are sure they will help to improve the paper. Responses to individual comments are provided below. Reviewer's comments are in italic.

2 Major Comments

C1) Heat storage is calculated as the residuum (P10, l. 26): the authors use the same notation (ΔQ) as above for the 'heat storage of the lake' (p9, l. 5). So, is this meant to yield the heat storage of the lake using the local energy balance? This is not appropriate for two reasons. First of all, the energy balance is

based on the turbulent fluxes (which can, in my opinion, be interpreted as reflecting the heat fluxes in the footprint, i.e. over water – of course, if the wind direction is accordingly). Net radiation, however, has a much smaller ‘field of view’ (a circle with a radius of maybe 2 m for a measurement height of 6 m), so that the albedo is that of the land surface, and the same is true for the longwave outgoing radiation. The considered energy balance, therefore, is reflecting (a combination of) two different surfaces – and the difference cannot be attributed to ‘anything’ (if not a careful disentangling of the differences between radiation conditions is performed). Second, even if the various sensors would see the same surface, the energy balance (measured at 6 m agl) cannot be expected to be closed. There will be mean advection (possibly even in the vertical – as this location is so close to a step change in surface conditions; water – land), storage (in the air layer between the surface and the sensors!), possibly even vertical flux divergence. Finally, for the local energy balance (still assuming that all the observations correspond to the same surface type), one would also need the ground heat flux. Give the importance of this term (P10, l. 30) the authors need definitely to do something about it

Thank you for this comment. Yes, heat storage was meant to be calculated from the local energy balance. Addressing the first point of the reviewer: it’s right that the sensor mounted at the station is located over land and thus, outgoing longwave and reflected shortwave radiation do not represent the water surface. We considered this problem using following approach:

To calculate reflected shortwave radiation, we used literature values of the albedo for the Dead Sea and additionally performed a short-term experiment, with radiation measurements directly over the water surface to confirm the literature values for our site. Stanhill (1987) calculated the albedo of the Dead Sea surface from ship measurements and reports values of 0.06 in the summer months and 0.09 in the winter months and an annual average of 0.07. He also reported albedo values from Kondrat’Ev (1969) for the latitude of the Dead Sea and the cloud cover observed in the northern part of the Dead Sea, which was 0.08 for November and 0.07 as an average annual albedo value. The results of the short-term experiment concurred well with the literature values and thus the annual average of 0.07 was used in our calculations. For the longwave outgoing radiation we used the Stefan-Boltzmann equation with a water surface emissivity of $\epsilon = 0.98$ (e.g. Konda et al. (1994)) and the surface water temperature calculated with the Monin-Obukhov approach. We also compared it with the results of the short-term experiment, where we found a good agreement. We will add a paragraph to the paper where we will explain this procedure.

The second point of the reviewer discussed the problem of energy balance closure. For every EC system there are uncertainties and a possible non closure of the energy balance through e.g. mean advection. The reviewer is right, that through calculating the heat storage as a residuum, the mentioned amount of heat storage on P10 l.30 also includes the possible non closure of the energy balance. It is definitely important to mention it in this context and we will add some comments on that to the text and to the discussion.

C2) Statistical model to estimate latent heat flux during offshore conditions: (P14, l. 8) Values up to 200 Wm⁻² ...: Interestingly, the estimated values are larger than the measured values (even on average!). Given the statistical model and the high wind speeds especially during the evening hours – in combination with presumed small water vapor pressure deficit - in spring and summer (Fig. 3), this suggests that the statistical model has possibly been used outside the conditions, for which it has been ‘constructed’. In other words, the statistical model is trained for cases of high wind speeds (but possibly not even as high as the downslope winds) in combination with (relatively) small water vapor pressure deficit while it is being used for high wind speed and large Δ_e . Since these estimated values are not only to fill some gaps in the measured time series but produce the largest values for characteristic times (i.e., after the evening transition), the authors should try to make a very strong case for these estimated latent heat fluxes. In this sense, the statistical model should not only be evaluated in the ‘classical sense’ (as it is being done – and very convincingly!), but also the question (hypothesis) should be addressed, whether the onshore (training) and offshore (application) conditions are comparable. In other words, how is Δ_e over land related to Δ_e over water? For this, potentially the two additional EC sites (p3, l. 17 - they are apparently available but not used in this study) could be employed. Similarly, the question should be addressed how the strong downslope winds are related to typical wind speed over water. The question here is therefore whether there is any suitable information (possibly from other studies or sources), which would support the hypothesis that the observed (offshore) wind speed can be used to estimate the wind speed over the [entire] Dead Sea.

Thank you for this comment. To assure, that the model is not applied outside the conditions for which it has

been 'constructed', the extreme values of offshore wind velocity and vapour pressure deficit, were not considered to calculate evaporation and it was always checked that data were not extrapolated. Extreme values in this case were considered to be the 1st and 99th percentile of the data and with this regulation wind velocity and vapour pressure deficit values, which were used to calculate evaporation were within the model boundaries. Evaporation values, which could not be calculated because wind velocity or vapour pressure deficit were outside the boundaries were treated as missing values and filled with the corresponding value of the median diurnal cycle. We know that this might lead to an underestimation of evaporation, but with this procedure we took care that training and application data were comparable and that the model was not used outside the training conditions.

The wind velocity measured at the station is a point measurement, so it is not valid for the entire Dead Sea water surface. However, the offshore wind velocities measured at the station are in our opinion suitable for the water area, and thus the fetch around the station. This is also confirmed by wind lidar measurements, which were performed during the DESERVE project. The measurements showed that westerly winds in the evening regularly reached several km over the lake without losing its strength. Furthermore other studies from Weiss et al. (1988) and Hecht and Gertman (2003), evaluated data acquired in the middle of the lake. They both observed westerly winds in their data and listed strong wind events with hourly averaged velocities between 8 to 9 m s⁻¹ in the study of Weiss et al. (1988) and 10 to 12 m s⁻¹ in the study from Hecht and Gertman (2003).

For the vapour pressure deficit we don't have data from the middle of the lake to directly compare it, but we made following observations: 1) We have seen in the data that the strong westerly winds are connected with high turbulence, and even rotor formation was observed. This means that vertical mixing and air mass exchange is enhanced and thus VDP decrease should be low. 2) The fetch of the station is around 600 m. In our opinion the decrease of VDP within such a distance is not very strong considering the turbulent mixing. 3) Evaporation has a stronger dependence on wind velocity than on VDP which makes the influence of VDP variations on the results weaker. But we agree with the reviewer that this is an uncertainty which has to be mentioned and we will discuss in the discussion chapter.

We also agree that the measured evaporation is only valid for the footprint of the measurement location and that it can vary from other areas of the water surface. Estimates of the entire Dead Sea evaporation are not possible by EC measurements, only by applying models. So for future work, the presented regression model can be used to estimate evaporation for the whole water area, when using model data (wind velocity and vapour pressure deficit) at multiple locations over the Dead Sea water surface.

C3) The Discussion Section as a whole is, first of all, more a summary than a discussion. Much of what has been stated before is repeated (and the 'discussion' consists to some extent in adding some literature values). The statistical model, for example, is repeated to be good enough (no discussion), rather than addressing potential difficulties (see major comment 2).

The 'problem' with the downslope winds (having a stronger wind speed that over the lake, and (probably much) larger water vapor pressure deficit is mentioned - but only mentioned to yield a 'slight overestimation'. Based on what is this called 'slight'? Is it 10% (and would 10% be slight)? Or 50%? (but only occurring during 30% of the hours? - and what is then slight?). I think this would be a discussion.

Another point that apparently needs discussion is the fact that the radiation measurement does not 'see' the same surface as the authors want to probe with their EC system, i.e. the lake surface. In fact, I think that either all the aspects, which include Rn have either to be removed from the paper, or an estimate has to be made to establish a method to estimate Rn over water from measured Rn over land (see major comment 1). The discussion then, would consist of the associated uncertainty and the potential impact on the interpretation of the empirical methods (i.e. their performance). Given the relatively large uncertainty in the statistical model, a useful contribution to the discussion would be to test the empirical relations for only a subset of days (for the 1 day averaging period, say), for which the impact of the statistical regression model is minimal (only a few or no estimated evaporation hours, mostly measured values). If then the comparison to the 4 empirical methods (Tab 6) would be robust, this would indeed be an indication that the conclusions regarding the appropriateness of the empirical relations are supported by data (not the statistical model). This last discussion, of course, would only make sense if the 'radiation problem' had somehow been overcome. Finally, an important point for the discussion seems to be that the empirical estimates are relatively good 'average estimates' (28 days) - but do only have reduced predictive (diagnostic, that is) skill for short time scales

Thank you for the detailed remarks about the discussions sections.

We agree that the discussion of the possible uncertainties of the applied method is probably too short. We will revise this section and will address also address the 4 points of the reviewer:

1) The first point of the reviewer refers to C2 and if the regression model was used 'outside' its boundaries. As already discussed this was not the case and so an overestimation does not take place but we will discuss the possible underestimation of evaporation, due to the fact that 'extreme' values of wind velocity and VDP were not used for calculations but instead a median value was used (see also answer to C2). Furthermore, we will discuss the question of overestimating evaporation due to the possibility that VDP decreases with increasing distance to the shoreline. As already mentioned in the answer to C2, in our opinion, the influence on the evaporation results is weak but nevertheless will be discussed in the revised version of the paper.

2) Regarding the net radiation measurements, we will explain our methods for calculating net radiation in section 3 and then discuss the possible uncertainties caused by the applied methods and their impact on the empirical methods in the discussion section.

3) We also want to thank the reviewer for the idea to test the empirical formulas on a subset of days to show that the regression model does not influence the results of the empirical relations. We can realise those tests and add the results to the paper.

4) The last point of the reviewer, refers to the reduced predictive skills of the empirical approaches on short time scales. We will discuss the reduced predictive skill of the empirical approaches, especially on the sub-daily time scales and that only the aerodynamic approach can reliably be applied to short time scales.

3 Minor comments

P5, l.4 I don't think I have ever heard of a Rototronic sensor that was a typo. The company is called Rotronic.

P5, l. 4 the height of measurement is not given for the radiation, precipitation and pressure sensors. Please specify.

The measurement height for radiation is 2 m. Precipitation and pressure at 1 m height. It will be added to the paper.

P5, l. 25 I agree with the present authors (in contrast to the other reviewer) that the appropriate conversion should be based on the water temperature and not the air temperature (the differences will be small, though). $\rho w'a'z$ is already an energy flux (i.e., the kinematic flux of absolute humidity), which is only converted into energy units by multiplication with L_v . The corresponding energy (enthalpy) is calculated at the location where the process of evaporation takes place and the relevant temperature is the lake surface temperature (whether this is called T_w , the very top of the water or T_s , the very bottom of the air, is the same). I don't think that the 'constancy of the fluxes' in the SL is a valid argument for using the air temperature (at 6 m agl in the present case) since the fluxes are only 'constant' within the SL, but not below (note that the SL has also a lower boundary, i.e. the laminar layer with a thickness of some millimeters over water wherein the turbulent fluxes are zero by definition – and rapidly change to their 'atmospheric surface value' at its top). The 'SL theory' (including the constant flux assumption) produces a temperature profile (e.g., eq. 11.12 in the text book of Arya (1988)), which is based on a 'surface temperature, T_o or T_s , which actually corresponds to the temperature at the height of the (thermal) roughness length. If we assume an (ideal) stratified SL, stable or unstable, and perfect measurements with a given 'surface' latent heat flux, each measurement height would have a (slightly) different L_v (because of the temperature profile) and hence a different latent heat flux – which is inconsistent with the 'constant flux'. In order to obtain the 'surface flux' from a measurement at any height (within the SL, of course), we must therefore relate the conversion into energy units to a common height, i.e. the thermal roughness length. The actual task therefore in determining L_v is to find the temperature at the height of the thermal roughness length. I think the water temperature is a better estimate for this than the temperature at measurement height. This 'surface latent heat flux' then can be used to assess how many mm of water had been evaporated (what is one of the primary goals in the present study). In any case, two comments have to be added: first, it is in fact potential temperature that has to be used (more precisely, virtual potential temperature) – again the differences (for a 6 m level) are negligible. Second, if L_v is estimated using eq. (3), latent heat fluxes will be in [kW m⁻²] (since L_v is in [kJ kg⁻¹]) and not comparable to the sensible heat fluxes from eq (2).

Thank you for your comment. For better comparison of eq. 3 it's changed to $LE = L_v \cdot \overline{w'a'} \cdot 1000$

P6, l.17 ... the mean vertical velocity: this could be mixed up with the mean vertical velocity over the averaging period that is zero in the 'double rotation approach'. In PF, it is the mean vertical velocity over the period that is used to define the plane.

Thank you for the comment. We will make this clearer in the text and will add '...over the period that is used to define the plane'.

P6, l.20...calculations at sites..
will be changed accordingly.

P6, l. 28 below 0.5 'what'?' (units)
0.5 is 50%. This will be changed.

P7, l.3 ...data, which...
will be changed accordingly.

P7, l. 11 is reasonably good
will be changed accordingly..

P7, l. 28 is built
will be changed accordingly.

P9, l. 5 on longer time scales (or: on a longer time scale)
will be changed accordingly.

P8, l. 11ff The presentation of the methods to estimate evaporation is somewhat difficult to comprehend. I try to exemplify this for the Energy balance method. First, one is referred to Tab. 1. The first mentioned aspect of this method is, what is difficult to obtain. I then try to check (find) F_n in Tab 1 (which is apparently difficult to obtain). F_n does not appear in the given equation (but it is 'explained' below in the list of symbols – even if it is not present in any of the given equations). So, I cannot at least judge how this variable appears in the full equation or is related to other variables that do so. The same with the other neglected variables. The 'result' is called V_0 (it has an 'X' in Tab 2). Estimating 'somehow' ΔQ produces then ' V_2 ' (why is it V_2 for this method, and V_1 for the first?). Anyway, it also has an 'X' in Tab 2 (why not a V_2 ?). And some of the other methods also have an 'X' in this table for ΔQ . Overall, I think the overview on the employed approaches should be presented in a much more concise manner. The reader should be able to judge what has actually been done.

P8, l. 15 to make the confusion complete, the 'third method' is then – not the third line in Tab 1 but the fourth.....

Thank you for this remark. We will go over this section and revise this part that it will be easier for the reader to follow and identify the different sensitivity tests which were performed.

P10, l. 15 on 26% of the days will be changed accordingly.

P10, l. 16 on about 57%... will be changed accordingly.

P10, l. 18 if the along-valley flow is northerly (with what I concur judging from Fig. 1) the lake breeze would be expected to be perpendicular (easterly on the western shore). Wouldn't this mean that, what was called a 'lake breeze' before (p10, l. 11) is rather a superposition of the along-valley flow and the lake breeze? Can the authors comment on that?

We don't think that it is a superposition of the described nocturnal along-valley flow and the lake breeze, as the data showed the following: (1) When we analyse data of the station further inland and a little bit south of the shoreline station, wind during daytime has a much stronger easterly direction. (2) when we look at the third station (which is in the north) data shows a south-easterly flow during daytime. (3) Other studies, e.g. Bitan (1974, 1976); Alpert et al. (1990), all show a lake breeze development during the day, but depending on the

location of the station the exact wind direction of the lake breeze changes. Alpert et al. (1990) analyzed data in the south and found a northerly lake breeze, whereas Bitan (1974) in the north found a south-easterly flow (which corresponds to ()).

So in our opinion the north-easterly flow shown at the shoreline is not a superposition with an along valley flow but rather caused by the specific geographic location of the station, meaning the shape of the shoreline and the nearby orography which modulates the lake breeze.

P10, l. 22 in the beginning
will be changed accordingly.

P0, l. 25 on individual days
will be changed accordingly.

P12, l. 4 on most of the days ...
will be changed accordingly.

P14, l. 2 what does 'uncorrected' mean? No Webb correction, etc.? Or do the authors refer to 'only measured, no estimated (with the multiple regression model) values'? (same in Fig. 4). I think the term 'uncorrected' is not appropriate.

uncorrected means only measured values. To avoid confusion we will rename it to "only measured"

P15, l. 1 Maximum values are reached....: see major comment 2.

as explained in the answer to comment 2, the values of the model should not overestimate evaporation, as it was only used in the reliable boundaries.

P15, l.11 the uncertainty...: so, how large was it found to be?

The uncertainty due to the gap filling method was estimated using the corresponding MAD of the used timestep of the respective month. Overall the mean MAD for the different months varied between 0.019 and 0.029 mm 30 min⁻¹. For the total annual evaporation amount the uncertainty due to gap filling resulted in 81.2 mm

P16, l. 10 evaporation rates of 5.1 mm d-1,... are measured: do these 'extreme days' contain any estimated values (using the statistical model)? How about the other 'large days' throughout the year?

Only 3 out of the 72 values for these 3 'extreme' days had to be estimated as northerly winds prevailed. The estimated values were also not the maximum values reached on these days. We can also check this for other 'large days'

P16, l. 14 not shown: I think that this 'case' (if it were shown) could serve as a good example to gain some confidence in the statistical model (see major comment 2). Tab 6 'MD' must be defined.

Thank you for this comment. Indeed this event serves as a good example to show the performance of the statistical model. As mentioned in the previous comment on these 3 days only 3 out of the 72 evaporation values were estimated. When the model is applied to estimate the evaporation of these days, we get quite a good agreement for day 1 and 3 the model underestimates evaporation by only 4% and 5%. But we also see that the model potentially underestimates the extreme values as on day 2 the value was underestimated by 18%. We will add the information about this case to the paper.

P17, l. 16 MD is probably mean difference, right? Anyway, the mean differences are not given in the table to demonstrate this (for V0 they are essentially the same ...).

Thank you for this comment. MD is mean difference. The mean difference for the 30 min interval is only given in the text (P.17, l.14) and not in the table to keep the table clear and better readable. To make the comparison easier we will add the value of the mean difference for the 1d interval to the text.

P17, l. 20 I suggest to start a new paragraph for the BREB method.
will be changed accordingly.

P17, l. 25 the largest

will be changed accordingly.

P17, l. 30 22%: judging from Fig. 7 this number is probably valid for a 28 d averaging period Yes that's right. It will be added to the text to make it clear and we will add another explanation, that the numbers for the other averaging times are comparable.

p17, l. 31 coefficients
will be changed accordingly.

p17, l. 31 improved the results: I do not really agree. Indeed, the correlation coefficients do somewhat increase (but look at Fig. 6 – both versions would probably serve as examples for ‘statistics 101’ students for data sets, for which a linear model is not appropriate). At the same time, slope and offset are getting worse (this is why we usually use different statistical measures....). In my view the results of V2 (as compared to V0) simply demonstrate that the calculation of heat storage is not appropriate (see major comment 1) we will revise this description.

p17, l. 34 11%: same as above (and also in the following) these values seem to apply for the 28 day averages
That's right, please see also answer to comment P17,l.30

p19, l. 7 up to 100%: I don't think I can see this from Fig. 7d
Thank you for pointing this out. Fig. 7d has to be compared to the measured daily evaporation amounts in Fig. 5. We will add the necessary reference.

p20, l. 16 with a heat storage term....: in fact, I only now understand what actually V6 does: it fits the Penman equation to account for the missing storage term, right? So, how is this fitting process being done? If it is fitted – as I assume – using the measured evaporation, it does not come as a surprise that a negligible mean difference results. The results from this exercise have to be discussed in this light (major comment 3).

Thank you for the comment. For the fitting it was assumed that the heat storage term equals the difference between the calculated evaporation amounts using the Penman equation V4 and the measured evaporation amounts and that it can be described as a linear function of the net radiation $\Delta Q = a \cdot R_n$. With this assumption the coefficient 'a' was derived and averaged over the measurement period, which resulted in $\Delta Q = 0.77 R_n$.

As already mentioned, we will add a more detailed part to the discussion to go into detail about the uncertainties, shortcomings and possibilities of the different methods.

p21, l. 8 due to the much higher
will be changed accordingly

p22, l. 12 The BREB, Priestly-Taylor and Penman... All these methods do not employ radiation (which has a different field of view – and does not represent the water surface, see major comment 1). I would hypothesize that this is, in the first place, the reason for their bad performance. Only if the Penman method is fitted to the data, it can also produce some reasonable results.

Thank you for the comment. As already explained for comment 1, we considered the different field of view of the radiation measurements. We still think that the strongest influence is the correct representation of the heat storage term, but of course there are uncertainties connected to the calculation of the net radiation.

References

Alpert, P., Abramsky, R., and Neeman, B. U. (1990). The prevailing summer synoptic system in Israel - Subtropical high, not Persian Trough. *Israel J. Earth Sci.*, 39(2/4):93–102.

Bitan, A. (1974). The wind regime in the north-west section of the Dead-Sea. *Arch. Meteorol. Geophys. Bioklim. Ser B*, 22(4):313–335.

Bitan, A. (1976). The influence of the special shape of the Dead-Sea and its environment on the local wind system. *Arch. Meteorol. Geophys. Bioklim. Ser B*, 24(4):283–301.

- Hecht, A. and Gertman, I. (2003). *Dead Sea meteorological climate*. In Nevo, E., Oren, A., and Wasser, S., editors, *Fungal Life in the Dead Sea*, pages 68–114. International Center for Cryptogamic Plants and Fungi, Haifa.
- Konda, M., Imasato, N., Nishi, K., and Toda, T. (1994). *Measurement of the sea surface emissivity*. *J. Oceanogr.*, 50(1):17–30.
- Kondrat'Ev, K. Y. (1969). *Radiation in the Atmosphere*. International Geophysics Series, Vol. 12, Academic Press, New York.
- Stanhill, G. (1987). *The radiation climate of the dead sea*. *J. Climatol.*, 7(3):247–265.
- Weiss, M., Cohen, A., and Mahrer, Y. (1988). *Upper atmosphere measurements and meteorological measurements on the dead sea*. Technical report, Ministry of Energy and Infrastructure (in Hebrew). 19pp.