Reply to comments of referee 1

We thank the referee for reading our manuscript on the development and validation of the dynamical wetlands extent scheme. We very much appreciate the very thorough review and valuable remarks. In the following paragraphs we want to comment on the review and answer the raised questions. We will repeat the specific comments of the referee (bold font) before every paragraph.

Specific remarks

I find the scope of the study is not clear. From an hydrological point of view, a distinction/categorization of the wetland types in two main classes could be done: i) the wetland generated by an increase of the water table depth from below, i.e. wetlands resulting from vertical water fluxes (peatlands, swamps, marsh; hereafter called "saturated wetlands") and 2) the floodplains, i.e. wetlands associated with a river and resulting from horizontal flows. The main interest of the paper is to try to simulate these two kinds of wetlands with a common approach and to test if this latter could reproduce with success the big patterns of the wetland distribution at the global scale. Some indications let think it is also the aim of the authors, e.g.: p409;l14-15 "Both parts share a joint water storage" with "both parts" referring to the land surface model part computing the vertical water fluxes and a state of the art river routing model.p412;l1-5: The authors do a distinction based on the topography between areas where the "water is distributed over a large plain" .vs. areas where a "new water volume is need to raise the water table". But this aim is not clearly stated in the draft. Thus, the reader wonders if the authors model floodplains, saturated wetlands or the both until the end of the Method section. I found the problematic is not well described in the draft.

Thanks to this comment, we see that a more elaborate paragraph on the rational of our model development is needed. The major scope of our study is the development of a simple hydrological scheme dedicated to represent the global distribution and extent variability of very different kinds of wetlands. The scheme is designed for the application in complex Earth System Models (ESMs) on global scale with medium to coarse resolutions, because we feel that the representation of surface water dynamics is – albeit important – not strongly developed in such models. While an explicit representation of wetland dynamics is necessary for the calculation of CH_4 emissions, we also expect an improved simulation of the hydrological cycle due to our scheme. From this objective several limitations arise. Of course, we strive for a realistic representation of wetland extent but nevertheless our approach needs to be simple. It should be easily implementable into different ESMs and should only require boundary data which is readily available on global scale. Thus, we want to minimize the necessity to recalibrate our model parameters for different applications or setups of the ESM in order to allow for future projections and hindcast experiments as well as for present day simulations. Thus, we restrict our scheme to the use of the general water balance terms, which are considered in all ESMs, and topographical data, which is globally available.

Furthermore, we decided not to differentiate the treatment of different wetland types but to use a general approach. This simplifies the scheme as the need to derive specific parameters sets for every type of surface water body is omitted. Indeed, considering our focus on hydrological processes and the global perspective of the approach, we argue that the different surface water bodies are very similar in that respect and only vary in the relation of the different water fluxes in their water balance as well as their topographic conditions. Both aspects are explicitly accounted for in our approach. Being restricted to hydrological indices only, we also lack the means to classify our simulated water body fractions into different type of wetlands or even separate between wetlands and lakes. We kept the term "wetland" though as wetlands represent the largest fraction of surface water bodies on the land surface. Nonetheless, we find your proposed separation of wetland types into saturated wetlands and floodplains very useful. It reflects our model structure with the separate modules for the computation of vertical and horizontal water fluxes and will clearly improve the comprehensibility of our method. Please note that in our approach both water balances are coupled. Thus, we do not model only the one or the other type but also allow for wetlands that are influenced by both regimes of water flow or vary seasonally between the dominance of either horizontal or vertical water fluxes. We will include this explanation in the beginning of our manuscript to avoid any confusion about our wetland definition.

Additionally, we do not distinct wetlands based on topography as the referee assumes. Instead, the sub-grid slope distribution is used for a continuous scaling of the wetlands water volume change.

In relation to the above comment, the authors need to be much clearer on their definition of the "wetland" term. For instance, it appears that, in addition to their no-distinction, the authors simulate only wetland with a free surface water. p409;l20: "Wetlands without surface water are not accounted for." (it is also not clear how they can distinguish wetlands with/without surface water; see next point). Also, the authors consider they can model lake because lake data are used for the validation (section 3.3).

Indeed, our explicit definition is missing and will be included in our manuscript revision: Due to the mentioned simplicity criteria, all land surface water is defined as "wetland" in our model. While we do compute the water balance for wetlands below and above the soil surface, our extent dynamics calculation is confined to wetlands with surface water. Their average water level is estimated as the quotient of surface water volume and extent.

We do not claim to model lakes explicitly as neither our used topography data is fine enough to resolve lake bathymetry nor is lake water volume properly initialized. However, we solve the overall grid cell water balance and compute the water which remains on the land surface of the grid cell. Thus, we assume to model water level variations that might be also visible in lake level data. We would prefer to also compare the water levels in wetlands but unfortunately the availability of long term water level records for wetlands is rather limited.

The no-accounting for the diversity of wetlands is not addressed in the discussion while it could explain some mismatches between the data and the model simulation. For instance, the importance of floodplains process is not discussed as a potential reason for disagreement with data in the Tropics (p424;l1-10).

We did touch these issue by dividing wetlands into artificial and natural wetlands to explain some mismatches in Europe and North America. Furthermore, we argue that the limitation of our model to the simulation of inundated wetlands explains its underestimation of wetland area in peatland dominated regions. However, we agree that the discussion should be extended to other wetland type specific features too and thank for making us aware of this shortcoming.

Also, in the discussion, the authors claim the mismatch in boreal regions could be explained by the fact these regions "are dominated by water logged peatlands" (p424;l14) which are not accounted for in the model. But do this wetland type appear in the Papa et al. data, which are based on the detection of inundated areas? The authors have to be more careful not only with the definition of wetland in their model but also in potential difference in the definition for each dataset they used. For instance, p415;l23: "a high std is found for the wetland observation indicating a considerable uncertainty between the observational data sets": no discussion is given about eventual differences in the accounted wetland (e.g. do all the databases account for lakes?).

Only inundated areas are considered by the Papa et al. (2010) data. However, we compare to the ensemble mean of all four observations which do include non-inundated peatlands and thus give a higher estimate than our model. However, we admit giving too less consideration to the differences between the wetland observation data sets as we focused only on the ensemble mean distribution derived from them. As indicated by the referee, the differences can be exploited to assess which types of wetlands can be simulated well by our model and which not. Therefore, in the revised manuscript we will give more background information about the observation data sets as well as focus more on the differences between simulation and the distinct observation data instead of using only the observation ensemble mean.

More important problem is related to the lack of some important informations in the manuscript. Some assumptions on which equations are based, the explanations of the choose of an equation as well as lot of details in the whole manuscript are missing or are not well described. This prevents to evaluate in an appropriate way the pertinence of the chosen methods as well as its potential success. Lot of points need to be clarified and more details

have to be given.
At first, the authors have to explain more the strategy and the computation of the different terms in the original version of the MPI-HM model. E.g. p410;l17: the authors should give some words about the methodology of Weedon et al., 2011; in particular because, the crude way in which ET is approached seems to explain a part of the mismatch with the data in the Tropics (p424;l5. Discussion section; "Following Weedon et al., 2011, we use globally constant parameters of short grass for the vegetation height and the surface resistance").

The standard Max Planck Institute – Hydrology Model (MPI-HM) consists of the Simplified Land Surface Scheme (SL-Scheme) (Hagemann and Dümenil Gates, 2003) and the Hydrological Discharge Model (HD-Model) (Hagemann and Dümenil, 1998, 1999; Hagemann and Dümenil Gates, 2001) which is a state of the art river routing model. The main purpose of the SL-Scheme is to derive surface runoff and drainage from the climate forcing and provide those to the HD-Model. The SL-Scheme uses a simple bucket scheme for the water balance calculations and neglects the energy balance. Thus, potential evapotranspiration (PET) has to be provided externally. As we used the WATCH Forcing Data by Weedon et al. (2011) as forcing for the SL-Scheme, we also adopted their method for PET calculation and used reference crop evapotranspiration (ET_0) as proxy for PET. Using globally constant parameters of short grass follows the FAO recommendation (Allen et al., 1998) and allows for the systematic comparison of ET_0 derived from different forcing data. The application of a reference method for PET calculation instead of solving the energy balance explicitly is very common for global hydrological models (see table 1 in Haddeland et al., 2011). We will point this out in our paper in a more explicit manner as recommended by the referee.

Then, the authors have also to be more explicit on how they adapt the different terms of MPI-HM to DWES (e.g. p410;l15: "when the wetland water table is below the soil surface both water fluxes are scaled according to the actual soil moisture content"; no equations/information are given about this scaling).

We did not provide it because this is done similar to the internal computation of the original SL-Scheme and HD-Model routines. We do not want to give a detailed description of the components of the MPI-HM because these are already published. Of course we will add the respective references (Hagemann and Dümenil, 1998, 1999; Hagemann and Dümenil Gates, 2001, 2003) in our revision and include a general paragraph about the model.

Also, we do not know how the authors could compute the the water table depth in the model. And yet, this variable seems to be used to scale some terms in the different equations (p410;l15) and allows to do the distinction between the accounted/noaccounted wetland (p409;l20). How is h in the Equation (5) computed?

The surface water table h is the solution of the wetland water balance. In every grid cell a potential wetland water balance is calculated that is different from the land fraction water balance. Here, we assume that no surface runoff and only minimal drainage occur because wetlands are usually located in topographical depressions and on low permeable soils. As long as the potential water balance solution results in a non-saturated soil moisture state, the grid cell is considered to be not suitable for wetland formation. However, when enough water is available to exceed the maximum soil moisture capacity, the dynamical wetland extent scheme (DWES) uses the potential wetland water balance to compute h as the water column above the soil surface. Together with the wetland extent A the wetland water volume V is calculated as $V = A \times h$. Next, the wetland extent dynamics routine calculates the new wetland extent depending on the relative change in water volume. As the water volume has to be conserved, h is modified by the new extent as $h = \frac{V}{A}$. As an additional constraint, the extent change is limited such that h cannot decrease while the wetland is growing and vice versa. Furthermore, the wetland extent is initialized with a minimum grid cell fraction that is the larger value of either 1.E-10 or the zero slope grid cell fraction. Of course, we will included this explanation in our revision.

The authors need to explain more their equation and the assumption on which they are based: - Eq 2: how can the authors explain the "z" exponential? There are no references to previous study and no justification. - e.g. Eq 3 and p411;l4: why "the lateral outflow is only computed for wetland"? Why is there no distinction between wetland and river flow for the outflow? - Eq (6) and Fig.3: unclear. Does the grey area on the Fig.3 (or the red curve on

the same figure) correspond to the cumulative distribution of the slope index? Also what is the advantage to use the slope rather than the topographic index as used in Kaplan et al., 2006 or in TOPMODEL approaches?

We apologize for not providing enough information to follow our reasoning. As the referee explicitly mentioned the equations 2, 3 and 6 we will limit our reply to those but generally deepen the explanation of our reasoning in the revised paper.

Eq. 2 ($I_{\mathrm{lat}} = I_{\mathrm{gb}} \times f^{z}_{\mathrm{wetl}}$) deals with the separation of lateral inflow from upstream grid cells (I_{gb}) into the wetland (I_{lat}) and river reservoirs of the HD-Model depending on the wetland fraction (f_{wetl}). To our knowledge, no observations are available (and probably not even measurable) about the distribution of river flow into all surface water bodies of a grid cell sized area (0.5°× 0.5°). However, even if such data were available for some regions we doubt that it could be used to derive a best fit approximation for the whole land surface. Thus, we tried several simple ideas to find a useful parametrization for this separation. Our basic reasoning was that the ratio of lateral inflow into the wetland depends on the amount of overall grid cell inflow and the wetland covered grid cell fraction. While it is trivial that a very small wetland will get almost no inflow and a grid cell size wetland will get it all, the associated relation in between these two extremes is not clear. We tested two simple approaches: a exponential function and a tanh function. These functions can be interpreted as follows. Exponents < 1 results in a large ratio of inflow already for small wetlands meaning that wetlands are usually close to the rivers and store a considerable amount of water even while being small. For exponents > 1 the wetland inflow increase is shifted to large wetlands indicating that water flow would be confined to river channels and bypass the wetlands. The tanh function would indicate a tipping point meaning that below a certain wetland fraction the inflow is confined to the river channel but above this fraction rivers cannot bypass wetlands anymore and wetlands gain more water from the grid cell inflow. An optimization method was used to find the best parametrization for the inflow partition based on the difference between simulated and observed river discharge (see Eq. 10 in the manuscript). We give more information about this in a later paragraph of our reply.

There seems to be a misunderstanding concerning the question about Eq. 3 and the lateral outflow. In our manuscript we only explain the modification done to the original formulation of reservoir outflow (of which Eq. 3 is a standard formulation, see e.g. Singh (1988)). Of course, there is a distinct calculation of wetland outflow and river outflow in the MPI-HM. While the same equation (Eq. 3) is used for both, the lag time k is parametrized differently as explained in the manuscript. Also, lateral outflow for wetlands is only calculated for those wetlands which are inundated as we do not consider horizontal water movement within the soil layer.

We use sub grid slope instead of the topographical index because we use it for a different purpose. The topographic index relates the drainage area for a certain point with its slope. If its distribution within a grid cell is related to the mean grid cell soil moisture, the sub grid distribution of soil moisture can be derived. The fraction where the local soil moisture then exceeds the soil field capacity is considered a wetland (e.g. Kleinen et al., 2012). While the approach is an elegant solution, we see one major problems in it. As the wetland fraction depends on the mean soil moisture, it follows that there is an upper boundary for the maximum water depth and wetland fraction. For the extreme case of a grid cell with zero slope no wetland can emerge because the mean soil moisture can obviously not exceed the maximum soil moisture capacity. However, flat regions appear to be more suitable for wetland formation. In order to deal with this problem, we introduced an explicit surface water storage on top of the soil moisture storage. Also, we do not apply the topographic index to redistribute the soil moisture in the grid cell. Instead, we use slope only as an resistance factor against the growing wetland and to calculate the flow velocity within the wetland. The bigger the slope is, the slower the wetland spreads over the grid cell and the more water does it loose to horizontal outflow. For this approach it is necessary to know how the sub grid slope is related to the grid cell fraction. Figure 3 of the manuscript shows the maximum slope that occurs within a given area fraction and the analytical approximation (red line == Eq. 6) which is used in the model.

The optimization is based on two steps. Three parameters are optimized (inflow exponent, flow coefficient, slope sensitivity); two of them are optimized in the first step during which the wetland extent is fixed; the last parameter is optimized when the dynamic in wetland extent is accounting for. I find the justification of the optimization process is not clear: what is the reason of the two-steps decomposition of the optimization? Why are static wetland used for the first step? What are the explanations to optimize the inflow exponent and flow

coefficient first?

There are two reason for the decomposition of the optimization. First, to limit the number of necessary simulations and second, to limit the degrees of freedom in the model. If we want to iteratively optimize three parameters at the same time and use 5 possible values for every parameter we need $5 \times 5 \times 5$ realizations of our model. If instead we could do it one after the other and fix the remaining parameters with sensible values in the mean time, we need only $5 + 5 + 5$ realizations. Furthermore, if these fixed parameters are well chosen they can give a better constraint to the free parameter.

In case of our three parameters the inflow scheme exponent as well as the flow velocity have to be optimized simultaneously because they depend on each other. The slope sensitivity is also influenced by the former two parameters and would also feed back to them via the wetland extent. However, we temporarily disabled this feedback loop and provided prescribed wetland extent instead. Thus, we provide optimal (respectively observed) boundary conditions and, thus, give a more realistic constraint to the parameter optimization than using simulated wetland extent. Furthermore, the parameters can be optimized separately with much less model realizations. If observations would be available to give a good (globally constant) constraint to either of the first two parameters, we would have even used a three step optimization, resulting in even better parameter constraints and less model realizations.

The reasoning for the stepwise optimization will also appear in our paper revision.

The first step of the optimization is unclear: - p413;l27: "were considered using a static version of the wetland extent scheme (SWES) with prescribed wetland fraction". Is a seasonality prescribed using Papa et al. data? - the first optimization step is not totally independent to the wetland database - how does the cost function (equation 10) vary depending on z and c? The authors has to give information about the found values for the cost function and how it varies depending to the values of the parameters couple (for instance, thanks to a figure). In fact, the results of this optimization step consists in only one sentence (p414;l22: "when comparing the different maps, the best agreement on a low cost value was found for z=2 and c=1.1")

It is true that the optimization is not independent of the wetland observations. However, we did not use them as a target that we want to match with the optimized model parameters. Instead, we want to limit the degrees of freedom during the optimization by prescribing the wetland extent with observations (and thus fix f_{wetl} in Eq. 2 and s (which depends on f_{wetl}) in Eq. 5). We conducted four simulation series with the four wetland observations as boundary data and systematically varied values of discharge coefficient and inflow separation scheme (both had to be optimized simultaneously as they influence each other). As a seasonal cycle is only available for the Papa et al. (2010) data, we used the maximum fraction of wetland cover for every grid cell to make their data as consistent as possible to the other wetland data sets. Then simulated and observed river discharge curves were compared because we judge the river discharge observations to be more reliable than global wetland extent data. The model performance was calculated using a cost function (Eq. 10 in the manuscript) that evaluated the differences in peak flow month and river flow seasonality for the simulation and the observation. The results are shown in figure 1. The most robust results are found for large exponents and a low discharge coefficient. However, the cost value is quite high indicating that all simulations agree that this parameter combination lead to a decreased model performance. The lowest cost value (1.67) is found for an inflow exponent $z = 2$ and a discharge coefficient $c = 1\ m^{1/3}s^{-1}$. At this point, the robustness of the result is still reasonable ($\sigma = 0.05$). Thus, two more refinements around this point were done. The medium sampling resolution showed the lowest cost value (1.66) at $z = 1.\overline{3}3$ $c = 2.0\ m^{1/3}s^{-1}$ but a slight decrease in robustness ($\sigma = 0.06$). Eventually, the best results were found in the fine resolved sampling for $z = 2$ and $c = 1.1\ m^{1/3}s^{-1}$ with a cost value of 1.66 and $\sigma = 0.05$.

In the final version of our paper we may include this results and provide the Fig. 1 in the annex.

The 2nd optimization step is also not well justified: p414;l26: "the river discharge simulation was not compared to observations anymore but directly to the range of discharge curves which were generated by the optimized SWES". Does it mean the optimization is not based on the reduction of the gap between model and data? What are the justifications of this?

We did this because two model simulations are better comparable to each other than to observations. Looking at the differences between the simulated and observed river discharge data we do not know which

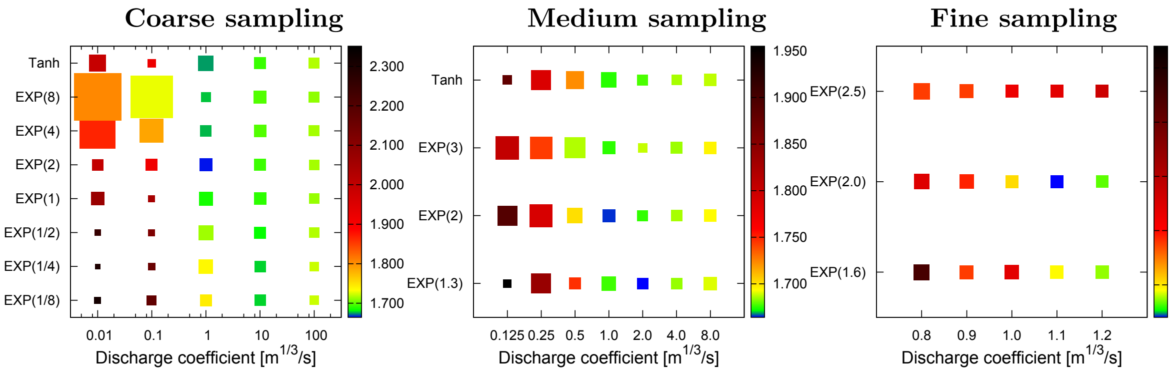

Figure 1: Optimization results for static wetland simulations with prescribed wetland extent. The color indicates the performance of the simulation in respect to river discharge observations (low values are best) and the size of the square indicates its robustness (inverse standard deviation, large squares are most robust). The sampling of parameter space of discharge coefficient and inflow parametrization is gradually refined from left to right.

part of the gap is due to wetland processes and which parts are related to other missing processes, other model parametrizations and effects of model resolution. As we did not want to compensate for such non wetland related errors, we had to design the cost function which evaluates only certain quantities of the river discharge curves and were restricted to certain river basins where we expect a good model performance. In principle, we could do the same optimization in the second step. However, thanks to the first optimization step we now possess simulation results which are optimized for the observed wetland fractions and thus are the "best" the model can produce. Thus, we did not only optimize the first two parameters but also created a benchmark for later model simulations. If we now compare the DWES simulations to this static one, we can be certain that all differences in the river discharge curves are solely caused by the parametrization of the slope sensitivity. Furthermore, we can evaluate it even on grid cell level instead of being restricted to a small number of river catchments.

A main problem of the validation is related to the use of water level in lakes. This supposes that DWES is able to simulate lake. How could the authors justify this? This remark has to be linked to the first point in the "Specific comments" section. A lake is commonly characterized by a small variability in its horizontal extent. Because no distinction is done between lake and wetland in the model, it could explain why "the DWES underestimates the range of these variations (i.e. vertical variability) for almost every locations" (pp419;l9). It is not clear how the water level is simulated in the model (see previous comment).

As previously explained we do not simulate lakes explicitly but they are included in our definition of wetlands. We would prefer to restrict our evaluation to the uses of wetland water level observations but unfortunately these are quite rare. Thus, we had to switch to lake level data. As were are aware that these data accounts only for a small part of the simulated surface water we focus our analysis on the correlation of simulated and observed seasonality. We expect the seasonality to be comparable in lakes and wetlands because two of their major water sources, rainfall and inflow from upstream areas, can be the same for both water bodies. The overall range of variability is only analysed to learn about the limitations of our approach. We agree with the referee about the reason for the underestimation of water level variability by the model. We will modify the respective paragraph in our manuscript to point out our reasoning more clearly.

The water level is the result of the water balance calculation (see reply to the question about Eq. 5).

A part of the validation is based on the comparison between the modelled/observed seasonality of wetland extent. To do so, they use the Papa et al., data (the only one dataset giving information about the wetland extent variability in time). One important result in boreal regions is the authors need to account for a snow mask as done in the Papa et al. data to match the 'observed' seasonality. p417;l14 "however, the satellites are not able to identify

wetlands below snow cover". How is this results linked with the fact that "the MPI-HM is not able to simulate the freezing of wetlands, snowfall would add directly to the wetland water storage and lead to a too high inflow during water. This behaviour is corrected by allowing for a virtual snow layer on top of the wetland" (p410;l9)? Also, do the authors use the same mask as the one used by Papa et al.? Or they compute themselves a snow mask? The comparison of the seasonality is done at very large scale (two latitude bands); it could be interesting at least to separate boreal, temperate and tropical latitudes.

We applied the monthly snow mask of Papa et al. (2010) for the analysis of seasonality. The reasoning for this it to account for missing data in the observations. It is not linked to our snow parametrization in the MPI-HM and does not match perfectly with the snow cover simulated by the MPI-HM. Its only effect on the model parametrization is that we cannot directly evaluate the effect of our virtual snow layer for wetlands because most relevant grid cells are mask out by the snow cover mask. However, the virtual snow layer effects the spring melt which we see in the simulated river runoff. Therefore, it is indirectly evaluated when analysing the river discharge of northern rivers.

The referee suggested to analyse the seasonality on more latitude bands and with a relative scale. As shown in figure 2 the relative seasonal variations of observed and simulated wetland fraction are similar in the low latitudes. For the high latitude distinct differences occur. However, if our data is corrected for snow covered areas both, the simulated and observed, seasonalities agree well. We will include this figure and discuss it more deeply in the final version of our paper.

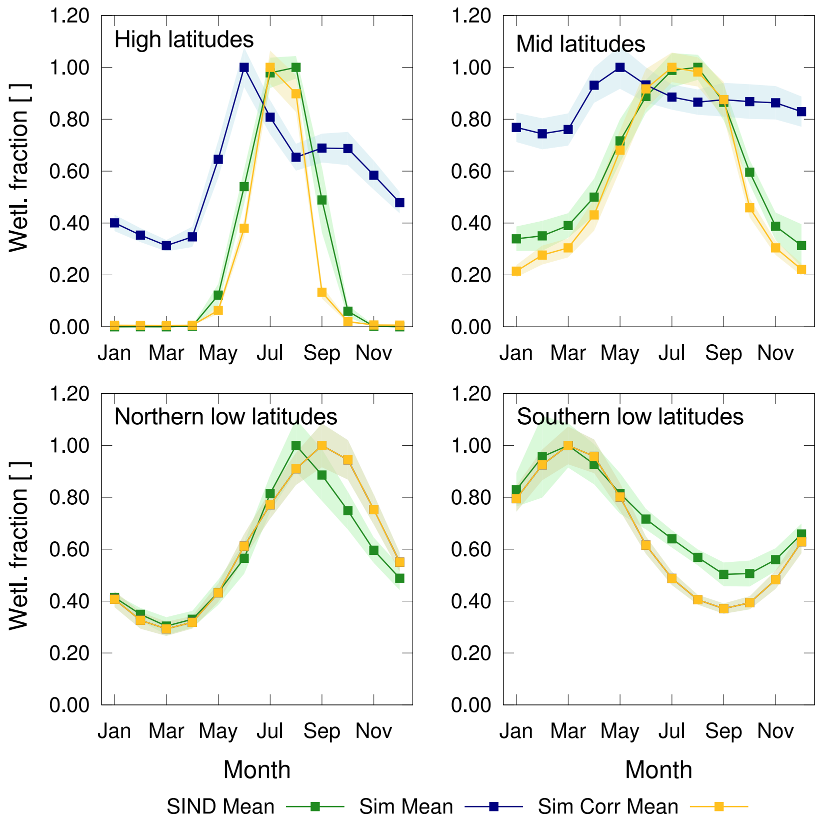

Figure 2: Mean seasonal variations of land surface wetland fraction for four different latitude bands. All curves are shown with a relative scale. The shaded areas indicate the yearly variations of wetland fraction.

Could the authors justify they do not compare the year-to-year variability of the simulated/"observed" wetland extent (this information is provided in the Papa et al. data). The study of the year-to-year variability could be a way to estimate the accuracy of the model to reproduce a good wetland extent sensitivity to the climate. The authors should justify their choice to not validate the year-to-year variability of the wetland extent.

We skipped the proposed analysis because the year-to-year variations in wetland extent is much smaller than its seasonal variation throughout the year. Additionally, only eight years (from 1993 to 2000) are available in our version of the Papa et al. (2010) data which are not enough to reveal a statistically robust wetland extent sensitivity in observation or simulation. Concerning the model accuracy and its sensitivity to short term climatic variations we argue that these are much better demonstrated by the model's representation of the seasonal wetland extent variations. A comment on this will be added to our revision.

The validation against riverflow is not clear due to not well-defined separation with the optimization (riverflows have been used for the optimization). Have some catchments been used for optimization and others for validation (p414;l18: "the analysis considered only those river catchments which include at least 40 grid-cell and have a similar area (+- 10%) in model and observations")?? Moreover, it seems the cost function between simulation/observation is not reduced while the same function have been used for the optimization (p420;l15: "as the river discharge has been used to optimize the ..., an improvement ... is expected" .vs. p420;l26: "small improvement of the discharge simulation which is almost balanced by those catchments with degraded results"). Finally, no figures is given which do not help for a better understanding of this draft part.

The purpose of this section is not the validation of our model against riverflow. Instead, we first investigate in which way the river discharge simulation is affected by the simulated wetland processes. We will move this paragraph to the discussion section and apologize for this confusion.

Secondly, we identified all catchments that are adequately represented in the model (larger than 40 grid cells and similar extent in model and observation). As the catchments differ strongly in size, location and their wetland fraction we expect the parameters to depend strongly on their selection. In order to avoid an accidental bias towards a special type of catchment all of them are used for the optimization. In order to judge the success of the optimization, we computed agreement of simulated river flow with observation (using the cost function). In this analysis we want to investigate whether there is a systematic bias in the model error indicating a more appropriate selection of catchments should be done. Finally, we found that the mean cost value is about the same when using the new model parametrization versus a control version of the MPI-HM. Obviously, the river discharge related parameters in the MPI-HM were already optimized very well and cannot profit from our new parametrization as long as globally constant parameters are used. We will move this second part to the methods sections as it still belongs to the optimization.

I found the position of the problematic relatively to the state of the art is not well described. For instance, some key studies (in particular the studies of Coe, 1997, 2000; see references below) are not quoted. p408;l28: "while these models usually lack an explicit surface water storage, this feature is included in a dynamic inundation model by Decharme et al.". What is the position of the current study relatively to Decharme et al.?

Thanks for pointing us to the additional studies by Coe and Krinner. We will include them into our introduction and emphasize the differences between the already existing models and ours much stronger. The studies of Decharme et al. (2008, 2011) and ours share the idea of a dedicated water storage for inundated areas. However, they focus on the explicit simulation of floodplains while we aim to use a more simple approach to represent floodplains as well as other types of wetlands.

The organization of the manuscript presents many problems: [. . .]

We agree with the restructuring proposed by the referee and will follow the suggestions. Especially the change in the method section (explaining the wetland dynamics first and then the water balance) and the shifted focus of the discussion (on actual changes in water balance components instead of possible feedbacks between earth system components) seem to be very sensible recommendations.

Minor/technical comments

All of the referees minor and technical comments are valuable for us and the correction will be done as proposed. Thus, here we will focus our reply to the comments which pose questions. The answers will also be mentioned in our revision.

p409;l19: is no spin-up performed to reach a steady state?
Yes it is and we apologize for not mentioning this. Of the overall simulation period from 1958 to 1999 the first five years are needed for spin up. Thus, only the period 1963 to 1999 is used for the analysis.

p413;l21: "furthermore, time series of river discharge are not available on grid cell scale, but only as integrals over catchments". It is not true; the riverflow are measured at gauging station
We will rewrite such sentences more carefully. What we meant is that river discharge observations are only available at certain locations (where gauging stations exist in reality) and not for every model grid cell. However, the measured river discharge is, of course, not only dependent on the actual location of the gauging station but represents an integral over the lateral water fluxes of the whole upstream area. We mention this to justify the use of point measurements for the optimization of global parameters.

p416;l17: "These simulation results demonstrate that the DWES is able to reproduce the large scale wetland patterns": the authors should be less affirmative. In particular when one of the following sentence (p416;l23) is "between 10N and 20S the DWES overestimates the wetland extent by a factor of three".
Here, we have to disagree with the referee because we are very specific about our affirmations. While we state clearly that our approach fails to reproduce the observed wetland extent in the low latitudes, we indeed showed that the wetland pattern (meaning the locations of the simulated wetlands) are reproduced well on the large scale.

p417;l21: "the timing coincides with the SIND data". It is difficult to say given the differences between the two curves. The authors should show the figure with a relative scale (between 0 and 1)?
We used a absolute scale because we neither wanted to hide the strong disagreement in the absolute numbers for the southern hemisphere nor the good agreement for the northern one. However, we agree that the focus of this analysis lies on the seasonal variations rather than on the extent and therefore followed this suggestion (see Fig. 2).

p417;l23: "not shown"; please show it
We will include the following figure and text:

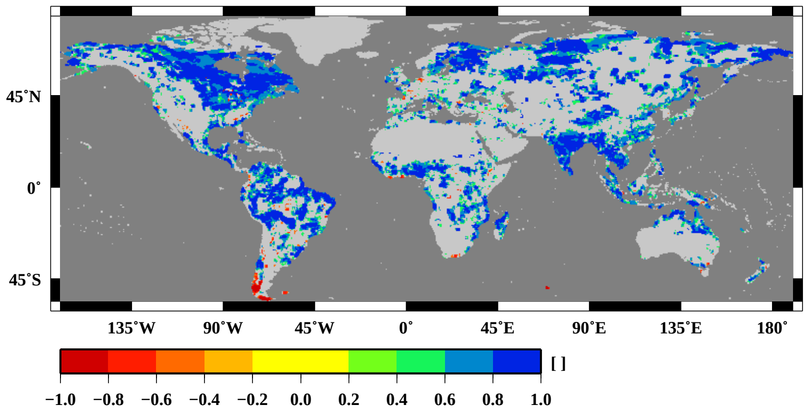

Figure 3: Temporal correlation between simulated and observed wetland extent climatology. Only grid cells with a correlation significance > 90% are shown.

Figure 3 displays the temporal correlation between the Papa et al. (2010) data and our simulation. Of course only grid cells could be considered which have wetlands in both data sets and the snow mask had to be applied, too. The correlation is generally positive and very high with the exception of a few anti-correlated time series mostly clustered at the southern tip of South America.

p422;l17 "as more water is evaporating, the runoff from the land surface is decreased". But there is an opposite effect: the increase of free surface water through the accounting for wetlands leads to increase the Dunne runoff? How do theses effects compensate each other?
We are afraid we have to disagree with the referee. Dunne runoff occurs because the moisture storage capacity of the soil is exceeded. If surface water bodies exist on top of the soil layer they provide an additional water storage and increase the storage capacity. Therefore, less Dunne runoff is occurring and this also happens later, thereby reflecting that water is stored in topographical depressions instead of being routed to the next river. This water is also subject to evaporation. Thus, the amount of water that would be available for Dunne runoff in a non-wetland simulation is to certain parts stored as surface water, evaporated from the free water surface of the wetland and routed to the next river via wetland outflow.
Of course, this is only true as long as the overall water input is identical for the non-wetland and the dynamical wetland simulation. This is guaranteed in our case because our model is not interactively coupled to the atmosphere but uses prescribed precipitation, temperature and PET forcing. In a coupled ESM simulation this effect might be different due to increased moisture recycling. We will add this information in our revised manuscript to illustrate this process more clearly.

p422;l7-8: "the larger vegetation skin reservoir of the land surface in the control runs evaporate water more easily than the wetland soils of the DWES simulation": why is the vegetation skin smaller in the simulation with DWES?
In the DWES wetlands replace the vegetated area of the grid cells to some extent. However, the skin reservoir evaporation is only computed for the non-wetland part of the vegetation. Thus, not the maximum skin reservoir capacity is larger in the control simulation but the grid cell area which provides skin evaporation. We will explain this issue more clearly in our revision.

p416;l7 "The model computes increased wetland fractions mostly for the same regions which are wetland focus regions in the observations " ?
We will rephrase this sentence to make it more easily understandable: We mean that the wetland observation data and our simulation results agree on the location of large scale wetland clusters.

p416;l19: "with deviations up to 10% of the grid-cell area": at which space scale?
at 0.5 degree horizontal resolution. We will add this information to the text.

p420;l10: "a minority of catchments show": please, give the numbers
We will include in our revision that 7 of 97 catchments show an increased amount of river discharge and an increased seasonality.

p426;l14: "we will aim to extent the DWES into the soil": what does it mean?
This sentence was poorly phrased and will be rewritten in the our revision. We mean that we want to enable our approach to calculate extent dynamics for wetlands without surface water, too.

References

Allen, R.G., L.S. Pereira, D. Raes, and M. Smith (1998). Crop evapotranspiration: Guidelines for computing crop water requirements. *FAO Irr. Drain. Paper* 56 Rome, Italy (cit. on p. 3).

Decharme, B., H. Douville, C. Prigent, F. Papa, and F. Aires (2008). A new river flooding scheme for global climate applications : Off-line evaluation over South America. *J. Geophys. Res. D Atmos.* 113(11), p. D11110. DOI: 10.1029/2007JD009376 (cit. on p. 8).

Decharme, B., R. Alkama, F. Papa, S. Faroux, H. Douville, and C. Prigent (2011). Global off-line evaluation of the ISBA-TRIP flood model. *Clim. Dyn.* Pp. 1–24. DOI: 10.1007/s00382-011-1054-9 (cit. on p. 8).

Haddeland, I., D.B. Clark, W. Franssen, F. Ludwig, F. Voß, N.W. Arnell, N. Bertrand, M. Best, S. Folwell, D. Gerten, S. Gomes, S.N. Gosling, S. Hagemann, N. Hanasaki, R. Harding, J. Heinke, P. Kabat, S. Koirala, T. Oki, J. Polcher, T. Stacke, P. Viterbo, G.P. Weedon, and P. Yeh (2011). Multimodel estimate of the global terrestrial water balance: Setup and first results. *J. Hydrometeorol.* 12(5), pp. 869–884. DOI: 10.1175/2011JHM1324.1 (cit. on p. 3).

Hagemann, S. and L. Dümenil (1998). A parametrization of the lateral waterflow for the global scale. *Clim. Dyn.* 14(1), pp. 17–31. DOI: 10.1007/s003820050205 (cit. on p. 3).

– (1999). Application of a global discharge model to atmospheric model simulations in the BALTEX region. *Nord. Hydrol.* 30(3), pp. 209–230 (cit. on p. 3).

Hagemann, S. and L. Dümenil Gates (2001). Validation of the hydrological cycle ECMWF and NCEP reanalyses using the MPI hydrological discharge model. *J. Geophys. Res. D Atmos.* 106(2), pp. 1503–1510. DOI: 10.1029/2000JD900568 (cit. on p. 3).

– (2003). Improving a subgrid runoff parameterization scheme for climate models by the use of high resolution data derived from satellite observations. *Clim. Dyn.* 21(3-4), pp. 349–359. DOI: 10.1007/s00382-003-0349-x (cit. on p. 3).

Kleinen, T., V. Brovkin, and R.J. Getzieh (2012). A dynamic model of wetland extent and peat accumulation: results for the Holocene. *Biogeosciences* 9(3), pp. 235–248. DOI: 10.5194/bg-9-235-2012 (cit. on p. 4).

Papa, F., C. Prigent, F. Aires, C. Jimenez, W.B. Rossow, and E. Matthews (2010). Interannual variability of surface water extent at the global scale, 1993-2004. *J. Geophys. Res. D Atmos.* 115(12), p. D12111. DOI: 10.1029/2009JD012674 (cit. on pp. 2, 5, 7–9).

Singh, V.P. (1988). Hydrological Systems: Rainfall-Runoff Modeling. Ed. by C. Fellows. 1 Englewood Cliffs, New Jersey: Prentice Hall, p. 480 (cit. on p. 4).

Weedon, G. P., S. Gomes, P. Viterbo, W. J. Shuttleworth, E. Blyth, H. Österle, J. C. Adam, N. Bellouin, O. Boucher, and M. Best (2011). Creation of the WATCH Forcing Data and its use to assess global and regional reference crop evaporation over land during the twentieth century. *J. Hydrometeorol.* 12 823—848. DOI: http://dx.doi.org/10.1175/2011JHM1369.1 (cit. on p. 3).