

OVERALL REPLY: We thank the Reviewer for the extensive review and take the opportunity to clarify some important points of our approach, while responding to the criticisms made. We provide detailed responses to specific comments below, but we thought it helpful to summarise the key point here: our manuscript introduces a new evapotranspiration component into a previously developed large-scale hydrological model for karst areas (Hartmann et al., 2015, GMD). We are therefore following an approach to large-scale hydrological modelling that has been widely used for climate change impact studies (e.g. Beyene et al., 2010; Sperna Weiland et al., 2012; Gosling et al., 2017). The Reviewer focuses on the use of land surface models to the same issue, which is an alternative approach currently taken to simulate climate and land cover change impacts on hydrological variables. However, we are not attempting to build a land surface model here, but rather to advance a hydrological model for large-scale applications. Indeed, we ourselves have argued in the past that these two communities should come closer together and learn from each other (Archfield et al., 2015, WRR).

We provide below a point-by-point reply to the comments of the Reviewer with additional references to previous studies that help us support our modelling choices (see list of references at the end of this report). We have found it difficult to address some of the Reviewers' specific comments, because he/she did not provide any reference to support his/her statements.

COMMENT 1: *The paper proposed by Sarrazin et al. aims at adding a new evaporation formulation to the recharge model VarKarst which specialises on the hydrology of karst systems. The aim of this development is to make the model suitable for exploring the impact of climate and land surface changes on these very sensitive hydrological structures. The main themes of these improvements are to be applicable at the large scale and to be parsimonious.*

I believe this model fails on both accounts for a simple reason, the authors have neglected the fact that evaporation is strongly controlled by the diurnal cycle of radiation and atmospheric processes. One of the main consequences of climate change is to modify the diurnal cycle at the surface and in the atmosphere. Thus the application of V2Karst to climate change is bound to produce unrealistic sensitivities. The model would be more parsimonious and more robust (because based on stronger physical grounds) if it would explicitly represent the diurnal cycle.

REPLY 1: Our model is in line with widely published approaches to climate change assessment using large-scale hydrological models (e.g. Beyene et al., 2010; Sperna Weiland et al., 2012; Gosling et al., 2017), that all neglect the diurnal cycle and apply hydrological models at a daily time step. We provide further details on this issue in the reply to the reviewer's COMMENT 3 (Penman-Monteith equation) and 9 (canopy interception). We agree that there is indeed a strong need for better comparison studies to understand how neglected processes (e.g. of diurnal cycles) affect climate change impact studies, but this is beyond the scope of the study presented here.

Moreover, the reviewer wrote "*One of the main consequences of climate change is to modify the diurnal cycle*". We agree that global average radiative forcing is projected to increase (IPCC, 2013; Van Vuuren et al., 2011), and average land surface air temperature is documented to have already increased globally (IPCC, 2013, Chapter 2 p187-188 for a global assessment, 2014, Chapter 23 p1275-1276 for an assessment for Europe) and is projected to further increase (IPCC, 2013, Chapter 12 p1062-1064 for a global assessment, 2014, Chapter 23 p1276 for an assessment for Europe). However, to our knowledge, changes in the diurnal cycle appear to be much more uncertain. Changes in temperature are more documented, possibly because the historical temperature record is more accurate than the other climate variables (Allen and Ingram, 2002).

We are aware of multiple studies that analysed the past changes in diurnal temperature range (i.e. difference between minimum and maximum daily temperature). A summary of these studies is provided in the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC, 2013). They overall indicate that decreases in diurnal temperature range have been observed, but these decreases were found to be smaller than changes in average temperature (IPCC, 2013, Chapter 2, p188). Moreover, some studies suggest that these apparent changes in diurnal temperature range may be attributed to non-climatic factors (IPCC, 2013, Chapter 2, p188). In this regard, a more recent study points out that the drivers of the past observed changes in diurnal temperature ranges are still not well understood (Davy et al., 2017). Additionally, it has been shown that climate models involved in the Coupled Model Intercomparison Project Phase 5 (CMIP5, scientific basis for the IPCC fifth Assessment Report) are reproducing poorly the past observed changes in diurnal temperature range (Lewis and Karoly, 2013), which suggests that future projections in diurnal temperature range have large uncertainties.

COMMENT 2: *Furthermore this enhancement of VarKarst neglects 30 years in the developments of land surface models. These models do not represent hydrological processes and even less karst systems, and are rightfully criticized for this. But they have specialized on the surface/atmosphere exchanges and in particular the simulation of evaporation, vegetation processes and infiltration. At no moment do the authors refer to developments in one of the three leading land surface models (JULES, ORCHIDEE and CLM) or their application to the 4 FLUXNET stations used here. A simple Google search would have shown to the authors that these open-access codes (Thanks in great part to GMD !) perform much better on these sites and do not require the tuning of so many parameters. Furthermore they are designed to be applicable at the large scale.*

I would recommend to reject the paper and encourage the authors to download one of the above mentioned land surface models and couple it to VarKarst. This would produce a model for these sensitive hydrological regions which is much more robust and produces more credible result for the impact of climate and land-cover changes. I am sorry to have to make such a harsh recommendation to GMD and in the following I will detail where I believe the basic assumptions of the authors to be wrong and where the usage of developments made for land surface models would help.

REPLY 2: For clarity, we have structured our reply in three parts: (1) we explain why we chose to develop V2karst as an evolution of a parsimonious hydrological model, rather than using more complex land surface models, (2) we attempt to compare the performance of V2Karst and land surface models at the four FLUXNET sites used in our study and (3) we specify the changes that we intend to introduce in the revised version of the manuscript to clarify these points. We do not discuss V2karst representations here, as we will explain them more specifically in REPLY 3 (Penman-Monteith equation), REPLY 6 (soil water balance), and REPLY 9 (interception).

1. Reasons for developing a parsimonious hydrological model

Land Surface Models are undeniably crucial tools because they include state-of-the-art scientific understanding of moisture and energy processes. However, parallel to the development of land surface models, a wealth of studies have drawn attention to the problem of dealing with model complexity in the context of natural systems. In fact, in natural systems, controlled experimentation to ascertain model formulations is not possible and model components and parameters tend to be poorly defined, especially at large-scales, because of a lack of data and knowledge (e.g. Young et al., 1996; Abramowitz et al., 2008; Beven and Cloke, 2012; IPCC, 2013, Chapter 9, pp790-791; Hong et al., 2017; Haughton et al., 2018). For example, Beven and

Croke (2012) highlight the fact that more complex models do not necessarily produce more robust predictions, presumably because of our lack of knowledge of natural processes and because of the uncertainty in estimates/observations of the variables needed to run such complex models. Therefore, regarding future modelling challenges, Beven and Cloke (2012) have argued that understanding which parameterisation may be more appropriate and assessing model uncertainties may be more relevant than further increasing the detail of process representation.

A recent study published in GMD (Haughton et al., 2018) argues that: “*In general, numerical LSMs [Land Surface Models] have become increasingly complex over the last 5 decades, expanding from basic bucket schemes to models that include tens or even hundreds of processes involving multiple components of the soil, biosphere, and within-canopy atmosphere. Model components may have been added on to existing models without adequate constraint on component parameters (Abramowitz, 2013) or without adequate system closure (Batty and Torrens, 2001). New component parameters may be calibrated against existing model components, leading to problems of equifinality (Medlyn et al., 2005), non-identifiability (Kavetski and Clark, 2011), and epistemological holism (Lenhard and Winsberg, 2010). These problems can often only be overcome by ensuring that each component is itself well constrained by data and numerically stable. As noted earlier, these conditions rarely exist for any given component.*” In fact, although land surface models have a strong physical basis, they also include many empirical functions that are typically difficult to constrain (Mendoza et al., 2015). More critically, the fact that land surface models include a large number of parameters (many of which hard-coded, as highlighted in Cuntz et al. (2016) or Mendoza et al. (2015)) hampers an exhaustive assessment of uncertainty and sensitivity of model predictions (Young et al., 1996). We further discuss the issue of parameter estimation in land surface models in REPLY 15. Moreover, the study by Haughton et al. (2018, GMD) shows that land surface models can be outperformed by simple empirical models, in line with the results of previous studies conducted within the Land Surface Model Benchmarking Evaluation Project (PLUMBER, Best et al., 2015).

Importantly, land surface models simulate a large range of different fluxes (e.g. sensible heat, latent heat, ground heat flux, radiation, runoff, CO₂) and the sensitivity of model parameters has been reported to vary depending on the simulated flux considered (Cuntz et al., 2016; Rosero et al., 2010; Rosolem et al., 2012). For instance, Cuntz et al. (2016) showed that a large number of parameters of the Noah land surface model are non-influential or have a very small influence on total simulated runoff. This means that parts of the land surface models may be simplified when the objective is to simulate hydrological variables. In this sense, Hong et al. (2017) highlighted the fact that model development should account for the model intended uses.

V2Karst aims to simulate seasonal and annual groundwater recharge, because these variables are appropriate to characterise the amount of renewable groundwater, and hence the amount of groundwater available to human consumption and ecosystems (e.g. Scanlon et al., 2006; Döll and Fiedler, 2008; Wada et al., 2012). As generally done in hydrological models (Table A1 of our manuscript), V2Karst focuses on solving the water balance, while it does not solve the energy balance, as land surface models do. In fact, V2Karst is not meant to be used for assessing the energy fluxes (radiation, sensible heat, latent heat and ground heat flux). An additional motivation for us not to solve the energy balance is that its inclusion increases model complexity and computational cost tremendously, which makes it difficult to perform a full uncertainty and sensitivity analysis to assess the adequateness of the different model components. By focusing on the water balance instead and by using parsimonious representations, we enable all

components of V2Karst to be subject to uncertainty and sensitivity analysis, as we have explained in Sect. 2.1 p4 L16- p5 L7 in our manuscript.

2. Performance of Land Surface Models at the four FLUXNET sites

We are not aware of studies that would allow us to directly infer that JULES, ORCHIDEE and CLM have better performance compared to V2Karst at the four FLUXNET sites, as stated by the reviewer. We found six studies in which the JULES, ORCHIDEE, CLM or Noah land surface models were applied at some of the FLUXNET sites we have used (Anav et al., 2010; Davin et al., 2011; Zhao et al., 2012; Kuppel et al., 2012; Van den Hoof et al., 2013; Chaney et al., 2016). In the following paragraphs, we explain in detail that either the results of these studies cannot be compared to our results, or that the performance of the land surface models appears to be similar to or slightly inferior to those of V2Karst.

The results of four of these studies cannot be compared with our results. The study by Chaney et al. (2016) does not specifically report performance results for any of the four FLUXNET sites. The study by Davin et al. (2011) analysed a full Regional Climate Model including the CLM model and did not present performance results regarding latent heat and soil moisture simulations, which are the two variables we analysed in our manuscript. The study by Kuppel et al. (2012) did not analyse the bias or the correlation coefficient between measured and simulated latent heat/evapotranspiration, while these two metrics were used in our analyses (Sect 4.1 of our manuscript). In fact, both metrics are important to characterise the hydrological performance of the recharge model, since the bias assesses the performance in reproducing the overall water balance, while the correlation coefficient assesses the consistency between the temporal pattern of simulated and observed latent heat/evapotranspiration. Finally, the study by Anav et al. (2010) did not report quantitative performance metrics for the individual sites.

In the two “comparable” studies (Zhao et al., 2012; Van den Hoof et al., 2013), land surface models show similar or slightly inferior performance compared to V2Karst, although it is difficult to make a fair comparison because the time periods analysed were not the same as in our manuscript. Zhao et al. (2012) assessed the squared correlation coefficient R^2 between monthly observed and simulated latent heat at the French 2 site (Puéchabon) using the ORCHIDEE model and local forcing data. They found a value of R^2 of 0.59, i.e. a correlation coefficient of $\sqrt{0.59} = 0.77$ (Zhao et al., 2012, fourth row in Figure 6). This is comparable to the performance of the V2Karst model, since we identified simulations for which the correlation coefficient between monthly simulated and observed ET was higher than 0.77 (Figure 4 in our manuscript). We wish to mention that there appears to be a mistake in the labels in Figure 6 in Zhao et al. (2012) and therefore we are not completely sure we have read the figure correctly. Finally, in Van den Hoof et al. (2013), the JULES model was tested at the German site (Hainich) using local forcing data. The bias between observed and simulated evapotranspiration for the best performing version of the model is around 29% (this percentage is inferred from Figure 2.b in Van den Hoof et al. (2013), which reports that the mean annual observed ET is 280 mm.y^{-1} and the simulated mean annual ET is 360 mm.y^{-1}). The bias is therefore above the limit of acceptance of 20% that we have used in our manuscript. Van den Hoof (2013) did not provide the monthly correlation values for individual sites, but we can infer that, for the German site, the monthly correlation coefficient is between 0.7 and 1 (Van den Hoof et al., 2013, Figure 5.b), which is comparable to our study.

3. Changes in the revised version of the manuscript

Based on the Reviewer's comment, we realise that in our manuscript we have not stated clearly enough our philosophy in developing V2Karst in line with other large-scale hydrological models. We will clarify this point in the introduction p4 L3-11 (in the paragraph that states the objectives of the study). We agree that for completeness our manuscript should also refer to the literature on land surface modelling. In the revised version of our manuscript we will thus include a discussion of the ET components of land surface models in Sect. 2.1. However, we would avoid including a comparison between the performance of the V2Karst model and other land surface models at FLUXNET sites because, as shown in our point (2) above, it is difficult to make a fair comparison between different studies that used different metrics and different time periods to evaluate the models.

COMMENT 3: *Rational to explicitly represent land cover properties :*

It is laudable for the authors to use the Penman-Monteith formulation for potential evaporation. But should they have paid attention to its derivation, they would have noted that it only provides potential evaporation over a infinitesimal time intervals as it assumes that atmospheric variables and surface states do not evolve through other processes. A constant $Rn(t)$ or $ra,can(t)$ over the course of the day is a very unsatisfactory assumption, especially under a changing climate. Because of very contrasted impact of changing atmospheric composition on long-wave and short-wave radiation, we can encounter the same Rn but with very different radiation balance, turbulent fluxes and surface temperatures. The authors will find in the literature a number of paper which examine the impact of climate change on the different potential evaporation formulation. They all recommend to use sub-diurnal solutions because of the modified diurnal dynamics.

REPLY 3: We are aware of the fact that the Penman-Monteith equation has been determined over an infinitesimal time step (e.g. Shuttleworth, 2012) and that, therefore, using sub-daily rather than daily time step for the calculation of the equation provides conditions that are closer to its theoretical derivation. Nevertheless, Penman-Monteith equation has been shown to be applicable at daily time step and is widely used at daily time step.

In particular, Allen et al. (2006) commented on applications of the Penman-Monteith reference crop framework, that was designed by the Food and Agriculture Organization of the United Nations (FAO) and the American Society of Civil Engineers (ASCE). The framework has standardised the use of the Penman-Monteith equation for both hourly and daily time step (Allen et al., 1998; Walter et al., 2005) with a focus on agricultural crop. Allen et al. (2006) stated the following: *"The favorable performance of the PM equation in many studies, when applied with 24-h (and even monthly) time steps, is somewhat surprising, since the formulation of the combination equation (combined energy balance and aerodynamic components) theoretically requires weather inputs on a nearly instantaneous basis. The general consistency and accuracy of the PM method for 24-h time steps speaks to the combination equation's robustness in estimating evaporative behavior given a particular set of meteorological conditions."* More recent papers still support both hourly and daily applications of the Penman-Monteith reference crop framework (e.g. Pereira et al., 2015). The daily framework has been included for instance in an R package for simulating evapotranspiration at daily time step (Guo et al., 2016).

The Penman-Monteith equation is widely used at daily time step. Among its applications, we can cite the Moderate Resolution Imaging Spectroradiometer (MODIS) Evapotranspiration product

(Mu et al., 2011; Running et al., 2017). Furthermore, the Penman-Monteith equation is implemented in the large-scale hydrological models PCR-GLOBWB, MacPDM and VIC (these models are reviewed in Table A1-A3 of our manuscript). These models have been applied at daily time step in climate change impact studies. For instance, Gosling and Arnell (2016) used MacPDM, Sperna Weiland et al. (2012) used PCR-GLOBWB and Beyene et al. (2010) used VIC at daily time step. These three hydrological models are also included in the Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP, Warszawski et al., 2014) that aims to assess the impact of climate change. In that context, daily simulations of PCR-GLOBWB, MacPDM and VIC are analysed for instance by Gosling et al. (2017).

Given all the above, we think that applying the Penman-Monteith equation at daily time step is one possible choice that is consistent with a wide range of published studies. The direct comparison between running models for hydrological projections at daily or sub-daily resolutions, while considering the uncertainties present in the different projections (as discussed in Beven and Cloke, 2012), would indeed be very interesting but beyond the scope of our study.

Our manuscript refers to other hydrological models that applied the Penman-Monteith equation at a daily time (Table A2). Additionally, in the revised version of the manuscript, p12 when discussing the Penman-Monteith equation, we will refer to Allen et al. (2006) (see above citation) to further justify the use of the Penman-Monteith equation at a daily time step.

COMMENT 4: *The parsimony of our representation of nature if not for us to choose. We have to prove that certain simplifications in the representation of surface processes are valid for the application we envisage. The authors aim to develop a model valid at the large scale, for climate and land surface change. Is it then reasonable to assume that over the course of a day ra_{can} does not change ? I think the development of land surface models has shown that one cannot neglect the diurnal dynamic of the opening of the stomata, the soil moisture stress or the dependence of stomatal resistance to atmospheric CO₂ concentration. If the authors believe that they have found a way to represent with a single daily value these complex processes and their interaction with the environment they should let the world know as it would allow land surface models to be simplified.*

REPLY 4: We explain in REPLY 1 and 2 why we have chosen to include a parsimonious process representation in V2Karst, which is in line with well-established large-scale hydrological modelling studies on climate change impacts. An in-depth comparison with land surface models would be interesting, but it is beyond the scope of our work here.

The evaluation of the Penman-Monteith equation, which requires the specification of the aerodynamic resistance ra_{can} , is well documented at a daily times step (see REPLY 3). Our representation does include soil moisture stress using a linear function of soil moisture which multiplies the potential evapotranspiration rate (Eq.(7) p10, and Eq.(8) p11 of our manuscript). This formulation is similar to the previous version of the model (VarKarst, Hartmann et al., 2015) and to other hydrological models (Table A2 of our manuscript).

V2Karst neglects the dependence of the stomatal resistance on atmospheric factors such as temperature, radiation, humidity and CO₂ concentration, as implemented for instance in Ball-Berry (Sellers et al., 1996) and Jarvis-Stewart (Jarvis, 1976; Stewart, 1988) schemes. Stomatal resistance is assessed as a function of a minimum stomatal resistance (which is a parameter of the model) and the leaf area index (Eq.(13) p13). This formulation is similar to the hydrological

model MacPDM (Table A2) that has been applied in climate impact studies (e.g. Gosling and Arnell, 2016) as already reported in REPLY 3. Compared to MacPDM, V2Karst also includes the dependence of the stomatal resistance on the leaf area index, although we recognise that this approach is still a strong simplification of stomatal processes. The reason for our choice comes from the fact that the value of the stomatal resistance is poorly characterised for large-scale applications, because few ground measurements of stomatal resistance are available (some are reported in Breuer et al. (2003) and Körner et al. (1995)). Moreover, the temporal variability of stomatal resistance makes its measurements particularly difficult to interpret (Breuer et al., 2003) and therefore to use in modelling applications. We have discussed the lack of ground measurements of stomatal resistance (and more generally of vegetation properties) in Sect. 2.1 p4 L20 and in Sect S1 of our supplementary material. The Ball-Berry and Jarvis-Stewart parameterisations of stomatal resistance significantly increase model complexity and in particular introduce several empirical parameters, whose values are uncertain because of the lack of stomatal resistance measurements. The constant value of the stomatal resistance implemented in V2Karst allows us to lump all the uncertainties around stomatal resistance estimation into a single parameter, that we can include in our uncertainty and sensitivity analysis. In this way, we can easily analyse the impact of the value of the stomatal resistance on simulated recharge. We obviously do not claim to “*have found a way to represent with a single daily value these complex processes and their interaction with the environment*”, but simply to have included in our model a simple approach to control the overall impact of this variable on recharge predictions.

The temporal behaviour of the stomatal resistance could be investigated in V2Karst in future studies. However, we believe that, as a first version of the model, such in-depth investigation is beyond the scope of this study.

COMMENT 5: *In their rational for their modelling strategy they only mention one land surface model : ISBA in its 1998 version. This is not up to date. Even ISBA has evolved since then and does not use any more a Jarvis type parametrisation. It now also uses a Ball-Berry type formulation which balances carbon uptake and transpiration. Please note that ISBA operates at sub-diurnal time steps.*

REPLY 5: The Reviewer is right, the description of the ISBA model is not up-to-date. In the revised version of the manuscript, we will refer to the new version of the ISBA model in Sect. 2.1 when briefly discussing the representation of ET in land surface models (as mentioned in REPLY 2).

COMMENT 6: *Soil water balance :*

The explanations of the evolution of moisture in the unsaturated zone is not very clear to me. It looks to me like a superposition of buckets with the addition of lateral flows. It has been the experience in the land surface model community that this simple representation of soil moisture limits the ability to simulate the impact of stresses on transpiration. This is particularly critical in semi-arid those encountered at 3 of the selected FLUXNET stations. What is the reasoning of the authors behind this simplification in the treatment of the unsaturated zone, apart from “parsimony” ?

REPLY 6: The evolution of moisture in the saturated zone in V2Karst is the same as in VarKarst (Hartmann et al., 2015, GMD). This representation is supported by physical reasoning and by previous testing of the VarKarst model.

As explained in Sect. 2.3.2 p8-9, percolation from a given soil layer (Soi_i) to the underlying soil layer (Soi_k) is equal to the saturation excess in Soi_i . Percolation from the deeper (third) soil layer to the epikarst layer is also equal to the third soil layer saturation excess. As explained p1 L10-11 and p8 L24 – p9 L3, the reasoning behind this simplification is that soils in karst areas typically have a high clay content (Blume et al., 2010), and therefore tend to have low unsaturated permeability (Clapp and Hornberger, 1978). Instead, saturated permeability is typically high because clay soil generally have cracks that can act as preferential flow pathways under wet conditions (Lu et al., 2016).

The representation of the process of lateral flow is a feature specific to the VarKarst model, which, together with the explicit representation of sub-surface sub-grid heterogeneity (using model vertical compartments), makes the model more appropriate for karst areas than other large models. The adequacy of the soil water balance representation in VarKarst has been tested in previous studies for previous versions of the model (Hartmann et al., 2012, 2015, 2017). In particular, Hartmann et al. (2015, 2017) showed that there is no systematic bias in the model predictions, and that recharge predictions produced by VarKarst are significantly higher than those produced by models that do not include karst processes. Further discussion of the representation of diffuse and concentrated flow in V2Karst and of the testing of the previous versions of the model are given in our response to the first review of this manuscript (REPLY 2 in ‘Response to Referee #1’ available at <https://www.geosci-model-dev-discuss.net/gmd-2017-315/>).

COMMENT 7: Evapotranspiration :

Only one vegetation type seems to be allowed per grid-box, is this correct ? Because of the strong heterogeneity of the distribution of vegetation, it has been the experience of the community that a larger number of plant functional types is needed per grid box. The strict minimum has been found to be a low and a high vegetation. This simplification will be critical for the application to larger domains and in particular in semi-arid regions where the competition of the various vegetation types for water is critical.

REPLY 7: We agree that, for large-scale applications, the model should be able to represent different vegetation types within a given simulation grid cell. The model can account for sub-grid heterogeneity in vegetation type using a ‘tile’ approach.

A ‘tile’ approach consists of subdividing each model grid cell in a number of independent units (tiles), each of which has a specific land cover (e.g. short or tall vegetation). The model can then be evaluated separately over each tile. The overall simulated fluxes for a given grid cell are computed as the area weighted average of the fluxes calculated over the tiles. The same approach is also used in other large-scale hydrological models, for instance in the Mac-PDM model (Gosling and Arnell, 2011) and in the VIC model (Bohn and Vivoni, 2016; Liang et al., 1994, <http://vic.readthedocs.io/en/master/Overview/ModelOverview/>).

We will add a sentence in Sect. 6.3 (that discusses large-scale application of V2Karst) to clarify the fact that sub-grid heterogeneity in vegetation types can be treated following a tile approach.

COMMENT 8: *Please explain here as well why the literature on vegetation modelling is not relevant for this model.*

REPLY 8: We are not sure which specific literature the reviewer is referring to. In Sect. 2.1 and Tables A1-A3 of our manuscript we have provided a detailed review of ET and vegetation modelling in other large-scale hydrological models. As mentioned in REPLY 2, we will also add a brief discussion of the ET components of land surface models in Sect. 2.1 of our revised manuscript.

We are aware of the fact that our manuscript is rather long and we have already been advised by another reviewer to shorten it, hence we would avoid adding further details on vegetation modelling besides Sect. 2.1.

COMMENT 9: *Canopy interception :*

This is another topic where the community has acquired a rich experience which could benefit the authors. The representation of canopy interception at different temporal and spatial scales has been fiercely debated in the early 90s. Thus a number of parametrisations were developed to take into account the spatial and temporal variability of interception. This would be relevant here. Do the authors believe that a rainfall event in the evening or in the morning produces the same interception loss ? Does a rainfall event of 10mm/h and 100mm/h produce the same interception ? So does the assumption of treating these processes averaged over the day have any implication on the sensitivity of V2Karst to climate change ? We know that rainfall intensity and possibly also the time of day at which precipitation will occur will change in a warmer climate.

REPLY 9: V2Karst implements a parsimonious daily interception model as done in other large-scale hydrological models.

In V2Karst, for each day, evaporation from the interception store is assessed over the vegetated fraction as the maximum value between daily precipitation, potential evapotranspiration and interception storage capacity. Our model is similar to the large-scale hydrological models WaterGap, PCR-GLOBWB and VIC (Table A3 in our manuscript). These models have been simulated at daily step to assess the hydrological impact of future changes in climate, for instance in Gosling et al. (2017) within the Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP, Warszawski et al., 2014), and in Döll et al. (2018) for WaterGap, in Sperna Weiland et al. (2012) for PCR-GLOBWB and Beyene et al. (2010) for VIC. Past studies have shown that interception can be reasonably represented at a daily time scale (Savenije, 1997; De Groen and Savenije, 2006; Gerrits et al., 2009).

We are aware of the fact that more sophisticated interception schemes have been developed, as reviewed for instance in (Muzylo et al., 2009). These more complex schemes include additional parameters compared to our formulation that includes only one parameter, namely the interception storage capacity. We chose to implement a simple interception scheme for large-scale applications because (1) the physical processes and atmospheric conditions driving evaporation from canopy interception are still poorly understood (Van Dijk et al., 2015), (2) ground measurements of evaporation from interception that would be necessary to constrain model parameters are not available (Fatichi and Pappas, 2017; Miralles et al., 2016), and (3) ground measurements of canopy resistance are also sparse and affected by large uncertainties (Van Dijk et al., 2015).

We have discussed the issues arising from lack of measurements in large-scale model applications in our model development rationale (Sect. 2.1 p4 L19-20) and in Sect S1 of our supplementary material. Similar to our representation of stomatal resistance (REPLY 4), since the representation of interception processes can only be poorly constrained for large-scale applications, we chose a simple interception scheme. In this way, we can easily include the uncertainty in interception representation in our analyses and gain a systematic understanding of the impact of this uncertainty on seasonal and annual recharge predictions (our variables of interest).

COMMENT 10: *May I point out at this stage that precipitation intensification has been observed at the sub-diurnal range. Daily mean rainfall has not yet been too much affected by climate change. On the other hand, hourly precipitation rates have been increasing faster than expected from the Clausius-Capleyron relation. Thus, the virtual experiments experiments proposed in section 4.3 are not relevant for climate change. The authors are referred to the wealth of literature published on this topic in the last few years.*

REPLY 10: We have actually found evidence suggesting that daily mean rainfall has been affected by climate change, and therefore analysing the impact of changes in daily precipitation (overall mean and temporal distribution), as done in our virtual experiment, is relevant.

The last IPCC assessment reported that many regions show a significant trend (positive or negative) in mean daily precipitation for the past period (IPCC, 2013 and particularly Fig. 2.33, reported as Figure R1 below). Additionally, a more recent study by Ye et al. (2016) found that an increase in mean daily precipitation intensity occurred in all seasons in the northern Eurasia region over the period 1966-2010. Regarding future climate, model projections indicate changes in seasonal precipitation and changes in the 95 percentile of daily precipitation, for instance in Europe (IPCC, 2014, Chapter 23, Section 23.2.2.2 and Fig. 23-2).

In the revised version of the manuscript, we will clarify the fact that daily precipitation is likely to change in the introduction section (Sect. 1).

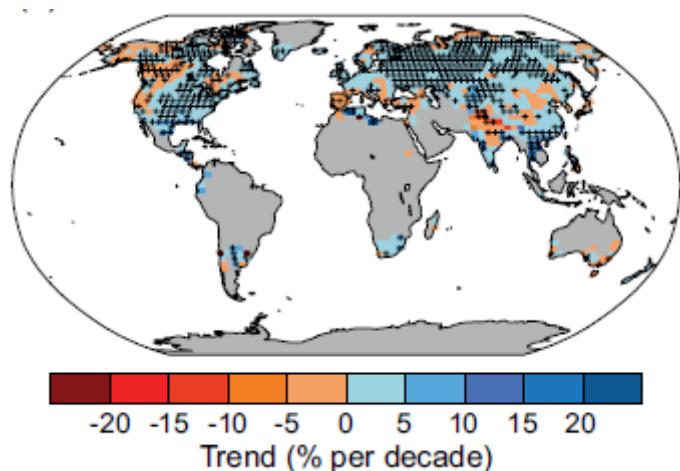


Figure R1. Trend in daily precipitation intensity over the period 1951-2010. Trends were calculated only for grid boxes that had at least 40 years of data during this period and where data ended no earlier than 2003. Grey areas indicate incomplete or missing data. Black plus signs (+) indicate grid boxes where trends are significant (i.e., a trend of zero lies outside the 90% confidence interval).

Source of the Figure: (IPCC, 2013, Figure 2.33)

COMMENT 11: *Transpiration from vegetated soil :*

Transpiration does not occur from the soils (as written in the paper) but from the stomata in the vegetation. This is not a negligible detail. Firstly the stomata only open when daylight is present and thus photosynthesis can occur. During the early afternoon, once the water in contact with the roots and within the plant has been evaporated, transpiration declines. This is caused by the slower diffusion of water within the soil which limits the supply. This is known to be a critical process for transpiration and which will be affected by higher CO₂ concentration which will lead plants to reduce the opening of their stomata. I guess these processes are neglected in the proposed model, why ? It would be a very interesting topic to see how this early afternoon depression of transpiration is affected by climate change for plants on karstic soils. It is bound to be different than on loamy soils for instance.

REPLY 11: We agree with the reviewer that “*transpiration does not occur from the soil*”. In the revised version of the manuscript, we will correct this mistake. More specifically, p11 L1 we will replace ‘Transpiration from vegetated soil’ by ‘Transpiration over the vegetated fraction’.

Regarding the reviewer’s comment “*I guess these processes are neglected in the proposed model, why ?*”, as explained in detail in REPLY 4, we neglect the dynamics of stomatal resistance because lack of observations does not allow constraining complex stomatal resistance schemes at a large-scale. We instead use a constant value of the stomatal resistance (scaled by the leaf area index) which enables us to easily analyse the effect of uncertain stomatal resistance on simulated recharge by including the stomatal resistance parameter in our sensitivity analysis.

“*How this early afternoon depression of transpiration is affected by climate change for plants on karstic soils*” is a very interesting research question but is beyond the scope of our research. Again, the aim of our modelling activity here is to (reasonably, given all data and knowledge limitations) assess the effects of climate and land cover changes on groundwater recharge at the spatial and temporal scale of interest (large/regional; seasonal/annual). Our objective is not to assess the effect of climate variations on every specific process involved in plant transpiration.

COMMENT 12: *Sorry, the assumption “... evaporation from interception is constant throughout the day ...” is not valid and will change with climate and land surface type.*

REPLY 12: We agree with the reviewer that the justification of Eq. (11) p12 L2-3 (fraction of the day with wet canopy) has been expressed incorrectly, ultimately misleading the interpretation of our proposed approach.

We do not actually make the strong assumption that “evaporation from interception is constant throughout the day”. We instead assess the fraction of the day with wet canopy as the fraction of available energy that was used to evaporate water from the interception store. This conceptual representation was proposed in Kergoat et al. (1998) and it was adopted in the dynamic vegetation model LPJ (Gerten et al., 2004; Murray, 2014) presented in Table A1-A3. In V2Karst we revised this formulation, which in the referenced studies uses the fraction of daytime with wet canopy (hence assuming that ET fluxes occur during daytime only), by using the fraction of the entire day, given that night time ET fluxes have been shown to be also important (e.g. Pearce et al., 1980; Sugita and Brutsaert, 1991; Kelliher et al., 1992).

In the revised version of the manuscript, we will clarify the point by replacing the statement p12 L2-3 (“The fraction of the day with wet canopy $t_{wet}(t)$ [–] is estimated by assuming that evaporation from interception is constant throughout the day …”) by “The fraction of the day with wet canopy $t_{wet}(t)$ [–] is estimated as the fraction of available energy that was used to evaporate water from the interception store …”.

COMMENT 13: *Parameter estimation :*

The proposed parameter estimation is difficult to interpret in view of the strong hypothesis made in the basic equations of the model. The 15 parameters of this model are so conceptual, i.e. far away from first physical principles, that indeed they can all be tuned. But given the large number of “tunable parameters” can it not be expected that the model can be made to match any dataset ? To me hveg, LAI(min,max) or z0 are not “tunable parameters” as they can either be measured or derived from turbulence theory.

REPLY 13: We agree with the Reviewer that (1) the ability to identify parameter sets that produce predictions close to observations is a necessary but not sufficient condition for the plausibility of the V2Karst model and (2) some parameters of V2Karst can be derived from a priori information. We think that we have accounted for these two points in our methodology, as we explain below.

Regarding (1), we agree that a model can produce reasonable predictions by ‘incorrectly’ activating processes to compensate for structural deficiencies. For this reason, in our evaluation of V2Karst, in addition to the comparison between observations and predictions (Sect. 4.1), we also performed two sensitivity analyses: one using measured forcing data (Sect. 4.2) and the other one using synthetic forcing data (Sect. 4.3). The aim of these analyses was precisely to assess whether the sensitivities of V2Karst outputs across sites are consistent with our understanding of the physical characteristics of those sites, and hence whether the model reproduces the observations for the right reasons by activating the appropriate controlling factors at the right place. In our manuscript, we have discussed the fact that the global sensitivity analysis applied to V2Karst showed a set of sensitivities that are interpretable in light of the different climatic conditions at the four FLUXNET sites (see Sect. 5.2.2 p22 and in Sect. 6.1 p25). Such use of sensitivity analysis as a tool for model diagnostic evaluation (verification of model structures) has been successfully applied already in several modelling studies (e.g. Rosero et al., 2010; Reusser and Zehe, 2011; Rosolem et al., 2012; Hartmann et al., 2013).

Regarding (2), we have actually used measurements at the FLUXNET sites to constrain the value of parameters hveg, LAI(min,max), Vr and Vsoi. As explained in Sect. 4.1, our parameter estimation strategy is based on the sequential application of ‘soft rules’ to identify plausible parameter values within an initial sample spanning over a wide range. Rules 1-4 select ‘plausible’ parameter sets based on their ability to produce model outputs consistent with observations, while the last rule (rule 5) selects the parameter sets that are consistent with a priori information about the sites. Constraining parameter values based on a priori information in the last step, instead of doing it in the first step as typical of other estimation strategies, provides us with a more stringent test of the adequacy of the model structure. In fact, if the parameter sets identified as plausible by Rules 1-4 (consistency with observations) are also plausible according to Rule 5 (a priori knowledge about the site), we can indeed conclude that the model ‘gives the right response for the right reason’. If instead some parameters sets identified as ‘plausible’ by the comparison with

observations (Rules 1-4) are then ruled out when including a priori information (Rule 5), this indicates some deficiencies in the model structure that are being ‘compensated for’ by tunable parameters, and points us to the model components that need improvement. In our experiment, presumably because too few data are available to constrain the simulations, Rules 1-4 result in little constraining of the parameter ranges (Figure 4, Sect. 5.1.1). However, importantly, we can identify parameter sets that satisfy all rules simultaneously and that are therefore consistent with a priori information at the site, while also producing predictions consistent with the available observations.

A similar strategy was used in previous studies, and more specifically in Hartmann (2015) to estimate the parameters of the VarKarst model and in Rosero et al. (2010) to estimate soil and vegetation parameters of the Noah land surface model.

In future large-scale applications of V2Karst, we envisage that parameters hveg, LAI(min,max), Vr and Vsoi can be a priori constrained based on grid-cell specific information about vegetation and karst landscape type (Vsoi was constrained based on karst landscape type in Hartmann et al. (2015)).

In this sense, in Sect. 6.3, we explain that some model parameters will be estimated using a priori information for the different simulation grid-cells (p27 L21 – p28 L2). In the revised version of the manuscript, we will clarify Sect 6.3 to better convey this point.

COMMENT 14: *Furthermore I find that the range of values explored for these parameters (Table 3 does not provide the limits for all 15 parameters) is much wider than realistic values I have observed.*

REPLY 14: Table 2 provides the wide ranges for all 15 parameters that were used to derive the initial parameter sample for our parameter estimation strategy. Table 3 provides the a priori ranges that were used for applying our soft rule 5. These ranges are defined only for those parameters for which site-specific a priori information was available.

All these ranges were determined in accordance with values found in the literature. References and detailed explanations are given in Sect. S3 of our Supplementary material (Table S11 and S12 p13-15).

COMMENT 15: *Land surface models also use the FLUXNET observations to “tune parameters”. But fewer parameters are adjusted and only those where the definition itself includes processes which are not modelled, i.e. are conceptual. Furthermore these parameters are specific to the plant functional type present at the FLUXNET station and then then transferred to the larger scale. This is the value of using vegetation classes in land surface models. A simple internet search for FLUXNET and the name of one of the leading land surface models, returns a large number of papers. Some where the models are simply validated and others where the observations are used to refine some vegetation parameters. The authors should have done that search during the development of their model.*

REPLY 15: We are aware of the fact that land surface models indeed typically tune a very limited number of model parameters. However, the results of recent studies suggest that excluding many parameters from calibration and uncertainty/sensitivity analysis raises a number of issues and that parameter estimation strategies of land surface models should be enhanced, as explained in the next paragraphs.

Among the parameters of land surface models that are not ‘tuned’, some are considered as ‘physical’ parameters and are commonly read from look-up tables (more specifically soil and

vegetation parameters are assigned based on soil texture and vegetation type respectively) or are set to site-specific values when measurements are available. Other parameters, which have an empirical basis, are set to constant values or are even hard-coded, i.e. are embedded in the model code as fixed values (e.g. coefficients to describe a particular shape for a curve in the model, such as the exponent used in the snow depletion curves for the melting season in the Noah land surface model as reported in Mendoza et al., (2015)). Hogue et al. (2006) highlighted the fact that many parameters in land surface models cannot be directly obtained from field measurements but through curve-fitting techniques, such as parameters controlling the soil hydraulic properties (derived for instance in Clapp and Hornberger, 1978).

Previous studies suggest that reading parameters from look-up table or setting them to fixed value are not satisfying strategies. Some studies have put into questions the physical meaning of vegetation and soil parameters in land surface models, and therefore the fact that they can be assigned from look-up tables or from site measurements. Rosero et al. (2010) analysed the sensitivity and optimal values of different vegetation and soil parameters for different version of the Noah land surface model along a precipitation gradient in the southern USA. Their results suggest that climate strongly controls the optimal soil and vegetation parameters values, which means that vegetation and soil parameters do not strictly represent vegetation and soil properties but also account for other properties. Hogue et al. (2006) analysed the differences in optimal values of different vegetation and soil parameters across five different land surface models (BUCKET, CHASM, BATS1e, BATS2, and Noah) and for five different vegetated sites in the USA. They found significant differences in optimal parameter values for the same parameters and the same site across models, which suggests that the vegetation and soil parameters have a different meaning across models. Therefore, they conclude that land surface models should be interpreted as simplified conceptual representations of natural systems, in which parameters are also conceptual representations of physical properties. The incommensurability between model parameters and physical properties can be explained by the scale mismatch between physical variables that can be actually measured and model parameters that represent cell average properties. An additional possible explanation for this incommensurability is that models implement governing equations that are well established to describe the behaviour of the system over small scales (e.g. Darcy's law), but that may not be valid for applications over larger domains (Kirchner, 2006).

Moreover, it has been shown that some empirical parameters that are typically fixed to constant values, and in particular some hard-coded parameters, have a significant impact on model predictions and therefore these parameters would need to be included in the parameter estimation strategy, i.e. to be 'tuned' (Cuntz et al., 2016; Mendoza et al., 2015).

The fact that land surface models include a large number of parameters that would potentially need to be 'tuned', and the limited availability of data to constrain model predictions at large-scales (as discussed in Sect. S1 of our supplementary material, and in REPLY 4 and 9 above) make the calibration of land surface models particularly challenging in large-scale applications. How to enhance parameter estimation of land surface models is indeed still an open question (e.g. Chaney et al., 2016).

COMMENT 16: *Conclusion :*

I am very sorry to have to write this review about the development of V2Karst. I know what a huge effort it is to develop a complex numerical model. As the authors are working in Britain, I would recommend that they look into the JULES land surface model. It is freely available and could be coupled to VarKarst to produce a very innovative tool which could indeed allow to explore the consequences of climate and land surface changes on water resources of karst aquifers. This need, to initiate a convergence between hydrological and land surface modelling, has been recognized by NERC and lead to the initiation of the HydroJULES program. The authors should contact the leaders of this program to obtain help.

REPLY 16: The V2karst V1.0 model is an extension of a large-scale hydrological model previously published in GMD (VarKarst, Hartmann et al., 2015) that aims to simulate groundwater recharge in karst areas under changing environmental boundary conditions. Hydrological models are widely applied to study groundwater recharge at large-scales. As we have stressed in REPLY 2 and 15, land surface models simulate many more fluxes than hydrological models (e.g. sensible heat, latent heat, ground heat flux, radiation, CO₂), which implies that they include many more parameters that are difficult to constrain in large-scale applications. Since our objective is to predict seasonal and annual recharge, which is the key variable of interest for water resources management in karst areas, and not to assess the other fluxes simulated by land surface models, we chose to build on and expand a hydrological model instead of using a land surface model. While we acknowledge that the approach proposed by the Reviewer may be an alternative route to achieve our goal, we hope we have provided a convincing rationale for our modelling choices and we would take the opportunity to revise our manuscript to clarify these points.

References to support our reply to the Reviewer

Abramowitz, G.: Calibration, compensating errors and data-based realism in LSMs, Presentation, 2013.

Abramowitz, G., Leuning, R., Clark, M. and Pitman, A.: Evaluating the performance of land surface Models, J. Clim., 21(21), 5468–5481, doi:10.1175/2008JCLI2378.1, 2008.

Allen, M. R. and Ingram, W. J.: Constraints on future changes in climate and the hydrologic cycle, Nature, 419(6903), 224–232, doi:10.1038/nature01092, 2002.

Allen, R. G., Pereira, L. S., Raes, D. and Smith, M.: Crop evapotranspiration: Guidelines for computing crop requirements, FAO Irrigation and Drainage Paper 56, Food and Agriculture Organization (FAO), Rome, Italy., 1998.

Allen, R. G., Pruitt, W. O., Wright, J. L., Howell, T. A., Ventura, F., Snyder, R., Itenfisu, D., Steduto, P., Berengena, J. and Yrisarry, J. B.: A recommendation on standardized surface resistance for hourly calculation of reference ETo by the FAO56 Penman-Monteith method, Agric. Water Manag., 81(1–2), 1–22, doi:10.1016/j.agwat.2005.03.007, 2006.

Anav, A., D'Andrea, F., Viovy, N. and Vuichard, N.: A validation of heat and carbon fluxes from high-resolution land surface and regional models, J. Geophys. Res. Biogeosciences, 115(G4), 13,281–13,235, doi:10.1029/2009JG001178, 2010.

Archfield, S. A., Clark, M., Arheimer, B., Hay, L. E., McMillan, H., Kiang, J. E., Seibert, J., Hakala, K., Bock, A., Wagener, T., Farmer, W. H., Andréassian, V., Attinger, S., Viglione, A., Knight, R.,

Markstrom, S. and Over, T.: Accelerating advances in continental domain hydrologic modeling, *Water Resour. Res.*, 51(12), 10078–10091, doi:10.1002/2015WR017498, 2015.

Batty, M. and Torrens, P. M.: Modeling complexity: the limits to prediction, *Cybergeo Eur. J. Geogr.*, doi:10.4000/cybergeo.1035, 2001.

Best, M. J., Abramowitz, G., Johnson, H. R., Pitman, A. J., Balsamo, G., Boone, A., Cuntz, M., Decharme, B., Dirmeyer, P. A., Dong, J., Ek, M., Guo, Z., Haverd, V., van den Hurk, B. J. J., Nearing, G. S., Pak, B., Peters-Lidard, C., Santanello, J. A., Stevens, L. and Vuichard, N.: The Plumbing of Land Surface Models: Benchmarking Model Performance, *J. Hydrometeorol.*, 16(3), 1425–1442, doi:10.1175/JHM-D-14-0158.1, 2015.

Beven, K. J. and Cloke, H. L.: Comment on “hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth’s terrestrial water” by Eric F. Wood et al., *Water Resour. Res.*, 48(1), W01801, doi:10.1029/2010WR010090, 2012.

Beyene, T., Lettenmaier, D. P. and Kabat, P.: Hydrologic impacts of climate change on the Nile River Basin: Implications of the 2007 IPCC scenarios, *Clim. Change*, 100(3), 433–461, doi:10.1007/s10584-009-9693-0, 2010.

Blume, H.-P., Brümmer, G. W., Horn, R., Kandeler, E., Kögel-Knabner, I., Kretzschmar, R., Stahr, K. and Wilke, B.-M.: Lehrbuch der Bodenkunde, Springer-Verlag, Berlin Heidelberg, doi: 10.1007/978-3-662-49960-3., 2010.

Bohn, T. J. and Vivoni, E. R.: Process-based characterization of evapotranspiration sources over the North American monsoon region, *Water Resour. Res.*, 52(1), 358–384, doi:10.1002/2015WR017934, 2016.

Breuer, L., Eckhardt, K. and Frede, H. G.: Plant parameter values for models in temperate climates, *Ecol. Modell.*, 169(2–3), 237–293, doi:10.1016/S0304-3800(03)00274-6, 2003.

Chaney, N. W., Herman, J. D., Ek, M. B. and Wood, E. F.: Deriving global parameter estimates for the Noah land surface model using FLUXNET and machine learning, *J. Geophys. Res.*, 121(22), 13,218–13,235, doi:10.1002/2016JD024821, 2016.

Clapp, R. B. and Hornberger, G. M.: Empirical equations for some soil hydraulic properties, *Water Resour. Res.*, 14(4), 601–604, doi:10.1029/WR014i004p00601, 1978.

Cuntz, M., Mai, J., Samaniego, L., Clark, M., Wulfmeyer, V., Branch, O., Attinger, S. and Thober, S.: The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model, *J. Geophys. Res.*, 121(18), 10,676–10,700, doi:10.1002/2016JD025097, 2016.

Davin, E. L., Stöckli, R., Jaeger, E. B., Levis, S. and Seneviratne, S. I.: COSMO-CLM2: A new version of the COSMO-CLM model coupled to the Community Land Model, *Clim. Dyn.*, 37(9–10), 1889–1907, doi:10.1007/s00382-011-1019-z, 2011.

Davy, R., Esau, I., Chernokulsky, A., Stephen, O. and Zilitinkevich, S.: Diurnal asymmetry to the observed global warming, *Int. J. Climatol.*, 31(1), doi:10.1002/joc.4688, 2017.

Van Dijk, A. I. J. M., Gash, J. H., Van Gorsel, E., Blanken, P. D., Cescatti, A., Emmel, C., Gielen, B., Harman, I. N., Kiely, G., Merbold, L., Montagnani, L., Moors, E., Sottocornola, M., Varlagin, A., Williams, C. A. and Wohlfahrt, G.: Rainfall interception and the coupled surface water and energy balance, *Agric. For. Meteorol.*, 214–215, 402–415, doi:10.1016/j.agrformet.2015.09.006, 2015.

Döll, P. and Fiedler, K.: Global-scale modeling of groundwater recharge, *Hydrol. Earth Syst. Sci.*, 12(3), 863–885, doi:10.5194/hess-12-863-2008, 2008.

Döll, P., Trautmann, T., Gerten, D., Muller-Schmid, H., Ostberg, S., Saeed, F. and Schleussner, C.-

F.: Risks for the global freshwater system at 1.5 °C and 2 °C global warming, *Environ. Res. Lett.*, 13(4), 23, doi:10.1088/1748-9326/aab792, 2018.

Fatichi, S. and Pappas, C.: Constrained variability of modeled T:ET ratio across biomes, *Geophys. Res. Lett.*, 44(13), 6795–6803, doi:10.1002/2017GL074041, 2017.

Gerrits, A. M. J., Savenije, H. H. G., Veling, E. J. M. and Pfister, L.: Analytical derivation of the Budyko curve based on rainfall characteristics and a simple evaporation model, *Water Resour. Res.*, 25(4), W04403, doi:10.1029/2008WR007308, 2009.

Gerten, D., Schaphoff, S., Haberlandt, U., Lucht, W. and Sitch, S.: Terrestrial vegetation and water balance - Hydrological evaluation of a dynamic global vegetation model, *J. Hydrol.*, 286(1–4), 249–270, doi:10.1016/j.jhydrol.2003.09.029, 2004.

Gosling, S. N. and Arnell, N. W.: Simulating current global river runoff with a global hydrological model: Model revisions, validation, and sensitivity analysis, *Hydrol. Process.*, 25(7), 1129–1145, doi:10.1002/hyp.7727, 2011.

Gosling, S. N. and Arnell, N. W.: A global assessment of the impact of climate change on water scarcity, *Clim. Change*, 134(3), 371–385, doi:10.1007/s10584-013-0853-x, 2016.

Gosling, S. N., Zaherpour, J., Mount, N. J., Hattermann, F. F., Dankers, R., Arheimer, B., Breuer, L., Ding, J., Haddeland, I., Kumar, R., Kundu, D., Liu, J., van Griensven, A., Veldkamp, T. I. E., Vetter, T., Wang, X. and Zhang, X.: A comparison of changes in river runoff from multiple global and catchment-scale hydrological models under global warming scenarios of 1 ??C, 2 ??C and 3 ??C, *Clim. Change*, 141(3), 577–595, doi:10.1007/s10584-016-1773-3, 2017.

De Groen, M. M. and Savenije, H. H. G.: A monthly interception equation based on the statistical characteristics of daily rainfall., 2006.

Guo, D., Westra, S. and Maier, H. R.: An R package for modelling actual, potential and reference evapotranspiration, *Environ. Model. Softw.*, 78, 216–224, doi:10.1016/j.envsoft.2015.12.019, 2016.

Hartmann, A., Lange, J., Weiler, M., Arbel, Y. and Greenbaum, N.: A new approach to model the spatial and temporal variability of recharge to karst aquifers, *Hydrol. Earth Syst. Sci.*, 16(7), 2219–2231, doi:10.5194/hess-16-2219-2012, 2012.

Hartmann, A., Wagener, T., Rimmer, A., Lange, J., Briemann, H. and Weiler, M.: Testing the realism of model structures to identify karst system processes using water quality and quantity signatures, *Water Resour. Res.*, 49(6), 3345–3358, doi:10.1002/wrcr.20229, 2013.

Hartmann, A., Gleeson, T., Rosolem, R., Pianosi, F., Wada, Y. and Wagener, T.: A large-scale simulation model to assess karstic groundwater recharge over Europe and the Mediterranean, *Geosci. Model Dev.*, 8(6), 1729–1746, doi:10.5194/gmd-8-1729-2015, 2015.

Hartmann, A., Gleeson, T., Wada, Y. and Wagener, T.: Enhanced groundwater recharge rates and altered recharge sensitivity to climate variability through subsurface heterogeneity, *Proc. Natl. Acad. Sci.*, 114(11), 2842–2847, doi:10.1073/pnas.1614941114, 2017.

Haughton, N., Abramowitz, G. and Pitman, A. J.: On the predictability of land surface fluxes from meteorological variables, *Geosci. Model Dev.*, 11(1), 195–212, doi:10.5194/gmd-11-195-2018, 2018.

Hogue, T. S., Bastidas, L. A., Gupta, H. V. and Sorooshian, S.: Evaluating model performance and parameter behavior for varying levels of land surface model complexity, *Water Resour. Res.*, 42(8), W08430, doi:10.1029/2005WR004440, 2006.

Hong, E.-M., Pachepsky, Y. A., Whelan, G. and Nicholson, T.: Simpler models in environmental studies and predictions, *Crit. Rev. Environ. Sci. Technol.*, 47(18), 1669–1712,

doi:10.1080/10643389.2017.1393264, 2017.

Van den Hoof, C., Vidale, P. L., Verhoef, A. and Vincke, C.: Improved evaporative flux partitioning and carbon flux in the land surface model JULES: Impact on the simulation of land surface processes in temperate Europe, *Agric. For. Meteorol.*, 181, 108–124, doi:10.1016/j.agrformet.2013.07.011, 2013.

IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by V. B. and P. M. M. Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, Cambridge University Press., 2013.

IPCC: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Part B: Regional Aspects, edited by V. R. Barros, C. B. Field, D. J. Dokken, M. D. Mastrandrea, K. J. Mach, T. E. Bilir, E. S. Kissel, M. Chatterjee, K. L. Ebi, Y. O. Estrada, R. C. Genova, B. Girma, A. N. Levy, S. MacCracken, P. R. Mastrandrea, and L. L. White, Cambridge University Press., 2014.

Jarvis, P. G.: The Interpretation of the Variations in Leaf Water Potential and Stomatal Conductance Found in Canopies in the Field, *Phil. Trans. R. Soc. Lond. B.*, 273(927), 593–610, doi:10.1098/rstb.1976.0035, 1976.

Kavetski, D. and Clark, M. P.: Numerical troubles in conceptual hydrology: Approximations, absurdities and impact on hypothesis testing, *Hydrol. Process.*, 25(4), 661–670, doi:10.1002/hyp.7899, 2011.

Kelliher, F. M., Whitehead, D. and Pollock, D. S.: Rainfall interception by trees and slash in a young *Pinus radiata* D. Don stand, *J. Hydrol.*, 131(1–4), 187–204, doi:10.1016/0022-1694(92)90217-J, 1992.

Kergoat, L.: A model for hydrological equilibrium of leaf area index on a global scale, *J. Hydrol.*, 212–213, 268–286, doi:10.1016/S0022-1694(98)00211-X, 1998.

Körner, C.: Leaf diffusive conductances in the major vegetation types of the globe, in *Ecophysiology of photosynthesis*, edited by E. D. Schulze and M. M. Caldwell, pp. 463–490, Springer Study Edition, vol 100, Springer, Berlin, Heidelberg., 1995.

Kuppel, S., Peylin, P., Chevallier, F., Bacour, C., Maignan, F. and Richardson, A. D.: Constraining a global ecosystem model with multi-site eddy-covariance data, *Biogeosciences*, 9(10), 3757–3776, doi:10.5194/bg-9-3757-2012, 2012.

Lenhard, J. and Winsberg, E.: Holism, entrenchment, and the future of climate model pluralism, *Stud. Hist. Philos. Sci. Part B - Stud. Hist. Philos. Mod. Phys.*, 41(3), 253–262, doi:10.1016/j.shpsb.2010.07.001, 2010.

Lewis, S. C. and Karoly, D. J.: Evaluation of historical diurnal temperature range trends in CMIP5 models, *J. Clim.*, 26(22), 9077–9089, doi:10.1175/JCLI-D-13-00032.1, 2013.

Liang, X., Lettenmaier, D. P., Wood, E. F. and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, *J. Geophys. Res.*, 99(D7), 14415–14428, doi:10.1029/94JD00483, 1994.

Lu, Y., Liu, S., Weng, L., Wang, L., Li, Z. and Xu, L.: Fractal analysis of cracking in a clayey soil under freeze-thaw cycles, *Eng. Geol.*, 208, 93–99, doi:10.1016/j.enggeo.2016.04.023, 2016.

Medlyn, B. E., Robinson, A. P., Clement, R. and McMurtrie, R. E.: On the validation of models of forest CO₂ exchange using eddy covariance data: some perils and pitfalls, *Tree Physiol.*, 25, 839–857, doi:10.1093/treephys/25.7.839, 2005.

Mendoza, P. A., Clark, M. P., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G. and Gupta, H.: Are we unnecessarily constraining the agility of complex process-based models?, *Water Resour. Res.*, 51(1), 716–728, doi:10.1002/2014WR015820, 2015.

Miralles, D. G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M. F., Hirschi, M., Martens, B., Dolman, A. J., Fisher, J. B., Mu, Q., Seneviratne, S. I., Wood, E. F. and Fernández-Prieto, D.: The WACMOS-ET project - Part 2: Evaluation of global terrestrial evaporation data sets, *Hydrol. Earth Syst. Sci.*, 20(2), 823–842, doi:doi.org/10.5194/hess-20-823-2016, 2016.

Mu, Q., Zhao, M. and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, *Remote Sens. Environ.*, 115(8), 1781–1800, doi:10.1016/j.rse.2011.02.019, 2011.

Murray, S. J.: Trends in 20th century global rainfall interception as simulated by a dynamic global vegetation model: Implications for global water resources, *Ecohydrology*, 7(1), 102–114, doi:10.1002/eco.1325, 2014.

Muzylo, A., Llorens, P., Valente, F., Keizer, J. J., Domingo, F. and Gash, J. H. C.: A review of rainfall interception modelling, *J. Hydrol.*, 370(1–4), 191–206, doi:10.1016/j.jhydrol.2009.02.058, 2009.

Pearce, A. J., Rowe, L. K. and Stewart, J. B.: Night time, wet canopy evaporation rates and the water balance of an evergreen mixed forest, *Water Resour. Res.*, 16(5), 955–959, doi:10.1029/WR016i005p00955, 1980.

Pereira, L. S., Allen, R. G., Smith, M. and Raes, D.: Crop evapotranspiration estimation with FAO56: Past and future, *Agric. Water Manag.*, 147, 4–20, doi:10.1016/j.agwat.2014.07.031, 2015.

Reusser, D. E. and Zehe, E.: Inferring model structural deficits by analyzing temporal dynamics of model performance and parameter sensitivity, *Water Resour. Res.*, 47(7), W07550, doi:10.1029/2010WR009946, 2011.

Rosero, E., Yang, Z. L., Wagener, T., Gulden, L. E., Yatheendradas, S. and Niu, G.-Y.: Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the Noah land surface model over transition zones during the warm season, *J. Geophys. Res. Atmos.*, 115(D3), D03106, doi:10.1029/2009JD012035, 2010.

Rosolem, R., Gupta, H. V., Shuttleworth, W. J., Zeng, X. and De Gonçalves, L. G. G.: A fully multiple-criteria implementation of the Sobol' method for parameter sensitivity analysis, *J. Geophys. Res.*, 117(D7), D07103, doi:10.1029/2011JD016355, 2012.

Running, S., Mu, Q. and Zhao, M.: MOD16A2 MODIS/Terra Net Evapotranspiration 8-Day L4 Global 500m SIN Grid V006, [Data set], , doi:10.5067/MODIS/MOD16A2.00, 2017.

Savenije, H. H. G.: Determination of evaporation from a catchment water balance at a monthly time scale, *Hydrol. Earth Syst. Sci.*, 1(1), 93–100, doi:10.5194/hess-1-93-1997, 1997.

Scanlon, B. R., Keese, K. E., Flint, A. L., Flint, L. E., Gaye, C. B., Edmunds, W. M. and Simmers, I.: Global synthesis of groundwater recharge in semiarid and arid regions, *Hydrol. Processes*, 20(15), 3335–3370, doi:doi.org/10.1002/hyp.6335, 2006.

Sellers, P. J., Randall, D. A., Collatz, G. J., Berry, J. A., Field, C. B., Dazlich, D. A., Zhang, C., Collelo, G. D. and Bounoua, L.: A revised land surface parameterization (SiB2) for atmospheric GCMs. Part I: Model formulation, *J. Clim.*, 9(4), 676–705, doi:10.1175/1520-0442(1996)009<0676:ARLSPF>2.0.CO;2, 1996.

Shuttleworth, W. J.: *Terrestrial Hydrometeorology*, John Wiley & Sons, Ltd, Chichester, UK., 2012.

Sperna Weiland, F. C., Van Beek, L. P. H., Kwadijk, J. C. J. and Bierkens, M. F. P.: Global patterns of change in discharge regimes for 2100, *Hydrol. Earth Syst. Sci.*, 16(4), 1047–1062, doi:10.5194/hess-16-1047-2012, 2012.

Stewart, J. B.: Modelling surface conductance of pine forest, *Agric. For. Meteorol.*, 43(1), 19–35, doi:10.1016/0168-1923(88)90003-2, 1988.

Sugita, M. and Brutsaert, W.: Daily evaporation over a region from lower boundary layer profiles measured with radiosondes, *Water Resour. Res.*, 27(5), 747–752, doi:10.1029/90WR02706, 1991.

Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J. F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J. and Rose, S. K.: The representative concentration pathways: An overview, *Clim. Change*, 109, 5–31, doi:10.1007/s10584-011-0148-z, 2011.

Wada, Y., Van Beek, L. P. H. and Bierkens, M. F. P.: Nonsustainable groundwater sustaining irrigation: A global assessment, *Water Resour. Res.*, 48(6), W00L06, doi:10.1029/2011WR010562, 2012.

Walter, I. A., Allen, R. G., Elliott, R., Itenfisu, D., Brown, P., Jensen, M. E., Mecham, B., Howell, T. A., Snyder, R., Eching, S., Spofford, T., Hattendorf, M., Martin, D., Cuenca, R. H. and WrightIvan, J. L.: The ASCE standardized reference evapotranspiration equation, ASCE-EWRI Task Committee Report, Environmental and Water Resources Institute of the American Society of Civil Engineers., 2005.

Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O. and Schewe, J.: The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP): Project framework, *Proc. Natl. Acad. Sci.*, 111(9), 3228–3232, doi:10.1073/pnas.1312330110, 2014.

Ye, H., Fetzer, E. J., Behrangi, A., Wong, S., Lambrightsen, B. H., Wang, C. Y., Cohen, J. and Gamelin, B. L.: Increasing daily precipitation intensity associated with warmer air temperatures over northern Eurasia, *J. Clim.*, 29(2), 623–636, doi:10.1175/JCLI-D-14-00771.1, 2016.

Young, P., Parkinson, S. and Lees, M.: Simplicity out of complexity in environmental modelling: Occam's razor revisited, *J. Appl. Stat.*, 23(2–3), 165–210, doi:10.1080/02664769624206, 1996.

Zhao, Y., Ciais, P., Peylin, P., Viovy, N., Longdoz, B., Bonnefond, J. M., Rambal, S., Klumpp, K., Olioso, A., Cellier, P., Maignan, F., Eglin, T. and Calvet, J. C.: How errors on meteorological variables impact simulated ecosystem fluxes: A case study for six French sites, *Biogeosciences*, 9(7), 2537–2564, doi:10.5194/bg-9-2537-2012, 2012.