Response to reviewers – details and changes to the manuscript

In the following tables, we provide a detailed response to the four reviewers' comments. When responses have numbered items we have kept the same numbers. Lines and section numbers refer to the revised manuscript (with tracked changes) attached to this response.

Reviewer 1:

Main points	Response
1.Sensitivity analyses performed on a case study in Cape Fear	We thank the reviewer for this comment and agree that the work on the Budyko framework is appealing. Because of the focus on ecosystem service decisions, we think that the uncertainties associated with the use of the Budyko theory at the pixel level need to be put in context of other model uncertainties so as to appropriately inform decisions. Change: We have clarified the scope of the paper both in the introduction and with an overview of the Methods section [l.97 on, and l. 180 on].
2.No assessment of ecosystem services	The reviewer suggests to emphasize the ecosystem service application. And, it is true that the focus of this paper is on the water yield component, as stated on p11003 ("The biophysical module, the focus of this paper, is based on the Budyko theory"…). At the same time, the emphasis on uncertainty quantification and assessment fits directly with the decision-making context of ecosystem services. Change: we clarified in the introduction the typical applications of the model for ecosystem service assessment [l90-96]
3.Broader implications of the paper (eco-hydrological relevance of the Cape Fear region and the relevance of the results in the Budyko framework)	We thank the reviewer for these comments and have expanded the methods and discussion sections so as to make the results more broadly interpretable and applicable. (e.g. in the introduction: l. 98 on; in the Methods: l.180 on; in the Discussion: l. 465 on; l. 583 on)

Reviewer 2:

Main points	Response
1. Main focus of the paper (including abstract).	As indicated above, we agree that the emphasis on the Budyko theory is appealing. At the same time, our study aims to present an uncertainty analysis of the model and thus has a broader focus. This includes insights into the model empirical parameters and the input parameter uncertainties.

	We understand the reviewer's particular interest in the spatial vs. lumped application, but this is only one aspect of the uncertainties associated with the model, which are of interest to both hydrologists and model users. We note that the structure of the paper is consistent with current approaches to uncertainty assessment: it includes sensitivity analyses, model comparison (lumped vs. pixel-based), and comparison with observations and calibration (cf. Refsgaard 2007 for review of uncertainty analyses).
2.Description of the Budyko theory inconsistent with Eq.1	We agree with the reviewer that the general description of the Budyko theory should be consistent with the expression used in the manuscript. Change: we have revised the text of the description (Section 2.1)
3.Description of the w parameter.	We agree that the explicit relationship between w and vegetation is found in the work by Zhang et al. (2001). In the 2004 paper, the same authors derive the equation (Eq.1 in our paper) from an analytical approach, with a parameter w called "catchment parameter". This parameter is related to vegetation, but also captures local geology or topography. Change: we have clarified the nature of the empirical parameter w (l. 132) ("ω characterizes the partitioning of precipitation between evapotranspiration and runoff, and is a function of climate and physical factors.")
4.Model formulation for distributed predictions	We agree that the presentation of the outputs may have been confusing for the readers. The model uses spatially-explicit inputs, but results are tested on aggregated values, at the catchment scale. Change: We have clarified this point. While the proposed model is capable of providing spatially explicit output, this paper focuses on aggregated yields, consistent with available measurements.
5.Presentation of the distributed predictions in one or more of your study catchments	We thank the reviewer for this suggestion, which will help readers that are unfamiliar with InVEST to picture outputs given by the model. Change: We propose to add a map of the distributed water yields (Figure 4).
6.Test the validity of the spatial patterns without the benefit of data	Indeed, we are unable to test the pixel-based outputs from the model. We do, however, test the aggregated water yields against observed data from ten subcatchments. We think this is now clearer. Changes: See changes in Point 4 above. We renamed the section (section 3.4): "Testing the spatially-explicit outputs against observed data".

Reviewer 3:

Main points	Response
1.Recent studies testing InVEST	Unfortunately there are not many other studies that have tested the model in a systematic way. The two studies highlighted are the only ones to our knowledge that have gone further in the model testing than comparing a

	single model output (average annual water yield) to observed data, for a single point. Changes: We clarified the introduction (l. 95-96: "In particular, they assess the effects of climate variables uncertainty, but do not examine the ability of the model to represent land use change.")
2.Main focus in abstract/paper	We agree that the original structure of the paper was confusing. We have now clarified the structure to highlight the different uncertainty analyses that are performed, using standard terms in the field: "sensitivity analyses, model comparison, testing against observed data". This helps emphasizing the scope of the paper. Changes: In addition to the changes in the paper structure, we have also clarified the scope of the paper in the introduction and in the overview of the methods section (see Reviewer 1 – Comment 1)
3.Paragraph explaining the increased demand for spatially explicit ES tools	We agree that this point could be made clearer. In addition to citing recent work by Guswa et al. (2014), where the demand for ecosystem services tools is analyzed in more details, we provide examples of typical applications of these tools. Change: In the second paragraph of the introduction (l.51), we added: "Typical applications of the model include the development of land planning policies, such as the delineation of priority areas for conservation or for agricultural development."
4.Consistency in terminology	We agree that the lack of consistency in terms is confusing to readers and apologize for overlooking this point. Changes: We have revised the text with a consistent terminology. We now use exclusively "water yield", "groundwater withdrawal", "crop factor", "spatially-explicit", and "lumped model". We also use the term "subcatchment" only, except when referring to the Cape Fear basin.
5.Paper structure	We thank the reviewer for his concrete suggestions on the paper structure. We agree that the parallel structure for methods, results and discussion will help readers understand the key points of the analyses. In particular, we clarify that some analyses do not require observed data (i.e. sensitivity analyses, and comparison between spatially-explicit and lumped models) , whereas the last part of the analyses rely on observed data . Changes: We propose the following structure, which may entail minor text revisions to keep the logical flow: 1. Introdution 2. Spatially-explicit InVEST annual water yield model 3. Methods 1.1 Cape Fear study area 1.2 Sensitivity analyses - K and Z - Climate inputs

	1.3 Comparison of spatially-explicit and lumped models 1.4 Testing the spatially-explicit model against observed data 4. Results 4.1 Sensitivity analyses 4.2 Comparison of spatially-explicit and lumped models 4.3 Testing the spatially-explicit model against observed data 5. Discussion 4.4 Sensitivity analyses 4.5 Comparison of spatially-explicit and lumped models 4.6 Model performance in calibrated and uncalibrated 4.7 Practical implications 6. Conclusion For memory, below is the previous structure: 1 Introduction 2 Methods 2.1 InVEST annual water yield model 2.2 Cape Fear study area 2.3 Sensitivity to Z and Kc 2.4 Comparison of distributed and lumped application of the water-balance model 2.5 Performance of the InVEST model 3 Results 3.1 Sensitivity of Water Yield to climate, Z, and Kc 3.2 Comparison of spatially explicit and lumped models 3.3 Performance of the InVEST model 4 Discussion 4.1 Sensitivity to Z and Kc 4.2 Comparison of spatially explicit and lumped models 4.3 Model performance in gauged and ungauged basins 4.4 Practical implications 5 Conclusion
Minor comments	The typo and the reference to climate input uncertainty were rectified.

Reviewer 4:

Main points	**Response**
1.Absolute discharges	We agree that the sensitivity to precipitation could be better illustrated with a graphical form. We believe that Table 2 provides sufficient information about the water balance for each subcatchment, with the predicted and observed discharge values being reported. Change: We have added the baseline run values to Table 2, and we revised Figure 3 to include sensitivity to precipitation
2.Sensitivity to precipitation (for Cape Fear and more generally)	The sensitivity to precipitation error is a very important factor when assessing model performance, and this point motivated our work on precipitation error assessment in the manuscript. We agree that the discussion would benefit from the extrapolation of this idea to other catchments

	Change: [l.583] we have elaborated on the expected sensitivity to precipitation, based on the example of arid climates.
3.Value of omega for lumped models	We thank the reviewer for this suggestion, reporting the values of omega will improve the clarity of the discussion. Change: we have added these values in the Results (Section 4.2) and discuss them in Section 5.2
Minor comments	Minor comments have been addressed or discussed above, in other reviewers' response.

Uncertainty analysis of a spatially-explicit annual water-balance model: case study of the Cape Fear ~~catchment~~basin, ~~NC~~North Carolina.

P. Hamel[1]
A. J. Guswa[2]

[1]Natural Capital Project, Woods Institute for the Environment, Stanford University, Stanford, CA, USA

[2]Picker Engineering Program, Smith College, Northampton, MA, USA

Abstract

There is an increasing demand for assessment of water provisioning ecosystem services. While simple models with low data and expertise requirements are attractive, their use as decision-aid tools should be supported by uncertainty characterization. We assessed the performance of the InVEST annual water yield model, a popular tool for ecosystem service assessment based on the Budyko hydrological framework. Our study involved the comparison of ten subcatchments ranging in size and land use configuration, in the Cape Fear basin, North Carolina~~, ranging in size and land use configuration~~. We analyzed the model sensitivity to ~~the eco-hydrological parameters~~climate variables and input parameters, and the ~~effect of~~structural error associated with the ~~extrapolating~~ use of the Budyko framework, ~~a~~ a lumped (catchment-scale) model ~~lumped~~ theory, ~~to~~ in a ~~fully distributed~~spatially-explicit ~~model~~way. Comparison of ~~the~~ model predictions with observations and with ~~a~~ the lumped ~~water balance~~ model predictions confirmed that the InVEST model is able to represent differences in land uses, and therefore in the spatial distribution of water provisioning services. Our results ~~also~~ emphasize the effect of climate input errors, especially annual precipitation, and errors in the eco-hydrological parameter Z, which are both comparable to the model structure uncertainties. ~~In practice, o~~Our case study supports the use of the model for predicting land use change effect on water provisioning, although its use for identifying areas of high water yield will be influenced by precipitation errors. ~~While the results are inherently local, analysis of the model structure suggests that many insights from this study will hold globally.~~ While some results are context-specific, our study provides general insights and methods to ~~Further work toward characterization of uncertainties in such simple models will~~ help identify the regions and decision contexts where the model predictions may be used with confidence.

1 Introduction

The interactions between hydrology and land-use and land-management decisions have received increased attention in recent years. The International Association of Hydrological Sciences (IAHS) recently declared this decade Panta Rhei – everything flows – to focus on the changing dynamics of the water cycle in connection with changing human systems (Montanari et al., 2013). Socio-hydrology has recently been proposed as a “use-inspired” discipline to focus on understanding the human-modified water cycle (Sivapalan et al., 2014). The quantification of water services, or the value that humans derive from natural processes, is also increasingly seen a means of elucidating the interactions between people and water. Examples of this approach abound globally: through its Grain-to-Green program, China incentivizes land-owners to convert annual crops to perennial species or natural forests (Liu et al., 2008). In South America, there now exist dozens of Water Funds, which invest in upstream conservation measures to ensure the downstream provision of clean water (Martin-Ortega et al., 2013). In the United States, Federal investments in water resources projects now require an assessment of impacts to ecosystem services (Council on Environmental Quality, 2013).

To quantify the impact of land-use and land-management decisions on ecosystem services, a number of tools have been developed by researchers and practitioners (Bagstad et al., 2013). Typical applications of these tools include the development of spatial planning policies, such as the delineation of priority areas for conservation or for agricultural development (Guswa et al., 2014). ~~Typical~~ These applications often ~~of these tools~~ i) occur in data-scarce environments, ii) require spatially-explicit information, at the scale of individual land holdings and parcels, and iii) ~~focus on the estimation of~~integrate a range of ecosystem services rather than focus on the precise quantification of a particular service. Accordingly, state-of-the-art models requiring extensive data and expertise are generally not appropriate for such applications. Instead, models for ecosystem-service valuation often focus on ease of use, using globally available data, accepting ~~spatially~~ spatially-explicit input and producing ~~spatially~~ spatially-explicit output, and limiting the model structure to key biophysical processes involved in land-use change (Guswa et al., 2014).

The InVEST annual water yield model was developed in line with this philosophy (Tallis et al., 2013). It includes a biophysical component, computing the provision of freshwater, or water yield, by different parts of the landscape, and a valuation component, representing the benefits of water provisioning to people. The biophysical module, the focus of this paper, is based on the Budyko theory, which has a long history and continues to receive interest in the hydrological literature (Budyko, 1979; Zhou et al. 2012; Zhang et al. 2004; Zhang et al. 2001; Donohue et al. 2012; Xu et al. 2013; Wang and Tang, 2014). The InVEST model applies a one-parameter formulation of the theory (Zhang et al., 2004) in a ~~semi-distributed~~spatially-explicit way. This raises two issues. First, application of the model to ungauged basins or to future land-use scenarios requires a methodology for determining the value of the model parameter from known characteristics of the climate and basin, since it cannot be determined via calibration. Second, the Budyko approaches have been developed for long-term water balances at the catchment scale, rather than at the scale of individual land parcels, which is required for ecosystem-service decisions. ~~the application of the water balance at the scale of individual patches of land, rather than the catchment scale for which the Budyko theory was developed, is uncommon in the literature~~. The effect of this change in spatial scale is unclear, and calls for a rigorous analysis of the model and its uncertainties~~ and their impact on ecosystem services assessments~~.

Uncertainty analyses remain rare or incomplete in ecosystem service~~s~~ assessments, where the focus is on analyzing trade-offs and valuation of multiple services, often at the expense of characterizing uncertainty of individual modeling components. For example, in reviewing the literature using the InVEST annual water yield model, we found the following common limitations: absence of or inadequate comparison with observed data, calibration of the model without prior identification of sensitive parameters, and lack of validation of the predictive capabilities in the context of land-use and land-cover (LULC) change (Bai et al., 2012; Nelson et al., 2010; Su and Fu, 2013; Terrado et al., 2014). To varying degrees, these limitations jeopardize the production of credible assessments of ecosystem services.

Recent work paved the way for understanding the uncertainties in the InVEST model predictions. Sánchez-Canales et al. (2012) analyzed the sensitivity of the model in their case study of the Llobregat catchment, in Spain. They found that the model was sensitive to climate variables, but less so to the Z parameter (see model description). Similarly, Boithias et al. (2014) and Terrado et al. (2014) reflect on the sensitivity of the model to climate inputs, and calibrate the model based on the climate parameters and return flows. ~~However, t~~The~~ir~~ conclusions of these studies are often context-specific and lack a quantitative estimate of the model structural uncertainties. In particular, they assess the effect of climate variables uncertainty, but do not examine the ability of the model to represent land use change.

This paper aims to extend this work by characterizing the uncertainty in the InVEST annual water-yield model, and assess its utility to inform ecosystem-service decisions. As indicated above, the focus on water services implies a focus on decisions related to land-use and land-management, thus requiring spatially explicit descriptions of the landscape and associated hydrologic parameters (Guswa et al., 2014). Ecosystem-service decisions may be based on spatially aggregated output (e.g., which landscape scenario provides the greatest water yield at the base of the catchment), or may require spatially explicit output (e.g., which parcels in this catchment are of highest priority for conservation). While the proposed model is capable of providing output to inform the latter, this paper focuses on the former, since typical measurements of water yield (streamflow) are inherently aggregated.~~ applied to watersheds in the Cape Fear region of North Carolina.~~ Using a case study in the Cape Fear region of North Carolina, our~~This~~ study i) quantifies the effect of parameter uncertainty on model outputs through sensitivity analyses; ii) compares the distributed application of the water balance to the catchment-scale application; and iii) quantifies the accuracy of calibrated and uncalibrated versions of the model by comparing model predictions to observations. From a practical standpoint, this work helps InVEST model users identify modeling uncertainties and proposes simple and replicable methods that can be used to quantify the reliability of water-service decisions~~their effect on water services~~.

2 InVEST annual water yield model

2.1 Background theory

The Budyko curve is a unique empirical function that relates the ratio of ~~actual to potential~~ actual evapotranspiration to precipitation (averaged over a catchment and over many years) to the ratio of ~~precipitation to potential~~ potential evapotranspiration to precipitation (Budyko, 1961)~~(Budyko, 1979)~~. The function is bounded by two limits – an energy limit in which actual evapotranspiration is equal to potential, and a water limit for which actual evapotranspiration is equal to precipitation. Due to spatial and temporal variability in climate forcing, the asynchronicity of water supply (P) and demand (PET), the imperfect capacity of the root zone to buffer that asynchronicity, and lateral redistribution of water within the catchment, the Budyko curve lies below those two limits (Figure 1).

To describe the degree to which long-term catchment water-balances deviate from the parameter-less Budyko curve~~theoretical limits~~, a number of scholars have proposed one-parameter functions that are similar ~~can replicate the Budyko curve~~ (e.g., Fu, 1981; Choudhury, 1999; Zhang et al., 2004; Wang and Tang, 2014). The InVEST water yield model employs the formulation by Zhang et al. (2004), which incorporates a catchment parameter, ω:

$$\frac{AET}{P} = 1 + \frac{PET}{P} - \left[1 + \left(\frac{PET}{P}\right)^{\omega}\right]^{1/\omega} \tag{1}$$

AET is the actual evapotranspiration (mm), P is precipitation (mm), PET is the potential evapotranspiration (mm). ω affects the partitioning of precipitation between evapotranspiration and runoff, and is a function of climate and physical factors. Larger values of ω indicate those basins that are more "efficient" in converting precipitation to transpiration, e.g., those with precipitation synchronous with PET and/or with deeper root zones. Gentine et al. (2011) and Troch et al. (2013) have shown that the natural co-evolution of vegetation, climate, and topography may lead to basins for which the effects implicitly captured by ω counter-balance each other, offering an explanation for the observed convergence of data close to ~~along~~ the original Budyko curve. The intent of the InVEST model, however, is to predict the effects of human-induced changes, i.e., to examine catchments for which natural co-evolution is disrupted.

2.2 Spatially-explicit application to land-use change

Model overview

To represent parcel-level changes to the landscape, InVEST represents explicitly the spatial variability in precipitation and PET, soil depth, and vegetation. The model is GIS-based, using rasters of climate and soil properties as inputs (see Tallis et al. 2013 for full details).

For vegetated land uses, InVEST applies the Zhang formulation in a ~~spatially~~ spatially-explicit way at the pixel scale (10 m to 100 m on a side):

$$\frac{AET_i}{P_i} = 1 + \frac{PET_i}{P_i} - \left[1 + \left(\frac{PET_i}{P_i}\right)^{\omega_i}\right]^{1/\omega_i} \tag{2}$$

In contrast to Equation 1, P, PET, w, and AET are all functions of the local position, indicated by the subscript i.

The parameter ω is further deconstructed to separate the effects of soil depth, rainfall frequency, and other factors, following an approach proposed by Donohue et al. (2012):

$$\omega_i = Z\frac{AWC_i}{P_i} + 1.25 \tag{3}$$

where AWC_i is the plant-available water content (depth), and Z is an empirical parameter. The constant, 1.25, in Equation 2 reflects the minimum value of ω corresponding to bare soil, following Donohue et al. (2012). In this representation, differences in land-use and land-cover affect both PET, through changes to the crop ~~coefficient~~factor, K_c, and Z, through changes to the root depth and plant-available water content.

For open water, wetlands, and urban land-uses, InVEST computes AET_i directly as a user-defined proportion of PET_i, with classical approaches such as the FAO 56 guidelines (Allen et al., 1998) or local knowledge used to determine the appropriate proportion (Tallis et al. 2014). The simple representation of these LULCs, compared to the vegetated land uses modeled with Equations 2 and 3, reflects the focus of the model on vegetation-dominated landscapes.

Total evapotranspiration from a catchment is computed as the sum of AET_i attributed to each cell, and water yield is obtained by subtracting this value from the total precipitation.

Selection of the Z parameter

The empirical constant Z captures catchment-wide characteristics of climate seasonality, rainfall intensity, and topography that are not described by the plant-available water content (AWC) and annual precipitation P. Given the empirical nature of the model, the value of the Z parameter remains uncertain. In this work, we examine the three methods for the determination of Z that are proposed in the InVEST user's guide (Sharp et al., 2014). The first draws upon recent work that suggests that Z is positively correlated with the average annual number of rain events per year, N, and that Z may be approximated by N/5 (Donohue et al., 2012). This implies that Z captures rainfall patterns, distinguishing between catchments with similar annual precipitation but different intensity. The second method relies on globally available estimates of ω (e.g. Liang and Liu, 2014; Xu et al., 2013). Z is inferred from these published values of ω by inverting Equation 2 with values of AWC and P averaged over the catchment. In the third method, Z is determined via calibration to streamflow data (see Section 2.5).

3 Methods

The goal of the InVEST model is not to reproduce observations with a high degree of accuracy and precision, but to provide reliable information to inform decisions. Therefore, utility or acceptability of the model should be couched in terms of relative uncertainty. That is, the uncertainty associated with the model (due to its simple structure or challenge of parameter identification) should be on par with or less than the irreducible predictive uncertainty that arises due to uncertainty in the forcing variables – in this case, precipitation and potential evapotranspiration. ~~Errors in hydrologic model predictions can be separated into three sources: the structural error associated with model formulation and scale, error in parameter selection, and error in the model inputs.~~ To assess ~~these~~ the relative importance of the three sources of error (structural error, parameter selection, climate variables), we applied the InVEST annual model to ten subcatchments in the Cape Fear basin, NC. Their co-location implies a similarity in climate and seasonality and facilitates a focus on variations in land-use, size and topography (Hrachowitz et al., 2013). ~~The~~ In the following sections, we ~~provide the description of~~describe the ~~case study~~study area, the methods for the sensitivity analyses~~, the~~ and uncertainty assessment of input parameters and forcing variables~~of input data errors, and the evaluation of model performance~~, and our approach to assess the structural error of the model: comparison with observations, and with the (classical) lumped model predictions.~~.~~

3.1 Cape Fear study area

The Cape Fear catchment is a 23,600 km^2 area in North Carolina. Its major land uses are forest (40%), wetland (15%), grassland (14%), and agriculture (12%), mainly in the lower parts of the catchment and including intensive swine and poultry farms. Urban and agricultural development has generated significant groundwater extraction throughout the catchment.

The climate is humid subtropical, with a precipitation average of ~1200 mm over the 2002-2012 study period (Table A1 in Appendix). This period was used for the analyses based on the longest period available for climate data, observed streamflow, and matching LULC map. The available precipitation data comprise the PRISM dataset (Gilliland, 2003) and a network of eight rain gauges maintained by the USGS (USGS, 2014). For our analyses, we use the PRISM data and two additional rasters interpolated from the USGS point data (rain gauges) via spline and inverse-distance weighting (IDW). The three input rasters (hereafter referred to as PRISM, IDW, and Spline) were used separately to compute the average precipitation over each of the ten subcatchments and assess the error introduced by the input data selection. The variability in average annual precipitation among the PRISM, IDW, and spline rasters (averaging 1118 mm, 975 mm, and 966 mm, respectively, Table 1) represents the uncertainty that may arise when precipitation data are limited, a situation that is common in many places around the world (McGlynn et al., 2012).

Potential evapotranspiration is represented by reference evapotranspiration ET_0 times a crop factor K_c (Tallis et al., 2013). Reference evapotranspiration (ET_0) was obtained from three sources: FAO data, representing a long-term average from 1961 to 1990 (FAO, 2000), MODIS data (Mu et al., 2012), and interpolation (IDW) from a network of thirteen weather stations maintained by the Climate Office of North Carolina (NCSU, 2014). These three sources indicate average annual PET for the Cape Fear region to be 1240 mm (FAO), 1160 mm (MODIS), and 1310 mm (NCSU). These climate data indicate an aridity index (P/PET) of approximately 0.9 for the Cape Fear ~~watersheds~~catchment. A summary of InVEST inputs is given in appendix (Table A1 and A2).

Streamflow observations were obtained from the USGS monitoring network (USGS, 2014). A total of ten stations with a minimum of ten years of data were used for the analyses (Figure 2 and Table 2). Subcatchments draining to each of these points were delineated based on the 30 m DEM.

~~Groundw~~Water withdrawal data were obtained from governmental agencies (NC Department of Environment and Natural Resources, 2014). Due to the lack of spatially explicit information for water withdrawals (reported by county, which do not follow the subcatchment boundaries), and on the magnitude of return flow, we represented their effect as homogeneous over the entire catchment. We think this decision has a limited effect on model testing since the value of water withdrawals is small compared to water yields (see Results). In addition, we explicitly accounted for this uncertainty by examining the effect of a 50% error on the water withdrawal – a magnitude consistent with the variance among the county withdrawals. The average withdrawal rate, 39 mm/year, was subtracted from the predicted water yields for comparison with observations.

3.2 Sensitivity analyses

Sensitivity to Z and K_c

Step one in our assessment of the InVEST model was a local sensitivity analysis of water yield to the Z parameter and the crop ~~coefficient~~factor, K_c, for forest – the dominant LU class. The sensitivity of the model to Z can also be interpreted as the sensitivity to AWC, when the raster values are varied homogeneously over the catchment, since these parameters play a similar role in the model structure (Equation 3).

As noted above, large uncertainties surround the selection of the Z parameter (Sharp et al., 2014). For what we term the "baseline" case, we set Z equal to one-fifth the number of rain days per year (Z = N/5). Based on historic precipitation data (SERCC, 2014), the average number of rain days per year is approximately 110, giving a value of Z of 22. We used this value as a baseline for all subcatchments, and allowed the parameter to vary between 1 and 30 for the sensitivity analyses. This range was estimated from Equation 3 used with extreme values of P and AWC found in our catchments, and extreme values of ω (2.1 and 3.75) found in the study by Zhang et al. (2004).

Forest was the dominant LULC in all basins, with its cover ranging from 43 to 72% of subcatchments. We therefore decided to use the crop factor K_c-forest for the sensitivity analyses, and a baseline value of 1 for K_c-forest was obtained from the FAO 56 guidelines (Allen et al., 1998). Uncertainties on this value are large since it remains difficult to provide accurate estimates of the actual evapotranspiration of forest (McMahon et al., 2013). We set the upper bound to 1.1, because values greater than this are unlikely (McMahon et al., 2013), and set the lower bound to 0.7.

For the two parameters, we performed sensitivity analyses with the ranges defined above. The results are presented as a change in predicted water yield compared to the baseline run, thus assessing absolute sensitivity. Precipitation and reference evapotranspiration used for these runs were from the PRISM (1118 mm) and the FAO (1240 mm) datasets, respectively (see Section 2.5 and Discussion for insights into the error introduced by climate data).

Sensitivity to climate inputs

To provide context for the uncertainty in the predictions of water yield from the InVEST model, we compared the prediction error to the uncertainty in water yield that arises from uncertainty in climate (i.e., variability in the rasters of P and ET_0). Uncertainties in climatic data and their impact on rainfall-runoff models are commonly cited in the literature (McGlynn et al., 2012; McMahon et al., 2013). To be an effective decision-support tool, errors attributed to model structure and parameter selection should be on par with or less than the irreducible error associated with uncertainty in the ~~inputs~~climate.

As illustrated in Table 1, the ~~mean~~ average precipitation differed significantly across subcatchments depending on the data source: the mean differences between the PRISM and USGS datasets, with the spline or IDW interpolation methods, respectively, were -14% and -13%. Catchment-by-catchment ~~The~~ differences ~~was~~ were more spatially heterogeneous with the spline method, with some subcatchments receiving less precipitation relative to the baseline (PRISM dataset) and others receiving more. The reference evapotranspiration data also showed significant differences across sources, although the FAO and Climate Office sources showed good agreement. The MODIS values were on average 22% higher ~~on average~~ than those from the other two sources (Table 1). Differences between the Climate Office and FAO data were also spatially variable, ~~being positive for some subcatchments and negative for others~~ranging from -8% to 5% across catchments.

To assess the uncertainty in water yield due to variability in climate inputs (precipitation and reference evapotranspiration), we examined the sensitivity of the baseline model results to spatially homogeneous increases and decreases in climate forcing. We considered climate inputs that are 10% greater and 10% less than the baseline, applied uniformly across the landscape.

3.3 Comparison of spatially-explicit and lumped models

Although the InVEST annual water yield model is based on the well-studied Budyko framework, it departs from its classical application by applying the partitioning model at the pixel scale. To our knowledge, the effect of the pixel-by-pixel calculation performed by InVEST has not been previously studied. In such an application, three issues arise related to lateral flows of water, the spatial variability in climate variables, and the co-variance of climate and soil in the prediction of the parameter omega.

In the catchment-scale application of Budyko-type models, lateral inflows and outflows across the catchment boundary are presumed negligible, resulting in a simple water budget based on catchment precipitation, evapotranspiration, and water yield. This assumption will not hold for a parcel-based application of equation 2. Thus, error in the catchment-scale water balance will arise by ignoring the excess water generated at one spot that is later evaporated at a downgradient location. Such explicit routing is not included in the InVEST model.

Additionally, even if lateral flows are negligible, applying the non-linear Budyko curve locally and aggregating the yield will lead to different results than applying equation 2 to average values of P and PET. The concave nature of function indicates that application over a range of climates will produce an average water yield that is higher than what would be predicted if applied at the catchment scale (Figure 1).

Finally, since local values of both available water content and precipitation combine to affect the local values of omega (equation 3), average values of omega from the spatially explicit model will be different from what one would obtain if average values of AWC and P were used to compute an average value of omega.

~~Therefore~~To investigate these effects, we compared the model predictions to those obtained by applying the ~~Zhang~~ lumped model (Zhang et al., 2004) at the catchment scale~~, therefore applying the Budyko framework in a more classical way~~. Application of ~~such a~~the lumped model requires a value of ω, which we derived from Equation 3 with average values of P~~, PET,~~ and AWC, and with Z set to the baseline value of 22, as would be done in a typical ungauged application. We thus obtained, for each subcatchment, an estimate of areal AET and water yields for the vegetated areas. AET for urban areas and wetlands was calculated separately, following the same method as InVEST, and total water yield was calculated as the area-weighted average of water yield from the vegetated and urban areas.

3.4 Testing the spatially-explicit model with observed data

To quantify the accuracy and precision associated with the InVEST water ~~-~~yield model, we assessed model performance by comparison with observed data for each of the ten subcatchments in the Cape Fear area. Our method aims to measure the aggregated value of water yields at the subcatchment scale, not to test whether the water yield predicted by each pixel is accurate. We measured performance with the model bias, i.e. the relative difference between predicted and observed water yields, and also with the subcatchment ranking by water yields. The ability of the model to predict ranking is important for applications where prioritization of areas of low and high water yields is needed (Guswa et al., 2014).

Uncalibrated model

We first examined the performance of the model when Z was determined without calibration. We ~~considered calculating~~calculated Z both from the number of rain days and from a global value of ω, to evaluate the appropriateness of these recommended methods. In addition to assessing overall model performance, we also assessed the correlation between model performance and the proportion of forest in the catchment. These analyses aimed to identify a potential bias that may be corrected by modifying the LULC-specific crop factor K_c.

Calibrated model

To separate the effects of error associated with model structure from error attributed to parameter estimation, we also determined the value of Z via calibration. We calibrated to individual subcatchments, identifying for each the Z value that resulted in a zero error in the water yield. We examined the similarity of Z values across the ten basins, allowing us to assess the robustness of the model structure since we expect Z to depend on larger-scale climate and geology and not on local-scale land-use. We also considered the performance of the model with a single value of Z applied to all subcatchments, determined by minimizing the average bias for all basins. This allowed us to assess the uncertainty in prediction of water yield due to model structure, i.e., the inherent uncertainty to applying equations 2 and 3 to different basins ~~even~~ when the parameter, Z, is chosen by best fit for the entire region.

4 Results

In the baseline case, we applied equation 2 and 3 in a spatially explicit way with a precipitation field from the PRISM data and potential evapotranspiration data from the FAO. The value of Z in equation 3 was set to 22, as mentioned above. In this baseline case, predicted water yields ranged from 163 mm to 322 mm across the ten subcatchments. Results are presented in Table 2.

4.1 Sensitivity analyses

Water yield predictions are very sensitive to climate inputs. The sensitivity is higher for precipitation than ET_0. Relative to the baseline case, ~~A~~a 10% increase in precipitation resulted in a 30% increase in water yield (Figure 3), while the same increase in ET_0 resulted in a 15% decrease in water yield.

In contrast to the climate variables, water yield is less sensitive to values of Z: for example, a change in Z from the baseline value of 22 to a value of 10 results in an increase in water yield of approximately 27% (Figure 3). However, given the large uncertainties in the Z parameter, potential errors in water yield can be large: for example, the water yield is 155% higher when Z is ~~at its minimum value~~set to 1, relative to the baseline case with Z=22. The sensitivity to Z is catchment-specific, as expected, since its effect on yield is modulated by AWC and P, both of which are spatially variable. In addition, the relative sensitivity of water yield to Z decreased with increasing values of Z and increased with increasing values of the aridity index (PET/P, results not shown).

The model was found to be more sensitive to K_c (Figure 3) with a 30% change in K_c resulting in a 41% change in the water yield. However, given the ~~expected~~ small range of K_c values, the effect of parameter uncertainty on the water yield prediction is lower than for Z.

4.2 Comparison of spatially-explicit and lumped models

Across the ten subcatchments, the water yields predicted by the spatially-explicit InVEST model were on average 10% lower than the outputs from the lumped ~~Zhang~~ model ~~(Table 2)~~. For eight of the ten catchments, the spatially explicit model predicted lower water yields than the lumped model, and differences ranged from from -24% to 14%. The two catchments for which the lumped model predicted lower water yield than the InVEST model were the Morgan Creek and Cane Creek catchments, which have the highest proportions of forest and the lowest proportions of urbanized area across the ten catchments (Table 2).

The ω values computed for the lumped model ranged from 4.29 to 6.25 across the ten catchments. These values are in the higher range of the values obtained by Zhang et al. (2004), as discussed in section 5.2.

4.3 Testing the spatially-explicit model with observed data

Uncalibrated model

Figure 4 shows the spatially-explicit output from the InVEST model. That figure is for illustrative purposes only; as indicated above, we aggregate the pixel values of water yield to the subcatchment scale to compare with observations. ~~Figure 4a presents~~Such comparison is presented in Figure 5a, where the Z-parameter for the InVEST model ~~predictions of water yield from the invest model when the Z-parameter~~ is determined from the number of rain days (Z = 22). Open triangles represent results from the InVEST model. To contextualize the error, gray bars represent the uncertainty in predicted water yield due to a 10% uncertainty in precipitation and black bars represent the uncertainty in water yield due to a 50% uncertainty in water withdrawals.

The performance of the model for ~~the~~ this baseline run ~~was~~ is fair. Across all basins, predicted water yields range from 163-322 mm/yr versus an observed range of 177-368 mm/yr. ~~, with the~~ The bias between predicted and observed values averag~~ing~~es -16% ~~for all~~across the ten subcatchments, ~~. This bias~~ rang~~e~~ing~~d~~ from -53% to -1%~~,~~. This indicates that the model structure combined with ~~implying t~~that this choice of Z leads to a systematic underestimation of water yield. With the exception of one catchment, the biases ranged from -25% to -1%. The outlier with an error of -53%, Rockfish catchment, is relatively small (237 km^2), and the observed water yield is also an outlier, being the highest in the dataset (367 mm). This area is also characterized by sandy soils; the plant available water content averages 0.11, compared to values between 0.17 and 0.20 for the other subcatchments. This suggests that the catchment may exhibit a unique behavior, which we will highlight in the following analyses. ~~Across all basins, predicted yields range from 163-322 mm/yr versus an observed range of 177-368 mm/yr.~~

Figure 5~~4~~b presents the ranking of catchments in terms of their observed and predicted water yields. Discarding the outlier catchment, the figure indicates that the model accurately predicts the high and low ranking catchments, while there is some dispersion in ranks for the five mid-range water yields, which vary from 236 mm/yr to 289 mm/yr.

~~For the second case, w~~When Z ~~is~~ iwas determined from published values of ω, the average value across the ten catchments was 6 (compared to 22 for the baseline case). ~~the m~~Model performance was not satisfying for this case, and model bias averaged 68%. ~~The Z value found for all subcatchments averaged 6, which results in a large model bias (averaging 68%).~~

Calibrated model

In the first approach to the calibration of Z, we determined the value for which the predicted water yield exactly matched the observations. In this case, values of Z range from 6 to 20 across the ten catchments. Not including the Rockfish catchment, the range is narrower (10-20) and the average across the nine remaining catchments is 14.5. ~~When Z is determined through calibration for each subcatchment, values of the parameter range from 6 to 20. The calibrated value of 6 was obtained for the Rockfish catchment; discarding that outlier catchment, values range from 10 to 20, averaging 14.5.~~ The narrow range of~~is~~ variability translates into relatively small changes in water yield – the average difference among the basins is 27%.

In a second approach, we determined a singe value of Z for all ten catchments ~~The single Z-value obtained~~ by minimizing the average ~~subcatchment~~ bias. This gives a value of ~~(~~Z=14, and ~~)~~ ~~is similar to these individual Z values. With this calibrated value,~~ the error in yield for all subcatchments ranges from -38% to 14% with a median value of -3%. Predicted water yields range from 183 mm/yr to 336 mm/yr versus an observed range from 177 mm/yr to 368 mm/yr. The open circles in Figure ~~4a~~ 5a presents ~~model~~ predictions from the calibrated model of water yield versus the observed values~~across the ten catchments~~. ~~Open circles represent results from the calibrated InVEST model, while black bars represent the uncertainty in yield due to a 50% uncertainty in water withdrawals. Gray bars represent the uncertainty in predicted yield due to a 10% uncertainty in precipitation.~~

Model bias is not correlated with forest cover (R^2=0.01), nor with any other LULC (Table 1). The absence of systematic bias suggests that K_c values are in a realistic range, with no significant error due to LULC parameter selection. No significant bias was detected with regard to catchment size, suggesting that this characteristic did not systematically influence the model predictions either.

5 Discussion

5.1 Sensitivity ~~to Z and Kc~~analyses

Variability in the Z parameter, which is linearly related to ω, results in a shift of the ~~Zhang~~ Budyko curve, which affects water yield predictions (Figure 1). Our results in Cape Fear suggest that the sensitivity of water yield to Z is low compared to the climate inputs, and decreases for larger values of Z (Figure 3). This is consistent with the ~~Zhang~~ lumped model for which the sensitivity to ω~~,~~ decreases with increasing values of ω (Figure 1). Due to this low sensitivity, small errors in estimating Z are likely to have limited impact on the reliability of water yield predictions. In particular, we note that the range investigated in the study (from 1 to 30) is greater than the typical uncertainty associated with Z: irrespective of the selection method, values less than 5 are unlikely.

The sensitivity to Z also provides a sense of the sensitivity to AWC, which is a function of the local ecohydrological properties: plant available water content, root depth and soil depth (cf. Sharp et al., 2014 for details). Examination of Equation 3 suggests that a relative change in Z has the same effect as a relative change in these ecohydrological parameters~~: a 50% error in these parameters, if assumed homogeneous over the catchment, will have the same response as a 50% error in Z. Given t~~. The ~~typical~~ confidence interval for these ~~measurable~~ physical parameters may be large but is reducible by measurements~~, the uncertainty on these parameters will have a smaller effect on model outputs than the uncertainty in Z~~.

When analyzing the model sensitivity to K_c, two things are to be considered. First, the K_c value affects only the portion of the landscape covered with forest, and this reduces its effect. Because total water yield is the sum of the yields from the different parts of the landscape, parameters affecting only a portion of the landscape will have a smaller effect. Second, it is worth noting that the K_c coefficient directly affects PET for a given LULC, since the latter is the product of K_c by ET_0. Examining the sensitivity of the model to K_c is therefore equivalent to a displacement along the ~~Zhang~~ Budyko curve, rather than a shift of this curve (Figure 1).

~~The results of t~~In summary, the sensitivity analyses showed that, for expected and reasonable ranges of parameter variability, precipitation and potential evapotranspiration have the greatest influence on water yield. These are followed by the parameter, Z, and then the crop coefficient, Kc. ~~indicate that embedded in the Zhang model is the concept that the dominant effects of land-use and land-cover change on water yield will be via the effects on K_c and PET rather than through changes to root depth and plant-available water content.~~

5.2 Comparison of spatially-explicit and lumped models

Comparison of the model predictions with the classical lumped model application suggests three insights. First, it provides a sense of the effect of the pixel-by-pixel application of the Budyko theory~~, which has not received much attention in the literature~~. Because of its non-linear nature, the average response of Equation 2 applied across the landscape in a spatially explicit way is not equivalent to the response of the function applied to the entire catchment, characterized by average parameters. Our results suggest that this discretization effect is not large for the Cape Fear subcatchments, with the difference between the lumped and explicit models ranging from -24% to +14%. This range is consistent with the typical errors expected from the application of simple empirical models. This point can be illustrated by the performance of the lumped model when compared with the observations: bias ranges from -36% to 29%. It is worth noting that the larger, positive biases (>22%), i.e. when the lumped model largely overestimated observed water yields, were obtained for the two subcatchments that had >25% urban cover, and the three basins with the least urban cover (Cane Creek, Rockfish, and Morgan Creek) had the largest underestimates of water yield. These results suggest that the contribution from urban areas was overestimated by the simple model.

The second point is related to the first one, focusing on the observation that water yields predicted by the spatially explicit model were consistently less than those predicted by the lumped model. As stated in the methods (Section 3.3), this difference can be expected from the differences in average climate values or average ω values, due to the non-linearity in Equation 2. In our case, the average ω values were high for the lumped model (ranging from 4.29 to 6.25). This indicates that the empirical expression for Z, developed for a lumped application (e.g., Donohue et al., 2012), may give values of Z (and, therefore, ω) that are too large for our case study, and this effect is emphasized when used in a spatially explicit model. Calibration of the model based on Z allows for correcting this error in the empirical expression, although further studies would be necessary to gain insights into the extrapolation of the Z parameter to spatially explicit models like InVEST.

~~Second~~Finally, the good agreement between the InVEST model and the lumped model allows to draw on the large body of work investigating the performance of the latter model. For example, Zhou et al. (2012) report a bias of less than 20% in a long-term study of 150 large basins worldwide; similarly, Zhang et al. (2004) report a mean absolute error of <60mm in their study of over 470 catchments worldwide, corresponding to a bias <10% for the majority of the catchments. Other local examples may be drawn by users to understand how the Budyko theory may apply locally (e.g. Yang et al., 2007 in China). Overall, there is a large ongoing effort to improve the parameterization and predictive use of the Budyko framework (Donohue et al., 2012; Liang and Liu, 2014; Xu et al., 2013). Future progress may therefore be used to refine the InVEST model interpretation in different geographic contexts. We note, however, that the agreement between the lumped model and the catchment model is context specific. As illustrated in Table 2, the differences between the lumped model and the InVEST model ~~will~~ vary ~~between~~ among catchments, such that extrapolation of the results from such studies will need to be done cautiously.

~~The final point is based on the observation that yields predicted by the spatially explicit model were consistently less than those predicted by the lumped model. This difference could be due to differences in mean parameter values or due to the non-linearity in Equation 2. Looking at Figure 1, the concave nature of the Zhang curve indicates that the average response over a range of climates will lead to lower evapotranspiration and higher yields than if the equation were applied to the mean climate. Similarly, application over a range of values of ω would also lead to higher yield than what is predicted using the mean yield (Figure 1). In this case, the lower yields predicted by the explicit model are due to differences in the mean values of ω between the lumped and explicit models. This indicates that the empirical expression for Z, developed for a lumped application (e.g., Donohue et al., 2012), may give values of Z (and, therefore, ω) that are too large when used in a spatially explicit model. Use of a smaller value of Z in the spatially explicit model would increase yield, although further studies would be necessary to gain insights into the extrapolation of the Z parameter to spatially explicit models like InVEST.~~

5.3 Model performance with and without calibration

Calibrated model

Our results indicate a fair performance of the calibrated model for multiple catchments ranging in size and LULC. The bias ranged from -38% to 14% for all subcatchments, and from -14% to 14% when discarding the Rockfish catchment. This narrow range suggests that the calibrated model was able to explain the variability in observed water yields. While it is possible that such variability is explained by climate more than LULC, this is not the case~~hypothesis is unlikely~~ in Cape Fear- since the average values of P and PET ~~on average they~~ varied by less than 3% between subcatchments (~~raster average for both P and ET_0,~~ Table 2).

Further consideration of the Z values obtained by calibrating it for each subcatchment ~~individual calibrations~~ provides insights into the interpretation of this parameter. With the exception of the Rockfish catchment, a ~~single~~ value between 10 and 20 ~~was~~ is able characterize the nine other subcatchments. This suggests that the parameter captures the topography and climate of the area, as intended by the model. The calibrated value of Z for the Rockfish catchment was much lower (Z=6), producing a higher water yield. This difference could be due to the inadequacy of Equation 3 to relate ω to soil characteristics (since the soils in the Rockfish catchment are particularly sandy). It could also be attributed to errors in the treatment of water withdrawals and return flows, especially since the entire subcatchment lies within Hoke County, which has minimal water withdrawals.

Despite the uncertainties around the outlier, the multi-catchment analyses allow~~e~~sd us to assess the model performance in representing LULC change. Use of the model for evaluation of LULC change is crucial in ecosystem service assessments, where scenarios analyses of LULC development are common (Guswa et al., 2014). Validating the use of models in such contexts is extremely challenging since it is rare for modelers to have sufficient pre- and post-LULC change data (Hrachowitz et al., 2013). In our study, the length of the precipitation and streamflow data did not allow conducting such temporal analyses. Regional analyses where space is substituted for time thus represent a powerful way to assess the ability of the model to capture differences in LULC configuration.

Uncalibrated model

Another important lesson from the analyses is that the calibrated Z value is relatively close to the baseline value, which was derived independently from the average annual number of rain events. Based on Figure 3, using one value or the other would result in a difference in water yield of approximately 10%. This error is small compared to other model uncertainties, suggesting that this method for determining Z is robust. The underprediction of water yield for ungauged catchments could be explained by errors in the precipitation raster, the Z parameter, and the treatment of water withdrawals. Based on Equation 2, the negative bias implies the underestimation of the precipitation data or overestimation of the Z coefficient. As already noted, errors in precipitation data are difficult to characterize. However, precipitation was more likely underestimated in this study since it did not include snowfall.

Conversely, the method relying on a constant ω value was not found satisfying for this case study, since it resulted in large overestimation of the water yields. Using ω=4, the Z value found for individual subcatchments ranged from 4 to 8, averaging 6, a value that results in a large model bias (averaging 68%).

With regard to relative water yield values, the model was able to predict subcatchment ranks fairly accurately (Figure 4b), which means that priority areas would be correctly identified. The uncertainties in ranking for medium water yield catchments (ranking from 3 to 6) could be partly explained by their similarity (observed water yields range from 236 mm to 278 mm) and the uncertainty in the water abstraction, as suggested by the overlapping error bars in Figure 4a. Interestingly, although these results were obtained with the calibrated value of Z, they are only slightly sensitive to the value of Z, since the ranking of subcatchments is largely maintained when the value of Z changes. The ranking of subcatchments based on the baseline run, for example, was identical to the one with Z=14.

5.4 Practical implications

In this final section, we discuss the results with a focus on practical implications for model users.

Our analyses suggest that the uncertainty introduced by *variability in the precipitation inputs* is high, comparable or higher than the uncertainty introduced by the parameter Z and the use of the lumped model theory on a pixel-by-pixel basis. Importantly, the sensitivity observed in Cape Fear (e.g. that a 10% change in precipitation may result in a 30% change in water yield) is specific to the climate: for example, in arid climates where evapotranspiration is water limited, an error in precipitation may have a lower effect on water yield since the precipitation surplus or deficit will be mostly converted to evapotranspiration. ~~This suggests that confidence intervals for climate data deserve particular attention (especially if interpolating local data from weather stations). The~~ In Cape Fear, comparison of three climate input data sources suggested that large errors may occur when using point data or datasets obtained with different modeling assumptions. These results confirm a wide body of research that highlight the importance of precipitation inputs for rainfall runoff models (McGlynn et al., 2012; Zhou et al., 2012) and in particular for the InVEST model (Boithias et al., 2014; Sánchez-Canales et al., 2012). Although it was outside the scope of this study to investigate which climate datasets are less prone to errors, our results also draw attention to spatially heterogeneous errors. If model users are interested in the relative ranking of subcatchments, the spatial distribution of errors should be specifically investigated (e.g. probability of a systematic bias in mountainous areas).

The model is not very sensitive to *uncertainty in Z* over a modest range (e.g., 14-22). This is consistent with the conclusions from Sánchez-Canales et al. (2012), who report a low sensitivity to Z in a Mediterranean catchment, for which Z varied between 7 and 9. Since the viable range of Z is quite wide, however, it is possible that large uncertainties in that parameter will translate to significant uncertainty in water yield (Figure 3). Our analyses provided further insights into the methods for Z selection and highlighted that the sensitivity of the model to Z decreased with increasing values of Z. Based on the examination of Equation 2, this property will apply generally. Therefore, in temperate climates where values of Z are high (based on the interpretation of Z as the number of annual rain events), the model outputs are likely to be less sensitive to this parameter.

Our study also presented a method to detect a *bias related to the LULC parameters*, when multiple observations are available in a catchment. Because K_c values are LULC-specific, the correlation between model performance and K_c values can be used to identify a possible error in the parameter and rectify the values accordingly. No bias was found in this study, bringing confidence in the ability of the model to capture the differences in LULC. We note that these correlation analyses rely on nested subcatchments that are not independent from each other, which decreases the significance of the relationship: five subcatchments are independent, while the other five partially overlap. However, proportions of forest cover varied widely between all subcatchments (from 43 to 72%), which justifies our interpretation of the analyses.

We conclude this section with a perspective on the model performance assessment, highlighting key *limitations in the calibration*/testing exercise. First, we note that some water transfers are missing in the model, including irrigation and water abstraction. The model represents agriculture in the same way that it does natural vegetation, and irrigation is not included explicitly. Second, in the Cape Fear catchment, the magnitudes of the water withdrawals are small but this aspect of the modeling may be improved in future applications. In particular, distinction between uses of groundwater (crop irrigation or drinking water) are necessary to account for the fate of water extraction: evapotranspiration in the case of irrigation water, or return flow to the river in the case of drinking water (e.g. Terrado et al., 2014). Additionally, performance was evaluated at the catchment scale. A potential benefit of a spatially explicit model, however, is the ability to predict patterns of water yield within a basin. To properly evaluate that capability, further work should focus on comparing the InVEST model to more sophisticated fully-~~-~~distributed models.

6 Conclusion

Our study aimed to assess the performance of the InVEST annual water yield, a tool that is gaining interest in the ecosystem services community. While such simple models with low requirements for data and level of expertise are needed for practical applications, greater attention should be paid to characterizing the modeling uncertainties. Our assessment of the potential input errors, sensitivity analyses and comparison with observations in the Cape Fear catchment add to this body of work. Key results of the analyses are as follow:

- In the Cape Fear catchment, the InVEST model was most sensitive to uncertainty in the precipitation forcing;
- Errors in climate input data may be significant and non-spatially homogeneous, resulting in uncertainties in the assessment of zones of high and low water yields;
- The study supports the recommendations for setting the Z parameter based on the number of rain events, or via calibration with available observed data;
- Based on the average bias and the explained variance in water yield among the subcatchments, the model performance was fair to high, suggesting that the effects of land-use and land-cover are adequately captured by the model;
- The errors potentially introduced by a pixel-level application of the Budyko theory will depend on catchment configuration; in Cape Fear, they remained small, comparable to the climate and ~~structural~~ parameter errors of the empirical model;
- Water abstractions and irrigation processes that are not represented in simple models may confuse the calibration exercise, especially in data scarce environments where the ecosystem services approach is gaining momentum.

~~While the sensitivity analyses results are inherently local, the methods outlined in this study provide a template that can be used in most InVEST model applications. The analyses do not~~

~~require hydrologic expertise and are facilitated by the model batch-processing capabilities.~~ ~~Since r~~Rigorous uncertainty analyses ~~are~~ have ~~currently~~ not been the norm in the ecosystem service~~s~~ community, but such ~~simple guidance~~work is essential to help users interpret models correctly to inform land-management decisions appropriately~~and conduct more robust assessment of the water-related ecosystem services~~.

Acknowledgements

We thank Rebecca Benner and Margaret Fields, from the Nature Conservancy, and Chris Cook for their support and help with data pre-processing.

References

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration - Guidelines for computing crop water requirements, FAO Irrigation and drainage paper 56. Rome, Italy.

Bagstad, K.J., Semmens, D.J., Waage, S., Winthrop, R., 2013. A comparative assessment of decision-support tools for ecosystem services quantification and valuation. Ecosyst. Serv. 5, 27–39.

Bai, Y., Zheng, H., Ouyang, Z., Zhuang, C., Jiang, B., 2012. Modeling hydrological ecosystem services and tradeoffs: a case study in Baiyangdian watershed, China. Environ. Earth Sci. 70, 709–718.

Boithias, L., Acuña, V., Vergoñós, L., Ziv, G., Marcé, R., Sabater, S., 2014. Assessment of the water supply:demand ratios in a Mediterranean basin under different global change scenarios and mitigation alternatives. Sci. Total Environ. 470-471, 567–77.

Budyko, M., 1961. The Heat Balance of the Earth's Surface. Washington D.C.

Donohue, R.J., Roderick, M.L., McVicar, T.R., 2012. Roots , storms and soil pores : Incorporating key ecohydrological processes into Budyko's hydrological model. J. Hydrol. 436-437, 35–50.

FAO, 2000. Global agro-ecological zones. FAO Land and Water Digital Media Series No. 11. Rome, Italy.

Gilliland, B.K., 2003. PRISM precipitation data, Technical Report 04-013. Idaho Water Resources Research Institute, University of Idaho.

Guswa, A., Brauman, K.A., Brown, C., Hamel, P., Keeler, B.L., Sayre, S.S., 2014. Ecosystem Services: Challenges and Opportunities for Hydrologic Modeling to Support Decision Making. Water Resour. Res. 50, 4535–4544.

Hrachowitz, M., Savenije, H.H.G., Blöschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W., Arheimer, B., Blume, T., Clark, M.P., Ehret, U., Fenicia, F., Freer, J.E., Gelfan, A., Gupta, H.V., Hughes, D. a., Hut, R.W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. a., Uhlenbrook, S., Wagener, T., Winsemius, H.C., Woods, R. a., Zehe, E., Cudennec, C., 2013. A decade of Predictions in Ungauged Basins (PUB)—a review. Hydrol. Sci. J. 58, 1198–1255.

Liang, L., Liu, Q., 2014. Streamflow sensitivity analysis to climate change for a large water-limited basin. Hydrol. Process. 28, 1767–1774.

Martin-Ortega, J., Ojea, E., Roux, C., 2013. Payments for Water Ecosystem Services in Latin America: A literature review and conceptual model. Ecosyst. Serv. 6, 122–132.

McGlynn, B.L., Blöschl, G., Borga, M., Bormann, H., Hurkmans, R., Komma, J., Nandagiri, L., Uijlenhoet, R., Wagener, T., 2012. A data acquisition framework for prediction of runoff in ungauged basins, in: Predictions in Ungauged Basins. IAHS.

McMahon, T.A., Peel, M.C., Lowe, L., Srikanthan, R., McVicar, T.R., 2013. Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: a pragmatic synthesis. Hydrol. Earth Syst. Sci. 17, 1331–1363.

Mu, Q., Zhao, M., Running, S.W., 2012. Brief Introduction to MODIS Evapotranspiration Data Set (MOD16).

NASS, 2013. National Agricultural Statistics Service - Cropland data layers 2012 [WWW Document]. URL http://www.nass.usda.gov/research/Cropland/SARS1a.htm

NC Department of Environment and Natural Resources, 2014. Reported daily water withdrawals [WWW Document]. URL http://www.ncwater.org/Water_Withdrawals/ResultsTabJS.php?wsrc=gw&fip=1&tab=data&byRegionType=basin&basinid=2

NCSU, 2014. NC Climate Office [WWW Document]. URL www.nc-climate.ncsu.edu/

Nelson, E., Sander, H., Hawthorne, P., Conte, M., Ennaanay, D., Wolny, S., Manson, S., Polasky, S., 2010. Projecting Global Land-Use Change and Its Effect on Ecosystem Service Provision and Biodiversity with Simple Models. PLoS One 5, e14327.

Sánchez-Canales, M., López Benito, A., Passuello, A., Terrado, M., Ziv, G., Acuña, V., Schuhmacher, M., Elorza, F.J., 2012. Sensitivity analysis of ecosystem service valuation in a Mediterranean watershed. Sci. Total Environ. 140–153.

SERCC, 2014. Number of days of precipitation Equal to or above 0.01 inches in selected cities in the Southeast, [WWW Document]. Southeast Reg. Clim. Cent. URL https://www.sercc.com/climateinfo/historical/meanprecip.html

Sharp, R., Tallis, H.T., Ricketts, T., Guerry, A.D., Wood, S.A., Nelson, E., Ennaanay, D., Wolny, S., Olwero, N., Vigerstol, K., Pennington, D., Mendoza, G., Aukema, J., Foster, J., Forrest, J., Cameron, D., Arkema, K., Lonsdorf, E., Kennedy, C., Verutes, G., Kim, C.K., Guannel, G., Papenfus, M., Toft, J., Marsik, M., Bernhardt, J., Griffin, R., Glowinski, K., Chaumont, N., Perelman, A., Lacayo, M., Mandle, L., Hamel, P., Chaplin-Kramer, R., 2014. InVEST 3.0 User’s Guide. Available at: http://ncp-dev.stanford.edu/~dataportal/invest-releases/documentation/current_release/ [accessed June 2014].

Su, C., Fu, B., 2013. Evolution of ecosystem services in the Chinese Loess Plateau under climatic and land use changes. Glob. Planet. Change 101, 119–128.

Terrado, M., Acuña, V., Ennaanay, D., Tallis, H., Sabater, S., 2014. Impact of climate extremes on hydrological ecosystem services in a heavily humanized Mediterranean basin. Ecol. Indic. 37, 199–209.

USGS, 2013a. Digital Elevation Model, The National Map [WWW Document]. URL http://nationalmap.gov/viewer.html

USGS, 2013b. STATSGO and SSURGO datasets [WWW Document]. URL http://water.usgs.gov/GIS/metadata/usgswrd/XML/ussoils.xml

USGS, 2014. USGS Real time water data for North Carolina. [WWW Document]. URL http://nc.water.usgs.gov/realtime/real_time_cape_fear.html

Xu, X., Liu, W., Scanlon, B.R., Zhang, L., Pan, M., 2013. Local and global factors controlling water-energy balances within the Budyko framework. Geophys. Res. Lett. 40, 6123–6129.

Yang, D., Sun, F., Liu, Z., Cong, Z., Ni, G., Lei, Z., 2007. Analyzing spatial and temporal variability of annual water-energy balance in nonhumid regions of China using the Budyko hypothesis. Water Resour. Res. 43, n/a–n/a.

Zhang, L., Dawes, W.R., Walker, G.R., 2001. Response of mean annual evapotranspiration to vegetation changes at catchment scale. Water Resour. Res. 37, 701–708.

Zhang, L., Hickel, K., Dawes, W.R., Chiew, F.H.S., Western, a. W., Briggs, P.R., 2004. A rational function approach for estimating mean annual evapotranspiration. Water Resour. Res. 40.

Zhou, X., Zhang, Y., Wang, Y., Zhang, H., Vaze, J., Zhang, L., Yang, Y., Zhou, Y., 2012. Benchmarking global land surface models against the observed mean annual runoff from 150 large basins. J. Hydrol. 470-471, 269–279.

All WWW documents accessed July 2014

Table 1. Precipitation and evapotranspiration in Cape Fear according to different data sources. Mean and standard deviation values are obtained from the 10 subcatchments. The relative difference between baseline data (i.e. PRISM and FAO sources, respectively, for P and ET_0), and the alternative data sources, is given as the mean and the range for the ten subcatchments.

	Annual P (mm)			**Annual ET_0 (mm)**		
	PRISM	*Spline*	*IDW*	*FAO*	*ClimOffice*	*MODIS*
Mean (± st. deviation)	1118 ±11	966 ±81	975 ±38	1200 ±18	1189 ±56	1459 ±19
Relative difference from baseline data (*mean difference and range)*		-14% [-23; 2]%	-13% [-17; -4]%		-1% [-8; 5]%	+22% [14; 24]%

Table 2: Summary of mean flow, precipitation, reference evapotranspiration, and land use characteristics of the ten study subcatchments. LULC classes shrubland, swine farm, open water and barren represetented ≤2% and are not reported here. Predicted mean flow values are results from the InVEST model with Z set to ~~14~~22 (the difference with the calibrated run, with Z=14, is shown in brackets). P and ET_0 are precipitation and reference evapotranspiration, respectively.

ID	**Name**	**Area (km^2)**	**Observed flow (mm)**	**Predicted flow[1] (mm)**	**P (mm)**	**ET_0 (mm)**	**%Forest**	**%Grassland**	**%Agriculture**	**%Pasture**	**%Wetland**	**%Urban**
2105769	CapeFear @Kelly	13,567	278	~~239~~208 [-31]	1112	1212	49	13	9	6	6	13
2105500	CapeFear @Tarheel	12,535	265	~~249~~218 [-32]	1109	1207	51	13	9	6	3	14
2102500	CapeFear @Lillington	8973	236	~~254~~225 [-29]	1110	1196	55	10	9	8	1	14
2104220	RockfishCR @Raeford	237	368	~~226~~174 [-53]	1118	1240	62	18	1	0	7	8
2102000	DeepRiver @Moncure	3727	250	~~248~~210 [-39]	1113	1203	58	9	7	11	0	11
2097314	NewHopeCR @Blands	197	357	~~336~~322 [-14]	1143	1199	49	5	2	2	3	39
2100500	DeepRiver @Ramseur	913	289	~~314~~287 [-27]	1112	1177	43	9	9	10	0	27
2096960	HawRiver @Bynum	3294	278	~~287~~264 [-23]	1110	1181	48	10	14	9	0	17
2097464	MorganCR @WhiteCross	22	177	~~201~~176 [-26]	1133	1198	72	7	10	5	0	5
2096846	CaneCR @OrangeCR	20	202	~~183~~163 [-20]	1123	1192	71	6	11	6	0	4

[1] In brackets, we report the difference in corrected water yield, in mm, between the baseline and calibrated runs (Z=22, and Z=14, respectively)

Table 3. Bias between the water yields obtained from the InVEST model (baseline value Z=22), the lumped ~~Zhang~~ model, and observed data. The average, minimum, and maximum bias values for all the subcatchments are reported. Note that comparison with observations discards the Rockfish subcatchment which was identified as an outlier (see text for details).

	Average	Min	Max
InVEST/Lumped model	-0.10	-0.24	0.14
InVEST/Observations	-0.16	-0.53	-0.01
Lumped model/Observations	0.04	-0.36	0.29

Figures

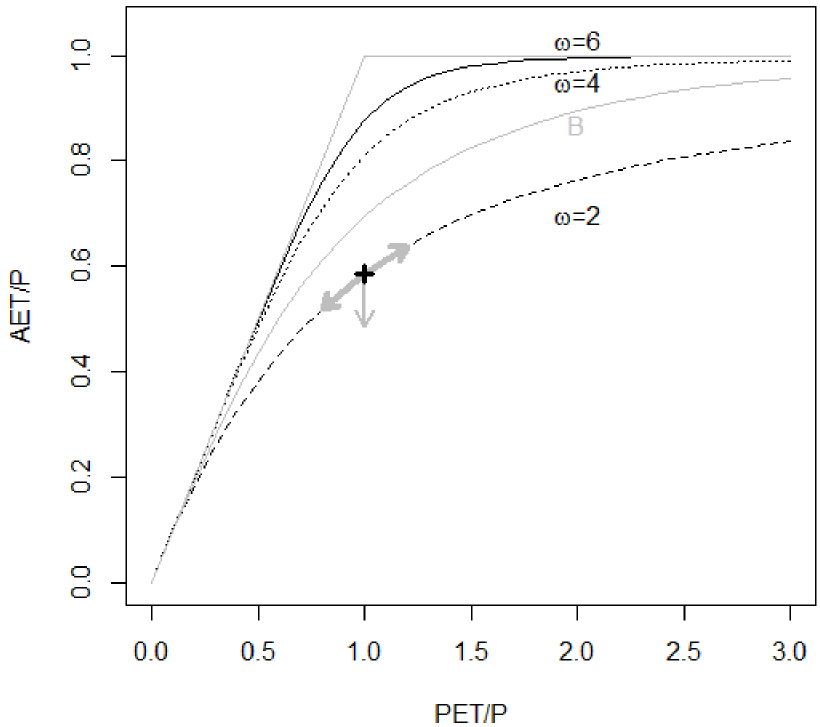

Figure 1. ~~Zhang~~Original ~~-~~Budyko curve ("B") and its variations used in the lumped model (Equation 1), shown for ω values of 2, 4, and 6. Grey lines represent the energy and water limits. Arrows illustrate the effect of a change in the climate forcing (thick arrows) and a change in the ω parameter, a function of Z, precipitation, and soil properties (thin arrow, see Equation 3 for details).

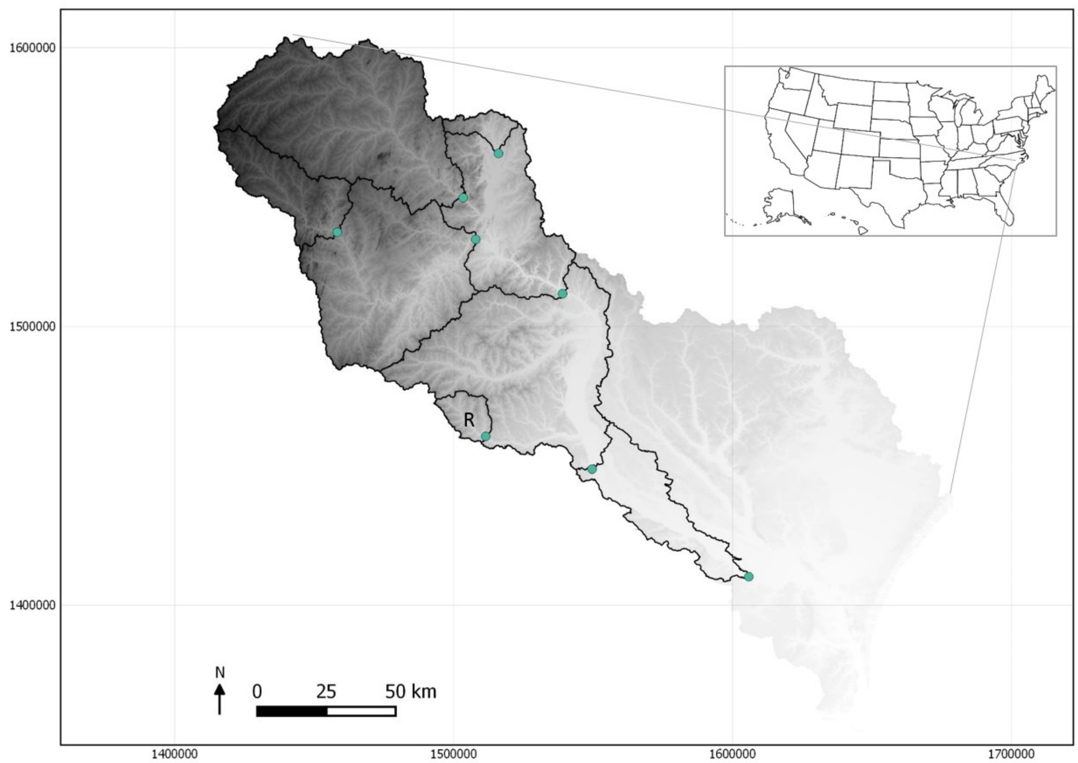

Figure 2. Cape Fear catchment showing locations of the stream gauges and subcatchments used in the study. The Rockfish catchment, discussed in the text, is indicated by a R.

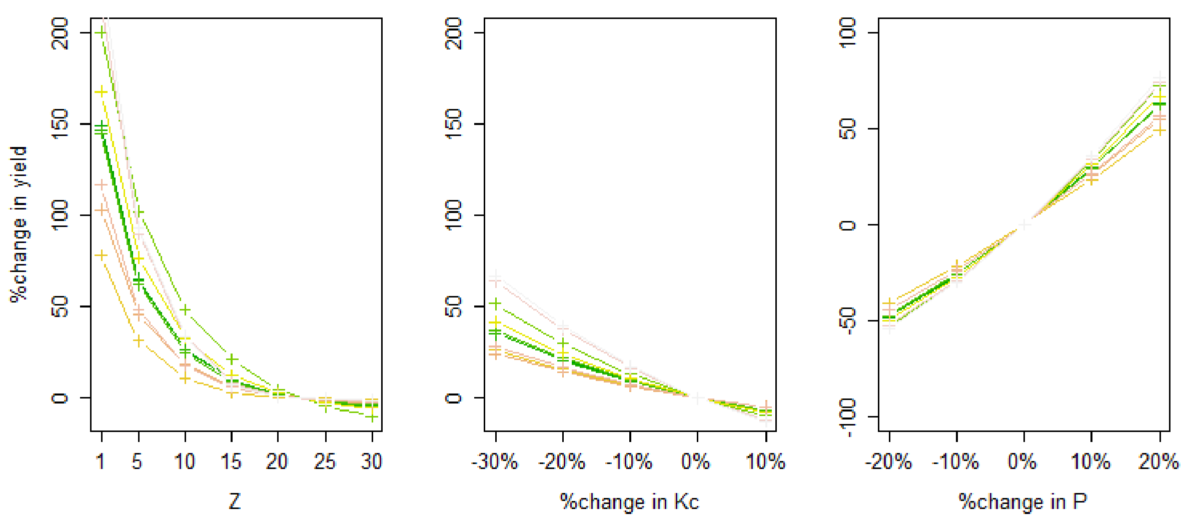

Figure 3. Sensitivity of the water yield output to the Z coefficient and crop coefficient for forest LULC (K_c).

Changes are relative to the baseline run (where Z=22 and K_c=1). On the left hand side plot, absolute Z values are plotted on the x-axis to facilitate the discussion on the Z coefficient. Each curve represents a subcatchment.

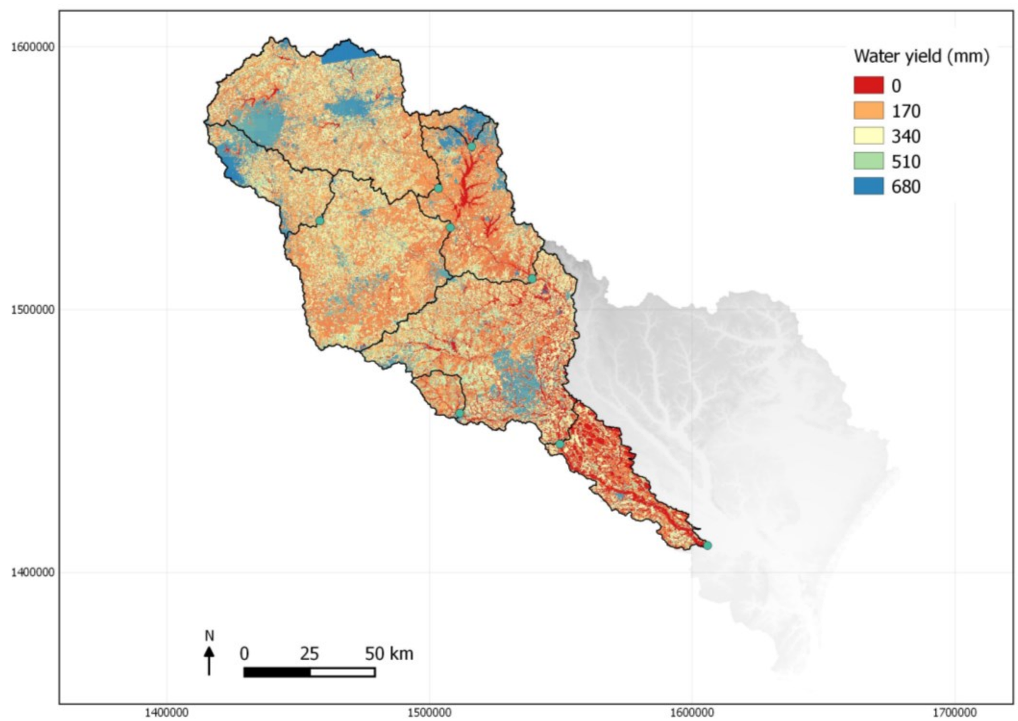

Figure 4. Spatially explicit output of the InVEST model, showing the water yield computed on a pixel scale. Model outputs are aggregated at the subcatchment scale, delineated by black lines, to be compared with observations at the gauging stations (green circles).

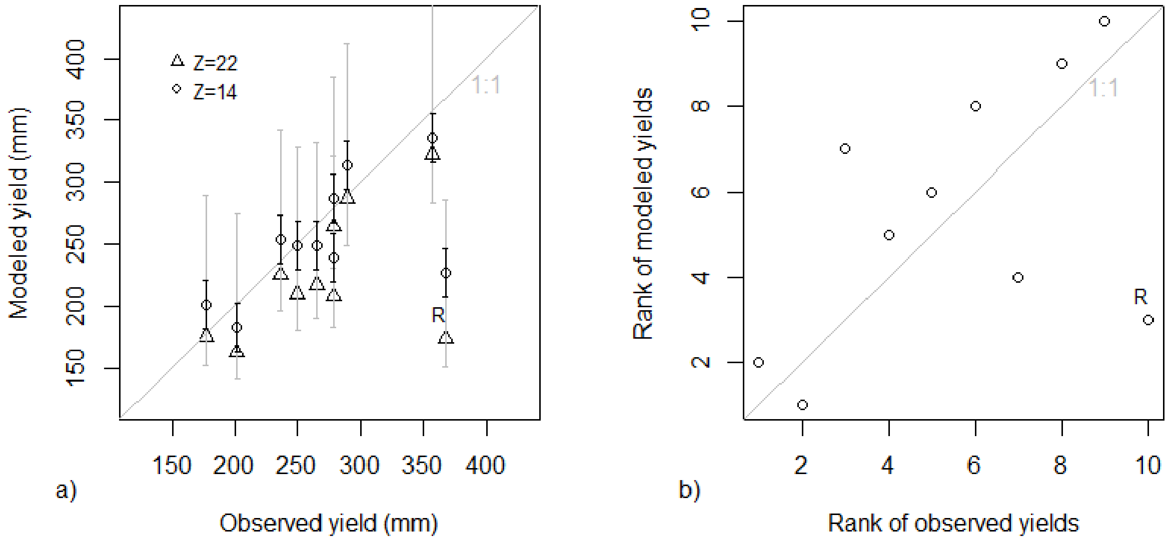

Figure 45. a) Comparison between modeled yields (corrected for water withdrawal) and observed yields, both for the baseline run (Z=22), and the calibrated run (Z=14). Black error bars represent the uncertainty on the value for water withdrawal, while grey bars represent a 10% error in the precipitation input. b) Comparison of subcatchment ranks. The outlier (Rockfish) subcatchment is noted with a R on each figure (see text for details).

Appendix:

Table A1. Data sources and statistics for model inputs. Raster statistics are for the entire Cape Fear catchment delineated in Figure 2.

Data	Type	Value (Mean and range)	Source	Range for sensitivity analyses
Precipitation	Raster	1180 mm [1030; 1450] mm	PRISM* (Gilliland, 2003) (USGS, 2014)	+/- 20%
Reference evapotranspiration	Raster	1240 mm [1160; 1310] mm	FAO* MODIS (Mu et al., 2012) Climate Office (NCSU, 2014)	+/- 10%
DEM	Raster	90 m [0; 250] m	(USGS, 2013a)	n.a.
LULC	Raster	Cf. Appendix	(NASS, 2013)	n.a.
Soil depth	Raster	1710 mm [0; 2110] mm	(USGS, 2013b)	n.a.
PAWC	Raster	0.18 [0.07; 0.52]	(USGS, 2013b)	n.a.
Root depth	Per LULC class	See Table A1	(Allen et al., 1998)	n.a.
K_c	Per LULC class	See Table A1	(Allen et al., 1998)	[- 30%; +10%]
Z	Constant	22*	(Sharp et al., 2014)	[1; 30]

* Indicates the data source used for the baseline run (see Section 3.2)

Table A2 – Biophysical table used for the baseline InVEST model run, giving the root depth and crop coefficient Kc for each Land use/Land cover (LULC) class (values from Allen et al, 1998)

LULC	Root depth (mm)	Kc
Ag-Corn	1500	0.75
Ag-other	1100	0.7
Grass	1100	0.9
Forest	5000	1
Wetland	na	1.1
Urban	na	0.4