Article (refereed) - postprint

Redhead, J.W.; Stratford, C.; Sharps, K.; Jones, L.; Ziv, G.; Clarke, D.; Oliver, T.H.; Bullock, J.M. 2016. **Empirical validation of the InVEST water yield ecosystem service model at a national scale**.

Empirical validation of the InVEST water yield ecosystem service model at a national scale

Redhead, J.W.[a,e]
Stratford, C.[a]
Sharps, K.[b]
Jones, L.[b]
Ziv, G.[c]
Clarke, D.[d]
Oliver, T.H.[a,1]
Bullock, J.M.[a]

[a] NERC Centre for Ecology and Hydrology, Maclean Building, Wallingford, Oxfordshire, OX10 8BB, UK.

[b] NERC Centre for Ecology and Hydrology, Deiniol Road, Bangor, Gwynedd, LL57 2UW, UK

[c] School of Geography, University of Leeds, Leeds LS2 9JT, United Kingdom

[d] Faculty of Engineering and Environment, University of Southampton, University Road, Highfield, Southampton, SO17 1BJ, UK

[e] Corresponding author: Email - johdhe@ceh.ac.uk

[1] Present address: School of Biological Sciences, Harborne Building, University of Reading, Reading, Berkshire RG6 6AS, UK

Abstract

A variety of tools have emerged with the goal of mapping the current delivery of ecosystem services and quantifying the impact of environmental changes. An important and often overlooked question is how accurate the outputs of these models are in relation to empirical observations. In this paper we validate a hydrological ecosystem service model (InVEST Water Yield Model) using widely available data. We modelled annual water yield in 22 UK catchments with widely varying land cover, population and geology, and compared model outputs with gauged river flow data from the UK National River Flow Archive. Values for input parameters were selected from existing literature to reflect conditions in the UK and were subjected to sensitivity analyses. We also compared model performance between precipitation and potential evapotranspiration data sourced from global- and UK-scale datasets. We then tested the transferability of the results within the UK by additional validation in a further 20 catchments.

Whilst the model performed only moderately with global-scale data (linear regression of modelled total water yield against empirical data; slope = 0.763, intercept = 54.45, R^2 = 0.963) with wide variation in performance between catchments, the model performed much better when using UK-scale input data, with closer fit to the observed data (slope = 1.07, intercept = 3.07, R^2 = 0.990). With UK data the majority of catchments showed less than 10% difference between measured and modelled water yield but there was a minor but consistent overestimate per hectare (86 m^3/ha/year). Additional validation on a further 20 UK catchments was similarly robust, indicating that these results are transferable within the UK. These results suggest that relatively simple models can give accurate measures of ecosystem services. However, the choice of input data is critical and there is a need for further validation in other parts of the world.

Keywords

UK, mapping, rainfall, evapotranspiration, river flow, land cover

1. Introduction

Ecosystem services are increasingly used to assess likely impacts of environmental change in societal and economic terms and to provide a rationale for conservation or environmental management (Tallis *et al.* 2008; Braat & de Groot 2012). However, to incorporate the ecosystem services concept into assessments and decision making, there is a requirement for accurate mapping and measurement of ecosystem services (Malinga *et al.* 2015). In some cases, this requirement has itself been incorporated into policy (European Commission 2011).

To meet this rising demand there has been a proliferation of methods and tools to map, quantify and value the provision of ecosystem services (Fisher, Turner & Morling 2009; Seppelt *et al.* 2011; Malinga *et al.* 2015). These vary in complexity from simple approaches based on maps and land use or habitat-based proxies to complex, process-based models (Seppelt *et al.* 2011). Ecosystem service tools have been designed and applied at widely varying geographic locations and both spatial and temporal scales. Potential users must thus choose which tools are most appropriate for their particular situation, and be aware of the limitations of these tools (Willcock *et al.* 2016). Recent reviews have identified that one of the key obstacles to successful ecosystem service mapping and implementation into decision making processes is the comparative scarcity of validation or measurements of uncertainty in many applications of ecosystem service models (Seppelt *et al.* 2011; Maes *et al.* 2012; Schulp *et al.* 2014; Malinga *et al.* 2015). Whilst it is frequently acknowledged that ecosystem service models function at best as reliable proxies, and at worst as crude estimates, the validation of the results of ecosystem service models against empirical measurements is comparatively rare (Seppelt *et al.* 2011; Vigerstol & Aukema 2011; Schulp *et al.* 2014; Hamel & Guswa 2015). Of those studies which do employ validation, many do so at a limited number of locations to check the performance of a model within their study region (e.g. Bai *et al.* 2013; Boithias *et al.* 2014; Terrado *et al.* 2014; Xiao *et al.* 2015). Whilst this is entirely sensible, the results of such local-scale validation are less likely to be transferrable to new locations and the regional or national scales at which ecosystem service models are most widely used (Martínez-Harms & Balvanera 2012) and most water resource planning takes place (Watts *et al.* 2015). Several studies have compared different ecosystem service models (e.g. Vigerstol & Aukema 2011; Cheaib *et al.* 2012; Rosenzweig *et al.* 2014; Dennedy-Frank *et al.* 2016), which gives some insight into the *uncertainty* surrounding the modelling of the service in question (see Hou, Burkhard and Müller (2013) concerning uncertainty in ecosystem service modelling) and the utility of the different models, but does not provide insight into the *accuracy* of each model in estimating ecosystem service delivery or representing the biophysical process underpinning the service.

This relative scarcity of large-scale validation means that, for many models, there is comparatively little information on either the accuracy of model outputs (Seppelt *et al.* 2011), or on the performance of models in different circumstances and locations, especially where the latter are in poorly-studied regions. There is also a lack of information on the requirements of the input data. In many cases, the availability and spatial coverage of data is inversely correlated with its resolution (Hijmans *et al.* 2005) and, potentially, its accuracy. Thus it is uncertain whether the most widely available data, even when used in a model which performs well under ideal circumstances, will produce sufficiently accurate results. Potential users are thus missing vital information on the performance of models, which they need if they are to make informed decisions on which tools to use and how best to employ them to provide accurate assessments for decision makers (Willcock *et al.* 2016). Validation also provides valuable feedback to ecosystem service model developers who are seeking to improve the accuracy, utility and efficiency of their models.

Hydrological services are particularly well suited to empirical validation, as the ecosystem processes which underpin them (e.g. runoff of water, nutrients and sediment) have physical expressions which can be directly measured at appropriate spatial and temporal resolutions (river flow, nutrient concentration and sediment load, respectively). In the UK, these measurements are undertaken by government bodies and are readily available for academic purposes (e.g. The National River Flow Archive, NRFA).

This study aims to validate a hydrological ecosystem service tool at the national scale, using widely available spatial data (of the sort available to most potential users and decision makers) for both model inputs and validation. We used a tool from a widely used, open-source ecosystem service modelling suite, InVEST (Integrated Valuation of Ecosystem Services and Tradeoffs, Sharp *et al.* 2015). Whilst InVEST tools have been widely used for a variety of research and planning applications (e.g. British Hydropower Association 2010; Bai *et al.* 2013; Bangash *et al.* 2013; Leh *et al.* 2013; Boithias *et al.* 2014; Terrado *et al.* 2014; Pessacg *et al.* 2015; Xiao *et al.* 2015), as with other ecosystem service models, comparatively few applications have employed an empirical validation of results at anything other than a local scale. Therefore our objectives are 1) to examine the sensitivity of the model to variation in the values of input parameters in a UK context; 2) to compare the performance of the model using two points on the spectrum of data availability and spatial coverage (global climatic data and UK specific climate data) by validation against empirical measurements; 3) to examine whether our results are transferable within the UK.

2. **Methods**

2.1. THE INVEST WATER YIELD MODEL

The InVEST suite of tools has been developed to enable decision makers to assess trade-offs among ecosystem services and to compare scenarios of change, for example in land use or climate (Sharp *et al.* 2015). To this end, InVEST comprises a set of models covering a wide variety of ecosystem services. The models are based on comparatively simple production functions, being intended to run quickly on a standard desktop computer and to take advantage of readily available data (Sharp *et al.* 2015). Although InVEST models are not designed to reproduce empirical observations, the water yield model is intended to quantify the relative yields of different catchments or sub-catchments, and be sensitive to modelled changes in drivers such as land use change or climate change. We would also suggest that, because the model produces figures of water yield which appear to have a high degree of numerical precision, and is freely available, it is important to test whether the results are accurate, as users may not always familiarise themselves with the intended use and limitations of the model before incorporating the results into the decision making process (see Willcock *et al.* 2016).

The InVEST water yield model (Hydropower/Water Yield, InVEST v3.2.0, Sharp *et al.* 2015) calculates annual water yield from a catchment, with the intended end use of reservoir hydropower production (Sharp *et al.* 2015). Although hydropower forms a relatively small contribution to the UK energy sector (DECC 2015), total annual water yield can be considered in the light of many potential services, including agricultural irrigation, provision of drinking water, hydropower and industrial abstraction. The UK is densely populated and has a large proportion of its land area under anthropogenic land uses. This leads to competition between demands for water, which is likely to intensify in the future due to population growth and climate change (Weatherhead & Knox 2000; Knox *et al.* 2009). Validated models of current and predicted future water yield, with clear estimates of their accuracy and uncertainty, are thus of great importance in strategic water resource planning (Watts *et al.* 2015). Therefore, in this study we focused on the biophysical output of water yield. As the InVEST model is compartmentalised into water yield, water consumption and hydropower valuation, we used the first two components only.

The model estimates the total annual water yield (*Y*) for each grid square (*x*) of the study catchment as total catchment annual rainfall (*P*) minus total catchment annual actual evapotranspiration (*AET*) (equation 1). The model assumes that, on an annual time step, all water falling as rainfall over a catchment, minus that which is evapotranspired, leaves the catchment. No distinction is made between surface and sub-surface water flow.

Eqn. 1 $$Y(x) = \left(1 - \frac{AET(x)}{P(x)}\right).P(x)$$

In practice, the measurement of annual actual evapotranspiration at the catchment scale is extremely difficult. Even plot scale evaluation requires highly specialised equipment, and plot and field scale methods to determine actual evapotranspiration are problematic to apply at the landscape scale (Evans *et al.* 2012). The InVEST approach relates AET to potential evapotranspiration (PET), which is easier to model, using the methodology developed by Budyko (1974) and later adapted by Fu (1981) and Zhang *et al.* (2004) (equation 2) where ω is an empirical parameter which defines the shape of the curve relating potential to actual evapotranspiration.

Eqn. 2 $$\frac{AET(x)}{P(x)} = 1 + \frac{PET(x)}{P(x)} - \left[1 + \left(\frac{PET(x)}{P(x)}\right)^{\omega}\right]^{1/\omega}$$

PET is estimated as the product of the reference evapotranspiration and the crop coefficient for each grid square. ω is related to the plant available water content (AWC), precipitation and the constant *Z* which captures the local precipitation pattern and additional hydrogeological characteristics (equation 3) (Sharp *et al.* 2015).

Eqn. 3 $$\omega = Z\,\frac{AWC(x)}{P(x)} + 1.25$$

For a more detailed description of the water yield model see Sánchez-Canales *et al.* (2012); Bangash *et al.* (2013); Hamel and Guswa (2015); Pessacg *et al.* (2015); and Sharp *et al.* (2015).

2.2. Model input parameters

The InVEST model requires five biophysical parameters as georeferenced rasters. These are root restricting layer depth (mm), plant available water content (AWC, as a proportion), average annual precipitation (mm), average annual potential evapotranspiration (PET, mm) and land use/land cover (LULC). We obtained these data from a variety of sources, with the aim of ensuring that the data were easily obtainable and free to license for at least academic use. These are the kind of data which are likely to be most widely used in a freely available tool such as InVEST, in terms of precision, spatial resolution and spatial coverage.

Root restricting layer depth and AWC were obtained from the European Soil Database (ESDB) version 2.0 (Panagos 2006; Panagos *et al.* 2012). Annual precipitation and reference evapotranspiration were obtained from several alternative sources. We used two pairings of these two variables, to compare model performance with data from two points on the spectrum of data availability and spatial coverage. First, we used global scale precipitation data from WorldClim (Hijmans *et al.* 2005) and PET from the CGIAR-CSI Global-Aridity and Global-PET Database (Zomer *et al.* 2007; Zomer *et al.* 2008). These are both freely available and have global coverage, at approximately 1km resolution. Secondly, we used UK Met Office UKCP09 precipitation data at 5km resolution (Perry & Hollis 2005; Jenkins, Perry & Prior 2008) and Met Office Rainfall and Evaporation Calculation System (MORECS) evapotranspiration data. These datasets are UK-specific and available to a wide variety of users under the UK's open government license. Where necessary, data were geoprocessed to meet the data formatting requirements of the InVEST model in ArcMap (v10.1, ESRI, Redlands, CA).

Where possible, all input data were limited to the same date range as the validation data (2000-2010, see below) and averaged across years, giving average annual precipitation and average annual PET. LULC data were obtained from the 25 m raster version of the UK Land Cover Map 2007 (LCM2007, Morton *et al.* 2011).

The InVEST model also requires several tabular values for each LULC class. These include whether the land cover class is vegetated or not, rooting depth and a plant evapotranspiration coefficient (*Kc*). This last is used to obtain potential evapotranspiration by modifying the reference evapotranspiration, which is based on a 15cm tall surface of actively growing, well-watered grass. We estimated these coefficients for LCM2007 broad habitat classes by matching class descriptions with those in Canadell *et al.* (1996), Allen *et al.* (1998) and Sharp *et al.* (2015). Further amendments were made to these values, to reflect the damp climate of the UK (Smethurst, Clarke & Powrie 2012). This generally resulted in raised crop coefficients and shallower rooting depths. The coefficients for urban and suburban areas were amended to reflect the approximate proportion of green space they typically contain (20% and 60% respectively). Coefficients for *Kc* and rooting depth for each LCM2007 broad habitat are given in Supplementary Material, Table S1. For arable land uses, actual evapotranspiration varies over the course of a year as crops are sown, grow and are harvested before the land is then re-cultivated. Evapotranspiration of growing crops also varies between crop plant species, crop condition and many other factors (Allen *et al.* 1998; Hulme, Rushton & Fletcher 2001). For UK crops and soil conditions, preliminary investigation and previous studies (Allen *et al.* 1998; Hulme, Rushton & Fletcher 2001) suggested a value of *Kc* close to one to best represent annual evapotranspiration from arable land.

The seasonality constant (*Z*) was estimated as 0.2*N, where N is the average number of rain days (> 1mm) per year over the study period (Donohue, Roderick & McVicar 2012; Hamel & Guswa 2015). N was estimated at approximately 150 from UK Met Office data (http://www.metoffice.gov.uk/climate/uk/datasets/), giving a value of 30 for *Z*.

Because the validation data (i.e. gauged annual yield, see below) are affected by any consumptive water use, it was important to account for this. The InVEST model uses a comparatively crude method of estimating consumptive water use, by assigning a value of annual consumption per hectare to each land cover class (Sharp *et al.* 2015). Because abstraction varies widely across the UK (DEFRA 2015), we split the LULC raster based on administrative regions, such that each LULC class-region combination had a unique value, allowing us to assign a suitable abstraction value from regional abstraction statistics (DEFRA 2015). We used only the values for abstraction for agricultural purposes (assigned to the arable LULC class) and public and industrial water supply (assigned to the urban/suburban LULC class), as most other uses (e.g. hydropower) do not consume water but return it to the catchment after use (Terrado *et al.* 2014). Public and industrial water supply may also return water after use, but in many cases it may be returned further downstream from the point at which it was abstracted, or to a different catchment.

2.3. Sensitivity analysis

We investigated the sensitivity of the model to variations in precipitation, PET, rooting depth, AWC, *Kc* and *Z* following Sánchez-Canales *et al.* (2012) and Hamel and Guswa (2015). The biophysical parameters, which are input as rasters, were varied by ± 10% and ± 20% applied uniformly across the raster. Sensitivity to *Kc* was examined by varying its value for the two dominant LULC classes across all catchments (arable and improved grassland), by the same proportions as the biophysical parameters. Sensitivity to *Z* was tested using values of zero, one, two, five and further increments of five up to 50. The model was run independently for each of these variations.

2.4. Validation data

For initial validation of the model and comparison of the two sets of climate datasets we selected 22 catchments in England, Scotland and Wales with widely varying land cover, rainfall, elevation, and geology (Fig. 1 and Supplementary Material, Table S2). All of these factors are likely to affect actual water yield and, potentially, the performance of the InVEST water yield model. Empirical measurements of gauged daily water flow were obtained from the National River Flow Archive (NRFA), which collates, quality controls, archives and disseminates hydrometric data from gauging station networks operated by government environmental bodies across the UK (Fry & Swain 2010). The catchments for each NRFA gauging station have been defined using the Centre for Ecology & Hydrology's Integrated Hydrological Digital Terrain Model (Morris & Flavin 1990). We calculated total gauged annual water yield for each catchment by summing gauged daily mean flow for each year from 2000 – 2010 and took the mean value across years. We analysed how well the modelled data predicted the empirical data using linear regression. This was done using total annual yield, but also the per hectare yield, i.e. the total yield divided by the catchment area, to remove trivial correlation caused by the large variation in area among the catchments. These calculations (and all subsequent data manipulation and statistics) were performed in R (v3.1.0, R Core Team 2014).

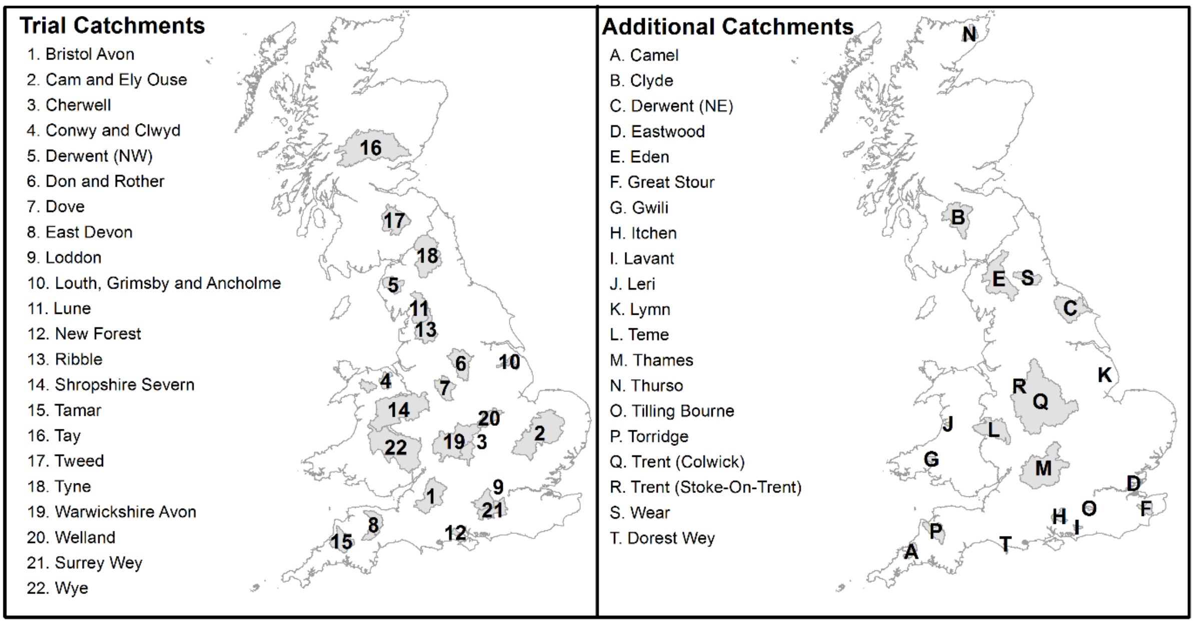

Fig. 1 Coastline of Great Britain overlain with the 22 trial catchments selected for testing the InVEST water yield model and the 20 additional catchments used for additional validation (grey shaded areas). See Supplementary Material, Table S2 for catchment characteristics.

2.5. Additional validation of the model

To ensure that our selected values of *Kc*, *Z* and input datasets for precipitation and PET did not simply 'calibrate' the model to the 22 trial catchments, (i.e. to check that the model performance obtained from the trial catchments was representative of UK catchments in general and thus that our results are transferable between UK catchments) we selected the climatic datasets and parameter values which resulted in the best fit to validation data for the original 22 catchments, and used these to run the model for a further 20 catchments (Fig. 1). Catchments were again defined from NRFA gauging station locations and were chosen to show wide variation in area, land cover and geology (Supplementary Material, Table S2).

3. Results

3.1. Sensitivity analysis

Modelled water yield was highly sensitive to changes in precipitation (Fig. 2A), with a 10% increase in precipitation resulting in an 11% -27% increase in water yield, and was somewhat less sensitive to variation in PET (Fig. 2B). Sensitivity to both precipitation and PET was highly catchment specific. With PET, in some catchments a 10% increase in PET resulted in a 14% decrease in water yield, while the mean decrease was only 5%. The model was relatively insensitive to rooting depth and AWC, with a 10% increase in either of these datasets resulting in a yield decrease of 0% - 3%. Sensitivity to

Kc was roughly similar to that for PET, which is unsurprising since the effect of the former in the model is to modify the latter, and was likewise catchment specific. In general, catchments were either 'highly sensitive', responding to variation in values of all input parameters with variation in water yield, or 'less sensitive', showing comparatively little variation in water yield with any variation in the values of model input parameters, although the latter still responded to percentage changes in precipitation with at least a corresponding percentage change in modelled yield.

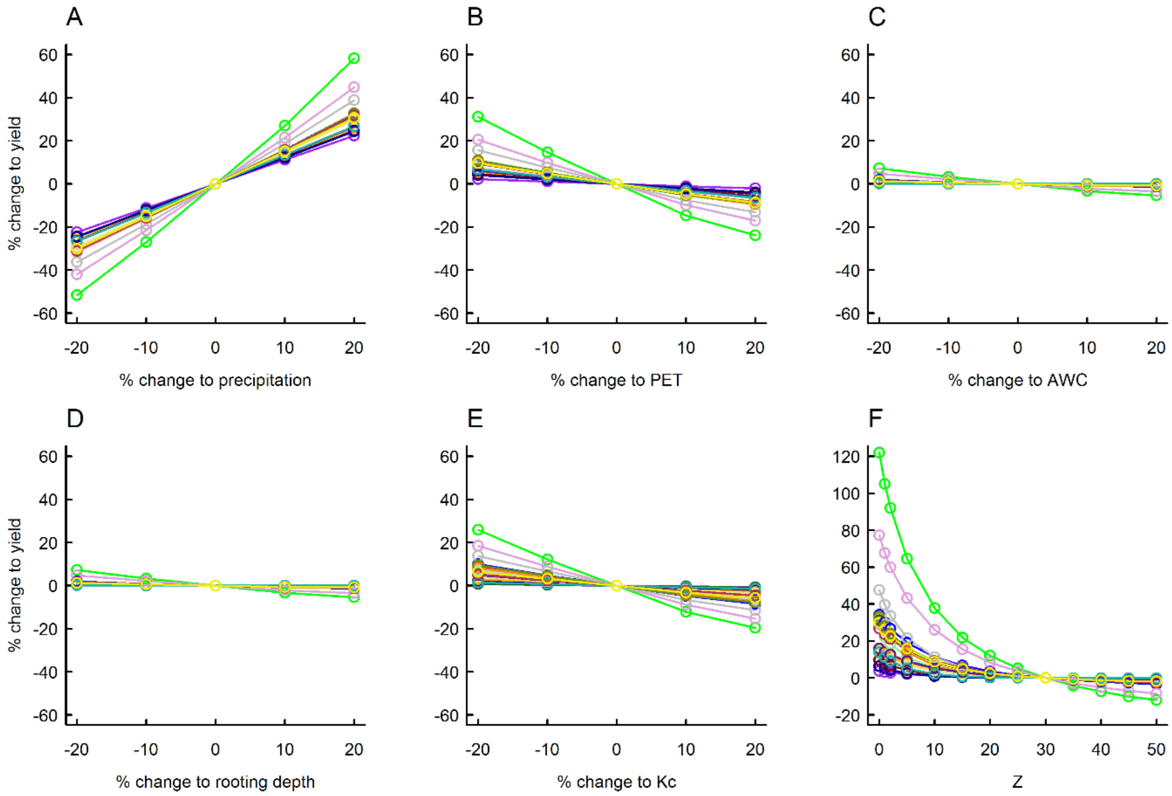

Fig. 2 Sensitivity of the InVEST water yield model to variation in the values of input parameters for 22 test catchments. Variation shown as percentage change relative to a 'baseline' run with *Z* = 30 using UKCP09/MORECS data. A coloured set of points with connecting line are shown for each catchment.

Catchment sensitivity to variation in precipitation (and thus in other model parameter values) was significantly positively correlated with mean PET (Pearson's $r = 0.555$, $p = 0.011$) and area of arable land (Pearson's $r = 0.759$, $p < 0.001$), and significantly negatively correlated with altitude (Pearson's $r = -0.676$, $p = 0.001$) and area of semi-natural habitat (Pearson's $r = -0.577$, $p < 0.008$). All of these catchment characteristics are significantly inter-correlated, such that catchments with higher mean altitude have a lower cover of arable land and a correspondingly higher cover of semi-natural habitat and a lower mean PET.

The sensitivity of the model to changes in the value of *Z* was also strongly catchment specific (Fig. 2F), as expected given the spatial variation in the biophysical variables which modulate the effect of

Z on water yield (Hamel & Guswa 2015). Because it is difficult to translate the sensitivity of the model to *Z* into an appropriate value of *Z* to use, the outputs from models with varying values of *Z* were compared to the validation data to identify which value of *Z* resulted in the best fit to the validation data. The results of this analysis (Fig. 3) showed that model fit (R^2) levelled off at $Z \approx 30$ (Fig. 3A), as did slope (Fig. 3B), whilst overestimation of per hectare water yield was also much reduced at values above 30 (Fig. 3C). This supports the value of $Z = 30$ for the model runs detailed below and hence the estimation of *Z* from mean annual rain days.

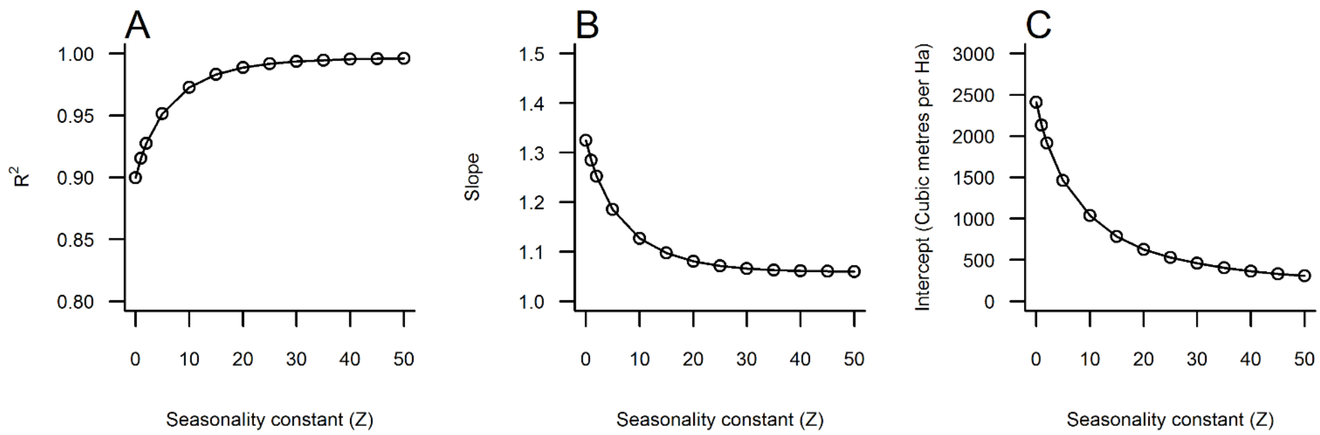

Fig. 3 Effects of varying the value of the seasonality constant (*Z*) on the relationship between modelled and gauged water yield for 22 test catchments, using the UKCP09/MORECS data. A) R^2 of linear regression between modelled and gauged catchment yield; B) Slope of linear regression between modelled and gauged catchment yield; C) Intercept of linear regression between modelled and gauged yield per hectare.

3.2. Model validation and comparison of climatic datasets

Both global- and UK-scale climate datasets resulted in estimated water yields which were strongly correlated with empirical yields obtained from NRFA gauged river flow (Fig 4, Table 1). The WordClim and CGIAR-CSI data performed less well than the UKCP09/MORECS datasets. Although R^2 values for models using the global input data were only slightly lower (e.g. 0.96 compared with 0.99; Table 1), the slope values for per hectare yield (including confidence intervals) were less than one (Table 1). Hence the global data led to considerable under-estimates (up to 45%) of water yield for catchments where the yield per hectare was high and to overestimates of water yield for those where it was low (Fig. 4B), leading to the intercept of 1443.63 m^3 per hectare per year. By contrast, the UKCP09/MORECS data led to more consistent and accurate estimates for total water yield when adjusted for consumptive abstraction (Table 1). When per hectare yield was considered, the

UKCP09/MORECS data gave good fits to the NRFA data ($R^2 = 0.949$), with a slope not significantly different from one and the intercept indicating a consistent but minor overestimate of 86.3 m^3 per hectare per year when adjusted for consumptive abstraction (Fig 4D).

Table 1. Results of linear regressions between InVEST modelled and empirical water yield for the 22 original catchments, using two input datasets for the precipitation and reference evapotranspiration parameters. Intercept, slope ± 95% confidence interval and R^2 are given for total catchment yield in millions of cubic metres and yield per hectare in cubic metres, for both raw water yield (R) and yield adjusted for consumptive abstraction (A)

		Total estimated yield			Estimated yield per hectare		
Input data	Abstraction	Intercept	Slope	R^2	Intercept	Slope	R^2
WordClim/ CGIAR-CSI	R	86.52	0.759 ± 0.072	0.958	1814.68	0.549 ± 0.131	0.781
	A	54.45	0.763 ± 0.068	0.963	1443.63	0.577 ± 0.127	0.810
UKCP09/ MORECS	R	35.13	1.066 ± 0.041	0.993	457.38	1.053 ± 0.115*	0.946
	A	3.07	1.069 ± 0.044	0.992	86.32	1.081± 0.114*	0.949

* Confidence intervals of slope include one

The mean percentage differences between gauged and modelled water yield were ± 23.36% (SE ± 4.40) for the WordClim/CGIAR-CSI data and ± 18.55% (SE ± 4.94) for the UKCP09/MORECS data. However, in both cases one catchment (Welland, labelled 20 on Fig.1) showed a percentage difference of over 100%. Although the difference between the mean percentage under/overestimates of the two datasets does not appear great, it is important to note that the mean is somewhat skewed by the few catchments for which the model performs particularly poorly, especially for the UKCP09/MORECS data. Median values show that for the UKCP09/MORECS data the majority of catchments had percentage differences between gauged and modelled water yield of less than 10% (median = 9.74%) whilst for the WordClim/CGIAR-CSI data more catchments vary by up to 20% (median = 17.19%).

Despite the significant correlations between catchment sensitivity to variation in the input parameter values and catchment characteristics, percentage under/overestimates of total water yield using the UKCP09/MORECS data did not show any significant correlations with catchment area, altitude, mean precipitation, mean PET, geology (i.e. base flow index) or land cover.

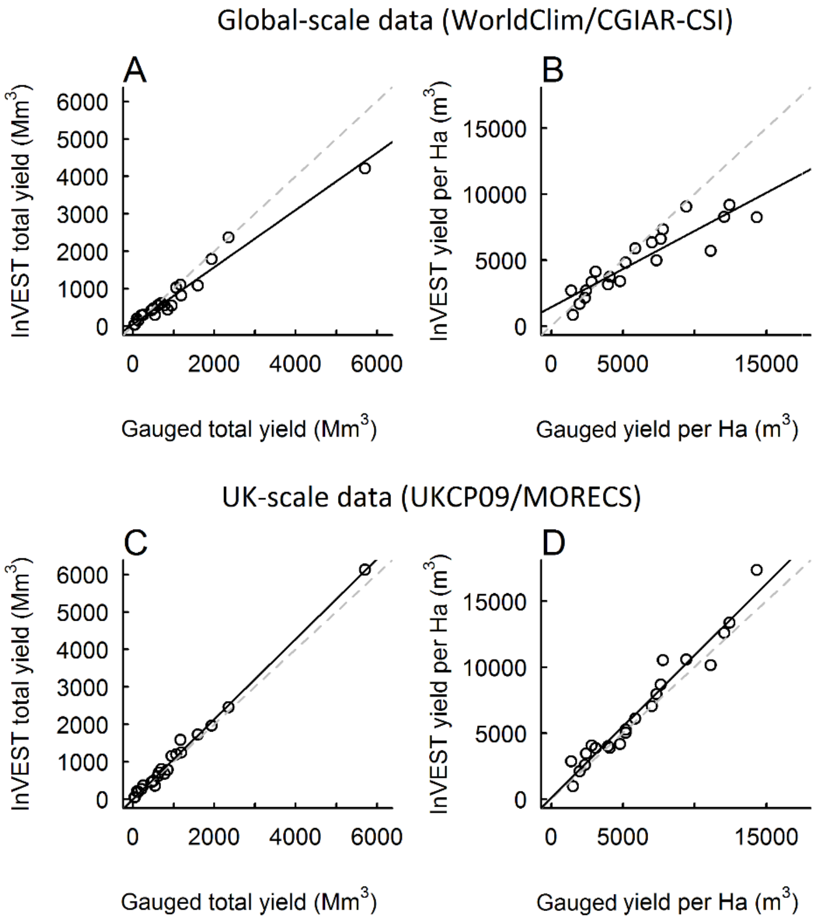

Fig. 4 InVEST modelled water yield (corrected for estimated consumptive abstraction) *vs* gauged water yield, using two input datasets for the precipitation and reference evapotranspiration parameters. A) Estimated total catchment yield in millions of cubic metres, using WordClim/CGIAR-CSI data; B) Estimated yield per hectare in cubic metres, using WordClim/CGIAR-CSI data; C) Estimated total catchment yield using UKCP09/MORECS data; D) Estimated yield per hectare using UKCP09/MORECS data. Grey, dashed line indicates a relationship with intercept = zero and slope = one.

3.3. Additional validation

Comparing modelled and gauged data for a further 20 catchments (using the UKCP09/MORECS dataset because this gave the best results for the original 22 catchments) showed very similar results to the original 22 catchments (Figure 5). Confidence intervals for the linear regression slopes again overlapped with one, and R^2 values were high for both total yield (intercept = 41.48, slope = 0.93 ± 0.13, R^2 = 0.92) and yield per hectare (intercept = 684.75, slope = 0.91 ± 0.23, R^2 = 0.87), suggesting that the data and parameters used to obtain the best results on the original catchments are likely to be applicable across a wide range of catchments within the UK.

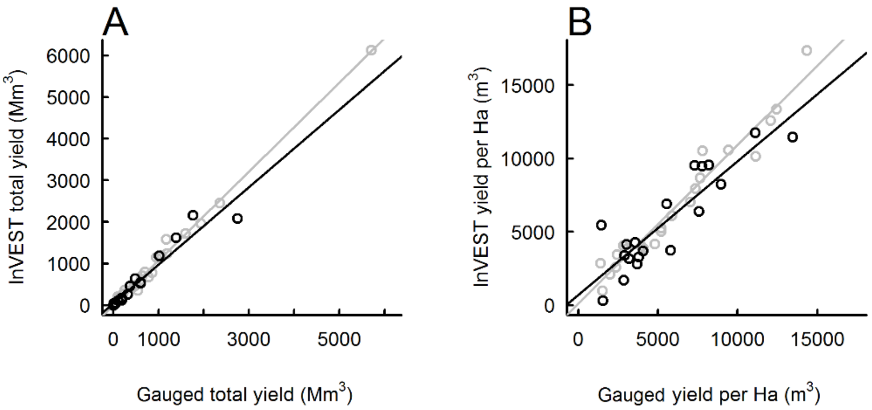

Fig. 5 InVEST modelled total water yield (millions of m^3) (A) and per hectare water yield (m^3/Ha) (B) against gauged water yield for the 22 original test catchments (grey symbols and line) and the 20 additional catchments (black symbols and line). Model run using UKCP09/MORECS dataset for precipitation and PET and corrected for estimated consumptive abstraction.

4. Discussion

Our results show that the InVEST water yield model can produce accurate estimates of water yield in UK river catchments. However, this accuracy is dependent upon careful selection of appropriate model parameters and input data, especially precipitation and PET to which the model is most sensitive. The input values used in this study are transferrable to other UK catchments (as seen by our additional validation using extra catchments). However, when the model is to be used elsewhere, we strongly advocate the trialling of different values for input parameters representing different environmental contexts and empirical validation wherever possible.

The InVEST model was initially designed to assess water availability for hydropower production. Hydropower forms a relatively small contribution to the UK energy sector (DECC 2015) and the spatial distribution of hydropower generation is very uneven, with most large hydropower schemes in large, upland catchments with abundant space for reservoirs. Of course, accurate assessments of water yield are also important for examining the current delivery of, and impact of future environmental change on, other ecosystem services including, water quality in terms of nutrients and sediment, drinking water and crop production. The demand for, and conflicts between, the latter two services are likely to increase with the effects of climate change both at a UK (Weatherhead & Knox 2000; Knox *et al.* 2009) and global scale (Döll 2002). As such, the InVEST water supply model has more general uses than for estimating hydropower. Although there is also the potential for ecosystem disservices from, for example, erosion and flooding, these are dependent on a wide range of additional factors (Brown & Damery 2002; CEH 2008).

4.1. The importance of validation

Several studies have sought to address the issue of validating ecosystem service models. However, many of these have been limited by the availability of suitable validation data. For example, Schulp et al. (2014) sought to undertake model validation for a variety of ecosystem services at a European scale, and whilst their results give an indication of the ability of different modelling approaches to predict spatial patterns of ecosystem service delivery, their need to use proxies in the absence of empirical ecosystem service measures prevented a quantitative assessment of model accuracy. Where data have been available**,** some studies at global or continental scales have compared results from the InVEST water yield model (Mendoza *et al.* 2011), or the Budyko modelling framework upon which it is based (Zhang *et al.* 2004; Zhou *et al.* 2012), to empirical observations. Such studies are useful in comparing models in terms of their ability to respond to global patterns of precipitation but do not provide the information necessary for users to assess whether the model is accurate at a national or regional scale, despite these being the scales at which the majority of ecosystem service mapping exercises are performed (Martínez-Harms & Balvanera 2012) and at which most strategic water resource planning takes place (Watts *et al.* 2015).

Other recent studies have used empirical validation data to assess the performance of the InVEST Water Yield model, but covering only single catchments or sub-catchments within a single river basin. In contrast to the present study, several of these studies have had the primary aim of quantifying spatial variation and predicted changes in water yield, with validation at a small number of points to check the reliability of their results, rather than a more general validation of the model across catchments (e.g. Bai *et al.* 2013; Boithias *et al.* 2014; Terrado *et al.* 2014; Xiao *et al.* 2015). Whilst this is entirely sensible, and such studies have found the InVEST model to be a good predictor of measured water yield, the results of such studies are not necessarily transferable to other locations or scales, especially where model inputs have been 'calibrated' to match the empirical data.

Comparatively few studies have had the explicit aim of investigating the model performance, uncertainty and sensitivity of the InVEST water yield model. These have largely been conducted in sub-catchments within single river basins which vary widely in area, climate and land cover (e.g. 4950 Km^2 Llobregat River basin, Catalonia, Spain (Sánchez-Canales *et al.* 2012); 23 600 Km^2 Cape Fear basin, North Carolina, USA (Hamel & Guswa 2015); 57 400 Km^2 Chubut River basin, Patagonia, Argentina (Pessacg *et al.* 2015)). However, their results are generally corroborated by our national scale analysis (UK land area = 241 930 km^2). These include high sensitivity to precipitation and, to a slightly lesser extent, to evapotranspiration data, as well as empirical support for setting *Z* from numbers of rain events per year (Hamel & Guswa 2015). Our results also corroborate those of these previous studies in demonstrating the substantial improvements in model performance which can be obtained by comparing alternative data sources, especially for those parameters which sensitivity analysis identifies as being major drivers of the model (Sánchez-Canales *et al.* 2012; Hamel & Guswa 2015; Pessacg *et al.* 2015). For example, Boithias *et al.* (2014) paid particular attention to obtaining precipitation and evapotranspiration data, because of the sensitivity analysis undertaken by Sánchez-Canales *et al.* (2012) in a similar catchment. As a result, they were able to obtain a good fit to their gauged data by relatively minor (± 10%) calibration of *Z*, *Kc* and water demand values (Boithias *et al.* 2014).

All of these studies using some form of validation, and the differences between them, support our suggestion that formal sensitivity analysis and, where empirical data are available, validation should be employed whenever the InVEST model is being used in new regions. Even a relatively small number of validation points from a range of locations can provide valuable insights into the accuracy of the model and the relative performance of different input datasets. If there are no validation data, our results suggest that datasets of the appropriate spatial scale (e.g. national rather than global) may perform better. The observed differences between our two pairs of input precipitation and PET datasets are probably due to several causes. The WorldClim and CGIAR-CSI data are annual averages calculated over the approximate period 1950-2010. The fact that they do not span the same date range as the validation data may explain some of their poor performance, although annual mean precipitation over England and Wales has not changed significantly over the date range (Jenkins, Perry & Prior 2008). Furthermore , WorldClim and CGIAR-CSI data are interpolated from data which are not spatially uniform in distribution, and thus vary spatially in the uncertainty around the given value of precipitation (Hijmans *et al.* 2005; Hamel & Guswa 2015; Pessacg *et al.* 2015). The UKCP09 data are also interpolated from a network of UK rain gauges, but at a much higher density of sampling points (Perry & Hollis 2005). Errors in the WorldClim precipitation data also tend to be highest in regions with high rainfall (Hijmans *et al.* 2005), such as the UK. Global-scale datasets like WorldClim are both widely used and readily available, so their relatively inconsistent performance across catchments, and the much better performance of the UKCP09 and MORECS data highlights the need for validation to select the most appropriate input data, or at least to assess model performance and the resultant confidence in the results if no other data are available. As might be expected, it appears that large improvements in model performance can be achieved simply by ensuring the input data are matched to the study region in terms of spatial and temporal scales.

The apparent trend for catchments to be consistently 'sensitive' or 'insensitive' to variation in the values of all model input parameters suggests that catchment characteristics (e.g. soil and bedrock characteristics) can strongly influence the degree to which errors in the input parameter values will affect the model outputs. Our results suggested that, in the UK, 'upland' type catchments, with low PET and high cover of semi-natural habitats are less sensitive than 'lowland' catchments with high PET and a higher cover or arable land (which has a high *Kc*). Pessacg *et al.* (2015) found that catchments with a higher cover of LULC classes with a high value of *Kc* were most sensitive, potentially giving a +150% change in modelled water yield in response to a +30% error in precipitation data, a response very similar to the most sensitive catchments in our results (see Fig. 2A). However, the good overall fit between modelled and measured data across catchments and the lack of significant correlation between model accuracy and catchment descriptors suggest that, when using UKCP09 and MORECS data, errors in the input datasets are comparatively minor, at least to the extent where they are not the major driver of differences between modelled and gauged water yield. The remaining model error is therefore likely to be due to limitations of the model or the validation data (see section 4.2) or more complex interplay between catchment characteristics. A productive area for further research could be more detailed investigation into the drivers of varying sensitivity between catchments, with the aim of using catchment descriptors as predictive variables in determining the impact of driving data on change in water yield.

4.2. Limitations of the model and validation data

Despite the good performance of the InVEST model when refined to account for water abstractions and using national input datasets, the accuracy of the modelled water yield values still varied to some extent between catchments (see Fig. 4C and 4D) and there was a slight but consistent overestimate of per hectare water yield. The InVEST water yield model contains several acknowledged limitations and simplifications (Sharp *et al.* 2015). These include the limited ability of the model to account for inter- or intra- annual variation in water supply. Many ecosystem services (irrigation, hydropower) and disservices (flooding) linked to water yield will be affected by the timing of water availability and peak flows, not just total annual yield. A further simplification is the lack of consideration of lateral and groundwater flows, such that effects of complex land use patterns or underlying geology remain unaccounted for (Sharp *et al.* 2015). Finally, the model handles consumptive water use in a very simplistic fashion, by allocating a per hectare value to each LULC. Although including per hectare estimates of consumptive abstraction did reduce overestimation of water yield and slightly improved model performance (Table 1), consumptive use is likely to vary widely between catchments and between different areas of the same LULC. In the UK (and in many other parts of the world), many large contributors to consumptive use are single point intakes. The use of reservoirs and water transfer schemes to regulate river flows for abstraction or flood prevention is common (Gibbins *et al.* 2001), can involve very large volumes of water (Davies, Thoms & Meador 1992; Boithias *et al.* 2014) and is indicated (but not quantified) in the NRFA catchment description metadata for several of the gauging stations used in this study (Fry & Swain 2010). Although the InVEST model structure does not directly account for point abstractions, where the locations of these are known, these can be represented as separate LULC classes, with corresponding consumptive water use values. Alternatively, the model outputs can be adjusted on a per catchment basis to account for known point source abstractions. However, such data can be hard to obtain due to regulatory restrictions in the UK water industry.

It is also worth noting that the empirical validation data themselves are also affected by issues of accuracy, many of which are not captured by the model (e.g. the InVEST water yield model does not distinguish between surface and sub-surface water flow). Measured river flows may be reduced by bypassing of the gauging station via flooding, canals or groundwater flow and either reduced or increased by catchment transfer, which may occur either consistently or only at times of particularly high or low flow. These factors are likely to result in measurements which accurately record the flow of water in the gauged channel, but not the true total water yield from the catchment of the gauging station. Catchments where a significant proportion of total water yield leaves via sub surface flow (or other routes) will show a considerable overestimate of total yield as gauged from stream flow. These issues are likely to affect many individual stations. For example, the severe model overestimation in the Welland catchment might be explained by the fact that it is comparatively small (707 km^2) and subject to high levels of abstraction to a reservoir. More seriously, gauging stations may be unable to record accurate readings of water flow over or under certain flow thresholds. Whilst at least one of these factors was present in the majority of catchments in this study (Fry & Swain 2010), these issues are unlikely to cause systematic bias because they are not consistent across catchments. For example, over half of the 42 catchments studied had factors affecting runoff documented by the NRFA which potentially offset one another (i.e. some factors likely to divert water flow from the river channel and others likely to increase it). The prevalence of these issues, along with the presence of outliers, does serve to illustrate the importance of incorporating local knowledge into decision making, alongside ecosystem service models and empirical validation, as stakeholders may often be able to provide information on processes not captured by the model which can help to explain or mitigate against poor model performance. Users of InVEST are strongly encouraged to involve stakeholders in scenario development and interpretation of model outputs (Sharp *et al.* 2015).

4.3. Conclusions

Ecosystem service models such as InVEST have the potential to provide a crucial underpinning to decision and policy making. However lack of robust testing limits their credibility. The work presented here demonstrates that the relatively InVEST simple water yield modelling framework can perform well as long as input data and parameters are representative of the spatial and temporal scale concerned. Care should be taken with application of these tools using indicative datasets at the global scale, and in the absence of more local scale data, empirical validation of model outputs becomes even more important. However, the need for ecosystem service models is driven by the fact that many parts of the world lack relevant empirical data (Crossman *et al.* 2013). Therefore, we firstly recommend that, where empirical data are available, models should be validated for locations in the region of interest and the effect of alternative parameter values or input data should be explored. Secondly, we recommend the application of sensitivity analyses to understand how model outputs vary across the region of interest, either in tandem with validation or, if validation data are not available, to understand uncertainty in model predictions. Finally, if no validation data are available, we advise exercising caution when interpreting model output values. For example, our results suggest that the InVEST water yield model could still be used to assess the rank order of catchments in terms of water yield or the direction of change in relation to scenarios of environmental change (e.g. Willcock *et al.* 2016) even where absolute values are less reliable.

Acknowledgements

Thanks to Matt Fry for advice on retrieval and analysis of NRFA data, Perrine Hamel for information and advice on the development and use of the InVEST water yield model and Rubab Bangash for assistance with researching suitable model parameters. This work was funded under National Capability funding from the Natural Environmental Research Council.

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. Food and Agriculture Organization of the United Nations, Rome, Italy.

Bai, Y., Zheng, H., Ouyang, Z., Zhuang, C. & Jiang, B. (2013) Modeling hydrological ecosystem services and tradeoffs: a case study in Baiyangdian watershed, China. *Environmental Earth Sciences,* **70,** 709-718.

Bangash, R.F., Passuello, A., Sanchez-Canales, M., Terrado, M., López, A., Elorza, F.J., Ziv, G., Acuña, V. & Schuhmacher, M. (2013) Ecosystem services in Mediterranean river basin: climate change impact on water provisioning and erosion control. *Science of the total environment,* **458,** 246-255.

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. *Science of the total environment,* **470–471,** 567-577.

Braat, L.C. & de Groot, R. (2012) The ecosystem services agenda:bridging the worlds of natural science and economics, conservation and development, and public and private policy. *Ecosystem Services,* **1,** 4-15.

British Hydropower Association (2010) England and Wales Hydropower Resource Assessment.

Brown, J.D. & Damery, S.L. (2002) Managing flood risk in the UK: towards an integration of social and technical perspectives. *Transactions of the Institute of British Geographers,* **27,** 412-426.

Budyko, M. (1974) Climate and Life. translated from Russian by Miller D H. *San Diego, CA: Academic*.

Canadell, J., Jackson, R., Ehleringer, J., Mooney, H., Sala, O. & Schulze, E.-D. (1996) Maximum rooting depth of vegetation types at the global scale. *Oecologia,* **108,** 583-595.

CEH (2008) *Flood estimation handbook*. Centre for Ecology and Hydrology, Wallingford.

Cheaib, A., Badeau, V., Boe, J., Chuine, I., Delire, C., Dufrêne, E., François, C., Gritti, E.S., Legay, M. & Pagé, C. (2012) Climate change impacts on tree ranges: model intercomparison facilitates understanding and quantification of uncertainty. *Ecology letters,* **15,** 533-544.

Crossman, N.D., Burkhard, B., Nedkov, S., Willemen, L., Petz, K., Palomo, I., Drakou, E.G., Martín-Lopez, B., McPhearson, T., Boyanova, K., Alkemade, R., Egoh, B., Dunbar, M.B. & Maes, J. (2013) A blueprint for mapping and modelling ecosystem services. *Ecosystem Services,* **4,** 4-14.

Davies, B.R., Thoms, M. & Meador, M. (1992) An assessment of the ecological impacts of inter-basin water transfers, and their threats to river basin integrity and conservation. *Aquatic Conservation: Marine and Freshwater Ecosystems,* **2,** 325-349.

DECC (2015) Environmental management – guidance. Harnessing hydroelectric power.

DEFRA (2015) Estimated abstractions from all sources except tidal by purpose and Environment Agency region: 2000 - 2013. (ed. DEFRA). London.

Dennedy-Frank, P.J., Muenich, R.L., Chaubey, I. & Ziv, G. (2016) Comparing two tools for ecosystem service assessments regarding water resources decisions. *Journal of Environmental Management,* **177,** 331-340.

Döll, P. (2002) Impact of Climate Change and Variability on Irrigation Requirements: A Global Perspective. *Climatic Change,* **54,** 269-293.

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

European Commission (2011) Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions: Our life insurance, our natural capital: an EU biodiversity strategy to 2020. Brussels.

Evans, J., McNeil, D., Finch, J., Murray, T., Harding, R., Ward, H. & Verhoef, A. (2012) Determination of turbulent heat fluxes using a large aperture scintillometer over undulating mixed agricultural terrain. *Agricultural and forest meteorology,* **166,** 221-233.

Fisher, B., Turner, R.K. & Morling, P. (2009) Defining and classifying ecosystem services for decision making. *Ecological economics,* **68,** 643-653.

Fry, M.J. & Swain, O. (2010) Hydrological data management systems within a national river flow archive. *Role of Hydrology in Managing Consequences of a Changing Global Environment.* (ed. C. Kirby), pp. 808-815. British Hydrological Society.

Fu, B.P. (1981) On the calculation of the evaporation from land surface (in Chinese). *Sci. Atmos. Sin.,* **5,** 23-31.

Gibbins, C.N., Soulsby, C., Jeffries, M.J. & Acornley, R. (2001) Developing ecologically acceptable river flow regimes: a case study of Kielder reservoir and the Kielder water transfer system. *Fisheries Management and Ecology,* **8,** 463-485.

Hamel, P. & Guswa, A.J. (2015) Uncertainty analysis of a spatially explicit annual water-balance model: case study of the Cape Fear basin, North Carolina. *Hydrological Earth System Science,* **19,** 839-853.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology,* **25,** 1965-1978.

Hou, Y., Burkhard, B. & Müller, F. (2013) Uncertainties in landscape analysis and ecosystem service assessment. *Journal of Environmental Management,* **127,** S117-S131.

Hulme, P., Rushton, K. & Fletcher, S. (2001) Estimating recharge in UK catchments. *IAHS PUBLICATION***,** 33-42.

Jenkins, G.J., Perry, M.C. & Prior, M.J. (2008) The climate of the United Kingdom and recent trends. Hadley Centre, Exeter, UK.

Knox, J., Weatherhead, K., Díaz, J.R. & Kay, M. (2009) Developing a strategy to improve irrigation efficiency in a temperate climate: a case study in England. *Outlook on AGRICULTURE,* **38,** 303-309.

Leh, M.D.K., Matlock, M.D., Cummings, E.C. & Nalley, L.L. (2013) Quantifying and mapping multiple ecosystem services change in West Africa. *Agriculture, Ecosystems & Environment,* **165,** 6-18.

Maes, J., Egoh, B., Willemen, L., Liquete, C., Vihervaara, P., Schägner, J.P., Grizzetti, B., Drakou, E.G., Notte, A.L., Zulian, G., Bouraoui, F., Luisa Paracchini, M., Braat, L. & Bidoglio, G. (2012) Mapping ecosystem services for policy support and decision making in the European Union. *Ecosystem Services,* **1,** 31-39.

Malinga, R., Gordon, L.J., Jewitt, G. & Lindborg, R. (2015) Mapping ecosystem services across scales and continents – A review. *Ecosystem Services,* **13,** 57-63.

Martínez-Harms, M.J. & Balvanera, P. (2012) Methods for mapping ecosystem service supply: a review. *International Journal of Biodiversity Science, Ecosystem Services & Management,* **8,** 17-25.

Mendoza, G., Ennaanay, D., Conte, M., Walter, M., Freyberg, D., Wolny, S., Hay, L., White, S., Nelson, E. & Solorzano, L. (2011) Water supply as an ecosystem service for hydropower and irrigation. *Natural Capital: Theory and Practice of Mapping Ecosystem Services***,** 53-72.

Morris, D.G. & Flavin, R.W. (1990) A Digital Terrain Model for Hydrology. *Proc 4th Int. Symposium on Spatial Data Handling*, pp. 250-262. Zurich.

Morton, D., Rowland, C., Wood, C., Meek, L., Marston, C., Smith, G. & Simpson, I.C. (2011) Final report for LCM2007 - the new UK land cover map. . pp. 112pp. NERC/Centre for Ecology and Hydrology.

Panagos, P. (2006) The European soil database. *GEO: connexion,* **5,** 32-33.

Panagos, P., Van Liedekerke, M., Jones, A. & Montanarella, L. (2012) European Soil Data Centre: Response to European policy support and public data requirements. *Land Use Policy,* **29,** 329-338.

Perry, M. & Hollis, D. (2005) The generation of monthly gridded datasets for a range of climatic variables over the UK. *International Journal of Climatology,* **25,** 1041-1054.

Pessacg, N., Flaherty, S., Brandizi, L., Solman, S. & Pascual, M. (2015) Getting water right: A case study in water yield modelling based on precipitation data. *Science of the total environment,* **537,** 225-234.

R Core Team (2014) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Müller, C., Arneth, A., Boote, K.J., Folberth, C., Glotter, M. & Khabarov, N. (2014) Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proceedings of the National Academy of Sciences,* **111,** 3268-3273.

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. *Science of the total environment,* **440,** 140-153.

Schulp, C.J.E., Burkhard, B., Maes, J., Van Vliet, J. & Verburg, P.H. (2014) Uncertainties in Ecosystem Service Maps: A Comparison on the European Scale. *PLoS ONE,* **9,** e109643.

Seppelt, R., Dormann, C.F., Eppink, F.V., Lautenbach, S. & Schmidt, S. (2011) A quantitative review of ecosystem service studies: approaches, shortcomings and the road ahead. *Journal of Applied Ecology,* **48,** 630-636.

Sharp, R., Tallis, H.T., Ricketts, T., Guerry, A.D., Wood, S.A., Chaplin-Kramer, R., 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., Vogl, A.L., Rogers, L. & Bierbower, W. (2015) *InVEST 3.2.0 User's Guide.* The Natural Capital Project, Stanford.

Smethurst, J.A., Clarke, D. & Powrie, W. (2012) Factors controlling the seasonal variation in soil water content and pore water pressures within a lightly vegetated clay slope. *Géotechnique,* **62,** 429-446.

Tallis, H., Kareiva, P., Marvier, M. & Chang, A. (2008) An ecosystem services framework to support both practical conservation and economic development. *Proceedings of the National Academy of Sciences,* **105,** 9457-9464.

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. *Ecological Indicators,* **37,** 199-209.

Vigerstol, K.L. & Aukema, J.E. (2011) A comparison of tools for modeling freshwater ecosystem services. *Journal of Environmental Management,* **92,** 2403-2409.

Watts, G., Battarbee, R.W., Bloomfield, J.P., Crossman, J., Daccache, A., Durance, I., Elliott, J.A., Garner, G., Hannaford, J. & Hannah, D.M. (2015) Climate change and water in the UK–past changes and future prospects. *Progress in Physical Geography,* **39,** 6-28.

Weatherhead, E.K. & Knox, J.W. (2000) Predicting and mapping the future demand for irrigation water in England and Wales. *Agricultural Water Management,* **43,** 203-218.

Willcock, S., Hooftman, D., Sitas, N., O'Farrell, P., Hudson, M.D., Reyers, B., Eigenbrod, F. & Bullock, J.M. (2016) Do ecosystem service maps and models meet stakeholders' needs? A preliminary survey across sub-Saharan Africa. *Ecosystem Services,* **18,** 110-117.

Xiao, Y., Xiao, Q., Ouyang, Z. & Maomao, Q. (2015) Assessing changes in water flow regulation in Chongqing region, China. *Environmental Monitoring and Assessment,* **187,** 1-13.

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 Resources Research,* **40,** W02502.

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. *Journal of Hydrology,* **470–471,** 269-279.

Zomer, R.J., Trabucco, A., Bossio, D.A., van Straaten, O. & Verchot, L.V. (2008) Climate change mitigation through afforestation/reforestation: A global analysis of hydrologic impacts with four case studies. *Agriculture Ecosystems & Environment,* **126,** 81-97.

Zomer, R.J., Trabucco, A., van Straaten, O. & Bossio, D.A. (2007) Carbon, land and water: A global analysis of the hydrologic dimensions of climate change mitigation through afforestation/reforestation. International Water Management Institute, Colombo, Sri Lanka:.