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

Article

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1016/j.scitotenv.2016.06.227

Publisher: Elsevier

www.reading.ac.uk/centaur

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

3   Redhead, J.W.[a,e]
4   Stratford, C.[a]
5   Sharps, K.[b]
6   Jones, L.[b]
7   Ziv, G. [c]
8   Clarke, D. [d]
9   Oliver, T.H.[a,1]
10   Bullock, J.M.[a]
11
12   [a] NERC Centre for Ecology and Hydrology, Maclean Building, Wallingford, Oxfordshire, OX10 8BB, UK.

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

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

15   [d] Faculty of Engineering and Environment, University of Southampton, University Road, Highfield,

16   Southampton, SO17 1BJ, UK

17   [e] Corresponding author: Email: johdhe@ceh.ac.uk, Tel: (+44) 1491 692538, Fax: (+44) 1491 692424

18   [1] Present address: School of Biological Sciences, Harborne Building, University of Reading, Reading, Berkshire

19   RG6 6AS, UK

20

21   **Abstract**

22   A variety of tools have emerged with the goal of mapping the current delivery of ecosystem services

23   and quantifying the impact of environmental changes. An important and often overlooked question

24   is how accurate the outputs of these models are in relation to empirical observations. In this paper

25   we validate a hydrological ecosystem service model (InVEST Water Yield Model) using widely

26   available data. We modelled annual water yield in 22 UK catchments with widely varying land cover,

27   population and geology, and compared model outputs with gauged river flow data from the UK

28   National River Flow Archive. Values for input parameters were selected from existing literature to

29   reflect conditions in the UK and were subjected to sensitivity analyses. We also compared model

30   performance between precipitation and potential evapotranspiration data sourced from global- and

31   UK-scale datasets. We then tested the transferability of the results within the UK by additional

32   validation in a further 20 catchments.

33   Whilst the model performed only moderately with global-scale data (linear regression of modelled

34   total water yield against empirical data; slope = 0.763, intercept = 54.45, R^2 = 0.963) with wide

35   variation in performance between catchments, the model performed much better when using UK-

36   scale input data, with closer fit to the observed data (slope = 1.07, intercept = 3.07, R^2 = 0.990). With

37   UK data the majority of catchments showed less than 10% difference between measured and

38   modelled water yield but there was a minor but consistent overestimate per hectare (86

39   m^3/ha/year). Additional validation on a further 20 UK catchments was similarly robust, indicating

40   that these results are transferable within the UK.   These results suggest that relatively simple

41   models can give accurate measures of ecosystem services. However, the choice of input data is

42   critical and there is a need for further validation in other parts of the world.

43   **Keywords**

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

45

46   **1. Introduction**

47   Ecosystem services are increasingly used to assess likely impacts of environmental change in societal

48   and economic terms and to provide a rationale for conservation or environmental management

49   (Tallis *et al.* 2008; Braat & de Groot 2012). However, to incorporate the ecosystem services concept

50   into assessments and decision making, there is a requirement for accurate mapping and

51   measurement of ecosystem services (Malinga *et al.* 2015). In some cases, this requirement has

52   itself been incorporated into policy (European Commission 2011).

53   To meet this rising demand there has been a proliferation of methods and tools to map, quantify

54   and value the provision of ecosystem services (Fisher, Turner & Morling 2009; Seppelt *et al.* 2011;

55   Malinga *et al.* 2015). These vary in complexity from simple approaches based on maps and land use

56   or habitat-based proxies to complex, process-based models (Seppelt *et al.* 2011). Ecosystem service

57   tools have been designed and applied at widely varying geographic locations and both spatial and

58   temporal scales. Potential users must thus choose which tools are most appropriate for their

59   particular situation, and be aware of the limitations of these tools (Willcock *et al.* 2016). Recent

60   reviews have identified that one of the key obstacles to successful ecosystem service mapping and

61   implementation into decision making processes is the comparative scarcity of validation or

62   measurements of uncertainty in many applications of ecosystem service models (Seppelt *et al.* 2011;

63   Maes *et al.* 2012; Schulp *et al.* 2014; Malinga *et al.* 2015). Whilst it is frequently acknowledged that

64   ecosystem service models function at best as reliable proxies, and at worst as crude estimates, the

65   validation of the results of ecosystem service models against empirical measurements is

66   comparatively rare (Seppelt *et al.* 2011; Vigerstol & Aukema 2011; Schulp *et al.* 2014; Hamel &

67   Guswa 2015). Of those studies which do employ validation, many do so at a limited number of

68   locations to check the performance of a model within their study region (e.g. Bai *et al.* 2013; Boithias

69   *et al.* 2014; Terrado *et al.* 2014; Xiao *et al.* 2015). Whilst this is entirely sensible, the results of such

70   local-scale validation are less likely to be transferrable to new locations and the regional or national

71   scales at which ecosystem service models are most widely used (Martínez-Harms & Balvanera 2012)

72   and most water resource planning takes place (Watts *et al.* 2015). Several studies have compared

73   different ecosystem service models (e.g. Vigerstol & Aukema 2011; Cheaib *et al.* 2012; Rosenzweig

74   *et al.* 2014; Dennedy-Frank *et al.* 2016), which gives some insight into the *uncertainty* surrounding

75   the modelling of the service in question (see Hou, Burkhard and Müller (2013) concerning

76   uncertainty in ecosystem service modelling) and the utility of the different models, but does not

77   provide insight into the *accuracy* of each model in estimating ecosystem service delivery or

78   representing the biophysical process underpinning the service.

79   This relative scarcity of large-scale validation means that, for many models, there is comparatively

80   little information on either the accuracy of model outputs (Seppelt *et al.* 2011), or on the

81   performance of models in different circumstances and locations, especially where the latter are in

82   poorly-studied regions. There is also a lack of information on the requirements of the input data. In

83   many cases, the availability and spatial coverage of data is inversely correlated with its resolution

84   (Hijmans *et al.* 2005) and, potentially, its accuracy. Thus it is uncertain whether the most widely

85   available data, even when used in a model which performs well under ideal circumstances, will

86   produce sufficiently accurate results.   Potential users are thus missing vital information on the

87 performance of models, which they need if they are to make informed decisions on which tools to

88 use and how best to employ them to provide accurate assessments for decision makers (Willcock *et*

89 *al.* 2016). Validation also provides valuable feedback to ecosystem service model developers who

90 are seeking to improve the accuracy, utility and efficiency of their models.

91 Hydrological services are particularly well suited to empirical validation, as the ecosystem processes

92 which underpin them (e.g. runoff of water, nutrients and sediment) have physical expressions which

93 can be directly measured at appropriate spatial and temporal resolutions (river flow, nutrient

94 concentration and sediment load, respectively). In the UK, these measurements are undertaken by

95 government bodies and are readily available for academic purposes (e.g. The National River Flow

96 Archive, NRFA).

97 This study aims to validate a hydrological ecosystem service tool at the national scale, using widely

98 available spatial data (of the sort available to most potential users and decision makers) for both

99 model inputs and validation. We used a tool from a widely used, open-source ecosystem service

100 modelling suite, InVEST (Integrated Valuation of Ecosystem Services and Tradeoffs, Sharp *et al.*

101 2015). Whilst InVEST tools have been widely used for a variety of research and planning applications

102 (e.g. British Hydropower Association 2010; Bai *et al.* 2013; Bangash *et al.* 2013; Leh *et al.* 2013;

103 Boithias *et al.* 2014; Terrado *et al.* 2014; Pessacg *et al.* 2015; Xiao *et al.* 2015), as with other

104 ecosystem service models, comparatively few applications have employed an empirical validation of

105 results at anything other than a local scale. Therefore our objectives are 1) to examine the

106 sensitivity of the model to variation in the values of input parameters in a UK context; 2) to compare

107 the performance of the model using two points on the spectrum of data availability and spatial

108 coverage (global climatic data and UK specific climate data) by validation against empirical

109 measurements; 3) to examine whether our results are transferable within the UK.

110 2. **Methods**

111 2.1. THE INVEST WATER YIELD MODEL

112 The InVEST suite of tools has been developed to enable decision makers to assess trade-offs among

113 ecosystem services and to compare scenarios of change, for example in land use or climate (Sharp *et*

114 *al.* 2015). To this end, InVEST comprises a set of models covering a wide variety of ecosystem

115 services. The models are based on comparatively simple production functions, being intended to

116 run quickly on a standard desktop computer and to take advantage of readily available data (Sharp

117 *et al.* 2015). Although InVEST models are not designed to reproduce empirical observations, the

118 water yield model is intended to quantify the relative yields of different catchments or sub-

119 catchments, and be sensitive to modelled changes in drivers such as land use change or climate

4

120 change. We would also suggest that, because the model produces figures of water yield which

121 appear to have a high degree of numerical precision, and is freely available, it is important to test

122 whether the results are accurate, as users may not always familiarise themselves with the intended

123 use and limitations of the model before incorporating the results into the decision making process

124 (see Willcock *et al.* 2016).

125 The InVEST water yield model (Hydropower/Water Yield, InVEST v3.2.0, Sharp *et al.* 2015) calculates

126 annual water yield from a catchment, with the intended end use of reservoir hydropower

127 production (Sharp *et al.* 2015). Although hydropower forms a relatively small contribution to the UK

128 energy sector (DECC 2015), total annual water yield can be considered in the light of many potential

129 services, including agricultural irrigation, provision of drinking water, hydropower and industrial

130 abstraction. The UK is densely populated and has a large proportion of its land area under

131 anthropogenic land uses. This leads to competition between demands for water, which is likely to

132 intensify in the future due to population growth and climate change (Weatherhead & Knox 2000;

133 Knox *et al.* 2009). Validated models of current and predicted future water yield, with clear estimates

134 of their accuracy and uncertainty, are thus of great importance in strategic water resource planning

135 (Watts *et al.* 2015). Therefore, in this study we focused on the biophysical output of water yield. As

136 the InVEST model is compartmentalised into water yield, water consumption and hydropower

137 valuation, we used the first two components only.

138 The model estimates the total annual water yield (*Y*) for each grid square (*x*) of the study catchment

139 as total catchment annual rainfall (*P*) minus total catchment annual actual evapotranspiration (*AET*)

140 (equation 1). The model assumes that, on an annual time step, all water falling as rainfall over a

141 catchment, minus that which is evapotranspired, leaves the catchment. No distinction is made

142 between surface and sub-surface water flow.

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

143 In practice, the measurement of annual actual evapotranspiration at the catchment scale is

144 extremely difficult. Even plot scale evaluation requires highly specialised equipment, and plot and

145 field scale methods to determine actual evapotranspiration are problematic to apply at the

146 landscape scale (Evans *et al.* 2012). The InVEST approach relates AET to potential evapotranspiration

147 (PET), which is easier to model, using the methodology developed by Budyko (1974) and later

148 adapted by Fu (1981) and Zhang *et al.* (2004) (equation 2) where ω is an empirical parameter which

149   defines the shape of the curve relating potential to actual evapotranspiration.

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

150   PET is estimated as the product of the reference evapotranspiration and the crop coefficient for

151   each grid square. ω is related to the plant available water content (AWC), precipitation and the

152   constant *Z* which captures the local precipitation pattern and additional hydrogeological

153   characteristics (equation 3) (Sharp *et al.* 2015).

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

154   For a more detailed description of the water yield model see Sánchez-Canales *et al.* (2012); Bangash

155   *et al.* (2013); Hamel and Guswa (2015); Pessacg *et al.* (2015); and Sharp *et al.* (2015).

156   2.2. MODEL INPUT PARAMETERS

157   The InVEST model requires five biophysical parameters as georeferenced rasters. These are root

158   restricting layer depth (mm), plant available water content (AWC, as a proportion), average annual

159   precipitation (mm), average annual potential evapotranspiration (PET, mm) and land use/land cover

160   (LULC). We obtained these data from a variety of sources, with the aim of ensuring that the data

161   were easily obtainable and free to license for at least academic use. These are the kind of data

162   which are likely to be most widely used in a freely available tool such as InVEST, in terms of

163   precision, spatial resolution and spatial coverage.

164   Root restricting layer depth and AWC were obtained from the European Soil Database (ESDB)

165   version 2.0 (Panagos 2006; Panagos *et al.* 2012). Annual precipitation and reference

166   evapotranspiration were obtained from several alternative sources. We used two pairings of these

167   two variables, to compare model performance with data from two points on the spectrum of data

168   availability and spatial coverage. First, we used global scale precipitation data from WorldClim

169   (Hijmans *et al.* 2005) and PET from the CGIAR-CSI Global-Aridity and Global-PET Database (Zomer *et*

170   *al.* 2007; Zomer *et al.* 2008). These are both freely available and have global coverage, at

171   approximately 1km resolution. Secondly, we used UK Met Office UKCP09 precipitation data at 5km

172   resolution (Perry & Hollis 2005; Jenkins, Perry & Prior 2008) and Met Office Rainfall and Evaporation

173   Calculation System (MORECS) evapotranspiration data. These datasets are UK-specific and available

174   to a wide variety of users under the UK's open government license. Where necessary, data were

175   geoprocessed to meet the data formatting requirements of the InVEST model in ArcMap (v10.1,

176   ESRI, Redlands, CA).

177 Where possible, all input data were limited to the same date range as the validation data (2000-

178 2010, see below) and averaged across years, giving average annual precipitation and average annual

179 PET. LULC data were obtained from the 25 m raster version of the UK Land Cover Map 2007

180 (LCM2007, Morton *et al.* 2011).

181 The InVEST model also requires several tabular values for each LULC class. These include whether

182 the land cover class is vegetated or not, rooting depth and a plant evapotranspiration coefficient

183 (*Kc*). This last is used to obtain potential evapotranspiration by modifying the reference

184 evapotranspiration, which is based on a 15cm tall surface of actively growing, well-watered grass.

185 We estimated these coefficients for LCM2007 broad habitat classes by matching class descriptions

186 with those in Canadell *et al.* (1996), Allen *et al.* (1998) and Sharp *et al.* (2015).   Further amendments

187 were made to these values, to reflect the damp climate of the UK (Smethurst, Clarke & Powrie

188 2012). This generally resulted in raised crop coefficients and shallower rooting depths. The

189 coefficients for urban and suburban areas were amended to reflect the approximate proportion of

190 green space they typically contain (20% and 60% respectively). Coefficients for *Kc* and rooting depth

191 for each LCM2007 broad habitat are given in Supplementary Material, Table S1. For arable land uses,

192 actual evapotranspiration varies over the course of a year as crops are sown, grow and are harvested

193 before the land is then re-cultivated. Evapotranspiration of growing crops also varies between crop

194 plant species, crop condition and many other factors (Allen *et al.* 1998; Hulme, Rushton & Fletcher

195 2001). For UK crops and soil conditions, preliminary investigation and previous studies (Allen *et al.*

196 1998; Hulme, Rushton & Fletcher 2001) suggested a value of *Kc* close to one to best represent

197 annual evapotranspiration from arable land.

198 The seasonality constant (*Z*) was estimated as 0.2*N, where N is the average number of rain days (>

199 1mm) per year over the study period (Donohue, Roderick & McVicar 2012; Hamel & Guswa 2015). N

200 was estimated at approximately 150 from UK Met Office data

201 (http://www.metoffice.gov.uk/climate/uk/datasets/), giving a value of 30 for *Z*.

202 Because the validation data (i.e. gauged annual yield, see below) are affected by any consumptive

203 water use, it was important to account for this. The InVEST model uses a comparatively crude

204 method of estimating consumptive water use, by assigning a value of annual consumption per

205 hectare to each land cover class (Sharp *et al.* 2015). Because abstraction varies widely across the UK

206 (DEFRA 2015), we split the LULC raster based on administrative regions, such that each LULC class-

207 region combination had a unique value, allowing us to assign a suitable abstraction value from

208 regional abstraction statistics (DEFRA 2015). We used only the values for abstraction for agricultural

209 purposes (assigned to the arable LULC class) and public and industrial water supply (assigned to the

210 urban/suburban LULC class), as most other uses (e.g. hydropower) do not consume water but return
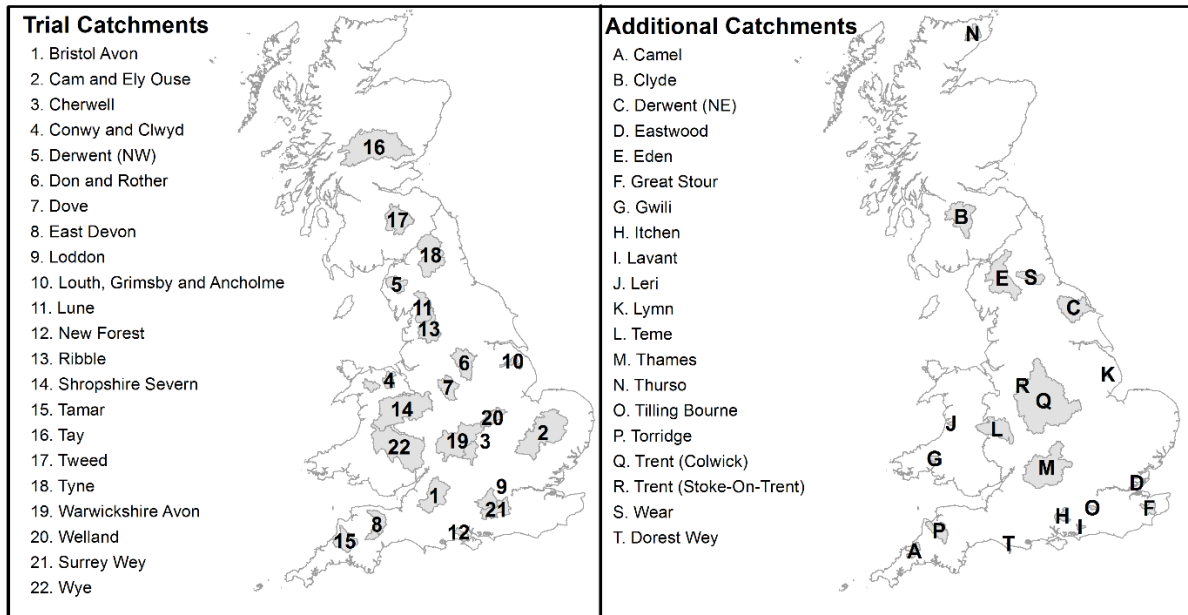
211 it to the catchment after use (Terrado *et al.* 2014). Public and industrial water supply may also

212 return water after use, but in many cases it may be returned further downstream from the point at

213 which it was abstracted, or to a different catchment.

214   2.3. SENSITIVITY ANALYSIS

215 We investigated the sensitivity of the model to variations in precipitation, PET, rooting depth, AWC,

216 *Kc* and *Z* following Sánchez-Canales *et al.* (2012) and Hamel and Guswa (2015). The biophysical

217 parameters, which are input as rasters, were varied by ± 10% and ± 20% applied uniformly across the

218 raster. Sensitivity to *Kc* was examined by varying its value for the two dominant LULC classes across

219 all catchments (arable and improved grassland), by the same proportions as the biophysical

220 parameters. Sensitivity to *Z* was tested using values of zero, one, two, five and further increments of

221 five up to 50. The model was run independently for each of these variations.

222   2.4. VALIDATION DATA

223 For initial validation of the model and comparison of the two sets of climate datasets we selected 22

224 catchments in England, Scotland and Wales with widely varying land cover, rainfall, elevation, and

225 geology (Fig. 1 and Supplementary Material, Table S2). All of these factors are likely to affect actual

226 water yield and, potentially, the performance of the InVEST water yield model. Empirical

227 measurements of gauged daily water flow were obtained from the National River Flow Archive

228 (NRFA), which collates, quality controls, archives and disseminates hydrometric data from gauging

229 station networks operated by government environmental bodies across the UK (Fry & Swain 2010).

230 The catchments for each NRFA gauging station have been defined using the Centre for Ecology &

231 Hydrology's Integrated Hydrological Digital Terrain Model (Morris & Flavin 1990). We calculated

232 total gauged annual water yield for each catchment by summing gauged daily mean flow for each

233 year from 2000 – 2010 and took the mean value across years. We analysed how well the modelled

234 data predicted the empirical data using linear regression. This was done using total annual yield, but

235 also the per hectare yield, i.e. the total yield divided by the catchment area, to remove trivial

236 correlation caused by the large variation in area among the catchments. These calculations (and all

237 subsequent data manipulation and statistics) were performed in R (v3.1.0, R Core Team 2014).

Trial Catchments
1. Bristol Avon
2. Cam and Ely Ouse
3. Cherwell
4. Conwy and Clwyd
5. Derwent (NW)
6. Don and Rother
7. Dove
8. East Devon
9. Loddon
10. Louth, Grimsby and Ancholme
11. Lune
12. New Forest
13. Ribble
14. Shropshire Severn
15. Tamar
16. Tay
17. Tweed
18. Tyne
19. Warwickshire Avon
20. Welland
21. Surrey Wey
22. Wye

Additional Catchments
A. Camel
B. Clyde
C. Derwent (NE)
D. Eastwood
E. Eden
F. Great Stour
G. Gwili
H. Itchen
I. Lavant
J. Leri
K. Lymn
L. Teme
M. Thames
N. Thurso
O. Tilling Bourne
P. Torridge
Q. Trent (Colwick)
R. Trent (Stoke-On-Trent)
S. Wear
T. Dorest Wey

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

259   *Kc* was roughly similar to that for PET, which is unsurprising since the effect of the former in the

260   model is to modify the latter, and was likewise catchment specific. In general, catchments were

261   either 'highly sensitive', responding to variation in values of all input parameters with variation in

262   water yield, or 'less sensitive', showing comparatively little variation in water yield with any variation

263   in the values of model input parameters, although the latter still responded to percentage changes

264   in precipitation with at least a corresponding percentage change in modelled yield.
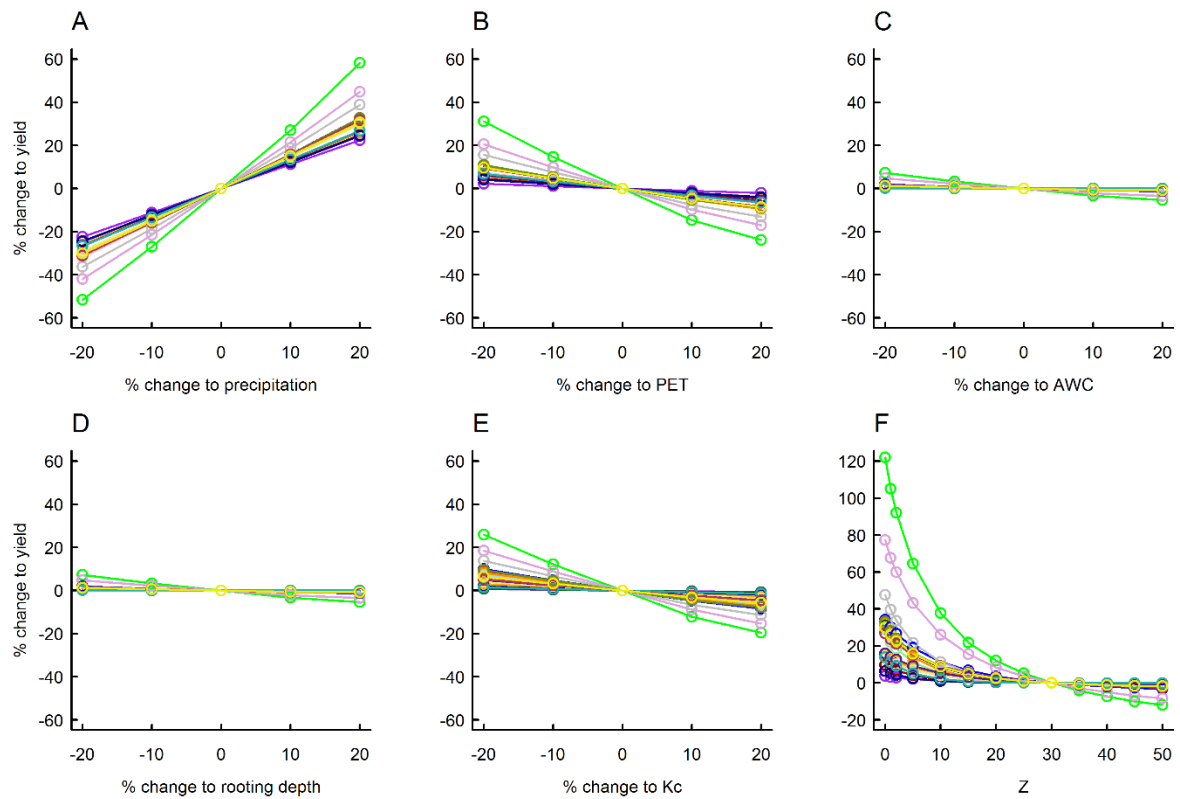
265

266   Fig. 2 Sensitivity of the InVEST water yield model to variation in the values of input parameters for 22

267   test catchments. Variation shown as percentage change relative to a 'baseline' run with *Z* = 30 using

268   UKCP09/MORECS data. A coloured set of points with connecting line are shown for each catchment.

269   Catchment sensitivity to variation in precipitation (and thus in other model parameter values) was

270   significantly positively correlated with mean PET (Pearson's $r = 0.555$, $p = 0.011$) and area of arable

271   land (Pearson's $r = 0.759$, $p < 0.001$), and significantly negatively correlated with altitude (Pearson's r

272    $= -0.676$, $p = 0.001$) and area of semi-natural habitat (Pearson's $r = -0.577$, $p < 0.008$). All of these

273   catchment characteristics are significantly inter-correlated, such that catchments with higher mean

274   altitude have a lower cover of arable land and a correspondingly higher cover of semi-natural habitat

275   and a lower mean PET.

276   The sensitivity of the model to changes in the value of *Z* was also strongly catchment specific (Fig.

277   2F), as expected given the spatial variation in the biophysical variables which modulate the effect of

278   *Z* on water yield (Hamel & Guswa 2015). Because it is difficult to translate the sensitivity of the

279   model to *Z* into an appropriate value of *Z* to use, the outputs from models with varying values of *Z*

280   were compared to the validation data to identify which value of *Z* resulted in the best fit to the

281   validation data. The results of this analysis (Fig. 3) showed that model fit (R^2) levelled off at *Z* ≈30

282   (Fig. 3A), as did slope (Fig. 3B), whilst overestimation of per hectare water yield was also much

283   reduced at values above 30 (Fig. 3C). This supports the value of *Z* = 30 for the model runs detailed

284   below and hence the estimation of *Z* from mean annual rain days.

285
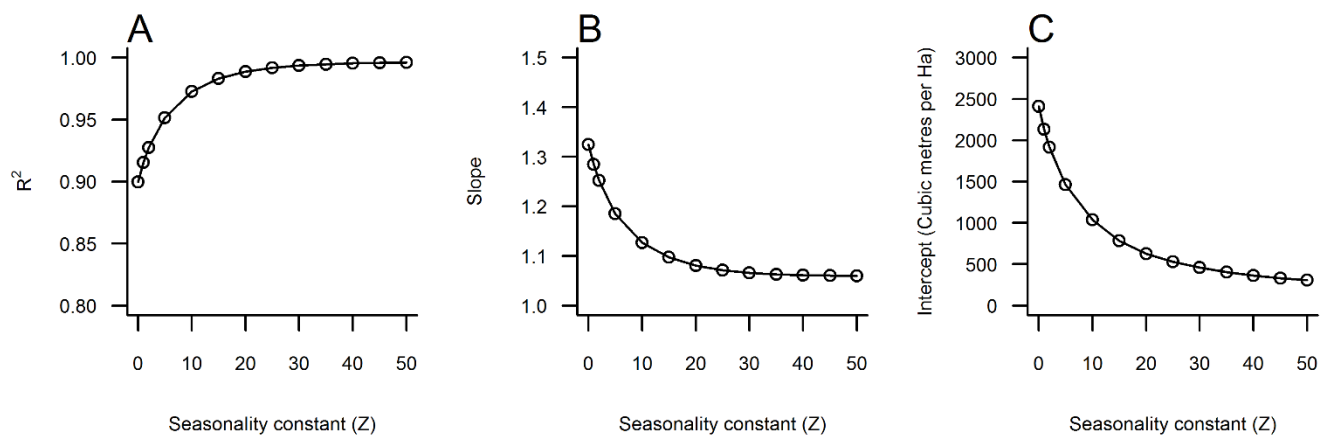
286

287   Fig. 3 Effects of varying the value of the seasonality constant (*Z*) on the relationship between

288   modelled and gauged water yield for 22 test catchments, using the UKCP09/MORECS data. A) R^2 of

289   linear regression between modelled and gauged catchment yield; B) Slope of linear regression

290   between modelled and gauged catchment yield; C) Intercept of linear regression between modelled

291   and gauged yield per hectare.

292   3.2. MODEL VALIDATION AND COMPARISON OF CLIMATIC DATASETS

293   Both global- and UK-scale climate datasets resulted in estimated water yields which were strongly

294   correlated with empirical yields obtained from NRFA gauged river flow (Fig 4, Table 1). The

295   WordClim and CGIAR-CSI data performed less well than the UKCP09/MORECS datasets. Although R^2

296   values for models using the global input data were only slightly lower (e.g. 0.96 compared with 0.99;

297   Table 1), the slope values for per hectare yield (including confidence intervals) were less than one

298   (Table 1). Hence the global data led to considerable under-estimates (up to 45%) of water yield for

299   catchments where the yield per hectare was high and to overestimates of water yield for those

300   where it was low (Fig. 4B), leading to the intercept of 1443.63 m^3 per hectare per year. By contrast,

301   the UKCP09/MORECS data led to more consistent and accurate estimates for total water yield when

302   adjusted for consumptive abstraction (Table 1). When per hectare yield was considered, the

303   UKCP09/MORECS data gave good fits to the NRFA data (R^2 = 0.949), with a slope not significantly

304   different from one and the intercept indicating a consistent but minor overestimate of 86.3 m^3 per

305   hectare per year when adjusted for consumptive abstraction (Fig 4D).

306   Table 1. Results of linear regressions between InVEST modelled and empirical water yield for the 22

307   original catchments, using two input datasets for the precipitation and reference evapotranspiration

308   parameters. Intercept, slope ± 95% confidence interval and R^2 are given for total catchment yield in

309   millions of cubic metres and yield per hectare in cubic metres, for both raw water yield (R) and yield

310   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

311   The mean percentage differences between gauged and modelled water yield were ± 23.36% (SE ±

312   4.40) for the WordClim/CGIAR-CSI data and ± 18.55% (SE ± 4.94) for the UKCP09/MORECS data.

313   However, in both cases one catchment (Welland, labelled 20 on Fig.1) showed a percentage

314   difference of over 100%. Although the difference between the mean percentage

315   under/overestimates of the two datasets does not appear great, it is important to note that the

316   mean is somewhat skewed by the few catchments for which the model performs particularly poorly,

317   especially for the UKCP09/MORECS data. Median values show that for the UKCP09/MORECS data

318   the majority of catchments had percentage differences between gauged and modelled water yield of

319   less than 10% (median = 9.74%) whilst for the WordClim/CGIAR-CSI data more catchments vary by

320   up to 20% (median = 17.19%).

321   Despite the significant correlations between catchment sensitivity to variation in the input

322   parameter values and catchment characteristics, percentage under/overestimates of total water

323   yield using the UKCP09/MORECS data did not show any significant correlations with catchment area,

324   altitude, mean precipitation, mean PET, geology (i.e. base flow index) or land cover.
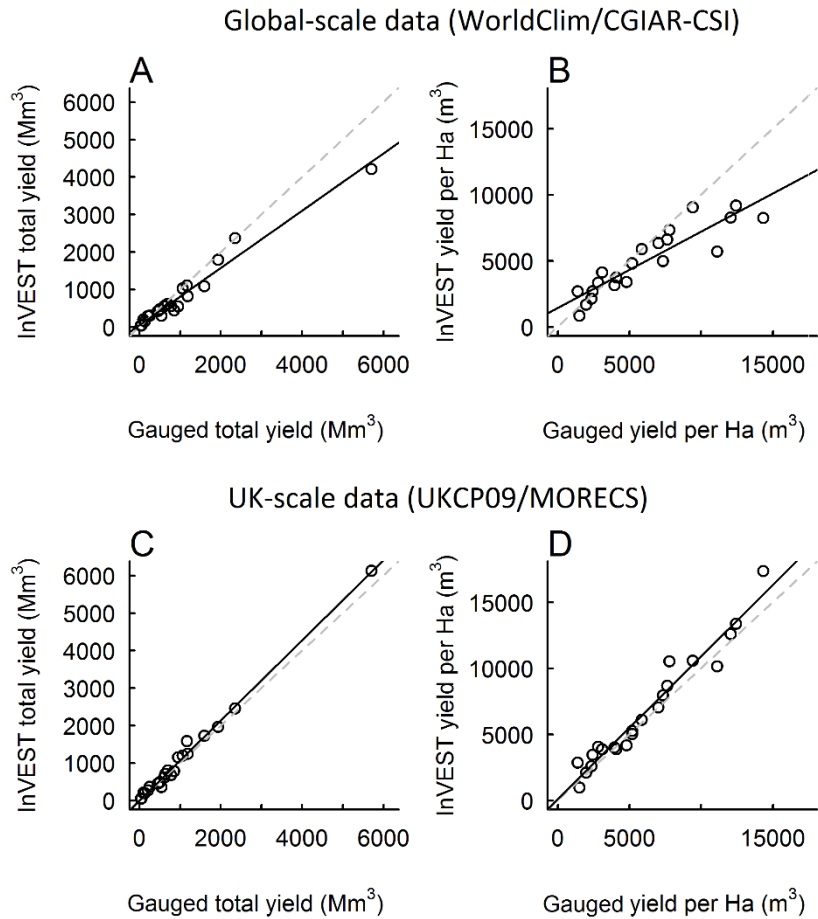
Global-scale data (WorldClim/CGIAR-CSI)

UK-scale data (UKCP09/MORECS)

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.
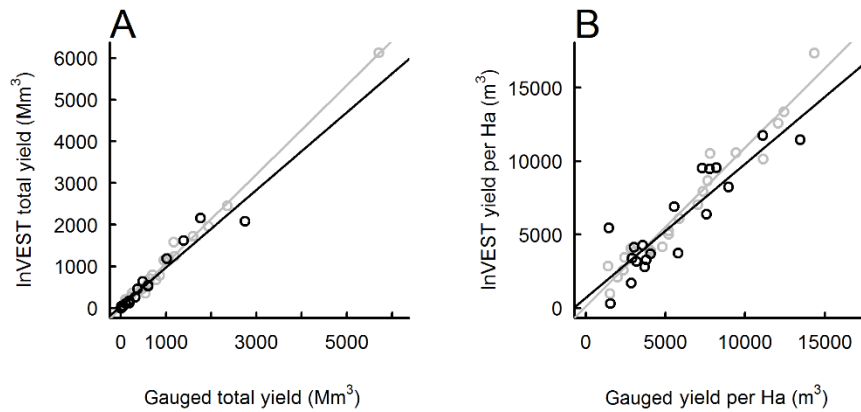
Fig. 5 InVEST modelled total water yield (millions of m³) (A) and per hectare water yield (m³/Ha) (B) against gauged water yield for the 22 original test catchments (grey symbols and line) and the 20 additional catchments (black symbols and line). Model run using UKCP09/MORECS dataset for precipitation 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).

14

4.1. The importance of validation

367 Several studies have sought to address the issue of validating ecosystem service models. However,

368 many of these have been limited by the availability of suitable validation data. For example, Schulp

369 et al. (2014) sought to undertake model validation for a variety of ecosystem services at a European

370 scale, and whilst their results give an indication of the ability of different modelling approaches to

371 predict spatial patterns of ecosystem service delivery, their need to use proxies in the absence of

372 empirical ecosystem service measures prevented a quantitative assessment of model accuracy.

373 Where data have been available, some studies at global or continental scales have compared results

374 from the InVEST water yield model (Mendoza *et al.* 2011), or the Budyko modelling framework upon

375 which it is based (Zhang *et al.* 2004; Zhou *et al.* 2012), to empirical observations. Such studies are

376 useful in comparing models in terms of their ability to respond to global patterns of precipitation but

377 do not provide the information necessary for users to assess whether the model is accurate at a

378 national or regional scale, despite these being the scales at which the majority of ecosystem service

379 mapping exercises are performed (Martínez-Harms & Balvanera 2012) and at which most strategic

380 water resource planning takes place (Watts *et al.* 2015).

381 Other recent studies have used empirical validation data to assess the performance of the InVEST

382 Water Yield model, but covering only single catchments or sub-catchments within a single river

383 basin. In contrast to the present study, several of these studies have had the primary aim of

384 quantifying spatial variation and predicted changes in water yield, with validation at a small number

385 of points to check the reliability of their results, rather than a more general validation of the model

386 across catchments (e.g. Bai *et al.* 2013; Boithias *et al.* 2014; Terrado *et al.* 2014; Xiao *et al.* 2015).

387 Whilst this is entirely sensible, and such studies have found the InVEST model to be a good predictor

388 of measured water yield, the results of such studies are not necessarily transferable to other

389 locations or scales, especially where model inputs have been 'calibrated' to match the empirical

390 data.

391 Comparatively few studies have had the explicit aim of investigating the model performance,

392 uncertainty and sensitivity of the InVEST water yield model. These have largely been conducted in

393 sub-catchments within single river basins which vary widely in area, climate and land cover (e.g.

394 4950 Km^2 Llobregat River basin, Catalonia, Spain (Sánchez-Canales *et al.* 2012); 23 600 Km^2 Cape Fear

395 basin, North Carolina, USA (Hamel & Guswa 2015); 57 400 Km^2 Chubut River basin, Patagonia,

396 Argentina (Pessacg *et al.* 2015)). However, their results are generally corroborated by our national

397 scale analysis (UK land area = 241 930 km^2). These include high sensitivity to precipitation and, to a

398 slightly lesser extent, to evapotranspiration data, as well as empirical support for setting *Z* from

399 numbers of rain events per year (Hamel & Guswa 2015). Our results also corroborate those of these

400   previous studies in demonstrating the substantial improvements in model performance which can

401   be obtained by comparing alternative data sources, especially for those parameters which sensitivity

402   analysis identifies as being major drivers of the model (Sánchez-Canales *et al.* 2012; Hamel & Guswa

403   2015; Pessacg *et al.* 2015). For example, Boithias *et al.* (2014) paid particular attention to obtaining

404   precipitation and evapotranspiration data, because of the sensitivity analysis undertaken by

405   Sánchez-Canales *et al.* (2012) in a similar catchment. As a result, they were able to obtain a good fit

406   to their gauged data by relatively minor (± 10%) calibration of *Z*, *Kc* and water demand values

407   (Boithias *et al.* 2014).

408   All of these studies using some form of validation, and the differences between them, support our

409   suggestion that formal sensitivity analysis and, where empirical data are available, validation should

410   be employed whenever the InVEST model is being used in new regions. Even a relatively small

411   number of validation points from a range of locations can provide valuable insights into the accuracy

412   of the model and the relative performance of different input datasets.   If there are no validation

413   data, our results suggest that datasets of the appropriate spatial scale (e.g. national rather than

414   global) may perform better. The observed differences between our two pairs of input precipitation

415   and PET datasets are probably due to several causes. The WorldClim and CGIAR-CSI data are annual

416   averages calculated over the approximate period 1950-2010. The fact that they do not span the

417   same date range as the validation data may explain some of their poor performance, although

418   annual mean precipitation over England and Wales has not changed significantly over the date range

419   (Jenkins, Perry & Prior 2008). Furthermore , WorldClim and CGIAR-CSI data are interpolated from

420   data which are not spatially uniform in distribution, and thus vary spatially in the uncertainty around

421   the given value of precipitation (Hijmans *et al.* 2005; Hamel & Guswa 2015; Pessacg *et al.* 2015). The

422   UKCP09 data are also interpolated from a network of UK rain gauges, but at a much higher density of

423   sampling points (Perry & Hollis 2005). Errors in the WorldClim precipitation data also tend to be

424   highest in regions with high rainfall (Hijmans *et al.* 2005), such as the UK. Global-scale datasets like

425   WorldClim are both widely used and readily available, so their relatively inconsistent performance

426   across catchments, and the much better performance of the UKCP09 and MORECS data highlights

427   the need for validation to select the most appropriate input data, or at least to assess model

428   performance and the resultant confidence in the results if no other data are available. As might be

429   expected, it appears that large improvements in model performance can be achieved simply by

430   ensuring the input data are matched to the study region in terms of spatial and temporal scales.

431   The apparent trend for catchments to be consistently 'sensitive' or 'insensitive' to variation in the

432   values of all model input parameters suggests that catchment characteristics (e.g. soil and bedrock

433   characteristics) can strongly influence the degree to which errors in the input parameter values will

434   affect the model outputs. Our results suggested that, in the UK, 'upland' type catchments, with low

435   PET and high cover of semi-natural habitats are less sensitive than 'lowland' catchments with high

436   PET and a higher cover or arable land (which has a high Kc). Pessacg et $al.$ (2015) found that

437   catchments with a higher cover of LULC classes with a high value of Kc were most sensitive,

438   potentially giving a +150% change in modelled water yield in response to a +30% error in

439   precipitation data, a response very similar to the most sensitive catchments in our results (see Fig.

440   2A). However, the good overall fit between modelled and measured data across catchments and the

441   lack of significant correlation between model accuracy and catchment descriptors suggest that,

442   when using UKCP09 and MORECS data, errors in the input datasets are comparatively minor, at least

443   to the extent where they are not the major driver of differences between modelled and gauged

444   water yield. The remaining model error is therefore likely to be due to limitations of the model or

445   the validation data (see section 4.2) or more complex interplay between catchment characteristics.

446   A productive area for further research could be more detailed investigation into the drivers of

447   varying sensitivity between catchments, with the aim of using catchment descriptors as predictive

448   variables in determining the impact of driving data on change in water yield.

449   4.2. LIMITATIONS OF THE MODEL AND VALIDATION DATA

450   Despite the good performance of the InVEST model when refined to account for water abstractions

451   and using national input datasets, the accuracy of the modelled water yield values still varied to

452   some extent between catchments (see Fig. 4C and 4D) and there was a slight but consistent

453   overestimate of per hectare water yield. The InVEST water yield model contains several

454   acknowledged limitations and simplifications (Sharp et $al.$ 2015). These include the limited ability of

455   the model to account for inter- or intra- annual variation in water supply. Many ecosystem services

456   (irrigation, hydropower) and disservices (flooding) linked to water yield will be affected by the timing

457   of water availability and peak flows, not just total annual yield. A further simplification is the lack of

458   consideration of lateral and groundwater flows, such that effects of complex land use patterns or

459   underlying geology remain unaccounted for (Sharp et $al.$ 2015). Finally, the model handles

460   consumptive water use in a very simplistic fashion, by allocating a per hectare value to each LULC.

461   Although including per hectare estimates of consumptive abstraction did reduce overestimation of

462   water yield and slightly improved model performance (Table 1), consumptive use is likely to vary

463   widely between catchments and between different areas of the same LULC. In the UK (and in many

464   other parts of the world), many large contributors to consumptive use are single point intakes. The

465   use of reservoirs and water transfer schemes to regulate river flows for abstraction or flood

466   prevention is common (Gibbins et $al.$ 2001), can involve very large volumes of water (Davies, Thoms

467   & Meador 1992; Boithias et $al.$ 2014) and is indicated (but not quantified) in the NRFA catchment

468   description metadata for several of the gauging stations used in this study (Fry & Swain 2010).

469   Although the InVEST model structure does not directly account for point abstractions, where the

470   locations of these are known, these can be represented as separate LULC classes, with

471   corresponding consumptive water use values. Alternatively, the model outputs can be adjusted on a

472   per catchment basis to account for known point source abstractions. However, such data can be

473   hard to obtain due to regulatory restrictions in the UK water industry.

474   It is also worth noting that the empirical validation data themselves are also affected by issues of

475   accuracy, many of which are not captured by the model (e.g. the InVEST water yield model does not

476   distinguish between surface and sub-surface water flow). Measured river flows may be reduced by

477   bypassing of the gauging station via flooding, canals or groundwater flow and either reduced or

478   increased by catchment transfer, which may occur either consistently or only at times of particularly

479   high or low flow. These factors are likely to result in measurements which accurately record the

480   flow of water in the gauged channel, but not the true total water yield from the catchment of the

481   gauging station. Catchments where a significant proportion of total water yield leaves via sub

482   surface flow (or other routes) will show a considerable overestimate of total yield as gauged from

483   stream flow. These issues are likely to affect many individual stations. For example, the severe

484   model overestimation in the Welland catchment might be explained by the fact that it is

485   comparatively small (707 km^2) and subject to high levels of abstraction to a reservoir. More

486   seriously, gauging stations may be unable to record accurate readings of water flow over or under

487   certain flow thresholds. Whilst at least one of these factors was present in the majority of

488   catchments in this study (Fry & Swain 2010), these issues are unlikely to cause systematic bias

489   because they are not consistent across catchments. For example, over half of the 42 catchments

490   studied had factors affecting runoff documented by the NRFA which potentially offset one another

491   (i.e. some factors likely to divert water flow from the river channel and others likely to increase it).

492   The prevalence of these issues, along with the presence of outliers, does serve to illustrate the

493   importance of incorporating local knowledge into decision making, alongside ecosystem service

494   models and empirical validation, as stakeholders may often be able to provide information on

495   processes not captured by the model which can help to explain or mitigate against poor model

496   performance. Users of InVEST are strongly encouraged to involve stakeholders in scenario

497   development and interpretation of model outputs (Sharp *et al.* 2015).

498   4.3. CONCLUSIONS

499   Ecosystem service models such as InVEST have the potential to provide a crucial underpinning to

500   decision and policy making. However lack of robust testing limits their credibility. The work

501 presented here demonstrates that the relatively InVEST simple water yield modelling framework can

502 perform well as long as input data and parameters are representative of the spatial and temporal

503 scale concerned. Care should be taken with application of these tools using indicative datasets at the

504 global scale, and in the absence of more local scale data, empirical validation of model outputs

505 becomes even more important. However, the need for ecosystem service models is driven by the

506 fact that many parts of the world lack relevant empirical data (Crossman *et al.* 2013). Therefore, we

507 firstly recommend that, where empirical data are available, models should be validated for locations

508 in the region of interest and the effect of alternative parameter values or input data should be

509 explored. Secondly, we recommend the application of sensitivity analyses to understand how model

510 outputs vary across the region of interest, either in tandem with validation or, if validation data are

511 not available, to understand uncertainty in model predictions. Finally, if no validation data are

512 available, we advise exercising caution when interpreting model output values. For example, our

513 results suggest that the InVEST water yield model could still be used to assess the rank order of

514 catchments in terms of water yield or the direction of change in relation to scenarios of

515 environmental change (e.g. Willcock *et al.* 2016) even where absolute values are less reliable.

References

522 Allen, R.G., Pereira, L.S., Raes, D. & Smith, M. (1998) Crop evapotranspiration. Guidelines for
523   computing crop water requirements. FAO Irrigation and Drainage Paper 56. Food and
524   Agriculture Organization of the United Nations, Rome, Italy.

525 Bai, Y., Zheng, H., Ouyang, Z., Zhuang, C. & Jiang, B. (2013) Modeling hydrological ecosystem services
526   and tradeoffs: a case study in Baiyangdian watershed, China. *Environmental Earth Sciences,*
527   **70,** 709-718.

528 Bangash, R.F., Passuello, A., Sanchez-Canales, M., Terrado, M., López, A., Elorza, F.J., Ziv, G., Acuña,
529   V. & Schuhmacher, M. (2013) Ecosystem services in Mediterranean river basin: climate
530   change impact on water provisioning and erosion control. *Science of the total environment,*
531   **458,** 246-255.

532 Boithias, L., Acuña, V., Vergoñós, L., Ziv, G., Marcé, R. & Sabater, S. (2014) Assessment of the water
533   supply:demand ratios in a Mediterranean basin under different global change scenarios and
534   mitigation alternatives. *Science of the total environment,* **470–471,** 567-577.

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

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

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

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

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

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

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

548 Crossman, N.D., Burkhard, B., Nedkov, S., Willemen, L., Petz, K., Palomo, I., Drakou, E.G., Martín-
549   Lopez, B., McPhearson, T., Boyanova, K., Alkemade, R., Egoh, B., Dunbar, M.B. & Maes, J.
550   (2013) A blueprint for mapping and modelling ecosystem services. *Ecosystem Services,* **4,** 4-
551   14.

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

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

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

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

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

563 Donohue, R.J., Roderick, M.L. & McVicar, T.R. (2012) Roots, storms and soil pores: Incorporating key
564   ecohydrological processes into Budyko's hydrological model. *Journal of Hydrology,* **436–437,**
565   35-50.

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

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

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

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

577 Fu, B.P. (1981) On the calculation of the evaporation from land surface (in Chinese). *Sci. Atmos. Sin.,*
578   **5,** 23-31.

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

582 Hamel, P. & Guswa, A.J. (2015) Uncertainty analysis of a spatially explicit annual water-balance
583   model: case study of the Cape Fear basin, North Carolina. *Hydrological Earth System Science,*
584   **19,** 839-853.

585 Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution
586   interpolated climate surfaces for global land areas. *International Journal of Climatology,* **25,**
587   1965-1978.

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

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

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

594 Knox, J., Weatherhead, K., Díaz, J.R. & Kay, M. (2009) Developing a strategy to improve irrigation
595   efficiency in a temperate climate: a case study in England. *Outlook on AGRICULTURE,* **38,**
596   303-309.

597 Leh, M.D.K., Matlock, M.D., Cummings, E.C. & Nalley, L.L. (2013) Quantifying and mapping multiple
598   ecosystem services change in West Africa. *Agriculture, Ecosystems & Environment,* **165,** 6-
599   18.

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

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

606 Martínez-Harms, M.J. & Balvanera, P. (2012) Methods for mapping ecosystem service supply: a
607   review. *International Journal of Biodiversity Science, Ecosystem Services & Management,* **8,**
608   17-25.

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

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

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

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

618 Panagos, P., Van Liedekerke, M., Jones, A. & Montanarella, L. (2012) European Soil Data Centre:
619   Response to European policy support and public data requirements. *Land Use Policy,* **29,**
620   329-338.

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

623 Pessacg, N., Flaherty, S., Brandizi, L., Solman, S. & Pascual, M. (2015) Getting water right: A case
624   study in water yield modelling based on precipitation data. *Science of the total environment,*
625   **537,** 225-234.

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

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

632 Sánchez-Canales, M., López Benito, A., Passuello, A., Terrado, M., Ziv, G., Acuña, V., Schuhmacher,
633   M. & Elorza, F.J. (2012) Sensitivity analysis of ecosystem service valuation in a
634   Mediterranean watershed. *Science of the total environment,* **440,** 140-153.

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

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

640 Sharp, R., Tallis, H.T., Ricketts, T., Guerry, A.D., Wood, S.A., Chaplin-Kramer, R., Nelson, E., Ennaanay,
641   D., Wolny, S., Olwero, N., Vigerstol, K., Pennington, D., Mendoza, G., Aukema, J., Foster, J.,
642   Forrest, J., Cameron, D., Arkema, K., Lonsdorf, E., Kennedy, C., Verutes, G., Kim, C.K.,
643   Guannel, G., Papenfus, M., Toft, J., Marsik, M., Bernhardt, J., Griffin, R., Glowinski, K.,

644   Chaumont, N., Perelman, A., Lacayo, M., Mandle, L., Hamel, P., Vogl, A.L., Rogers, L. &
645   Bierbower, W. (2015) *InVEST 3.2.0 User's Guide.* The Natural Capital Project, Stanford.

646 Smethurst, J.A., Clarke, D. & Powrie, W. (2012) Factors controlling the seasonal variation in soil
647   water content and pore water pressures within a lightly vegetated clay slope. *Géotechnique,*
648   **62,** 429-446.

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

652 Terrado, M., Acuña, V., Ennaanay, D., Tallis, H. & Sabater, S. (2014) Impact of climate extremes on
653   hydrological ecosystem services in a heavily humanized Mediterranean basin. *Ecological*
654   *Indicators,* **37,** 199-209.

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

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

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

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

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

667 Zhang, L., Hickel, K., Dawes, W.R., Chiew, F.H.S., Western, A.W. & Briggs, P.R. (2004) A rational
668   function approach for estimating mean annual evapotranspiration. *Water Resources*
669   *Research,* **40,** W02502.

670 Zhou, X., Zhang, Y., Wang, Y., Zhang, H., Vaze, J., Zhang, L., Yang, Y. & Zhou, Y. (2012) Benchmarking
671   global land surface models against the observed mean annual runoff from 150 large basins.
672   *Journal of Hydrology,* **470–471,** 269-279.

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

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

679