Using air photos to parameterise landscape predictors of channel wetted width

B. G. Rawlins[a], L. Clark[b] and D. S. Boyd[b]

[a]*British Geological Survey, Keyworth, Nottingham NG12 5GG, UK*, [b]*School of Geography, University of Nottingham, University Park, Nottingham NG7 2RD, UK*

Correspondence: B. G. Rawlins. E-mail: bgr@bgs.ac.uk

keywords: hydrological source, catchment area, gauging, orthophoto, prediction, wetted width

ABSTRACT: We investigated which landscape and climate-related data (including information on hydrological source of flow) were statistically significant predictors of channel wetted width (WW) across a sizeable (2200 km^2) region of the UK. This was conducted specifically when flow was less than mean daily flow (MDF) and where channels are in a near natural state. Orthorectified air photos at 25 cm spatial resolution were used to measure WW, with the magnitude of the errors in these measurements quantified. We used flow information from local gauging stations to ensure that channels were below MDF for the days on which the air photos were captured. The root mean squared difference between the field and air photo measurements of WW (n=28 sites) was small (0.14 m) in comparison to median WW (3.07 m).

We created points along sections of channels visible in air photos and used a terrain model to create drainage catchments for these points and computed their catchment area (CA). We selected a subset of points (n=472) and measured their WW from air photos, and computed landscape-related data for each of their catchments (mean slope, mean annual rainfall, land cover type, elevation) and also mean BFIHOST, a quantitative index relating to hydrological source of flow. We used a linear mixed model to predict WW by including the landscape data (including CA$^{0.5}$) as fixed effects, plus a spatial covariance function estimated by residual maximum likelihood (REML) to determine unbiased estimates of the predictors. There was no evidence for retaining the spatial covariance function. With the exception of land cover, all the predictors were statistically significant and accounted for 76% of the variance of WW. When CA$^{0.5}$ alone was used as a predictor it captured 54% of the variance. The vast majority of this difference was due to inclusion of an interaction between CA and hydrological source of flow (BFIHOST). As catchment area increases, those channels with larger mean catchment BFIHOST values (greater proportion of baseflow contribution) have narrower WW by comparison to those with smaller mean BFIHOST for the same CA. Improved predictions of channel WW (based on our findings) could be used in channel restoration.

1 Introduction

The effective restoration of stream or river channels following various modifications requires an understanding of natural channel morphology (Thorne *et al.*, 1996), including the morphology of channel cross sections and channel wetted width (WW). For channels in bedrock, scaling relationships have been established where bankful width (BW) or wetted width (WW) is a function of discharge or catchment area (CA) with exponents of between 0.3 and 0.5 (Whipple, 2004; Faustini *et al.*, 2009).

For nations with a large number of gauged rivers (such as the United Kingdom or New Zealand), power-law relationships have been developed to predict at-a-station hydraulic geometry based on data from gauging sites (Booker, 2010) or at natural river sections (Booker and Dunbar, 2008) under differing flow conditions. In a study based on measurements of discharge and hydraulic geometry at 3600 stations across England and Wales, Booker and Dunbar (2008) concluded that 'hydraulic geometry (including WW) is driven by catchment area rather than natural geomorphological variations in the streamwise direction, but that geomorphological variation can still have a major impact on channel structure'. In a study across the conterminous United States, Faustini *et al.* (2009) found that CA (with exponents of between 0.22 and 0.38) explained between 36 and 77% of the variation in BW and that this varied according to region. Channel substrate is also likely to influence hydraulic geometry; bedrock channels support much higher wall stress than gravel channels (Finnegan *et al.*, 2005) so the former will have narrower channels than the latter at the same discharge. Other factors which are known to account for variations in WW or bankful channel width include elevation, channel slope (Whittaker, 2007), hydrological source, land cover type and climate (Booker, 2010; Faustini *et al.*, 2009). In landscapes where few gauged measurements are available, it is necessary to use other sources of landscape-related data to predict WW. These sources of data may include digital terrain models (for the calculation of CA, slope and elevation) or maps of soil, geology, land cover, and climate-related information. Rather than relying solely on functions of CA, other sources of landscape information may explain substantially more of the variation in channel WW.

In the United Kingdom, one source of information relating to hydrological source of river flow is referred to as BFIHOST. It is a dataset derived from a combination of information on catchment baseflow index (Gustard *et al.*, 1992) and associated maps classified by the hydrology of their soil types and substrates (HOST) (Boorman *et al.*, 1995). Baseflow index (BFI; Institute of Hydrology, 1980) is the long-term ratio of baseflow to total stream flow, representing slower contributions to river flow and is often strongly related to catchment geology. There is a BFIHOST value (on a scale of zero to one) for every 1km^2 of the terrestrial landscape of the British Isles. A value of one implies that river flow is entirely related to groundwater sources (no runoff contributions), whilst a value of zero implies all flow is from shallow runoff. To our knowledge no studies to date have attempted to account for differences in channel WW using information such as BFIHOST which is related to hydrological source.

Remote sensing data are increasingly used to estimate hydraulic features of surface water bodies; for example, Bjerklie *et al.* (2005) developed a method from a combination of air photos and synthetic aperture radar images to estimate river discharge for various channels in the USA. In a more recent study, methodologies for the extraction of channel (bankful) widths based on freely available high-spatial resolution imagery and digital elevation models were demonstrated by Fisher *et al.* (2013); the authors did not estimate channel WW which (we consider) may be more effective based on a manual procedure. Where the view of a channel is unimpeded from above, the resolution of air-photos is now sufficiently fine (25 cm pixel resolution) to make accurate estimates of channel WW across the landscape. Such snapshot, instantaneous images are collected without regard to recent variations in antecedent rainfall or discharge. The majority of channels recorded in these images (for temperate climates of the United Kingdom) relate to discharges below mean daily flow (MDF; Smakhtin, 2001), avoiding the highest flows when channel WW is likely to be larger, and more variable. Mean daily flow, computed from long-term time series of discharge measurements, is heavily influenced by infrequent flood events leading to strongly skewed flow distributions (mean flows are typically much greater than median flows). We wanted to investigate which sources of climate and landscape-related data could be used to make accurate predictions of downstream channel WW measured by air photo when channel discharge is below MDF across a landscape with varied topography, geology, geomorphology and mean annual rainfall. Although we cannot include flow information as a predictor of WW because our sites do not coincide with gauging stations, we wanted to ensure that any significant effects due to variations in flow conditions were minimised. Large fluctuations in discharge across the study area (at the time of air photo capture) could introduce bias to our predictors for WW. We can use data from local gauging stations for the period over which the air photos were captured to check whether flow in local channels were less than MDF.

In general it is not possible to use air photos to measure channel WW across an entire landscape or region because vegetation will likely cover some sections of channels. We cannot assume that a set of sample data (i.e. estimates of WW) are independent random variables; using ordinary least squares (OLS) linear regression could lead to biased estimates of a predictor. Such sample data will likely exhibit varying degrees of spatial clustering which can also lead to bias in predictors if estimated by OLS. To overcome these limitations, we can adopt a model-based analysis where we assume the variable is a realization of a random process. We can ensure that estimates of the coefficients for any set of landscape predictors of WW are unbiased if we fit the model using residual maximum likelihood (REML) whilst accounting for spatial clustering in the sample data (Lark and Cullis, 2004).

The first aim of our study was to determine the magnitude of errors between field-based and air photo meaurements of WW for flows less than MDF. If these errors were sufficiently small, the second aim was to determine which landscape and climate-related data were statistically significant predictors of channel WW for these flow conditions for a region of the British Isles which encompasses broad variations in climate, land cover, geology and geomorphology. In particular, we wished to determine whether including information on variations in the hydrological source of flow (BFIHOST) was a significant predictor of channel WW.

2 Study region and Methods

2.1 Study region

The study region is an area of 2200 km^2 (20 km $\times$ 110 km) covering part of north Wales and western England (Figure 1). It was selected to encompass a broad range of bedrock lithology, topography and land cover types. Elevation is greatest to the west (>1000 m) and declines towards the east to around sea level (Figure 1a). The region has a temperate, maritime climate with mean annual rainfall varying from greater than 4000 mm in the west (Snowdonia National Park) to 650 mm in the east of the study region. There are a series of west to east changes in bedrock geology from Ordovician slate, through Silurian Gritstone, to Permo-Triassic Sandstone then Mudstone, and also halite (in the eastern most part of the study region). Based on maps from the British Geological Survey, there are extensive Quaternary glacial till deposits covering (by area) around 50% of the central part of the study region, increasing to 100% to the east. During the Holocene, uplift or subsidence rates across the British Isles are unlikely to have been sufficiently large (<2 mm yr^{-1}; Shannan and Horton, 2002) to have had a major impact on adjustments to channel width.

We have no quantitative information relating to variations in stream substrate for the study region; bedrock channels are common in the low order streams of upland settings in the Snowdonia National Park to the west, whilst alluvial stream beds are prevalent to the east with its extensive cover of Quaternary deposits and weaker rock types. Using vector data extracted from Ordnance Survey MasterMap for inland water channels we calculated a declining trend in average drainage density (length of channel per unit area) from the west (mean 5.3 km km^{-2}) to the east (mean 3.8 km km^{-2}) of the study region. The spatial distribution of BFIHOST values (Figure 1b) shows that the soils and rock types to the west are more runoff-dominated that those to the east, but there is a substantial degree of local complexity in this pattern relating to hydrological sources.

2.2 Wetted width data

Air photos: The channel networks for the study region were extracted directly from the 'inland water' layer of Ordnance Survey MasterMap (©Ordnance Survey) as vector features in the GIS package ArcMap^{TM} (ESRI). To determine whether the view of a channel section was impeded in the air photo, the channel networks were overlain in the GIS above 25 cm pixel aerial photos for all the region (©UKP/Getmapping). These air photos are colour (RGB) orthophotos derived from vertical stereo photography, and were captured on four dates across the study region: 11 May 2009, 01 June 2009, 24 April 2010 and 11 October 2011. We visually identified those sections of each channel vector which were not visible in the air photos, and these were removed from a copy of the vector file. We used ArcToolBox^{TM} (ESRI) to create a series of points along each of the remaining channel vectors at 1 km intervals. We refer to these subsequently as unimpeded points.

We used the ArcHydro extension and a 5 m resolution Digital Surface Model (NEXTMap Britain elevation data from Intermap Technologies, Intermap, 2009) to create drainage catchments upstream of all the unimpeded points of the channel network (n=2324). We created a set of catchment polygons so we could estimate catchment properties from other landscape data (see section 2.5). We then computed the area of the catchment draining to each of these points and also transformed the values by taking their square root. We chose to apply a minimum threshold CA of 1 km^2 for use in our study because we considered that the errors associated with air photo based estimates for the smallest catchments (i.e. $<$1 km^2) could lead to false inferences. There were 1255 unimpeded points with a CA $<$1 km^2. We sorted the unimpeded points by CA and used a routine in the R Environment (R Development Core Team, 2012) to randomly select 50 points within each decile of the ordered distribution. This procedure ensured we selected a subset of channel locations that encompassed a broad range of CA, which is typically a significant predictor of channel WW. We then made estimates of WW at each of these locations from the air photos using ArcMap^{TM} (ESRI). All the estimates were made by the same person. After experimenting with a range of scales to view the air photos, we found the optimum scale to view the images varied between 1:200 and 1:250; dependent on local conditions. The wetted channel was defined subjectively by the same person after having viewed the colour contrast between the water surface and either the adjacent exposed bed material or channel bank. At 28 of of the 500 point locations, there was insufficient colour contrast to accurately define one or both edges of the wetted channel, so these locations were excluded and our final dataset consisted of 472 estimates. The orientation of the cross-section at which we estimated width was determined by adding a temporary linear feature (approximately 10 m long) along the centre of the channel. The wetted width was estimated perpendicular to this linear feature. To account for some of the local variation we estimated the channel WW at three distinct positions around each point; first precisely at the point location on the channel, and also 50 cm upstream and downstream, in each case adding a temporary linear feature (10 m in length) to define the orientation of the cross-section. We computed the average of these three values and used this as the estimated WW. Tree roots are known to have an impact on channel morphology (Keller and Swanson, 1979) so our sample data, which exclude sites where trees are close to the bank, may be somewhat biased and we cannot account for this effect.

Field measurements: We selected a subset of sites (Figure 1) for field-based measurement of channel WW and located them using a handheld GPS with an accuracy of ± 1 m. We had limited resources to undertake field based measurements; to avoid large travel distances between individual sites we used a subset of stream sites across a smaller part (150 km^2) of the study region where we established there was a large range in catchment areas (between 1 and 70 km^2); we selected 28 sites in this region spanning the full variation in catchment area. We recognize that ideally we would have made measurements at more locations across the entire study area. The WW measurements were undertaken on 3rd November 2012 using a tape measure during which flow in the channels appeared to be low (below MDF) across this part of the study region. We measured WW by stretching a tape measure across the full width of the wetted channel. We computed the differences (or errors) between the field and air-photo WW measurements at each site, and also the root mean squared error (RMSE) and bias (whether the sum of the differences were substantially more positive or more negative) using the following formulae. For RMSE:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{z}_i - z_i)^2} \tag{1}$$

where z_i is the measured field width (in metres) and $\widehat{z}_i$ is the width estimated from the air photo (also in metres), and where n is the total number of sites. We calculated the mean error (ME – or bias) of the differences:

$$\text{ME} = \frac{\sum_{i=1}^{n}(\widehat{z}_i - z_i)}{n} \tag{2}$$

using the same notation. We also computed the standard deviation of the error (SDE) which is a measure of precision (after removal of the mean error).

$$\text{SDE} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\widehat{z}_i - z_i - \text{ME})^2} \tag{3}$$

2.3 Gauged flow data

In our analyses we used mean gauged daily river flow data for three gauging stations within the study region from the National River Flow Archive (http://www.ceh.ac.uk/data/nrfa/). The names of the stations are Cwm Llanerch (grid reference (SH802581), River Alyn (grid reference SJ336541), Wistaston Brook (grid reference SJ674552; Figure 1). We examined flow variations based on available data for three calendar years (dates between 1st January 2009 and 31st Dec 2011) which encompasses the four days on which the air photos were captured (Figure 2) across the region. The discharge measurements and their corresponding percentiles on a flow duration curve of these data over this three year period are shown for each station in Table 1, and also the mean discharge for each station for the same period. For the four dates on which air photos were captured, these data show that on each date, flow in each of these channels was below MDF measured for the 3-year period. We therefore feel justified in assuming that estimates of channel WW on these dates from the air photos relate to channel states where flow was less than MDF.

The flow data for the day when our field measurements of channel WW were made (3^{rd} November 2012) will not be released by the National River Flow Archive (www.ceh.ac.uk/data/nrfa/) for a further six months (September 2013), so we cannot provide the associated percentiles on a flow duration curve for these channels on this date as part of our study.

2.4 Landscape and climate data

BFIHOST: We extracted the 1 km grid values for the BFIHOST data for the study region. We used the catchment polygons referred to above to calculate the arithmetic mean catchment BFIHOST value (cBFIHOST) for the catchment upstream of each of the 472 selected points. We used the cBFIHOST values in the subsequent statistical analysis.

Local and mean catchment slope: We used the 5 m Digital Surface Model (Intermap, 2009) to compute the slope (in degrees) at each of the 472 channel locations. We also computed the arithmetic mean slope for the upstream catchment area using all slope values within the catchment polygons.

Mean annual rainfall: We used data for mean annual rainfall (1961–1990; mm) on a 5 km grid available from the Met Office (UK). We converted the grid values to point locations at the centre of each 5-km grid and calculated the mean of the value in each grid. If none of the rainfall points fell inside a catchment, we used the value of the nearest point location as the mean annual rainfall value for that catchment.

Land cover: We extracted a grid showing the dominant land cover class in each 1 km^2 pixel from the Land Cover Map 2007 (Fuller et al., 2000) for Great Britain. We then extracted the codes for the dominant land cover class for each of the catchment polygons, and used this code as a classification for land cover. The dominant classes were semi-natural grassland (44%), improved grassland (38%), cultivated land (4%), mountain-heath and bog (2%), with other smaller land cover types forming the remainder.

2.5 Statistical analysis

In this study we used both linear models and the linear mixed model (LMM) to explore which landscape and climate-related data could account for the variation in channel WW. Our sample data exhibits a degree of clustering (Figure 1) which can lead to bias in predictors if estimated by a linear model using OLS. To overcome these limitations, we used the LMM where we assume the variable is a realization of a random process. The coefficients for any set of landscape-related predictors (fixed effects) of WW are unbiased if the model of the spatial dependence of the error variation (an autocorrelated Gaussian variable) are fit using REML; this model of the spatial dependence accounts for spatial clustering in the sample data. Here we are referring to spatial clustering in the positions of the locations in coordinate space, we have not accounted for the locations in relation to their positions on the stream network. This implementation of the LMM has been described thoroughly in previous studies and the reader should consult the paper by Lark and Cullis (2004) for a complete description.

We used the ANOVA function in the R environment (R Development Core Team, 2012) based on model outputs from the LMM's to test whether there was evidence to include each of the fixed effects based on comparing the log-likelihood ratio statistics before and after their inclusion in the model. We also tested after inclusion of the statistically significant fixed effects whether there was evidence for inclusion of a spatial covariance function. This may be one of several authorized functions (Webster and Oliver, 2007). We used the LME function in the R package NLME (Pinheiro *et al.*, 2009) which fits LMMs and has an option for including a spatial covariance structure (fitted by REML). If our data for channel WW were strongly skewed it can present problems for geostatistical analysis because a variogram calculated from such data may be strongly biased. To investigate this further, we fit a simple, least squares model between $CA^{0.5}$ (predictor) and the estimates of WW (predictand) data; the residuals were close to normally distributed (skewness coefficient = 1.03) so we chose to undertake all our analyses on the original, untransformed data.

We formed a LMM model by including a series of fixed effects (the overall mean, $CA^{0.5}$, cBFIHOST, local (channel) slope, elevation, catchment slope, rainfall and dominant land cover class); with the exception of land cover class all the predictors were statistically significant ($P<0.05$). We then updated the LMM by including spherical and exponential spatial covariance functions fitted by REML. We tested the statistical significance of the additional predictor using the ANOVA procedure in the R Environment (R Development Core Team, 2012); there was no evidence for inclusion of the spatial covariance functions ($P>0.05$). Finally we used an OLS regression model to estimate coefficients for the various landscape-related predictors. We used the stepwise regression function STEPAIC (Venables and Ripley, 2002) which tests the inclusion of predictors based on the Akaike information criterion; the k-value (multiple of the degrees of freedom for penalty) was 2 and the mode of stepwise search was forwards and backwards. The set of statistically significant predictors which we subsequently refer to as the 'full model' were: $CA^{0.5}$, cBFIHOST, local (channel) slope, elevation, catchment slope and rainfall. We computed summary statistics and examined histograms of the residuals which exhibited some positive skew (skewness coefficient=0.89). We concluded that the skewness was dominated by a few outliers because the octile skewness (Brys *et al.*, 2003) was small (octile skewness coefficient=0.007).

We made a quantitative comparison between WW estimated from air photos and each model for the prediction of channel WW from catchment characteristics using: i) the full model), and ii) only $\mathrm{CA}^{0.5}$ as a predictor (we refer to this as the 'CA model'). We computed the adjusted coefficient of determination (R^2_{adj}) for both the CA and full models. We also used the two models to make predictions at all the sites, and from these computed the root mean squared error (of prediction) (RMSE) across all sites using Equation (1) where z_i is the observed width from air photos in metres and $\widehat{z}_i$ is its predicted value, and n is the number of sites. We also calculated both ME and SDE using Equations (2, 3) respectively, based on this notation.

3 Results and their interpretation

A set of summary statistics for the field-based measurements of channel WW (n=28) are shown in Table 2, with statistics for the differences between these measurements and those made from the air photos. The WW estimates encompass a broad range (0.72–13.3 m) with a median of 3.07 m. The RMSE between the field-based and air photo measurements was small (0.14 m) in comparison to the median width, and there was very little bias (0.026 m) between the measurements and air photo estimates. We therefore consider that estimates from 25 cm pixel air photo are sufficiently accurate to undertake a more comprehensive statistical analysis of landscape and climate-related predictors of WW.

Summary statistics of WW at the 472 sites across the study region, and for the landscape plus climate-related data for each upstream catchment are shown in Table 3; these data are also supplied as as supplementary online material associated with the published paper. The maximum difference between each of the three separate local estimates of channel WW from the air-photos are shown in Figure 3. This shows that the maximum differences tend to increase with increasing channel width, and in most cases the differences are small (< 1 m), but that in some cases the differences are substantial (> 2 m) which suggests that taking the average of three measurements would likely be an effective strategy to account for some of the local variations in WW in the images.

There is a significant degree of spatial clustering in the selected sites (Figure 1) which reflects a combination of downstream channel associations and the distributions of aerial obstructions which prevent clear aerial views of the channel. The WW estimates encompass a broad range of channel size (0.49–28 m) with a median value (3.7 m) which is similar to that of the 28 sites where field measurements were undertaken (3.07 m; Table 2). The median upstream CA from the sites of WW measurement was 5.5 km^2 and the largest catchment was 89 km^2. As one would expect, the range of mean catchment BFIHOST values (cBFIHOST; 0.28–0.59) was smaller than the range for the 1 km^2 values across the study region (0–0.93; Figure 1) because averaging across catchments reduces the variation.

The results from fitting linear models by OLS are shown in Table 4. This model suggests that all the various landscape and climate-related predictors of WW are statistically significant for prediction of WW (P-values<0.05). In order of decreasing importance these were: $CA^{0.5}$ > $CA^{0.5}$:cBFIHOST > cBFIHOST > rainfall > local slope > catchment slope > elevation.

The adjusted-R^2 values for these two linear models (full model and CA model) were 0.76 (76%) and 0.54 (54%), respectively. The vast majority of the difference in the proportions of variance explained was due to the inclusion of the interaction term ($CA^{0.5}$:cBFIHOST). Figure 4 shows the measured and predicted WW values for the two models; the CA model clearly underpredicts the WW for the widest channels (>13 m) by comparison with the full model. The CA model also consistently overpredicts WW for the narrowest channels (<2 m), whilst the predictions errors from the full model are more evenly distributed. Across the intermediate range of channel widths (2–13 m), the predictions errors based on the full model are generally less than those of the CA model, but the differences in error predictions are less apparent than for both the larger and smaller channels. There is a substantial difference in the RMSEP and for channel WW based on the two models; 1.80 m and (CA model) and 1.34 m (full model). The SDE values for the full and CA model were 2.01 and 2.72 m respectively. This suggests that including information from other landscape predictors, but particularly hydrological source of flow, could substantially reduce errors in estimating channel WW, for flow states less than MDF, across complex landscapes.

Figure 5 shows how the interaction between $CA^{0.5}$ and cBFIHOST accounts for WW; this plot was generated using the VISREG2D function in the VISREG package (Breheny and Burchett, 2012). As catchment area increases, those channels with larger mean catchment BFIHOST (cBFIHOST) values have narrower channel WW by comparison to those with smaller cBFIHOST for the same catchment area. We infer that channel morphology responds to the source and type of flow; those channels with greater proportions of flow derived from groundwater or slower throughflow (more permeable bedrock and soils) have narrower, and also likely, deeper channel profiles by contrast to those channels where hydrographs have more flashy responses dominated by shallow runoff.

4 Discussion

Our statistical analysis suggests that including information on hydrological source of flow can significantly improve predictions of channel WW across a complex landscape. The BFIHOST values provide an estimate of the relative magnitude of baseflow contributions to channels (based on the hydrology of soil and geology) for each 1 km^2 of the landscape. However, we cannot be certain that the mechanism through which cBFIHOST accounts for channel WW is entirely related to hydrological source because there are many other factors that control channel WW, including substrate type, slope (Finnegan *et al.*, 2005) and sediment supply (Liebault and Piegay, 2001) which may also relate closely to cBFIHOST values. To make clear inferences on the precise mechanism through which cBFIHOST accounts for the variation in channel WW, further research is required that incorporates quantitative data relating substrate type to channel WW using a similar landscape scale analysis as reported here.

Analyses of long-term flow series data in relation to flow on dates of air-photo acquisition from the three gauging stations across the study region (summarised in Table 1) show that there were considerable differences between gauged flow percentiles on the same date. For example, on 11^{th} of October Cwm Lanerch and Wistaston Brook had flow equivalent to the 69^{th} and 25^{th} percentiles on their respective FDC. We cannot assume that all relative flows were the same across the study area; both local climate (particularly rainfall) and catchment characteristics (including size, geology and land use) will result in differences in runoff on specific dates in relation to long-term flow quantities. Wetted width is flow dependent and therefore sensitive to the relative flow at which width was observed. Our model estimates a single width based on the state of flow on the observation date (which was likely below long-term mean flow). Although our width estimates from air photos were made on one occasion from four possible dates (with associated variations in flow between sites) our model demonstrated reasonable performance in accounting for WW. This suggests that the hydraulic geometry of channels in this region have profiles which are more rectangular than either 'V' or 'U'-shaped because in the latter the rate of change in WW would be substantial at lower flows.

For much of the globe where there is currently no information relating to hydrological source of flow (such as BFI values), it may be possible to develop a classification system using land cover (vegetation) and geology to estimate such an index. A recent study demonstrated that a lithological classification can account for a substantial proportion of the variation in BFI for a chalk basin in England (Bloomfield *et al.*, 2009). The study by Gustard (1993) suggested that prediction of BFI based on a classification for a single country was not as successful when extended to larger regions such as Western Europe. Analyses using data from 114 catchments in Victoria (Australia; Lacey and Grayson, 1998) showed that the most important factor for predicting BFI was a series of 12 classes comprising combinations of geology and vegetation; a regression on the class means accounted for 84% of the variation in BFI. The authors observed that if a catchment consisted of both a single rock and vegetation type, the mean BFI for other catchments of this type provide a reasonable estimate, but they recommended that the approach needed further testing in catchments with mixed classes.

The proportion of sedimentary rock in a catchment was also found to be a significant predictor of BFI for a study of 164 catchments in Victoria (Nathan and McMahon, 1992). Data for vegetation or land cover types are now widely available at reasonable resolutions at the global scale (Ramankutty *et al.*, 2008; Zhu and Wallter, 2003), whilst geological map data is available for the globe at coarse scales (Hartmann and Moosdorf, 2010) and finer resolutions (1:1 000 000) in more intensively surveyed areas (see http://www.onegeology.org). Although indices for BFI have been developed in western Europe based on soil groups and drainage classes (Gustard, 1993), the current lack of global soil data at a sufficiently fine resolution (e.g. <1:1 million scale) suggests that an approach based on combining geology and vegetation classes will likely be more comprehensive as it would encompass a greater range of gauged catchments (required for estimating BFI values) across the globe. It would then be possible to make a comprehensive assessment of hydrological source in controlling channel WW.

Our findings suggest that the hydrological properties of both soils and bedrock geology (or other features which relate to them) are a significant factor in determining channel WW. Although considerable research has demonstrated how flow and sediment transport influence channel WW, it is not clear how differences in hydrological source would influence WW at flows greater than mean flow. It may be possible to relate the dates of air photo capture and local gauging station flow information to explore this relationship in more detail.

In our analysis, we used 25 cm pixel resolution air photos to measure channel WW for a small region of Wales (and part of England). Air photo coverage at this resolution (and BFIHOST values) are available for all the British Isles so it would be possible to extend our analysis to determine how the relationships we identified vary at a regional scale, whilst ensuring flow less than MDF conditions prevailed (based on local gauging station data) on the date of the airborne survey. We undertook our estimates of channel WW from air photos in a GIS using a manual approach. It may be possible to automate the extraction of WW estimates from colour infra red (CIR) air photos which are available at 50 cm pixel resolution across the British Isles, and use super-resolution mapping approaches to measure subpixel waterline boundaries (Foody *et al.*, 2005). The magnitude of errors in measuring channel WW from an automated extraction procedure would need to be compared against estimates from both finer resolution (25 cm pixel) air photos and field-based measurements.

Based on river habitat surveys across England and Wales between 2007 and 2008 (Environment Agency, 2010), channels across around 80% of our study region have been reported to be close to a 'near-natural' state, or not subject to major physical modification. Some of the unexplained variability in channel WW likely results from past or on-going engineering interventions plus and/or channel maintenance. However, only 11% of the rivers across England and Wales were classified as 'near-natural' and it is unlikely there is sufficient local information on engineered modifications for this to be incorporated into predictions of channel WW. The inclusion of BFIHOST data in predictions of the 'natural' wetted width of a channel could help to improve channel restoration or rehabilitation design.

It is increasingly recognised that freshwater channels account for a sizeable proportion of the carbon dioxide (CO_2) flux to the atmosphere from the terrestrial carbon cycle (Butman and Raymond, 2011). Accurate predictions of the WW of small rivers are therefore required because CO_2 evasion rates are greater from the surfaces of smaller (compared to larger) water bodies because the former generally have more turbulent flow (Vachon *et al.*, 2010). Any attempt to compute the quantity of CO_2 evasion from the surfaces of streams at the landscape-scale may be improved if channel WW (and therefore surface area) can be predicted more accurately by including data on hydrological source of flow.

5 Summary and Conclusions

We used historical gauged flow data from three stations encompassing the four dates upon which air photos were captured across our study region for part of Wales and western England (with varying geology, geomorphology and climate) to demonstrate that flow was likely less than MDF on each date (of photo capture). Across the entire study region (2200 km^2), WW estimates at 472 sites based on air photos encompass a broad range of widths (0.49–28 m) with a median value (3.7 m). A linear mixed model fitted to the air photo-based channel WW estimates (predictand) with a range of landscape and mean annual rainfall data as predictors (fixed effects) showed that there was no evidence for inclusion of a spatial covariance function. One of these predictors (BFIHOST) is related to hydrological source of river flow.

By comparing field-based measurements and air photo (25 cm pixel resolution) estimates of channel WW at 28 sites for part of our study region (channel WW range 0.72–13.3 m), we showed that the root mean squared difference was small (0.14 m) and there was very little bias (0.026 m) between the two sets of observations.

We fit a linear regression model by ordinary least squares to predict channel WW and showed that the following were all statistically significant predictors (in order of decreasing importance): CA$^{0.5}$ > CA$^{0.5}$:cBFIHOST > cBFIHOST > rainfall > local slope > catchment slope > elevation. We refer to this as the full model. The adjusted-R^2 values for two linear models for prediction of WW (full model and another with only CA$^{0.5}$ as a predictor) were 0.76 (76%) and 0.54 (54%), respectively. The vast majority of the difference in the proportions of variance explained was due to the inclusion of the interaction term (CA$^{0.5}$:cBFIHOST). The RMSEP and SDE values for the full model were 1.34 m and 2.01 m respectively, substantially smaller than the equivalent statistics for the CA model (1.80 m and 2.72 m).

Information relating to hydrological source of flow (such as BFIHOST), could substantially reduce errors in estimating channel WW (below MDF) across complex landscapes. In this region, as catchment area increases, those channels with larger mean catchment BFIHOST (cBFIHOST) values have narrower channel WW by comparison to those with smaller BFIHOST for the same catchment area.

Acknowledgements

We wish to thank Murray Lark, Ben Marchant, Colin Thorne and Michael Ellis for helpful comments relating to analyses of the WW data and interpretations relating to them, and two anonymous reviewers who provided helpful comments on previous versions of the manuscript. We acknowledge the Centre for Ecology and Hydrology for providing the BFIHOST data and the 1-km pixel Land Cover Map 2007 data under licence. The source of the gauging station flow data was the National River Flow Archive (UK). The 25 cm air photos were used under the Pan Government Agreement from UKP/Getmapping (Licence No. UKP2006/01). This publications contains Ordnance Survey data ©Crown Copyright and database rights 2012 (Licence no. 100021290). This paper is published with the permission of the Executive Director of the British Geological Survey (Natural Environment Research Council).

References

Boorman DB, Hollis JM, Lilly, A. 1995. *Hydrology of soil types: a hydrologically based classification of the soils of the United Kingdom.* Institute of Hydrology Report Number 126, Wallingford.

Bjerklie DM, Moller D, Smith, LC, Dingman, SL. 2005. Estimating discharge in rivers using remotely sensed hydraulic information. *Journal of Hydrology* **309**: 191–209.

Bloomfield JP, Allen DJ, Griffiths KJ. 2009. Examining geological controls on low-flow index (BFI) using regression analysis: An illustration from the Thames Basin, UK. *Journal of Hydrology* **373**: 164–176.

Booker DJ. 2010. Predicting wetted width in any river at any discharge. *Earth Surface Processes and Landforms* **35**: 828–841.

Booker DJ, Dunbar MJ. 2008. Predicting river width, depth and veolcity at ungauged sites in England and Wales using multilevel models. *Hydrological Processes* **22**: 4049–4057.

Breheny P, Burchett W. 2012. visreg: Visualization of regression models. R package version 1.1-1. http://CRAN.R-project.org/package=visreg

Brys G, Hubert M, Struyf A. 2003. A comparison of some new measures of skewness. In *Developments in robust statistics*, eds Dutter R, Filzmosre P, Gather U and Rousseeuw PJ, Physica-Verlag Heidelberg, 98–113.

Butman D, Raymond PA. Significant efflux of carbon dioxide from streams and rivers in the United States. *Nature Geoscience* **4**: 839–842.

Environment Agency, 2010. *River Habitats in England and Wales: current state and changes since 1995-96*. Environment Agency, Bristol. pp 29.

Faustini J M, Kaufmann PR, Herlihy AT. 2009. Downstream variation in bankfull width of wadeable streams across the conterminous United States. *Geomorphology* **108**: 292–311.

Finnegan NJ, Roe G, Montgomery DR, Hallet, B. 2005. Controls on the channel width of rivers: Implications for modeling fluvial incision of bedrock. *Geology* **33**: 229–232.

Fisher GB, Bookhagen B, Amos CB. (in press) Channel planform geometry and slopes from freely available high-spatial resolution imagery and DEM fusion: Implications for channel width scalings, erosion proxies, and fluvial signatures in tectonically active landscapes. *Geomorphology*. DOI.org/10.1016/j.bbr.2011.03.031

Foody GM, Muslim AM, Atkinson PM. 2005. Super-resolution mapping of the waterline from remotely sensed data. *International Journal Of Remote Sensing* **26**: 5381–5392.

Fuller RM, Smith GM, Sanderson JM, Hill RA, Thomson AG. 2002. The UK Land Cover Map 2000: Construction of a parcel-based vector Map from satellite images. *Cartography Journal* **39**: 15–25

Gustard A. 1993. *Flow regimes from international experimental and network data, volume 1: Hydrological Studies.* Institute of Hydrology, Wallingford, UK.

Gustard A, Bullock A, Dixon JM. 1992. Low flow estimation in the United Kingdom. Wallingford, Institute of Hydrology, 88pp.

Hartmann J, Moosdorf N. 2012. The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry Geophysics Geosystems* **13**: Q12004. DOI:10.1029/2012GC004370.

Intermap, 2009. NEXTMap Britain. Intermap. http://www.intermap.com/nextmapbritain.

Institute of Hydrology, 1980. *Low flow studies.* Institute of Hydrology, Wallingford, UK.

Keller EA, Swanson FJ. 1979. Effects of large organic material on channel form and fluvial processes. *Earth Surface Processes and Landforms* **4**: 361–380.

Lacey GC, Grayson RB. 1998. Relating baseflow to catchment properties in south-eastern Australia. *Journal of Hydrology* **204**: 231–250.

Lark RM, Cullis BR. 2004. Model-based analysis using REML for inference from systematically sampled data on soil. *European Journal of Soil Science* **55**: 799–813.

Liebault F, Piegay H. 2001. Assessment of channel changes due to long-term bedload supply decrease, Roubion River, France. *Geomorphology* **36**: 167–186.

Nathan RJ, McMahnon, TA. 1992. Estimating low flow characteristics in ungauged catchments. *Water Resources Management* **6**: 85–100.

Pinheiro J, Bates D, DebRoy S, Deepayan S, the R Development Core Team. 2012. *nlme: Linear and Nonlinear Mixed Effects Models.* R package version 3.1-103

R Development Core Team, 2012. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Ramankutty N, Evan AT, Monfreda C, Foley JA. 2008. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochemical Cycles*, **22**: GB1003.

Shennan I, Horton B. 2002. Holocene land- and sea-level changes in Great Britain. *Journal of Quaternary Science* **17**: 511–526.

Smakhtin VU. 2001. Low flow hydrology: a review. *Journal of Hydrology* **240**: 147–186.

Vachon D, Prairie YT, Cole JJ. 2011, The relationship between near-surface turbulence and gas transfer velocity in freshwater systems and its implications for floating chamber measurements of gas exchange. *Limnology and Oceanography* **55**: 1723–1732.

Thorne CR, Allen RG, Simon A. 1996. Geomorphological river channel reconnaissance for river analysis, engineering and management. *Transactions of the Institute of British Geographers* **21**: 469–483.

Venables WN, Ripley BD. 2002. *Modern Applied Statistics with S.* Fourth Edition. Springer, New York.

Webster R, Oliver MA. 2007. *Geostatistics for Environmental Scientists*, 2nd edition. Wiley, Chichester.

Whipple KX. 2004. Bedrock rivers and the geomorphology of active orogens. *Annual Reviews of Earth and Planetary Science* **32**: 151–185.

Whittaker AC, Cowie PA, Attal M, Tucker GE, Roberts GP. 2007. Bedrock channel adjustment to tectonic forcing: Implications for predicting river incision rates. *Geology* **35**: 103–06.

Zhu Z, Wallter E. 2003. Global Forest Cover Mapping for the United Nations Food and Agriculture Organization Forest Resources Assessment 2000 Program. *Forest Science* **49**: 369–380.

List of Figures and Captions

Figure 1 Maps of the study region and the distribution of sites (n=472) at which channel wetted widths were measured: a) elevation, b) BFIHOST values. The red discs are sites where wetted widths were measured from air photos, the green discs (n=28) are sites where field measurements of wetted width were also undertaken. The three gauging stations referred to in the text are shown as orange (Cwm Llanerch), yellow (River Alyn) and blue squares (Wistaston Brook), respectively. [Supplied in colour for online publication]

Figure 2 Daily flow measurements at three gauging stations in the study region (Figure 1) between 2009 and 2011 : a) River Alyn, b) Wistaston Brook, and c) Cwm Llanerch. The vertical red lines shows the dates on which air photos were captured across the study region. Supplied in colour for online publication.

Figure 3 Scatterplot of the maximum difference between three separate measurements (metres) of channel WW at each stream site (n=472) against the average of the three measurements at the same site.

Figure 4 Scatterplot of measurements (from air photos) and predictions of channel WW for linear models with only catchment area as a predictor (CA model; red open discs) and all statistically significant predictors (full model; blue open discs). [Supplied in colour for online publication]

Figure 5 Visualization of the interaction between between catchment area ($CA^{0.5}$) and cBFIHOST and its effect on channel wetted width (m) for the study region. The greyscale shading shows the regression model predictions of wetted width for different combinations of catchment area and cBFIHOST values.

Table 1 Discharge measurements and corresponding percentiles on a flow duration curve (FDC) for dates over a 3 year period at the three gauging stations (Figure 1) on the four dates (date 1=11 May 2009, date 2=01 June 2009, date 3=24 April 2010 and date 4=11 October 2011) when air photos were captured across the study region.

Station	Cwm Llanerch				River Alyn				Wistaston Brook			
Date	1	2	3	4	1	2	3	4	1	2	3	4
Discharge (m^3s^{-1})	3.2	2.8	1.8	14.9	0.68	0.62	1.1	0.41	0.47	0.3	0.37	0.28
Percentile on FDC	23	18	13	69	47	32	59	33	61	32	46	25
[a]Mean flow for station	17.1				1.8				0.55			

[a] mean for daily observations over three year period

Table 2 Summary statistics for field-based measurements (n=28) of channel wetted width[a] and the same statistics calculated using the absolute differences (n=28) between the field measurements and the estimates of wetted width based on air photos for the same sites.

	Field measurement wetted width (m)	Absolute difference between wetted widths (field measurement and air-photo; m)
Minimum	0.72	0.01
Mean	3.97	0.11
Median	3.07	0.09
St. Dev.	2.92	0.09
Maximum	13.3	0.34
Skewness	1.65	0.93
[b]RMSE	–	0.142
Bias	–	0.026
SDE	–	0.140

[a] all channels have an upstream catchment area >1 km^2 - see text

[b] root mean squared error (see text)

Table 3 Summary statistics for measurements of wetted width at 472 sites and the associated catchment or site-related data.

	[a]Width (m)	Catch. Area (km^2)	Elevation (m)	cBFIHOST	Site slope (°)	Catch. slope (°)	[b]Rainfall (mm)
Minimum	0.490	1.00	5.60	0.28	0.092	1.90	660
Mean	4.80	14.0	170	0.47	3.70	6.20	1300
Median	3.7	5.5	170.0	0.5	2.7	5.7	1000
Max	28.0	89.0	550	0.59	18.0	8.2	4100
St. Dev.	4.0	19.0	100	0.11	3.50	1.60	750
Skewness	2.20	2.24	0.33	-0.67	1.60	-1.30	1.80

[a] channel wetted width measured from air photo

[b] mean annual rainfall

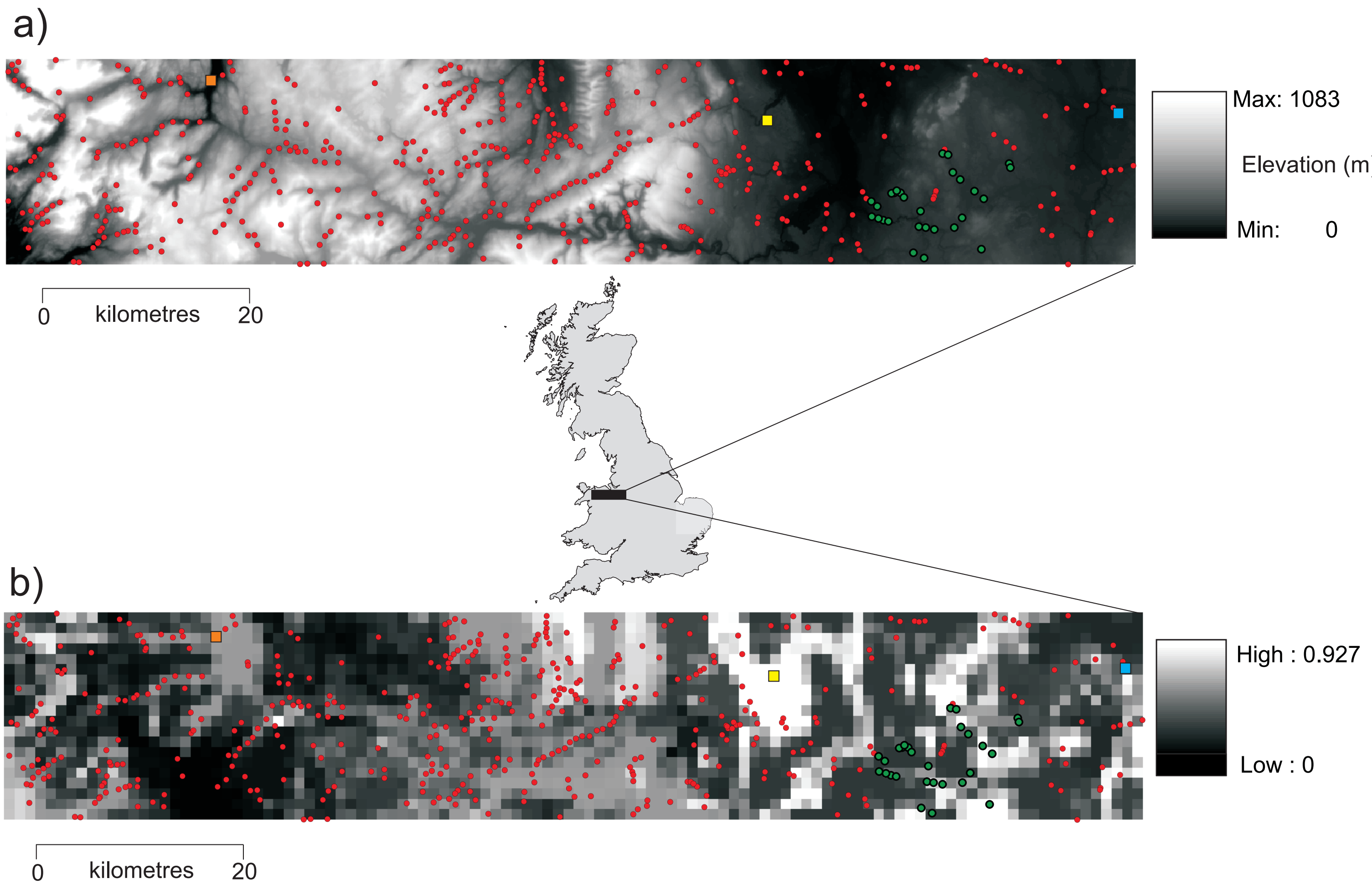
Table 4 Results of the linear models fitted by ordinary least squares

	Estimate	Std. Error	t-value	P-value
Intercept	-7.56	1.04	-7.24	$< 1.88 \times 10^{-12}$
[a]Catch. area$^{0.5}$	3.77	0.19	19.9	$< 2 \times 10^{-16}$
cBFIHOST	10.8	1.72	6.33	$< 2 \times 10^{-16}$
Catch. area$^{0.5}$:cBFIHOST	-4.96	0.39	-12.8	$< 2.0 \times 10^{-16}$
[b] Rainfall	8.6×10^{-4}	1.8×10^{-4}	4.73	$< 2.0 \times 10^{-16}$
Elevation	-0.002	9.8×10^{-4}	-2.27	0.0234
Catch. slope	0.27	0.07	3.71	0.002
Local slope	0.12	0.028	4.21	$< 2 \times 10^{-16}$

[a] Catchment area$^{0.5}$

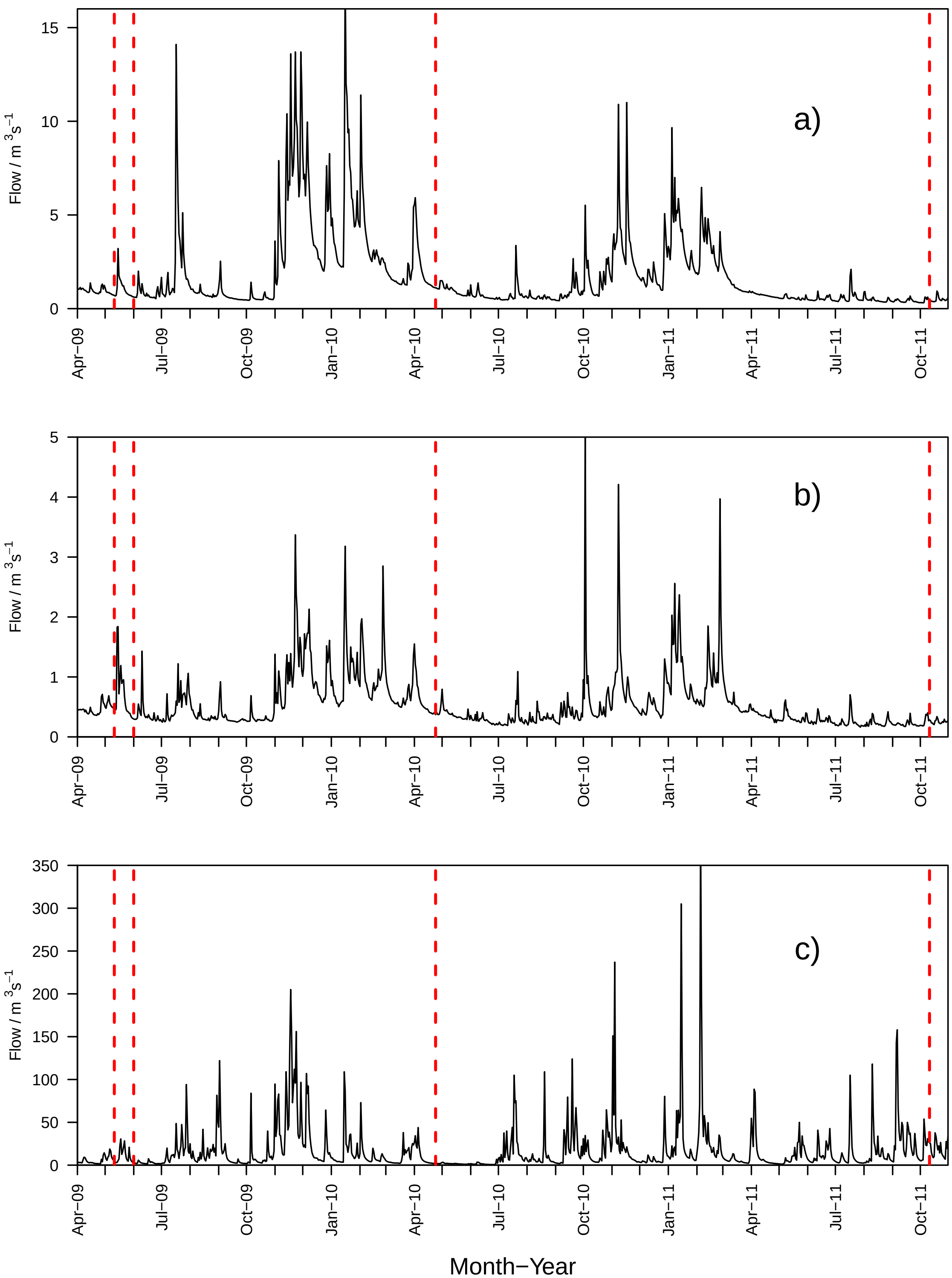
[b] mean annual rainfall (mm)

Figure 1:



1.pdf

Figure 2:



a)
b)
c)
Flow / m 3s−1
Month−Year
Apr−09
Jul−09
Oct−09
Jan−10
Apr−10
Jul−10
Oct−10
Jan−11
Apr−11
Jul−11
Oct−11

Figure 3:

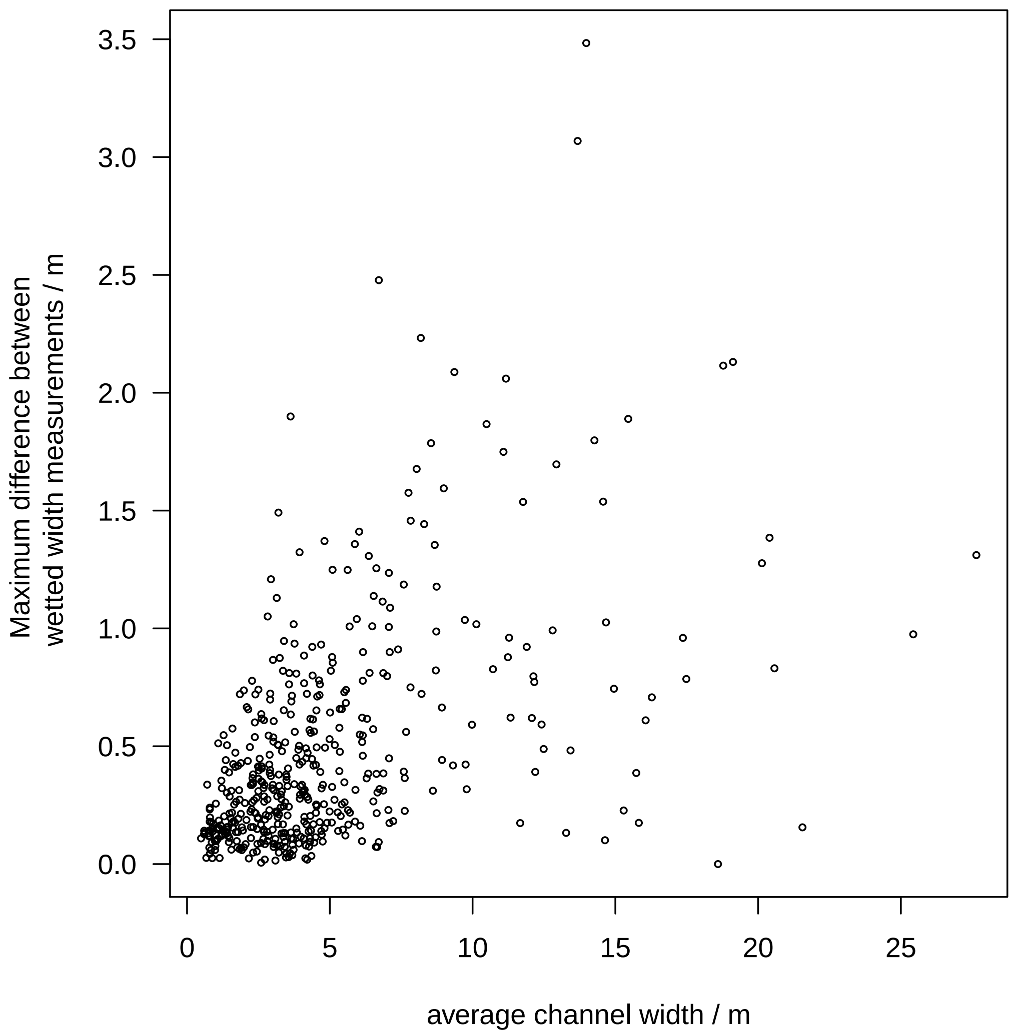
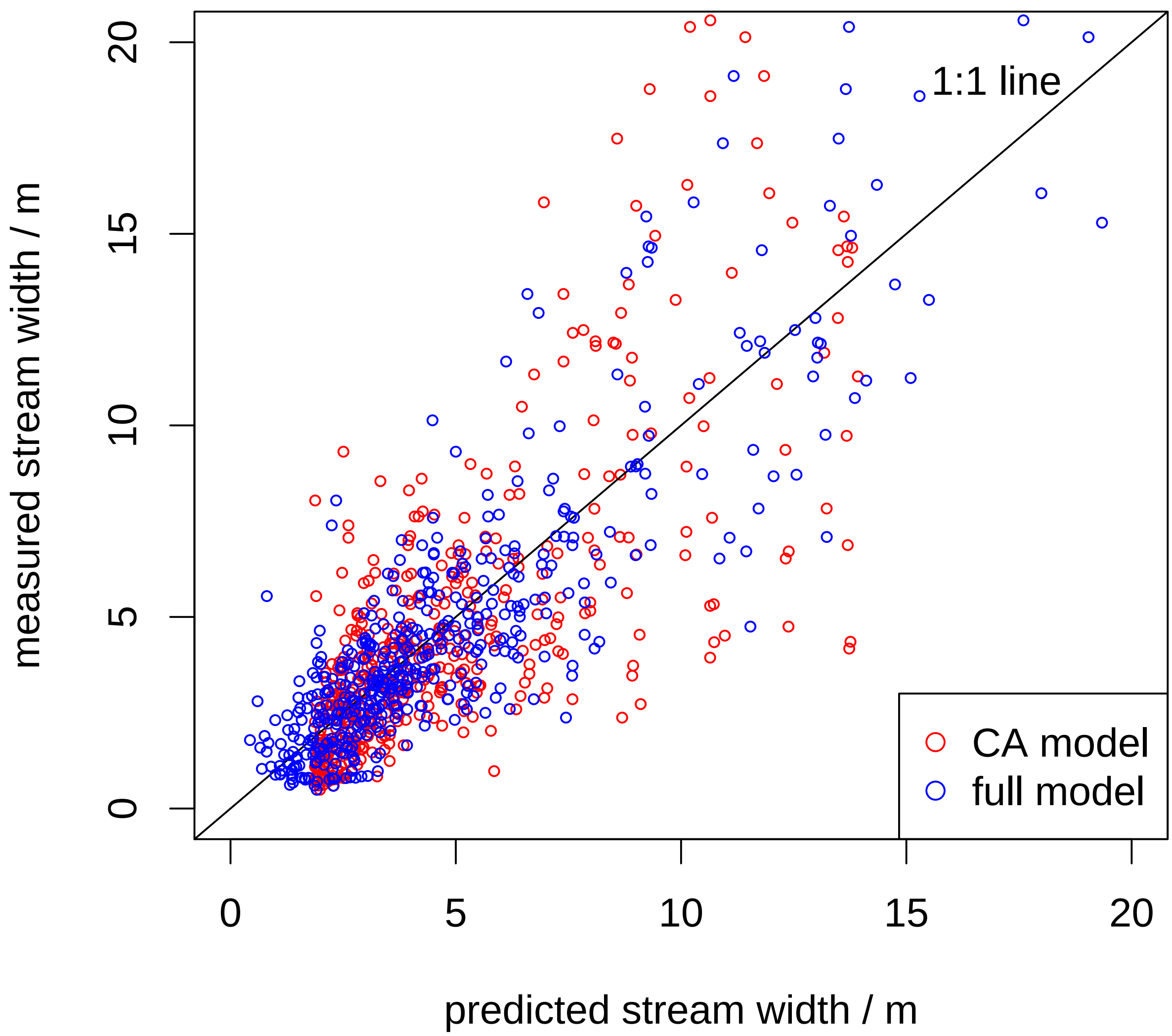
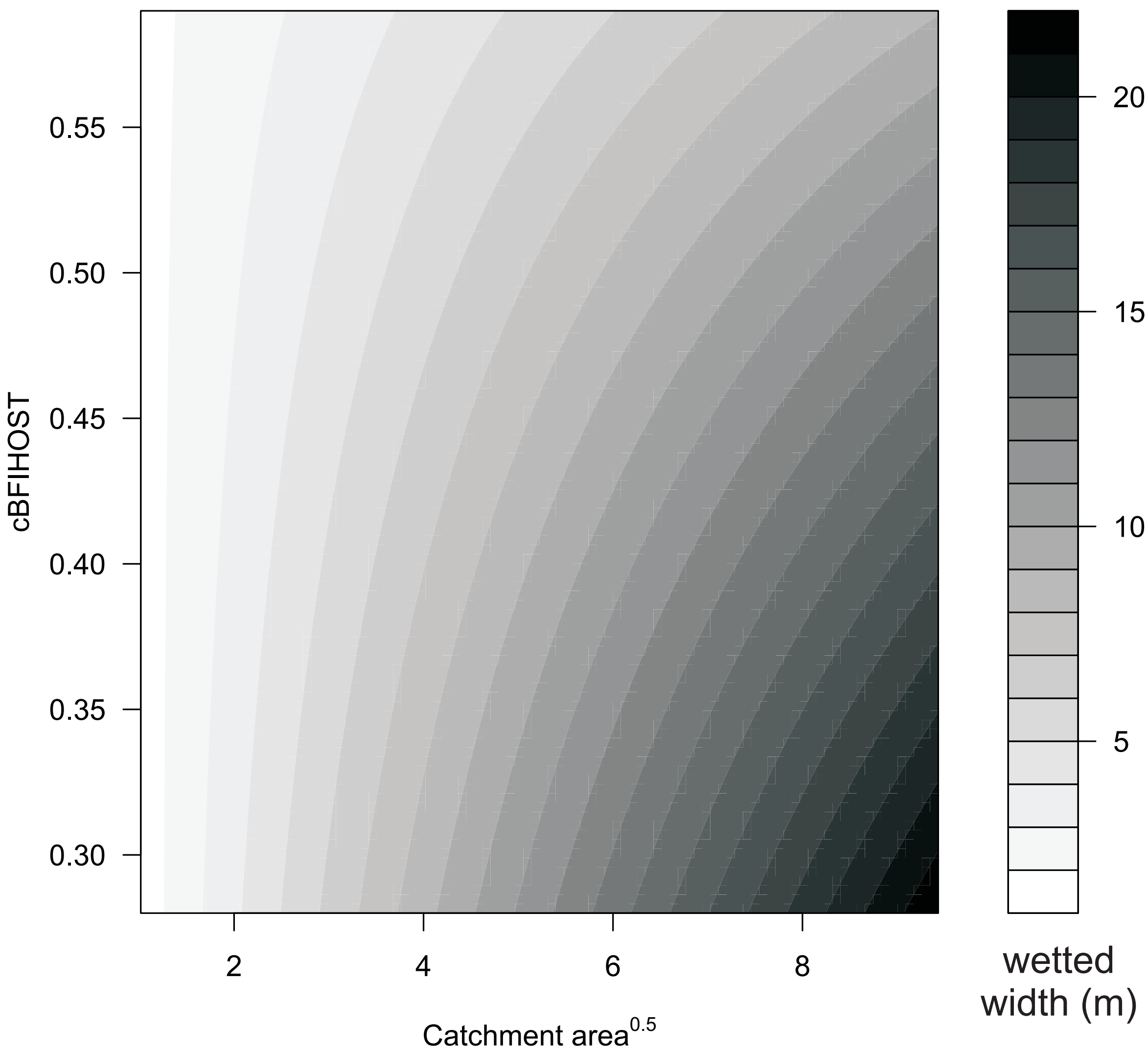

Maximum difference between
wetted width measurements / m
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
5
10
15
20
25
average channel width / m

Figure 4:



measured stream width / m
predicted stream width / m
0
5
10
15
20
1:1 line
CA model
full model

Figure 5:


0.55
0.50
0.45
0.40
0.35
0.30
cBFIHOST
2
4
6
8
Catchment area$^{0.5}$
20
15
10
5
wetted
width (m)