View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by MURAL - Maynooth University Research Archive Library

Stoch Environ Res Risk Assess (2011) 25:123–138
DOI 10.1007/s00477-010-0444-6

ORIGINAL PAPER

Links, comparisons and extensions of the geographically weighted regression model when used as a spatial predictor

Paul Harris · Chris Brunsdon · A. Stewart Fotheringham

Published online: 3 December 2010
© Springer-Verlag 2010

Abstract In this study, we link and compare the geographically weighted regression (GWR) model with the kriging with an external drift (KED) model of geostatistics. This includes empirical work where models are performance tested with respect to prediction and prediction uncertainty accuracy. In basic forms, GWR and KED (specified with local neighbourhoods) both cater for nonstationary correlations (i.e. the process is heteroskedastic with respect to relationships between the variable of interest and its covariates) and as such, can predict more accurately than models that do not. Furthermore, on specification of an additional heteroskedastic term to the same models (now with respect to a process variance), *locally-accurate* measures of prediction uncertainty can result. These heteroskedastic extensions of GWR and KED can be preferred to basic constructions, whose measures of prediction uncertainty are only ever likely to be *globally-accurate*. We evaluate both basic and heteroskedastic GWR and KED models using a case study data set, where data relationships are known to vary across space. Here GWR performs well with respect to the more involved KED model and as such, GWR is considered a viable alternative to the more established model in this particular comparison. Our study adds to a growing body of empirical evidence that GWR can be a worthy predictor; complementing its more usual guise as an exploratory technique for investigating relationships in multivariate spatial data sets.

Keywords Heteroskedastic · Local uncertainty · Relationship nonstationarity

P. Harris (✉) · A. S. Fotheringham
National Centre for Geocomputation, National University of Ireland, Maynooth, Co. Kildare, Ireland
e-mail: Paul.Harris@nuim.ie

C. Brunsdon
Department of Geography, University of Leicester, Leicester, UK

1 Introduction

A routine problem in spatial statistics is that of accurate prediction of an attribute at locations where no measurements have been taken and accurate estimates of the uncertainty surrounding such predictions. Spatial prediction is not only valuable to researchers who attempt to model spatial processes, but also to policy makers who need to plan and manage the outcomes of spatial processes at different spatial scales. Often overlooked in model specification is the importance of allowing: (a) data relationships and (b) data variability, to vary across space. In essence (and depending on the nature of the spatial process), incorporating the former specification has the potential to improve prediction accuracy, whilst incorporating the latter has the potential to improve prediction uncertainty accuracy. In this respect, a key aim of this study is to demonstrate the utility of basic (specification (a) only) and heteroskedastic (specifications (a) and (b)) geographically weighted regression (GWR) models in relation to comparable kriging with an external drift (KED) models.

The GWR models are evaluated in the context that KED is a best linear unbiased predictor (BLUP) and this statistical property ensures that it would be frequently the predictor of choice. The intention here is not to show that GWR is superior to KED, but that it may provide an alternative that in some circumstances is worth consideration for other reasons. The use of GWR as a predictor has

attracted much attention, where it has been empirically and *favourably* compared to: (i) alternative regressions (e.g. Zhang et al. 2005; Gao et al. 2006; Bitter et al. 2007; Kupfer and Farris 2007) or (ii) kriging (Páez et al. 2008). In contrast, GWR has been empirically and *unfavourably* compared to: (a) alternative regressions (Wheeler and Waller 2009; Salas et al. 2010) or (b) kriging (Lloyd 2010). Simulated data comparisons between GWR and kriging can be found in Harris et al. (2010a), where kriging is again favoured. Thus in context, our study is aimed at adding to the empirical evidence for using GWR as a predictor in relation to KED, but now we extend the empirical evaluation by calibrating heteroskedastic models, so that estimates of prediction uncertainty are evaluated. To our knowledge, only Harris et al. (2010a) has evaluated such measures, but for basic models only.

We now describe in detail the GWR and KED models, where critical linkages between them are stated and discussed. Models are then empirically evaluated using a freshwater acidification critical load data set for Great Britain, where relationships between contextual catchment data and critical loads are known to vary across space and where accurate local uncertainty outputs are vital to an informed management strategy for the protection of freshwaters. Although the results of our analyses may further understanding of the critical load process, the study has a firm methodological emphasis on the prediction techniques themselves.

2 Prediction methods

Both GWR and KED can be defined using $Z(\mathbf{x}) = m(\mathbf{x}) + R(\mathbf{x})$, where the random function $Z(\mathbf{x})$ is decomposed into a mean $m(\mathbf{x})$ and residual $R(\mathbf{x})$ component. Here $\mathbf{x}$ is any spatial location (observed or unobserved) and $z(\mathbf{x}_i)$ is the data with $i = 1, \ldots, n$. Both GWR and multiple linear regression (MLR), model $m(\mathbf{x})$ assuming $R(\mathbf{x})$ is a stationary random function with $\mathrm{E}\{R(\mathbf{x})\} = 0$ and $\mathrm{VAR}\{R(\mathbf{x})\} = \boldsymbol{\Sigma}$, where the elements of the diagonal $(n \times n)$ matrix $\Sigma = \sigma^2\mathbf{I}$ reflect zero spatial autocorrelation. In GWR, spatially-varying relationships between the dependent variable and its covariates are accounted for via locally weighted MLR fits that are calibrated from spatial data.

For KED, $m(\mathbf{x})$ is modelled with MLR, but now the elements of $\boldsymbol{\Sigma}$ reflect a structured covariogram $C(\mathbf{h})$, where $\mathbf{h}$ is the separation distance vector $\mathbf{h} = \mathbf{x}_i - \mathbf{x}_j$. As is standard practice, the elements of $\boldsymbol{\Sigma}$ are found from the variogram $\gamma(\mathbf{h})$ via the relationship $C(\mathbf{h}) = \sigma^2 - \gamma(\mathbf{h})$. Accordingly, $\boldsymbol{\Sigma}$ is a function of variogram parameters and can be denoted by $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, where for this study, $\boldsymbol{\theta}$ is a variogram parameter vector consisting of a nugget variance c_0; a structural variance c_1; and a correlation range a. The KED model can be calibrated in an implicit or explicit form (e.g. Bailey and Gatrell 1995), where the former caters for spatially-varying relationships, via the use of local neighbourhoods (as now the MLR component fit is also local; e.g. Wackernagel 2003, pp. 283–299). In this respect, this particular form of KED acts as a direct alternative to GWR.

For KED and GWR, prediction uncertainty at a point location can be accounted for via a conditional distribution defined by the prediction and its variance (i.e. an assumption of multivariate normality of $\mathbf{Z}$ is taken for both models). However for basic constructions, this approach is rarely recommended as the prediction variances are unlikely to vary in accordance with the variability in the local data used to provide the predictions (with respect to kriging see, Journel 1986; Switzer 1993; Heuvelink and Pebesma 2002; Harris et al. 2010b). As such, we adapt KED using the locally varying (variogram) sills methodology presented in Isaaks and Srivastava (1989, pp. 516–523), where variance non-stationarity is dealt directly and simply via a local correction of the KED variances. We call this heteroskedastic KED model, H-KED and as will be seen, we can locally-correct the KED variances in a number of different ways. Similarly for GWR, we address its stationary residual variance by calibrating a heteroskedastic version (H-GWR) (Fotheringham et al. 2002), which allows the residual variance to vary across space. Both H-KED and H-GWR should provide both *globally*- and *locally-accurate* prediction variances.

2.1 Geographically weighted regression (GWR)

For k independent covariates $y_1, y_2, \ldots y_k$, the MLR model can be written as $\mathbf{Z} = \mathbf{Y}\boldsymbol{\beta} + \mathbf{R}$, where $\mathbf{Z}$ is the $(n \times 1)$ sample (dependent) data vector, $\mathbf{Y}$ is the $(n \times k)$ covariate matrix, $\boldsymbol{\beta}$ is a $(k \times 1)$ vector of unknown parameters, and $\mathbf{R}$ is a $(n \times 1)$ residual vector. The ordinary least squares (OLS) parameter estimates $\hat{\boldsymbol{\beta}}$ are found from $\hat{\boldsymbol{\beta}} = \left(\mathbf{Y}^{\mathrm{T}}\mathbf{Y}\right)^{-1}\mathbf{Y}^{\mathrm{T}}\mathbf{Z}$, and $\hat{z}_{MLR}(\mathbf{x}) = \mathbf{y}(\mathbf{x})^{\mathrm{T}}\hat{\boldsymbol{\beta}}$ is the MLR prediction at $\mathbf{x}$ (where $\mathbf{y}(\mathbf{x})$ is a $(k \times 1)$ vector of covariates at $\mathbf{x}$). The corresponding GWR model has parameter estimates $\hat{\boldsymbol{\beta}}(\mathbf{x})$ found from $\hat{\boldsymbol{\beta}}(\mathbf{x}) = \left(\mathbf{Y}^{\mathrm{T}}\mathbf{W}(\mathbf{x})\mathbf{Y}\right)^{-1}\mathbf{Y}^{\mathrm{T}}\mathbf{W}(\mathbf{x})\mathbf{Z}$ where $\mathbf{W}(\mathbf{x})$ is a $(n \times n)$ diagonal matrix of spatial weights. The GWR prediction at $\mathbf{x}$ is:

$$\hat{z}_{GWR}(\mathbf{x}) = \mathbf{y}(\mathbf{x})^{\mathrm{T}}\hat{\boldsymbol{\beta}}(\mathbf{x}) \tag{1}$$

and the GWR prediction variance[1] at $\mathbf{x}$ is estimated using:

$$\sigma^2_{GWR}(\mathbf{x}) = \mathrm{VAR}\{\hat{z}(\mathbf{x}) - z(\mathbf{x})\} = \hat{\sigma}^2[1 + \mathrm{S}(\mathbf{x})] \tag{2}$$

Here $\mathbf{S}(\mathbf{x}) = \mathbf{y}(\mathbf{x})^{\mathrm{T}}[\mathbf{Y}^{\mathrm{T}}\mathbf{W}(\mathbf{x})\mathbf{Y}]^{-1}\mathbf{Y}^{\mathrm{T}}\mathbf{W}^2(\mathbf{x})\mathbf{Y}[\mathbf{Y}^{\mathrm{T}}\mathbf{W}(\mathbf{x})\mathbf{Y}]^{-1}\mathbf{y}(\mathbf{x})$ and the residual variance is taken as $\hat{\sigma}^2 = \mathrm{RSS}/(n - \mathrm{ENP})$,

[1] For corresponding MLR prediction variances use: $\mathrm{S}(\mathbf{x}) = \mathbf{y}(\mathbf{x})^{\mathrm{T}}[\mathbf{Y}^{\mathrm{T}}\mathbf{Y}]^{-1}\mathbf{y}(\mathbf{x})$.

where RSS is the residual sum of squares and ENP is the effective number of parameters of the GWR fit (Leung et al. 2000).

For our case study, the weighting matrix in GWR is specified using an (isotropic) exponential kernel, which was chosen from preliminary experimentation with GWR fits using a number of kernel types. For this continuous, distance-decay kernel, weights at location $\mathbf{x}$ accord to $w(\mathbf{x}) = \exp(-\mathrm{d_i}/\mathrm{b})$ where the bandwidth parameter is the distance b and d_i is the distance between $\mathbf{x}$ and a sample location i. An optimal bandwidth is found with an adaptive technique[2] using *leave-one-out* cross-validation, where the root mean squared prediction error (RMSPE) is calculated for a range of bandwidths and the bandwidth that gives the minimum RMSPE is considered optimal.

2.2 Heteroskedastic GWR (H-GWR)

A H-GWR model is calibrated by replacing the global $\hat{\sigma}^2$ in expression 2 with a local estimate $\hat{\sigma}^2(\mathbf{x})$. The model has an iterative form, where at each iteration step, a re-weighted GWR fit is found (Fotheringham et al. 2002, pp. 80–82). The re-weighting is based on a mean smoothing over the observed squared residuals, which is used to provide estimates of $\sigma^2(\mathbf{x})$. In particular, the usual geographic weight at $\mathbf{x}$ is multiplied by the inverse of $\hat{\sigma}^2(\mathbf{x})$, where this second weighting corrects for local heteroskedasticity and is analogous to the use of weighted least squares (WLS) in MLR to stabilise a non-constant residual variance. The algorithm is applied with updated estimates of $\hat{\boldsymbol{\beta}}(\mathbf{x})$ and $\hat{\sigma}^2(\mathbf{x})$, until an acceptable level of convergence is reached. As the parameter estimates $\hat{\boldsymbol{\beta}}(\mathbf{x})$ are updated (i.e. $\hat{\boldsymbol{\beta}}(\mathbf{x}) = \hat{\boldsymbol{\beta}}_{\mathrm{U}}(\mathbf{x})$), then the H-GWR prediction at $\mathbf{x}$ is also updated to give:

$$\hat{z}_{HGWR}(\mathbf{x}) = \mathbf{y}(\mathbf{x})^{\mathrm{T}}\hat{\boldsymbol{\beta}}_{\mathrm{U}}(\mathbf{x}) \tag{3}$$

and the final $\hat{\sigma}^2(\mathbf{x})$ replaces $\hat{\sigma}^2$ in expression 2 to give this prediction variance at $\mathbf{x}$:

$$\sigma^2_{HGWR}(\mathbf{x}) = \hat{\sigma}^2(\mathbf{x})[1 + \mathrm{S}(\mathbf{x})] \tag{4}$$

Thus H-GWR has the potential to improve prediction accuracy, as well as prediction uncertainty accuracy over a basic GWR model. The bandwidth (and kernel function) specified with H-GWR is the same as that found optimally with basic GWR. The same bandwidth and kernel function is also used to smooth the observed squared residuals.[3] The local mean smoother used, is given in Sect. 2.4, expression 7.

2.3 Kriging with an external drift (KED)

For parameter estimation in KED, we use restricted maximum likelihood (REML), to first identify relatively unbiased estimates of $\boldsymbol{\Sigma_\theta}$ and then in turn, relatively unbiased estimates of $\boldsymbol{\beta}$ via generalised least squares (GLS) (e.g. Schabenberger and Gotway 2005, pp. 259–263). Here, and after some initial experimentation, we specify an (isotropic) exponential variogram model-type only, i.e., $\gamma(h) = c_0 + c_1(1 - \exp(-h/a))$. Thus $\hat{\boldsymbol{\beta}}_{GLS} = \left(\mathbf{Y}^{\mathrm{T}}[\boldsymbol{\Sigma}_\theta]^{-1}\mathbf{Y}\right)^{-1}\mathbf{Y}^{\mathrm{T}}[\boldsymbol{\Sigma}_\theta]^{-1}\mathbf{Z}$ provides the parameter estimates for the MLR component and the KED prediction at $\mathbf{x}$ is:

$$\hat{z}_{KED}(\mathbf{x}) = \mathbf{y}(\mathbf{x})^{\mathrm{T}}\hat{\boldsymbol{\beta}}_{GLS} + \boldsymbol{\sigma}_\theta^{\mathrm{T}}[\boldsymbol{\Sigma}_\theta]^{-1}\left(\mathbf{Z} - \mathbf{Y}\hat{\boldsymbol{\beta}}_{GLS}\right) \tag{5}$$

where $\boldsymbol{\sigma_\theta}$ is a $(n \times 1)$ vector of spatial covariances between residuals at $\mathbf{x}$ and the sample locations. The KED variance at $\mathbf{x}$ is:

$$\begin{aligned}\sigma^2_{KED}(\mathbf{x}) = &\left\{\hat{\sigma}^2 - \boldsymbol{\sigma}_\theta^{\mathrm{T}}[\boldsymbol{\Sigma}_\theta]^{-1}\boldsymbol{\sigma}_\theta\right\} \\ &+ \left\{\left(\mathbf{y}(\mathbf{x}) - \mathbf{Y}^{\mathrm{T}}[\boldsymbol{\Sigma}_\theta]^{-1}\boldsymbol{\sigma}_\theta\right)^{\mathrm{T}}\left(\mathbf{Y}^{\mathrm{T}}[\boldsymbol{\Sigma}_\theta]^{-1}\mathbf{Y}\right)^{-1}\right. \\ &\left.\times\left(\mathbf{y}(\mathbf{x}) - \mathbf{Y}^{\mathrm{T}}[\boldsymbol{\Sigma}_\theta]^{-1}\boldsymbol{\sigma}_\theta\right)\right\}\end{aligned} \tag{6}$$

and the KED weights at $\mathbf{x}$ are $\boldsymbol{\lambda}_{KED}(\mathbf{x}) = [\boldsymbol{\Sigma_\theta}]^{-1}\boldsymbol{\sigma_\theta}$. Observe that $\hat{\sigma}^2$ is the estimate of the residual variogram sill $(c_0 + c_1)$, where c_0 and c_1 are partial sills.

To find $\hat{z}_{KED}(\mathbf{x})$ and $\sigma^2_{KED}(\mathbf{x})$, an implicit solution is adopted when KED is specified with local neighbourhoods. Neighbourhood size is chosen optimally using the same *leave-one-out* cross-validation procedure as that used to find the bandwidth in GWR.[4] Technically, local residual variogram parameters should be estimated that are specific to $\mathbf{x}$, but instead, the parameters of the global residual variogram are retained. When KED is approximated in this manner, it is no longer a BLUP (Chilès and Delfiner 1999, p. 201). However, the approximation usually has a minimal effect on (overall) prediction accuracy and is often used to reduce computational burden. The approximation is not so easily justified from a prediction uncertainty viewpoint as the KED variance depends strongly on the variogram, except that reliable (automatic) local variography is rarely viable (Schabenberger and Gotway 2005, pp. 425–426).

[2] In this case the bandwidth is a nonlinear parameter, which reflects a fixed local sample size that exerts the greatest influence on each local regression fit.

[3] This is a pragmatic modelling decision where an alternative would be to find an optimal bandwidth at each iteration step of the H-GWR fit. Further work could address this issue.

[4] In this respect, both GWR and KED are optimised for prediction accuracy only. For KED, a more succinct approach for neighbourhood selection is described in Rivoirard (1987), but is not used here as it would compromise our GWR to KED comparison.

2.4 Heteroskedastic KED (H-KED)

For our H-KED model, we first model some simple relationship between the local means and variances (of the raw data) and then transfer this relationship to locally correct the KED variances in relation to their predictions. This approach implicitly assumes that (unseen) local variograms vary proportionally across space, which in turn entails that they only differ from the global variogram by some re-scaling factor. As the KED variances are primarily a function of the variogram,[5] then they too are assumed proportional to (the sill of) the global variogram. Consequently, the global variogram and the KED variances can be locally adjusted by the same re-scaling factor. Ignoring local variogram proportionality when applying KED does not affect prediction accuracy, but does result in an over-estimation of the KED variance at locations where a local variogram's sill would be smaller than that defined by the global variogram (and vice versa). Problems with the H-KED approach centre on a tenuous assumption of variogram proportionality. An alternative would be to calibrate KED with the true local variograms (see above), but H-KED is much simpler, as only one (not multiple) variograms are needed.

In this study, the following H-KED procedure is adopted. Firstly, KED predictions and *relative* KED variances are found at target sites, via the use of a variogram model standardised to a sill of one. To find the H-KED variances, the *relative* KED variances need to be multiplied by local variance estimates at the same sites. Here the local variance estimates are found indirectly via a modelled relationship to local means. Thus (initial) local mean and local variance estimates are found (at the calibration sites, say) and then this (local statistic) data is used to calibrate some regression, where the variance is equated to some function of the mean. This regression is then solved at target sites, where the KED predictions are used as the local mean estimates so that the local variance estimates are found. Our (initial) local mean $m(\mathbf{x})$ and variance $s^2(\mathbf{x})$ estimates are found in a geographically weighted (GW) form using:

$$m(\mathbf{x}) = \sum_{i=1}^{n} w_i z(\mathbf{x}_i) \Big/ \sum_{i=1}^{n} w_i \tag{7}$$

$$s^2(\mathbf{x}) = \sum_{i=1}^{n} w_i (z(\mathbf{x}_i) - m(\mathbf{x}))^2 \Big/ \sum_{i=1}^{n} w_i \tag{8}$$

where the weights w_i accord to the same exponential kernel function defined before. Both the GW mean and GW variance calculations are specified using the same (adaptive) bandwidth. This bandwidth is chosen by judgement, where first, an optimal bandwidth for the GW mean is found using *leave-one-out* cross-validation. This bandwidth is then increased if it is considered too small to be used for the GW variance estimates (or vice versa), where the plot of GW means against GW variances is considered too scattered (or too smooth). This procedure is conducted so that simple but *true* relationships with this local statistic data are favoured, which in turn, promotes parsimonious regressions.[6]

2.5 Linkages between the methods

A. For KED, it is assumed that the residual process is worth modelling, whereas for GWR the residual process is assumed random. Here GWR strives to model spatial patterns through variation in regression coefficients (only) and as such, tends to use more parameters to describe the mean structure than KED does. On balance, KED tends to be the more complex model.
B. Model calibration for GWR and KED is influenced by one kernel or one variogram function, respectively. The kernel function in GWR *directly* weights data, whereas the variogram in KED *indirectly* weights data. For GWR, dependent and covariate data are weighted at each calibration point, whereas for KED, only dependent data are weighted (as the trend can be filtered out via the use of constraints, e.g. see Goovaerts 1997).
C. Weights in KED will not decay uniformly from a calibration point as in GWR, since accounting for spatial dependence leads to desirable weighting effects (i.e. information/screening/relay) that accord to the data's spatial configuration (Chilès and Delfiner 1999, p. 205). Thus spatial configuration will directly affect any performance comparison between GWR and KED.[7]
D. For KED in local neighbourhoods (KED-LN), each local MLR fit relates to GWR specified with a box-car kernel (i.e. weights accord to $w(\mathbf{x}) = 1$ if $d_i \leq \tau$ and $w(\mathbf{x}) = 0$ if $d_i > \tau$, where τ is the bandwidth). Differences still exist as the former uses GLS estimation locally, whilst the latter uses OLS estimation locally.
E. It follows from point D, the only instance when GWR can be considered a special case of KED-LN is when GWR is specified with a box-car kernel and KED-LN is specified with a nugget variogram (provided the same bandwidth/neighbourhood is specified); both of which generally do not happen in practice. This is analogous to an equivalence of MLR and KED in a global neighbourhood (KED-GN).

[5] Observe that the first part of the KED variance in expression 6 represents the kriging variance of the residuals and the second part is a consequence of estimating the MLR trend component.

[6] Clear relationships can also be promoted by replacing the GW variance estimates with GW standard deviation (SD) estimates (and adapting the whole H-KED procedure accordingly).

[7] Laslett (1994) provides a general view on this issue with respect to kriging versus splines.

F. With respect to each model's matrix algebra (and the use of box-car kernels and nugget variograms aside), the WLS weights matrix $\mathbf{W}(\mathbf{x})$ of GWR does not equate to the GLS variance–covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\mathbf{x})$ of KED-LN ($\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\mathbf{x})$ refers to specific rows and columns of $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ at $\mathbf{x}$ according to the size of the neighbourhood). Here, the diagonal elements of $\mathbf{W}(\mathbf{x})$ are nonstationary, whilst those of $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\mathbf{x})$ are stationary. Furthermore, $\mathbf{W}(\mathbf{x})$ always has zero off-diagonal entries, whereas those of $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\mathbf{x})$ are never all non-zero. For these reasons, the trend component of KED-LN (or KED) cannot equate to some GWR model.

G. The critical difference between GWR and KED is how each model is able to use information in relation to modelling spatially-varying relationships. For GWR specified with a continuous distance-decay kernel all sample data is used and a different set of regression parameters can be found at target locations. If KED uses all the data, the same set of regression parameters result at target locations. Thus KED can only model spatially-varying relationships, via data subsets. However when local subsets are used, GWR is not limited to a box-car form (as in KED-LN), since GWR can use a distance-decay kernel, such as the bi-square (i.e. $w(\mathbf{x}) = (1 - (d_i/\tau)^2)^2$ if $d_i \leq \tau$ and $w(\mathbf{x}) = 0$ if $d_i > \tau$).

H. Point G enables GWR to be a more flexible (and robust) model. This is particularly pertinent when covariates are not continuous and/or locally collinear, since KED-LN will fail across a range of neighbourhoods due to matrix instability (e.g. Deutsch and Journel 1998, p. 71), whilst GWR specified with some distance-decay kernel is usually able to circumvent such problems. For instances where collinearity problems cannot be adequately addressed with basic GWR, the ridge or lasso GWR models of Wheeler (2007, 2009) would be a more formal way of dealing with them. These variants of GWR are also known to improve prediction accuracy over basic GWR in such data situations. Presumably, the trend component of KED-LN could be similarly adapted, if collinearity problems occur.

I. Finally, our chosen H-GWR and H-KED models are deliberately simple. Here, we could specify our H-GWR model using the same approach as that used in H-KED and similarly, our H-KED model could follow a similar approach to that used in H-GWR. Instead, we calibrate heteroskedastic models that are likely to be familiar to each model's respective audience.

3 Case study: data and model assessment diagnostics

3.1 Case study data

Our case study data is a multivariate freshwater acidification critical load data set covering Great Britain (CLAG Freshwaters 1995).[8] The size and scale of this data suits nonstationary modelling, where a previous investigation with GWR provided evidence of space-varying relationships between contextual catchment data and critical load (Harris et al. 2010c). Here three continuous covariates and one class covariate are available. The continuous covariates are termed weighted geological sensitivity (Wt.GSP), weighted soil buffering capacity (Wt.SBCP) and weighted soil critical load (Wt.SCLP). The nominal nine-class covariate is termed dominant land cover (LC9D). Table 1 summarises the range of values that the weighted covariates can take according to an expected acid buffering capacity. Thus low critical load values would be expected to correspond to low Wt.GSP/Wt.SBCP/Wt.SCLP values (and vice versa). The land cover covariate is described in Table 2. Origins of the catchment data can be found in Kernan et al. (1998, 2001).

Table 1 Continuous weighted catchment covariates

Buffering capacity	Wt.GSP	Wt.SBCP	Wt.SCLP	Acid sensitivity
Low	1.0	10.0	0.1	High
↓	↓	↓	↓	↓
High	4.0	80.0	4.0	Low

Table 2 Nine-class land cover covariate

LC9D class	Description
1	Water and built/bare ground
2	Mown/grazed turf
3	Meadow/verge/semi-natural
4	Tilled land
5	Deciduous woodland
6	Coniferous woodland
7	Lowland semi-natural grass/moor
8	Upland semi-natural grass/bog moor
9	Upland semi-natural shrub moor

An unbalanced (or preferential) critical load sampling campaign entails that *declustered* data sets are used to ensure unbiased estimates of any (global) moment or model parameter. As such, (for this and previous studies), two data subsets were found which comprise of a spatially representative (*declustered*) data set of 497 sites for model calibration and a spatially representative (*declustered set-aside*) data set of 189 sites for model validation. This data post-processing also ensured that the distribution summary statistics for the calibration and validation data sets are almost identical. In summary, the calibration and validation data were chosen so that a well-judged comparison of model form is possible (see Fig. 1).

[8] Critical load data can be found at http://critloads.ceh.ac.uk/index.htm (last accessed 10 January 2009).

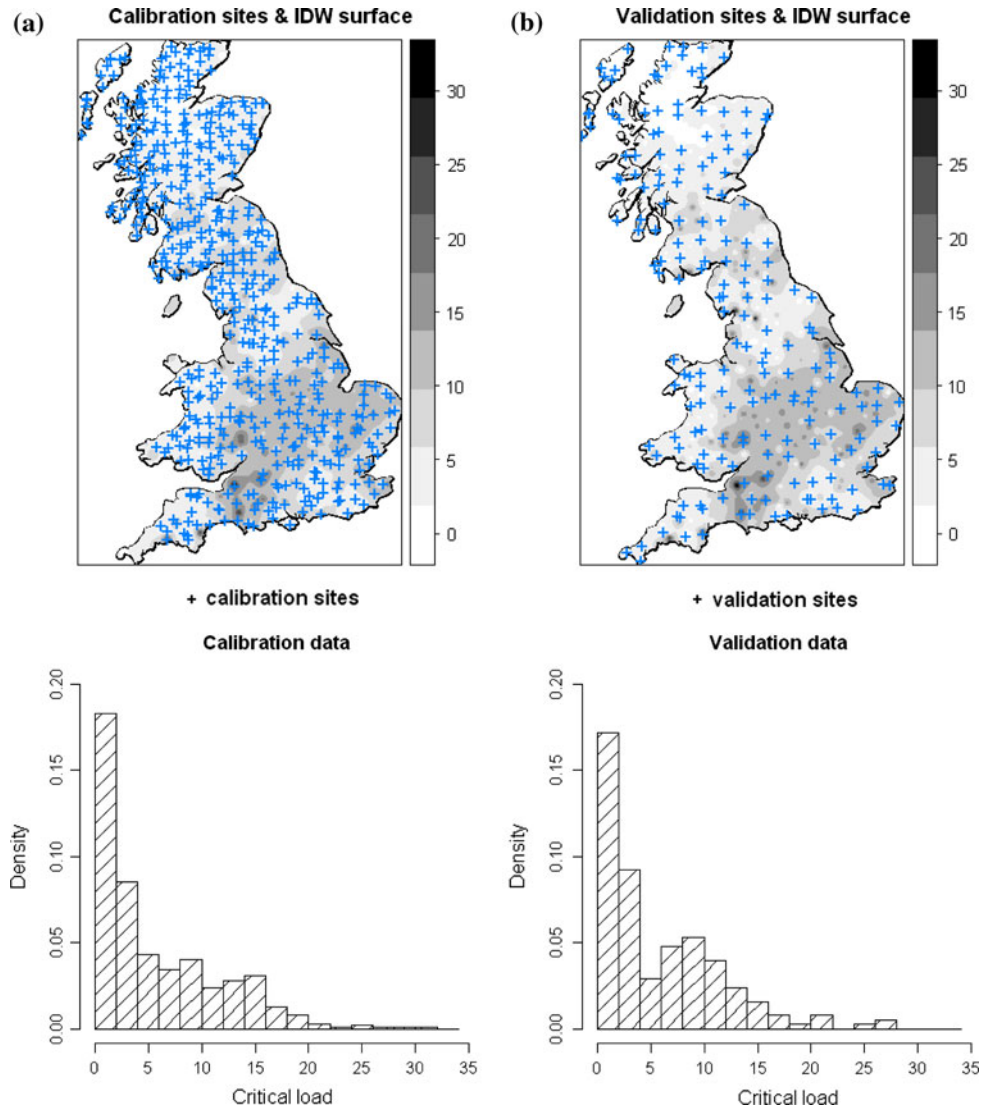

Fig. 1 Location and distribution of critical load data: **a** model calibration and **b** model validation. Maps shown with IDW fit to all critical load data for context (note that the Orkney and Shetland Islands in the far NE of Great Britain have no sampled sites). Units for critical load data are in keq H^+ ha^{-1} $year^{-1}$

Observe that we aim to predict site-specific threshold data. Further work (not presented here) would compare the resultant critical load predictions (and their estimates of uncertainty) with their corresponding deposition (contaminant) prediction model outputs; in order to find the (all important) critical load exceedance (critical load minus deposition) data. That is, apply the critical load concept at un-sampled freshwater sites (Nilsson and Grennfeld 1988). This dual and interactive prediction methodology (e.g. Van Meirvenne and Goovaerts 2001) should ultimately provide accurate estimates of critical load exceedance risk. For freshwater sites where this risk is high, acidification and associated environmental damage is likely. Such sites can then be targeted and managed accordingly. For Great Britain, the granite regions of Scotland and Wales are particularly affected by acidification.

It is not our intention to generalise the results from this study. Instead our aim is to present and interpret the results in the context of this particular data set. In our view, an empirical evaluation has merit and the concept of 'external objectivity' (Matheron 1989, p. 38) can be invoked, where the worth of a given predictor can be gauged by its performance in the 'long run' through an increasing number and variety of applications. For generalised results, simulated data comparisons can be more fruitful, where some initial steps have been undertaken in Harris et al. (2010a).

3.2 Model assessment diagnostics

For actual $z(\mathbf{x}_j)$ and predicted $\hat{z}(\mathbf{x}_j)$ data; model prediction accuracy is measured by: (a) the mean prediction error, $\mathrm{MPE} = (1/N)\sum_{j=1}^{N}\{z(\mathbf{x}_j) - \hat{z}(\mathbf{x}_j)\}$; $\mathrm{RMSPE} = \sqrt{(1/N)\sum_{j=1}^{N}\{z(\mathbf{x}_j) - \hat{z}(\mathbf{x}_j)\}^2}$; and the mean absolute prediction error, $\mathrm{MAPE} = (1/N)\sum_{i=1}^{N}|z(\mathbf{x}_j) - \hat{z}(\mathbf{x}_j)|$; where $N = 189$. Rank-based diagnostics are also found. Here the mean rank

of a model is defined as: $\mathrm{MRK}_m = (1/N)\sum_{j=1}^{N} r_{mj}$, where r_{mj} is the rank of the mth model for accurately predicting the jth observation. At each validation point, models are ranked according to the smallest absolute prediction error. The SD of the ranks is $\mathrm{SDRK}_m = \left[(1/(N-1))\sum_{j=1}^{N}\left(r_{mj} - \mathrm{MRK}_m\right)^2\right]^{0.5}$. Models that perform well should have a small mean rank and a small SD of ranks.

In a global-sense, prediction uncertainty accuracy can be measured using the mean squared deviation ratio, $\mathrm{MSDR} = (1/N)\sum_{j=1}^{N}\left(\left\{z(\mathbf{x}_j) - \hat{z}(\mathbf{x}_j)\right\}^2 \Big/ \sigma_P^2(\mathbf{x}_j)\right)$. Here $\mathrm{MSDR} < 1$ implies that the prediction variances $\sigma_P^2(\mathbf{x}_j)$ tend to over-estimate the squared prediction errors (and vice versa). In a local-sense, prediction confidence interval (PCI) accuracy can be assessed using coverage probabilities. For example, if symmetric 95% PCIs were calculated at each validation site, then a correct modelling of local uncertainty would entail there is a 0.95 chance that the actual value $z(\mathbf{x}_j)$ falls within the interval. If a coverage probability is found for a range of symmetric PCIs (say from a 1% to a 99% PCI in increments of 1%) and the results plotted against the probability interval p, then an accuracy plot is found. Accuracy plots can be summarised by the G-statistic, defined as $G = 1 - \int_0^1 [3a(p) - 2]\left[\bar{\xi}(p) - p\right]\mathrm{d}p$, where $\bar{\xi}$ is the fraction of actual values falling in the PCI, and a value of 1 is sought. The indicator function $a(p)$ is defined as

$$a(p) = \begin{cases} 1 & \text{if } \bar{\xi}(p) \geq p \\ 0 & \text{otherwise} \end{cases},$$

which implies that twice the importance is given to deviations when $\bar{\xi}(p) < p$. For cases where two models provide similar accuracy plots, one model can be preferred if its PCI widths that contain the actual value are smaller (i.e. more precise). Here the corresponding PCI width plots can be constructed and compared. For details use Goovaerts (2001).

4 Case study: analysis and results

Our first objective is to assess model prediction accuracy in relation to: (a) covariate subset selection; (b) the nature of any structure shown in the KED variogram; and (c) the nature of relationship nonstationarity between critical load and covariates. Here we only calibrate MLR, GWR, KED-GN and KED-LN models, where findings from this preliminary analysis should provide insight into the differences between GWR and KED-LN when predicting with space-varying relationships. This analysis is conducted using a *leave-one-out* approach, where the results are also used to provide a covariate subset for our second study objective, where we calibrate both basic and heteroskedastic models for prediction at the validation sites. In this focused analysis, we investigate prediction and prediction uncertainty accuracy for: MLR, GWR, KED-GN, KED-LN, H-GWR and H-KED.

4.1 Preliminary analysis

An AIC-led investigation presented in Harris et al. (2010c) indicated that: Critical Load = f(Wt.GSP, Wt.SBCP, LC9D2, LC9D3, LC9D4) provides the most parsimonious MLR fit; whereas a GWR model of this form: Critical Load = f(Wt.GSP, Wt.SBCP, Wt.SCLP, LC9D2, LC9D3, LC9D4) provides the most parsimonious local (and overall) fit (i.e. Wt.SCLP only appears to be important locally). This prior work also indicated that Wt.GSP and Wt.SCLP have the strongest nonstationary relationships with critical load. For this study, we build on such results and immediately discard the other six land class covariates. The retained land class covariates all have a positive relationship to critical load and to further simplify our analysis, we aggregate these covariates to form a single land class covariate, termed Arable. This covariate is similarly defined in the comparable critical load studies of Kernan et al. (1998, 2001) and the aggregation loses little in predictive information.

Therefore using our reduced covariate data set, we can now calibrate (a manageable) fifteen models for each of MLR, GWR, KED-GN and KED-LN (i.e. 60 models in total) and assess their prediction accuracy via RMSPE. Each of the fifteen models relates to a unique combination of covariates. In addition, we report the relative structural variability (RSV, where $\mathrm{RSV} = (c_1/(c_1 + c_0)) \times 100\%$) and the correlation range of each KED variogram. Here residual spatial dependence is viewed as strong when RSV values tend to 100% in conjunction with a long correlation range. Table 3 and Fig. 2 summarise the results of this exercise, where models are ranked according to the strength of spatial dependence found in their respective KED variogram. This ranking is based on a judged evaluation of the RSV and correlation range data (and is not intended to be definitive). All models are calibrated using covariate data with a small random error addition (between 0.0001 and 0.001). This was only necessary so that all 15 KED-LN models could be calibrated. The effect of this error addition (for models that could be calibrated with or without the error) was only observed in the fourth decimal place of the resultant RMSPE data.[9] This problem with KED-LN directly relates to point H of Sect. 2.5, where the nature of our covariate data results in exact (or near exact)

[9] The random error addition was not required for our focused analysis, as our chosen KED-LN model could be reliably calibrated without it.

Table 3 Calibration details for 60 model fits (see also Fig. 2)

Covariate subset	Model name	GWR Bandwidth	KED-LN Neighbourhood (%)	Variogram RSV	Variogram range	Variogram rank
All covariates	ALL	0.020	10	100.0	7.3	1
Wt.GSP + Wt.SBCP + Arable	G-SB-A	0.095	40	100.0	7.4	2
Wt.GSP + Wt.SCLP + Arable	G-SCL-A	0.022	8	100.0	7.5	3
Wt.GSP + Arable	G-A	0.034	22	100.0	7.8	4
Wt.SBCP + Arable	SB-A	0.034	57	100.0	9.0	5
Wt.SBCP + Wt.SCLP + Arable	SB-SCL-A	0.009	10	29.3	27.4	6
Wt.GSP + Wt.SBCP	G-SB	0.060	28	12.7	47.2	7
Wt.GSP + Wt.SBCP + Wt.SCLP	G-SB-SCL	0.022	26	12.1	58.9	8
Wt.SBCP	SB	0.022	93	20.8	78.6	9
Wt.SBCP + Wt.SCLP	SB-SCL	0.012	34	19.2	84.4	10
Wt.GSP + Wt.SCLP	G-SCL	0.022	10	18.0	113.7	11
Wt.GSP	G	0.022	10	20.8	116.0	12
Wt.SCLP + Arable	SCL-A	0.008	6	27.8	60.9	13
Wt.SCLP	SCL	0.008	10	32.0	152.0	14
Arable	A	0.007	12	40.4	61.0	15

A variogram rank of 1 refers to relatively weak spatial dependence

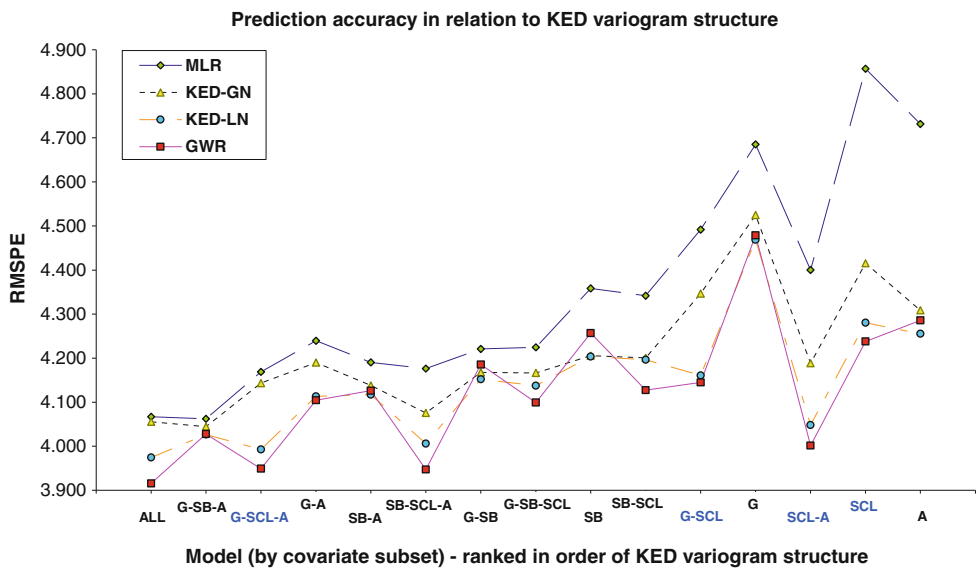

Fig. 2 Leave-one-out RMSPE data for 60 model fits. Models are ranked by the structure seen in the KED variogram, where models to the left relate to relatively weak spatial dependence. Covariate subset labels are explained in Table 3

local collinearity, which occurs more frequently as the neighbourhood is reduced.

From Table 3, the KED variograms with little structure (low ranks) tend to coincide with a well-informed trend component, which is expected. Observe that the first five ranked KED variograms essentially depict pure nugget (random) variation, as the correlation range is extremely small. In fact, all 15 KED variograms were estimated with a high value of c_0; reflecting a high level of underlying noise in the critical load data. This small scale random variation is expected from known flaws in the measurement and sampling strategy (Harris et al. 2010c). In summary, spatial dependence in all fifteen residual processes is not particularly strong. Also from Table 3, the size of the GWR bandwidth or KED-LN neighbourhood can indicate which data relationships are more nonstationary than stationary. Here Wt.GSP, Wt.SCLP and Arable appear to have a nonstationary relationship with critical load, whilst the large bandwidths/neighbourhoods associated with Wt.SBCP, suggest a stationary one.

From Fig. 2, some interesting and important results emerge. Here each model's prediction accuracy is plotted in relation to any structure seen in a model's corresponding KED variogram. Firstly, the most accurate set of models is

when all four covariates are specified. Here all four models perform similarly and as such, little is gained in specifying a spatial model in preference to MLR. This strong prediction performance coincides with little structure seen in the KED variogram; and in general, as the covariate subset is reduced, KED structure improves, but prediction accuracy weakens. A few GWR and KED-LN models still perform well however, with a reduced covariate data set. This is a model misspecification issue where seemingly important spatial effects can be (partially) attributable to missing covariates. In all fifteen model sets, MLR is always the poorest predictor, whereas in general, GWR is the most accurate predictor followed by KED-LN and KED-GN.

As would be expected, all models tend to perform in a similar manner if critical load to covariate relationships are stationary. This effect tends to occur when Wt.SBCP is included as a covariate (suggesting strength in this stationary relationship to critical load). In such cases, KED-GN (or possibly MLR) can be preferred to KED-LN or GWR. Conversely, strongly dissimilar model performances are found when some combination of Wt.GSP, Wt.SCLP and Arable are included as covariates. In these cases, MLR always performs poorly whilst GWR and KED-LN always perform relatively well (and should be preferred).

For KED-GN, its prediction improvement over MLR depends on the strength of structure seen in the KED variogram, where as would be expected, improvement is greatest when the KED variogram structure is strong. Although KED-GN does not account for nonstationary relationships, any short fall in the performance of its MLR trend component can be partially accounted for in the modelling of its residual component. In this respect, KED-GN often performs well with respect to GWR. This effect can be considered analogous to the usual analytical impasse when separating first- from second-order effects.

In the context of this study, we are most concerned with models fitted using a covariate subset where relationships are strongly nonstationary. This is apparent when the

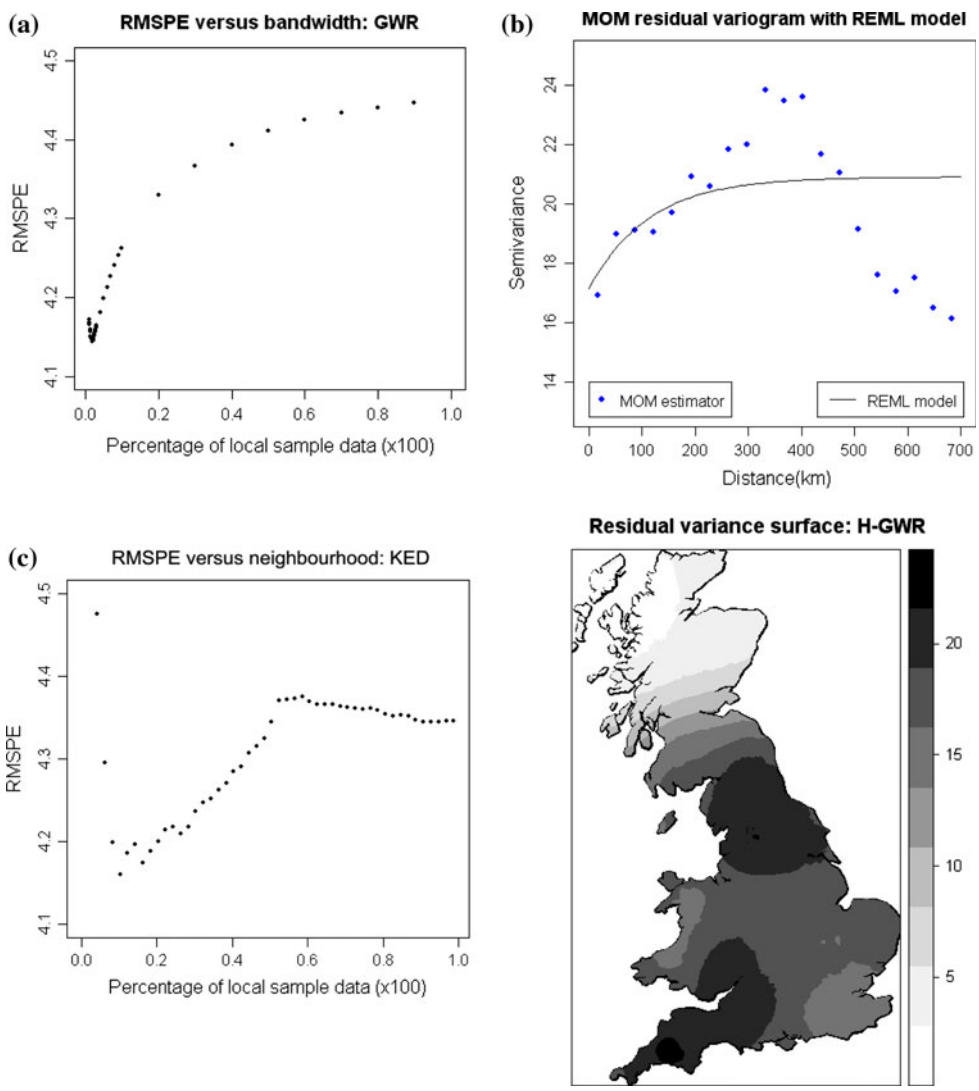

Fig. 3 **a** Bandwidth function for GWR (and H-GWR) calibration (shown with selective filtering of bandwidth size); **b** REML variogram fit for KED calibration; fit shown with method of moments (MOM) residual variogram estimator for context only; **c** KED neighbourhood function; **d** surface of $\hat{\sigma}^2(\mathbf{x})$ for H-GWR

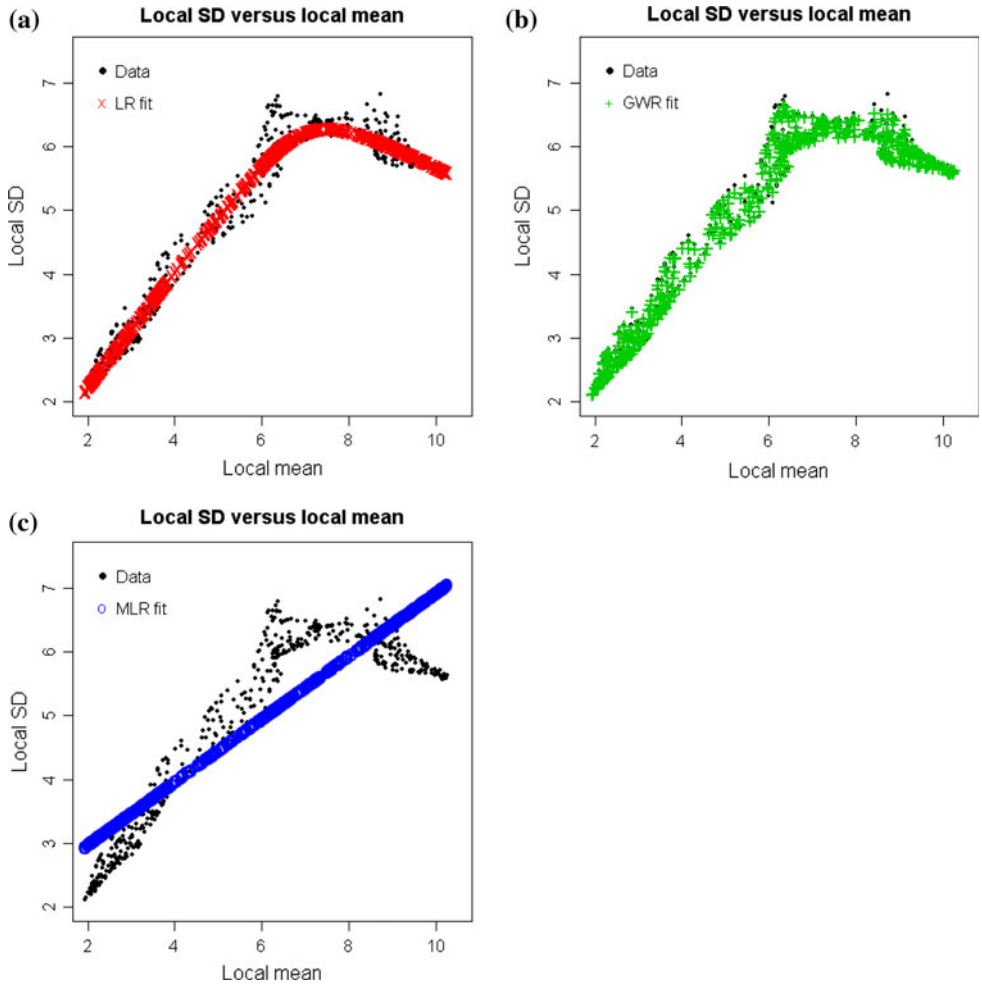

Fig. 4 Goodness of fit plots for the local mean and local SD regressions of the H-KED models: **a** LR; **b** GWR; and **c** MLR

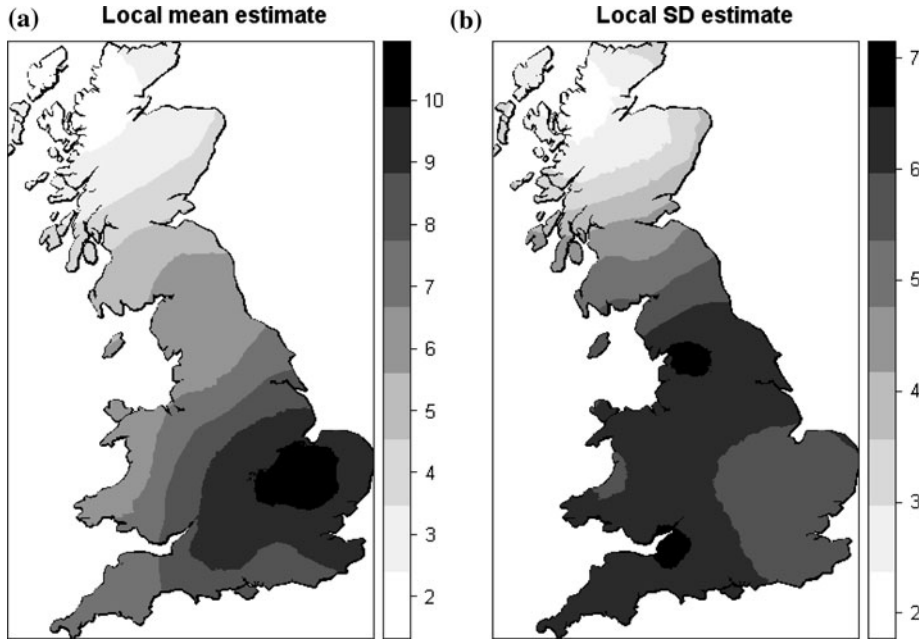

Fig. 5 Local **a** mean and **b** SD surfaces for H-KED models

difference in prediction performance between (i) MLR and GWR and (ii) KED-GN and KED-LN is greatest. Here four covariate subsets immediately stand out: (a) Wt.GSP + Wt.SCLP + Arable; (b) Wt.GSP + Wt.SCLP; (c) Wt.SCLP + Arable; and (d) Wt.SCLP only (these are highlighted in Fig. 2). In all four cases, either a KED-LN or a GWR model would be our chosen critical load predictor. Observe that the strength of structure seen in KED variogram does not play a role in differentiating between a KED-LN and GWR model. This empirically confirms the linkages between these models set out in Sect. 2.5, where GWR with anything other than a box-car kernel cannot be a special case of KED-LN or its trend component. If this were the case, differences in performance between the two models should increase as the KED variogram depicts greater structure. This effect is not apparent, not only across the four chosen covariate subsets, but also across all fifteen covariate subsets.

Finally, we can choose a covariate subset for a focused analysis where models are calibrated to predict at the validation sites. Here we can choose any one from the four covariate subsets above, as only in these cases would we consider applying KED-LN or GWR over a stationary counterpart. Choosing from these four covariates subsets should not be dismissed as a fabrication to promote our model comparison, as it is common for some covariates to be unavailable at some sites (Kernan et al. 1998, 2001). As such, we choose Wt.GSP + Wt.SCLP as our covariate subset, for no other reason than models using this covariate subset have been calibrated in related studies and thus an extended comparison is possible.

4.2 Focused analysis: model calibration

4.2.1 GWR and KED

The bandwidth function for our chosen GWR models is given in Fig. 3a. The function is well-behaved where a clear minimum is reached at a bandwidth of 2.2%. The exponential variogram model and the neighbourhood function for our chosen KED models are given in Fig. 3b, c, respectively. The variogram parameters are estimated at: $c_0 = 17.15$, $c_1 = 3.76$ and $a = 114$ km. The neighbourhood function is fairly well-behaved, where a size of 10% or 16% can be taken as optimal. The latter is chosen at it provides better prediction accuracy results at the validation sites.

4.2.2 H-GWR

To gauge residual variance nonstationarity with the GWR models, a surface of $\hat{\sigma}^2(\mathbf{x})$ for the fifth and final iteration of the H-GWR model is given in Fig. 3d. High residual variability appears in areas of northern and SW England, which can be attributable to pockets of (high) outlying critical load data. Residual variability is lowest in northern Scotland and directly relates to an area of low-valued critical loads and low critical load variation (see Fig. 5a, b). At the validation sites $1.86 \leq \hat{\sigma}^2(\mathbf{x}) \leq 21.94$ which can be compared to a global estimate of $\hat{\sigma}^2 = 14.08$ squared units used in the basic GWR model. Thus an assumption of residual variance stationarity in basic GWR can be viewed as rather strong.

4.2.3 H-KED

For the H-KED model, KED-LN is chosen for adaptation, since it predicts more accurately than KED-GN. Scatterplots for the (initial, GW) local mean and in this instance, local SD estimates (at the 497 calibration sites) are given in Fig. 4a–c, where a clear tail-off in linearity at high mean values directly relates to an area of unusually low critical load variability in SE England (cf. the corresponding surfaces in Fig. 5a, b). To model this nonlinearity, two nonparametric regressions to the mean and SD data are considered: (a) a local regression (LR) fit and (b) a GWR fit. For LR, nonstationarity and nonlinearity is modelled in attribute-space, whilst GWR models nonstationary and nonlinearity in geographic-space. In this instance, the LR fit is specified with a tri-cube weight function (together with a linear polynomial) and an optimal bandwidth is found using an AIC procedure (for details use Loader 2004). The GWR fit is specified with an exponential kernel and an optimal bandwidth is similarly found using AIC (for details use Fotheringham et al. 2002, pp. 61, 96). AIC procedures are preferred to cross-validation as they reduce instances of over-fitting, which can result in poor local SD estimates (including negatives) when the LR or GWR models are calibrated at the validation sites (especially in

Table 4 Prediction and prediction uncertainty accuracy at validation sites

Model	MPE	RMSPE	MAPE	MRK_m	$SDRK_m$	1-MSDR	*G*-statistic
MLR	0.13	4.47	3.22	3.48	1.64	0.007	0.942
GWR	0.09	4.05	2.82	2.98	1.10	−0.066	0.931
H-GWR	0.38	4.08	2.81	2.86	1.31	−0.062	0.957
KED-GN	0.10	4.09	2.87	3.00	1.47	0.110	0.913
KED-LN	0.03	3.91	2.72	2.68	1.39	0.204	0.892
H-KED-LR	0.03	3.91	2.72	2.68	1.39	−0.267	0.929
H-KED-GWR	0.03	3.91	2.72	2.68	1.39	−0.072	0.898
H-KED-MLR	0.03	3.91	2.72	2.68	1.39	0.344	0.887
H-KED-DV	0.03	3.91	2.72	2.68	1.39	0.485	0.857

All models use CL = *f*(Wt.GSP, Wt.SCLP) as the trend function

Calibration data mean used as a predictor gives MPE, RMSPE and MAPE values of −0.04, 5.61 and 4.59, respectively

extrapolation situations). In both cases, fairly well-defined minimums were reached. Figure 4a, b depict the LR and GWR fits to the mean and SD data. An AIC-defined MLR fit to the same data is also given for context (Fig. 4c); and is retained to calibrate its own (naïve) H-KED model. Here, GWR is the best fitting model (R^2 values are 0.99, 0.97 and 0.76, for GWR, LR and MLR, respectively) even though it is not the most complex (ENP values are 160.4, 492.0 and 2, respectively).

Thus the described LR, GWR and MLR models are solved at the validation sites using KED predictions as local mean estimates. This procedure provides local SD estimates, which are first squared and then, multiplied by relative KED variances in order to find the locally corrected KED variances. For LR, GWR and MLR, the following models result: H-KED-LR, H-KED-GWR and H-KED-MLR, respectively. In addition, a fourth (second naïve) H-KED model is calibrated where local variance estimates are found directly at the validation sites (i.e. simply using expression 8 with the calibration data and a bandwidth of 7%). This model is named H-KED-DV. It is envisaged that H-KED-KV will perform poorly in relation to the other three H-KED models as its local variance estimates do not benefit from a link (via the local statistic regression fits) to the spatial process as whole and will tend to a reflect an under-smoothed estimate of local variability (see discussions given in Switzer 1993).

4.3 Focused analysis: prediction accuracy

Prediction accuracy diagnostics for our (now nine) study models is given in Table 4, where KED-LN and the four H-KED models will provide the same results. Reassuringly, all models predict better than the calibration mean and diagnostics indicate that KED-LN is the most accurate predictor whilst MLR is the least accurate. Both GWR and H-GWR perform with merit. Observe that KED-GN performs little different to a GWR model, even though it does not account for relationship nonstationarity. This is a result of effects that have been described in Sect. 4.1, but also reflects KED-GN's BLUP status. Actual versus predicted scatterplots for each model (not shown) all indicated a high smoothing bias, where no one model could be considered a particularly good predictor of critical load.

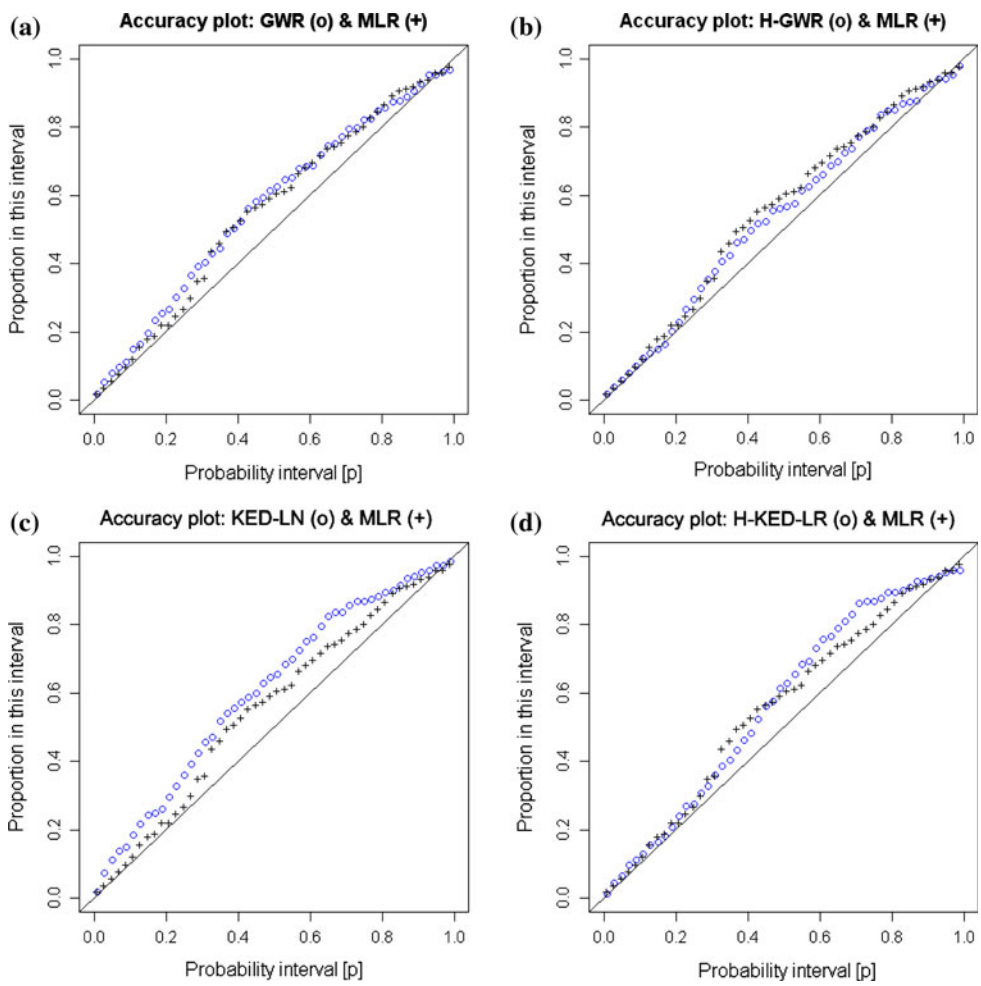

Fig. 6 Accuracy plots for: **a** GWR; **b** H-GWR; **c** KED-LN; and **d** H-KED-LR (all shown with naïve MLR model)

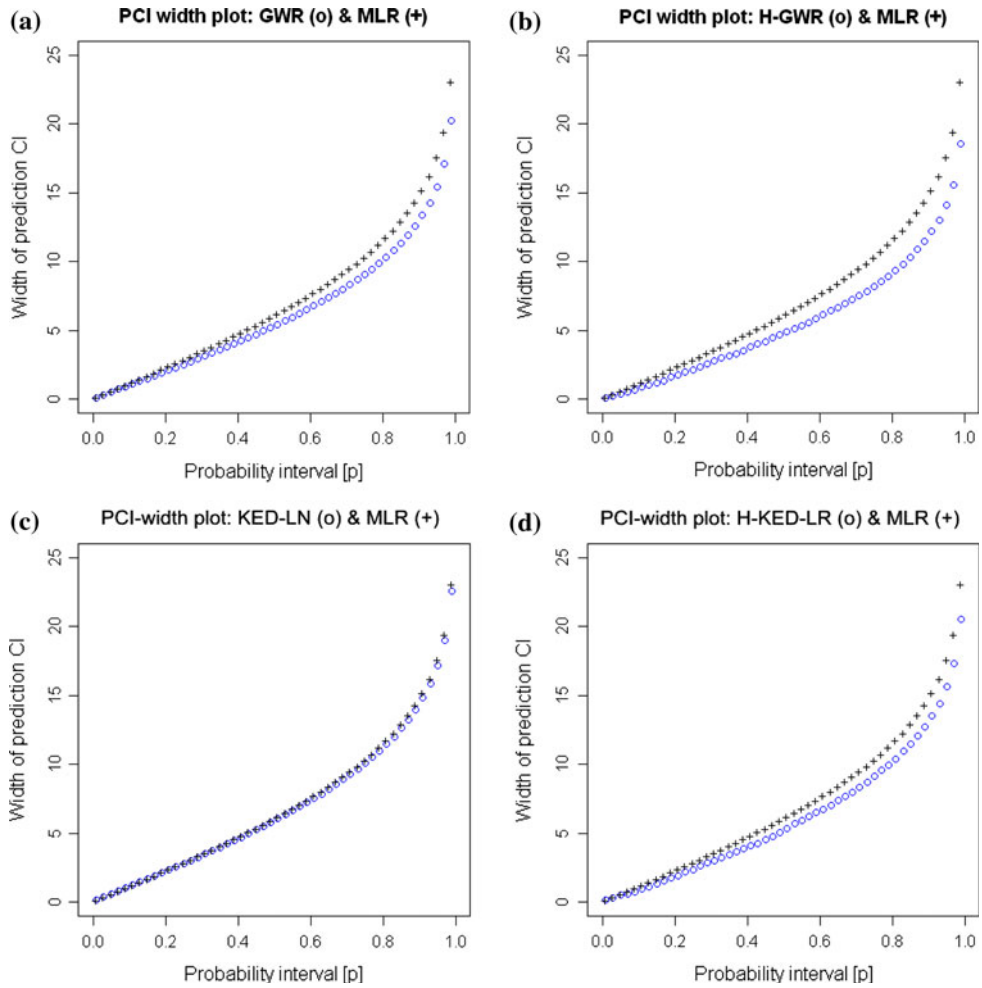

Fig. 7 PCI-width plots for: **a** GWR; **b** H-GWR; **c** KED-LN and **d** H-KED-LR (all shown with naïve MLR model)

4.4 Focused analysis: prediction uncertainty accuracy

Diagnostics measuring prediction uncertainty accuracy are given in Table 4. Models that perform well in a global-sense (via 1-MSDR i.e. MSDR bias) are MLR, H-GWR, GWR and H-KED-GWR. In a local-sense via the *G*-statistic, H-GWR, MLR, GWR and H-KED-LR perform well. For choosing an H-KED model, H-KED-LR and H-KED-GWR both out-perform the naïve H-KED-MLR and H-KED-DV models. Hence this provides some value to the use of nonlinear fits to local mean and SD data when applying this locally corrective kriging variance approach. Accuracy and PCI-width plots for the model pairs: (i) GWR/H-GWR and (ii) KED-LN/H-KED-LR[10] are given in Figs. 6 and 7; and are shown with the results from the naïve MLR model for context. For each of our chosen models, its accuracy and PCI-width plot should be viewed in conjunction. In this respect, H-GWR is clearly the best performer, as its accuracy plot is (above and) closest to the 45° line and its PCI-widths are consistently the smallest. Rather surprisingly, the next best performer is (basic) GWR. As expected, both basic models are out-performed by their heteroskedastic counterpart. Observe that MLR's good performance with respect to PCI accuracy is simply a consequence of wider PCI-widths (i.e. PCI precision is poor). Further locally orientated assessments are possible by relating the absolute (actual) prediction errors $\left|z(\mathbf{x}_j) - \hat{z}(\mathbf{x}_j)\right|$ to the (estimated) prediction standard errors $\sigma_P(\mathbf{x}_j)$. This data should have a MLR fit with a slope of one that passes through the origin. Here, all basic predictors perform poorly, whereas all heteroskedastic models perform relatively well. Scatterplots of this data, for the same model pairs as before are presented in Fig. 8; where both H-GWR and H-KED-LR are able to provide variability in their $\sigma_P(\mathbf{x}_j)$ distribution, and in doing so, these estimates can correlate with the actual prediction errors. However for both heteroskedastic models, $\sigma_P(\mathbf{x}_j)$ is often under-estimated, resulting in fairly weak correlations. In summary, H-GWR is viewed as the best overall performer and H-KED-LR (or H-KED-GWR) is viewed as the best KED model, with respect to prediction uncertainty accuracy.

[10] The H-KED-LR and H-KED-GWR models are effectively interchangeable, where the former is now chosen as demonstration.

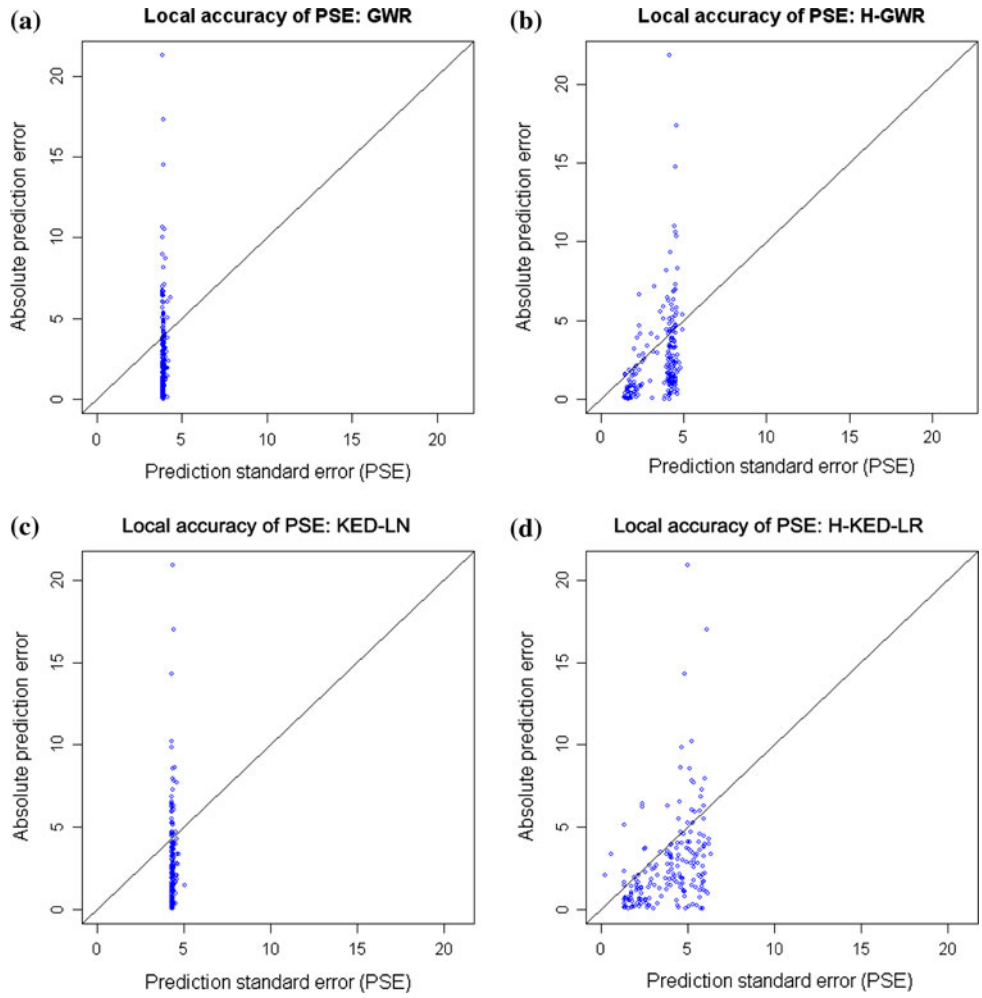

Fig. 8 Actual (absolute) prediction errors versus the estimated prediction standard errors for: **a** GWR; **b** H-GWR; **c** KED-LN; and **d** H-KED-LR. Correlations are 0.02, 0.34, 0.08 and 0.39, respectively

Finally, the spatial pattern of each model's prediction variances can be investigated. Here the prediction variances for KED-LN and GWR exhibit little spatial variation, whilst those for H-KED-LR and H-GWR (Fig. 9) do; reflecting the nonstationary modelling decisions taken. As is well-known, the KED-LN variances will depend largely on data geometry, as large KED-LN variances tend to lie where sampling is sparse. Conversely, GWR prediction variances do not depend on data geometry and instead vary according to the nature of local relationships (i.e. they reflect uncertainty in local attribute-space). The GWR model also provides lower prediction variances than KED-LN. Both heteroskedastic models provide low prediction variances in northern Scotland, reflecting an area of low-valued critical loads and low critical load variation. H-GWR prediction variances tend to mimic that of the residual variance surface of Fig. 3d and are high in areas of northern and SW England. In relation to the H-KED-LR model, H-GWR tends to provide a prediction variance distribution with a lower spread. The H-KED-LR variances are similarly high in areas of SW England, but unlike the H-GWR model, high prediction variances are also found in areas of eastern England.

5 Discussion and conclusions

In this study, we have linked GWR to KED (with local neighbourhoods) in both basic and heteroskedastic forms. Linkages have been discussed from a model construction viewpoint and via the outcomes of an empirical analysis. Both models cater for nonstationary relationships between a variable of interest and its covariates. However crucial differences exist, especially with respect to their use of information when modelling such relationships. Here GWR can be calibrated using all the sample data, whilst KED cannot.

On balance, our empirical work (with a freshwater acidification critical load data set) suggests that a basic KED model should be preferred with respect to prediction accuracy, but that a heteroskedastic GWR model should be preferred with respect to prediction uncertainty accuracy. For this particular process, models that account for both

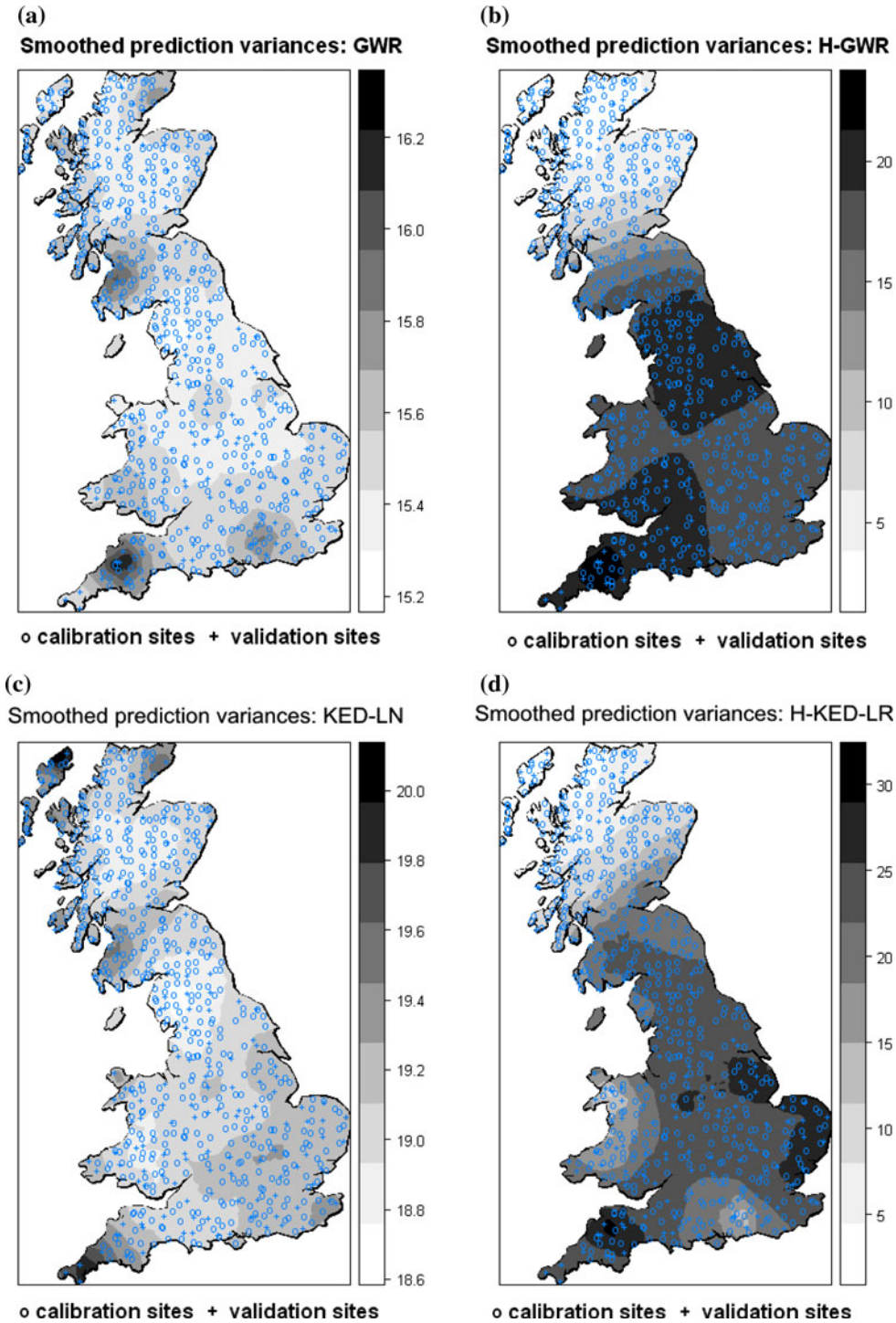

Fig. 9 Prediction variance surfaces for: **a** GWR; **b** H-GWR; **c** KED-LN; and **d** H-KED-LR. All surfaces are found using the same smoothing function

relationship and variance heteroskedasticity should be preferred to models that do not. Overall there is little to choose between GWR and KED, and in this respect, GWR (the less recognised predictor) has performed with merit. For GWR to be routinely used as a spatial predictor, it not only needs to provide tolerably accurate predictions, but also needs a form that can provide accurate prediction variances. In this respect, our study has demonstrated a GWR methodology to achieve this, which has not been applied in a prediction context before.

As always, empirical work cannot be readily generalised, but it is surmised that GWR will prove similarly worthy in other spatial prediction problems. It is envisaged that future work could: (a) investigate more deeply the nature of the GWR prediction variances; (b) include (and possibly adapt) a Bayesian spatially-varying coefficient

model (e.g. Gelfand et al. 2003; Waller et al. 2007) or the local cokriging models of Pereira et al. (2002) for modelling nonstationary relationships; and (c) find ways to generalise (and confirm) our results using both real and simulated data sets, where issues of sample size and configuration are also incorporated.

Acknowledgments Research presented in this paper was funded by a Strategic Research Cluster grant (07/SRC/I1168) by the Science Foundation Ireland under the National Development Plan. The authors gratefully acknowledge this support. Thanks are also due to the first author's PhD studentship and S. Juggins at Newcastle University; and to M. Kernan at University College London for providing the case study data.

References

Bailey TC, Gatrell AC (1995) Interactive spatial data analysis. Wiley, New York

Bitter C, Mulligan GF, Dall'erba S (2007) Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method. J Geogr Syst 9:7–27

Chilès JP, Delfiner P (1999) Geostatistics—modelling spatial uncertainty. Wiley, New York

Deutsch CV, Journel AG (1998) GSLIB geostatistical software library and user's guide. Oxford University Press, New York

Fotheringham AS, Brunsdon C, Charlton ME (2002) Geographically weighted regression—the analysis of spatially varying relationships. Wiley, Chichester

CLAG Freshwaters (1995) Critical loads of acid deposition for United Kingdom Freshwaters. Critical Loads Advisory Group. Subreport on Freshwaters. ITE, Penicuik

Gao X, Asami Y, Chung C (2006) An empirical evaluation of spatial regression models. Comput Geosci 32:1040–1051

Gelfand AE, Kim HJ, Sirmans CJ, Banerjee S (2003) Spatial modeling with spatially varying coefficient processes. J Am Stat Assoc 98:387–396

Goovaerts P (1997) Geostatistics for natural resources evaluation. Oxford University Press, New York

Goovaerts P (2001) Geostatistical modelling of uncertainty in soil science. Geoderma 103:3–26

Harris P, Fotheringham AS, Crespo R, Charlton M (2010a) The use of geographically weighted regression for spatial prediction: an evaluation of models using simulated data sets. Math Geosci 42:657–680

Harris P, Charlton M, Fotheringham AS (2010b) Moving window kriging with geographically weighted variograms. Stoch Environ Res Risk Assess 24:1193–1209. doi:10.1007/s00477-010-0391-2

Harris P, Fotheringham AS, Juggins S (2010c) Robust geographically weighed regression: a technique for quantifying spatial relationships between freshwater acidification critical loads and catchment attributes. Ann Assoc Am Geogr 100(2):286–306

Heuvelink GBM, Pebesma EJ (2002) Is the ordinary kriging variance a proper measure of interpolation error? In: Hunter G, Lowell K (eds) The fifth international symposium on spatial accuracy assessment in natural resources and environmental sciences. RMIT University, Melbourne, pp 179–186

Isaaks EH, Srivastava RM (1989) An introduction to applied geostatistics. Oxford University Press, New York

Journel AG (1986) Geostatistics: models and tools for the earth sciences. Math Geol 18:119–140

Kernan MR, Allott TEH, Battarbee RW (1998) Predicting freshwater critical loads of acidification at the catchment scale: an empirical model. Water Air Soil Pollut 185:31–41

Kernan MR, Haliwell RC, Hughes MJ (2001) Predicting freshwater critical loads from catchment characteristics using national datasets. Water Air Soil Pollut Focus 1:415–435

Kupfer JA, Farris CA (2007) Incorporating spatial non-stationarity of regression coefficients into predictive vegetation models. Landsc Ecol 22:837–852

Laslett GM (1994) Kriging and splines: an empirical comparison of their predictive performance in some applications. J Am Stat Assoc 89:391–409

Leung Y, Mei C, Zhang W (2000) Statistical tests for spatial nonstationarity based on the geographically weighted regression model. Environ Plann 32:9–32

Lloyd CD (2010) Nonstationary models for exploring and mapping monthly precipitation in the United Kingdom. Int J Climatol 30:390–405

Loader C (2004) Smoothing: local regression techniques. In: Gentle J, Härdle W, Mori Y (eds) Handbook of computational statistics. Springer-Verlag, Heidelberg

Matheron G (1989) Estimating and choosing: an essay on probability in practice. Springer, Berlin

Nilsson J, Grennfelt P (eds) (1988) Critical loads for sulphur and nitrogen. Nordic Council of Ministers, Copenhagen

Páez A, Long F, Farber S (2008) Moving window approaches for hedonic price estimation: an empirical comparison of modelling techniques. Urban Stud 45:1565–1581

Pereira MJ, Soares A, Rosario L (2002) Characterization of forest resources with satellite spot images by using local models of co-regionalization. In: Kleingeld WJ, Krige DG (eds) Geostatistics 2000 Cape Town. Geostatistical Association of Southern Africa

Rivoirard J (1987) Teacher's Aide—two key parameters when choosing the kriging neighbourhood. Math Geol 19(8):851–856

Salas C, Ene L, Gregoire TG, Næsset E, Gobakken T (2010) Modelling tree diameter from airborne laser scanning derived variables: a comparison of spatial statistical models. Remote Sens Environ 114:1277–1285

Schabenberger O, Gotway C (2005) Statistical methods for spatial data analysis. Chapman & Hall, London

Switzer P (1993) The spatial variability of prediction errors. In Geostatistics Troia 92 (A. Soares, Ed.) pp. 261-272

Van Meirvenne M, Goovaerts P (2001) Evaluating the probability of exceeding a site-specific soil cadmium contamination threshold. Geoderma 102:75–100

Wackernagel H (2003) Multivariate geostatistics: third completely revised edition. Springer-Verlag, Berlin

Waller LA, Zhu L, Gotway CA, Gorman DM, Grunewald PJ (2007) Quantifying geographic variations in associations between alcohol distribution and violence: a comparison of geographically weighted regression and spatially varying coefficient models. Stoch Environ Res Risk Assess 21:573–588

Wheeler D (2007) Diagnostic tools and a remedial method for collinearity in geographically weighted regression. Environ Plann 39:2464–2481

Wheeler D (2009) Simultaneous coefficient penalization and model selection in geographically weighted regression: the geographically weighted lasso. Environ Plann 41:722–742

Wheeler D, Waller L (2009) Comparing spatially varying coefficient models: a case study examining violent crime rates and their relationships to alcohol outlets and illegal drug arrests. J Geogr Syst 11:1–22

Zhang L, Gove JH, Heath LS (2005) Spatial residual analysis of six modelling techniques. Ecol Modell 186:154–177