Stoch Environ Res Risk Assess (2010) 24:1193–1209
DOI 10.1007/s00477-010-0391-2

ORIGINAL PAPER

Moving window kriging with geographically weighted variograms

Paul Harris · Martin Charlton · A. Stewart Fotheringham

Published online: 13 April 2010
© Springer-Verlag 2010

Abstract This study adds to our ability to predict the unknown by empirically assessing the performance of a novel geostatistical-nonparametric hybrid technique to provide accurate predictions of the value of an attribute together with locally-relevant measures of prediction confidence, at point locations for a single realisation spatial process. The nonstationary variogram technique employed generalises a moving window kriging (MWK) model where classic variogram (CV) estimators are replaced with information-rich, geographically weighted variogram (GWV) estimators. The GWVs are constructed using kernel smoothing. The resultant and novel MWK–GWV model is compared with a standard MWK model (MWK–CV), a standard nonlinear model (Box–Cox kriging, BCK) and a standard linear model (simple kriging, SK), using four example datasets. Exploratory local analyses suggest that each dataset may benefit from a MWK application. This expectation was broadly confirmed once the models were applied. Model performance results indicate much promise in the MWK–GWV model. Situations where a MWK model is preferred to a BCK model and where a MWK–GWV model is preferred to a MWK–CV model are discussed with respect to model performance, parameterisation and complexity; and with respect to sample scale, information and heterogeneity.

Keywords Geostatistics · Kriging · Nonstationary · Nonparametric · Variogram

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

1 Introduction

A fundamental problem in geostatistics is that of accurate prediction and associated measures of prediction uncertainty. Spatial prediction is not only valuable to scientists who attempt to model spatial processes, but also to policy makers who need to plan, assess risk and manage the outcomes of spatial processes at different spatial scales. Often overlooked in a geostatistical analysis is the importance of accurate measures of prediction uncertainty. It is this aspect of geostatistical prediction that the kriging models of our study particularly focus on, and for which we aim to demonstrate the utility of our hybrid geostatistical-nonparametric approach.

In the classical geostatistical framework, uncertainty at a target location can be accounted for in three distinct ways by: (i) a conditional distribution defined by the kriging prediction and kriging variance from (any form of) kriging; (ii) a conditional distribution from a specific nonlinear method such as indicator kriging (IK) (Journel 1983) or disjunctive kriging (DK) (Matheron 1976); or (iii) some conditional simulation approach (Journel 1989). Commonly, a method from approach (i) or (iii) is overlooked in favour of one from approach (ii). A standard method such as simple kriging (SK) from approach (i) is not recommended as kriging variances are unlikely to vary in accordance with the variability in the local data used to provide the kriged predictions. That is, the kriging variance is a function only of the variogram model and the spatial configuration of the data and, given a global variogram model (as in SK), it can only vary locally in this sense. For point prediction, a method from approach (iii) is not recommended as it will simply produce broadly similar results to a much simpler method of approach (ii) on which it is usually based (Goovaerts 2001). However, for this study

the concept of approach (i) is retained, but methods are adapted to allow the variance and/or the variogram to vary across space. This crucial adaptation provides the necessary conditional distributions that are comparable with those found from a method of approach (ii), especially when modelling some heterogenic (or nonstationary) spatial process.

Models that cater for such nonstationarity of the variance and/or variogram include: Box–Cox kriging (BCK) (Kitanidis and Shen 1996), kriging with locally varying sills (Kls) (Isaaks and Srivastava 1989), kriging with an interpolation variance (Kiv) (Yamamoto 2000), kriging within strata (StK) (Stein et al. 1988), kriging within classes (CK) (Cressie 1993), multiple population indicator kriging (MPIK) (Parker 1991) and moving window kriging (MWK) (Haas 1990). All models have the potential to improve estimates of prediction uncertainty, whereas only BCK, StK, CK, MPIK and MWK are also capable of improving predictions. The BCK, Kls and Kiv models address only nonstationarity of the variance and are the simplest to implement. The Kls model does this directly by locally correcting the kriging variance, whereas BCK uses a data transform to achieve the same objective. Commonly, both Kls and BCK (with respect to its back-transforms) assume a proportional relationship between local means and local variances. The Kiv model counters the data-independent properties of the kriging variance by calculating an alternative that depends on local data. The Kls, BCK and Kiv models all rely on one (global and stationary) variogram fit.

The StK, CK and MPIK models are examples of kriging within partitions. For StK the partitioning is spatial, for CK it usually relates to classified contextual data and for MPIK the partitioning is based on breakpoints in the histogram. Commonly, partitioned methods use separate variograms for each partition. For StK, this can account fully for variogram nonstationarity, but for CK and MPIK variogram nonstationarity will depend on the spatial layout of the partitions used. In practise, CK and MPIK will account for only a limited form of variogram nonstationarity. For MWK, a variogram is determined at every target location in this fully-automatic, continuous and locally-adaptive nonstationary variogram technique. The MWK model reduces the work-load of the analyst over a StK, CK or MPIK model, but at a cost of reduced control over the model calibration.

Existing geostatistical-nonparametric hybrids that account similarly for variogram nonstationarity include those where local variogram model parameters found from an initial application of StK (Fuentes 2001; Fuentes et al. 2003) or MWK (Haas 2002; Pardo-Igúzquiza et al. 2005) and are then kernel smoothed (i.e. KS–StK or KS–MWK models, respectively). This use of a kernel weighting function allows a *global* model of spatial dependence to be defined that represents the spatial process as a whole. Methods defined using partitions or moving windows can only provide a collection of individual local models of spatial dependence (Sampson et al. 2001). Smoothed parameters can also be robust to outlying data that can affect local variogram parameter estimation at a given location.

For the hybrid MWK-based model of this study, a kernel weighting function is not used to smooth *local variogram parameters* (as in KS–StK or KS–MWK), but is used instead to smooth the *individual* (*empirical*) *semivariances* of each lag interval according to the distance of these paired values from a target location. Thus, at any target location, a geographically weighted variogram (GWV) is estimated and modelled, creating the MWK–GWV technique. The expected benefit of a GWV is that it does not rely on limited local information to be *local* and reliable, as that does a local classic variogram (CV) of standard MWK (i.e. MWK–CV), which is estimated using only data within some circular window centred at a target location. In fact, a GWV can benefit from exactly the same information as the global variogram estimator used in say, BCK. The first published account of a GWV can be found in Johannesson and Cressie (2004) where it was used in a purely exploratory fashion. The natural extension and use of a GWV with kriging was not pursued and in this respect, the MWK–GWV model of this study is novel. To demonstrate the utility of MWK–GWV, its performance is compared empirically with MWK–CV, BCK (taken as a benchmark nonstationary variance model) and SK (the naïve model) using four example environmental datasets.

2 Four datasets

All four datasets are assumed to be representative of a heterogeneous spatial process that should benefit from a nonstationary modelling approach. Datasets were chosen to cover a broad range of environmental processes, where the data sizes for model calibration could range from the relatively small ($n = 78$) to the relatively large ($n = 808$). Crucially, two of these datasets were chosen blindly, with little or no prior knowledge of their actual properties. Experience suggested the other two datasets would benefit from nonstationary modelling. All four datasets are freely available.

2.1 Chosen blindly

- The first and second datasets are taken from the spatial interpolation comparison (SIC) exercises of 1997 and 2004 (e.g. see Dubois 2008) organised through the

AI-GEOSTATS website (http://www.ai-geostats.org). Here the first dataset is for SIC97 and consists of daily rainfall (wet deposited radioactivity) data in Switzerland, split into a model calibration data subset of 100 values and a model validation data subset of 367 values.

- The second dataset is for SIC2004 and consists of natural ambient radioactivity (gamma dose rates) measured in Germany, split into a model calibration data subset of 200 values and a model validation data subset of 808 values. Both SIC datasets are available in the R gstat package (Pebesma 2004).

2.2 Chosen with some prior knowledge

- The third dataset is the 'Calcium content in soil samples' dataset (Ca20) which is available in the R geoR package (Ribeiro and Diggle 2001). This dataset is split into a model calibration data subset of 78 values and a model validation data subset of 100 values. This dataset covers three distinct regions, each with different physical properties. Thus, a nonstationary model may suit the dataset as a whole.
- The fourth dataset is a freshwater acidification critical load dataset covering Great Britain (CLAG-Freshwaters 1995). This dataset is split into a model calibration data subset of 189 values and a model validation data subset of 497 values. The size and scale of this dataset should suit a nonstationary modelling approach. Applications of (different) nonstationary models to these data can be found in Harris and Brunsdon (2010). Data similar to those used in this study can be found at http://critloads.ceh.ac.uk/index.htm.

2.3 Eight data situations

Simple interpolation maps for the four datasets are shown in Fig. 1, where each map is shown with its calibration and validation sites. To enhance the comparisons of this study and to assess the effect (if any) of sample size and/or sample variation on model calibration, the roles of the calibration and validation data subsets are also reversed. This effectively provides four more datasets (i.e. eight data situations in total). Here the calibration/validation datasets given above are referred to as 1S, 2S, 3S and 4S (for *small* calibration data subsets of size 100, 200, 78 and 189, respectively); and calibration/validation datasets referred to as 1L, 2L, 3L and 4L correspond to *large* calibration data subsets of size 367, 808, 100 and 497, respectively.

3 Methods

3.1 Background to MWK

A standard algorithm such as ordinary kriging (OK) is itself a moving window technique. However, when OK is specified with local neighbourhoods, variograms are not found locally as they strictly should be, but instead the global variogram is retained (i.e. an assumption of quasi-stationarity, see Journel and Huijbregts 1978, pp. 33–34). Haas (1990) was the first to counter this approximation and actually fit a variogram within each OK neighbourhood. Haas (1990, 1996) proceeded to develop the MWK approach incorporating lognormal kriging, regression kriging, and cokriging versions. Other notable exponents of MWK include the OK version of Walter et al. (2001); the universal kriging (UK) version of Lloyd and Atkinson (2002); the UK (KS–MWK) version of Pardo-Igúzquiza et al. (2005); and the IK version of Cattle et al. (2002). Applications of MWK include those to: acid deposition data (Haas); soil data (Walter, Cattle); elevation data (Lloyd); and meteorological data (Pardo-Igúzquiza), where sample scale *and* size are commonly large (i.e. situations that warrant *and* enable the adoption of a MWK model).

For MWK an improvement in prediction accuracy is often not sufficiently strong enough to warrant its adoption by itself. Here, it is useful to refer to the empirical studies of Walter et al. (2001), Lloyd and Atkinson (2002), Paciorek and Schervish (2006), where prediction using a stationary measure of spatial dependence was just as accurate. Instead, it is MWK's ability to provide kriging variances that reflect local uncertainty accurately that is important. This latter property of MWK has been demonstrated in Van Tooren and Haas (1993), Walter et al. (2001) for example; and this improvement alone can justify its use. Commonly, the decision of whether or not to apply MWK over a stationary counterpart (such as BCK) depends on a trade-off between: (a) many ill-fitted local variograms, but with (potentially) more accurate model outputs and (b) a well-fitted (and understood) global variogram with possibly less accurate model outputs. As MWK is inherently more complex and approximate, there must be good reason to favour it. The MWK model will be of little worth if local variograms are only marginally different from each other and from the global variogram.

3.2 SK, BCK, MWK–CV and MWK–GWV: general specifications

For simplicity, the underlying kriging algorithm for BCK, MWK–CV and MWK–GWV is chosen as SK. It is, however, a straightforward extension to construct these models

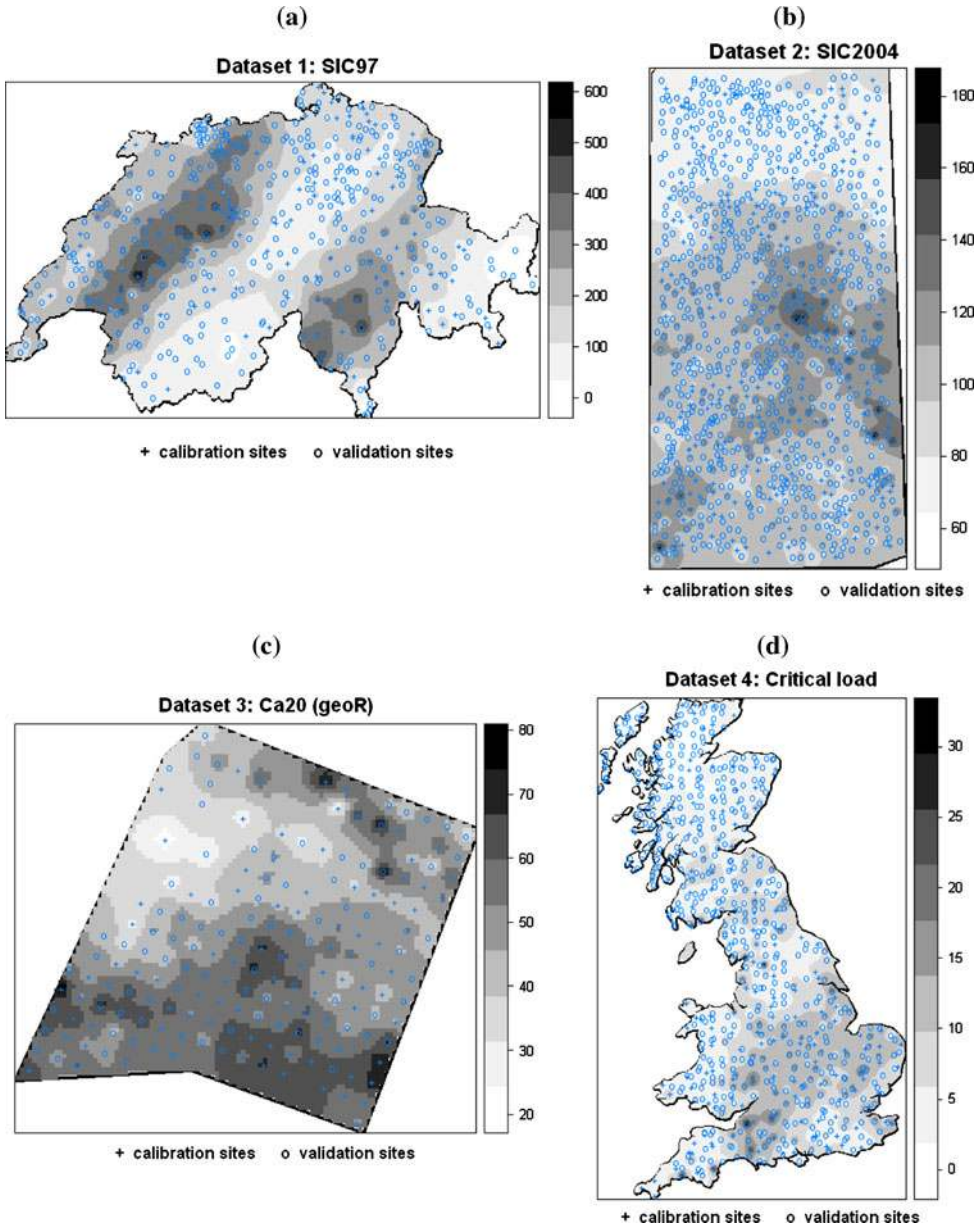

Fig. 1 IDW mean surfaces of the full datasets, shown with small calibration and large validation data subsets: **a** dataset one, SIC97; **b** dataset two, SIC2004; **c** dataset three, Ca20; **d** dataset four, critical load. Full datasets are of size 467, 1008, 178 and 686, respectively

using OK instead. To avoid any distraction from this study's primary objectives, all models are designed for a general robustness without being over-complicated. This entails that the MWK models are not constructed from individual BCK (or alternative nonlinear) models. It also entails that the use of non-constant mean functions and alternative variogram estimators are similarly avoided in all model constructions. Furthermore, the kriging neighbourhood is always taken as the same size as that used in the corresponding variography (i.e. no quasi-stationarity decisions are made in any model). As MWK necessitates a fully automated variographic approach, the simpler SK and BCK models (with their single, stationary variograms) are calibrated in a similar fully automated manner. Thus, with the given modelling constraints in place, it is likely that the performance of all models in the study could be improved upon if some judged interaction between the analyst and model specification were allowed. However, adhering to these constraints provides a means of comparing models objectively in a relative sense.

3.3 SK, BCK and MWK–CV: characterisation of spatial dependence

For SK, BCK and MWK–CV, a two-stage procedure is adopted where the variogram is first estimated and then modelled to get a smooth representation of spatial dependence. For SK and BCK this procedure is done once and

globally, whereas for MWK–CV it is done many times and locally. For simplicity, only omni-directional (isotropic) estimators are considered. Therefore, let $z(\mathbf{x}_1),\ldots,z(\mathbf{x}_n)$ denote observations from a constant mean spatial process at locations $\mathbf{x}_1,\ldots,\mathbf{x}_n$ and following Johannesson and Cressie (2004, pp. 27–30), the CV estimator at any location $\mathbf{u}$ can be defined as:

$$2\hat{\gamma}(h;\mathbf{u}) = \frac{1}{|N(h;\mathbf{u})|} \sum_{(i;j)\in N(h;\mathbf{u})} \left[z(\mathbf{x}_i) - z(\mathbf{x}_j) \right]^2; \quad h \in [0] \tag{1}$$

where $N(h;\mathbf{u})$ is a set of pairs of indices such that $(i, j) \in N(h;\mathbf{u})$ if $\|\mathbf{x}_i - \mathbf{x}_j\| \approx h$ (where this approximation reflects some user-specified tolerance on the lag distance h), and $\mathbf{x}_i$, $\mathbf{x}_j$ are both in the neighbourhood of $\mathbf{u}$. To interpret an estimated variogram, nugget and Matérn models are specified, which are defined as:

$$\gamma_N(h) = c_0 + c \tag{2}$$

$$\gamma_M(h) = c_0 + c\left[1 - \frac{1}{2^{\nu-1}\Gamma(\nu)}\left(\frac{h}{a}\right)^{\nu} K_\nu\left(\frac{h}{a}\right)\right]; \quad \text{for } h \in [0] \tag{3}$$

where $\gamma(h) = 0$ for $h = 0$. Here a is a correlation range parameter; ν is a smoothing parameter; K_ν is a modified Bessel function; Γ is the gamma function; and c_0 and c are partial sills that reflect small- and large-scale variation, respectively. The Matérn function offers much flexibility, where the higher is the value of ν, the smoother is the process. For this study, ν is not estimated (or fitted) but is fixed beforehand at 0.5, 1 or 1.5. This effectively provides four (not two) variogram models to choose from in total.

For each kriging technique, all four variogram models are fitted automatically using the weighted least squares (WLS) approach of Zhang et al. (1995). Here sensible heuristics are used to define: (a) the lag intervals of the CV estimator; (b) the minimum number of pairs allowed at the first lag of the CV estimator; (c) the truncation of the CV estimator at long (unreliable) lag distances; and (d) the model starting parameters for the WLS fit. Such variographic specifications are chosen so as to minimise the occurrence of a poor (or failed) WLS fit and are crucial in determining a good kriging model performance (see Müller 1999; Lahiri et al. 2002). One variogram model is chosen that provides the smallest weighted sum of squares (WSS) and the corresponding parameters are used to solve an SK system of equations. Optimal model parameterisation by a maximum likelihood (ML) approach was discounted from the outset as it does not suit the MWK comparisons of this study. In any case, CV/WLS variogram fits have been the usual choice for most MWK studies (e.g. Haas 1990, 1996).

3.4 BCK: transforms and back-transforms

A simple form of variance nonstationarity can be addressed by transforming the scale of measurement so that the sample distribution is approximately Gaussian with a stationary or stable variance. One such method is BCK, which transforms the data according to:

$$z_t = \begin{cases} \frac{z^{\lambda}-1}{\lambda} & \lambda \neq 0 \\ \ln(z) & \lambda = 0 \end{cases},$$

where z_t is the transformed value and z is the value to be transformed. The parameter λ is estimated using ML based on an assumption that the transformed data are Gaussian. Variogram inference and kriging with transformed data is the same as for raw data. In this respect, BCK is not strictly nonlinear but is instead linear kriging of a nonlinear transform of the data. A key problem with BCK (and related methods) is a bias when back-transforming to the original scale (Cressie 1993, pp. 135–138; Chilès and Delfiner 1999, pp. 190–193). Thus, the following *corrected* back-transforms are defined (Kitanidis and Shen 1996).

When $\lambda = 0$, an unbiased back-transform for a prediction at location $\mathbf{u}$ is taken as, $\hat{z}_{BCK}(\mathbf{u}) = \beta^{\alpha}[\beta^2 + \alpha\sigma^2_{BCKT}(\mathbf{u})]$; where $\beta = 1 + \lambda\hat{z}_{BCKT}(\mathbf{u})$; $\alpha = (1/\lambda) - 2$; $\hat{z}_{BCKT}(\mathbf{u})$ is the SK prediction in transformed-space; and $\sigma^2_{BCKT}(\mathbf{u})$ is the SK variance in transformed-space. Alternatively when $\lambda = 0$, an unbiased back-transform results in $\hat{z}_{BCK}(\mathbf{u}) = \exp\left[\hat{z}_{BCKT}(\mathbf{u}) + \sigma^2_{BCKT}(\mathbf{u})/2\right]$. An unbiased back-transform for the corresponding kriging variance is $\sigma^2_{BCK}(\mathbf{u}) = \sigma^2_{BCKT}(\mathbf{u})\left[\sigma^2_{BCKT}(\mathbf{u})/8 + \beta^2\right]$ if $\lambda = 0.5$; $\sigma^2_{BCK}(\mathbf{u}) = \exp\left[2\hat{z}_{BCKT}(\mathbf{u}) + \sigma^2_{BCKT}(\mathbf{u})\right] \times \left[\exp\sigma^2_{BCKT}(\mathbf{u}) - 1\right]$ if $\lambda = 0$; and $\sigma^2_{BCK}(\mathbf{u}) = \beta^{[(2/\lambda)-2]}\sigma^2_{BCKT}(\mathbf{u})$ for all other values of λ. These analytical back-transforms are sensitive to values of the kriging variance, and as a consequence, poor or unusual BCK outputs can often be attributed to this.

3.5 MWK–CV: window size

Probably the most important decision in MWK–CV is window size. That is, at what scale can the usual stationarity decisions for SK be safely made? Here a cross-validation approach is adopted as this ensures some objectivity in model comparison across the four study datasets. It is, however, not easy to decide on the form that the objective function should take. For example, should an optimal window size reflect MWK–CV prediction accuracy, MWK–CV prediction accuracy of uncertainty or some mixture of both? In this respect, Haas (2002) suggests a weighted sum of model bias statistics where the weights are user-specified according to relative priorities.

In a similar vein, an optimal window size for this study's MWK–CV models is found by first calibrating models across

a broad range of window sizes. Here a minimum window size is set at $n = 35$ (which is consistent with previous studies, e.g. Haas 1990; Pardo-Igúzquiza et al. 2005) and once set, window size is increased in 10% data increments. Secondly, three of the four model bias statistics given in Sect. 4 (rel-RMSE, *G*-statistic and Error-Corr) are calculated for each MWK–CV fit. Models are then ranked according to their average ranking performance for each model bias statistic (i.e. to reflect a mixture of prediction accuracy and prediction uncertainty accuracy). The MWK–CV model with the best ranking is taken to have the optimal window size.

3.6 MWK–GWV: GWV estimation

For the GWV estimator, a weighting function is used to smooth the individual (empirical) semivariances (of each lag interval) according to the distance of these paired values from a target location. The expected benefit of a GWV is that it is not so reliant on limited local information to be local and reliable as that for a local CV. On the downside, the resultant MWK–GWV model adds another layer of complexity to the MWK approach. Empirical variograms have been weighted before; either to address clustering bias due to preferential sampling or to address statistical homogeneity for paired values at each lag interval (Omre 1984; Richmond 2002; Rivoirard 2002). Goovaerts et al. (2005) applied a weighted variogram to cancer mortality rate data, where paired values are weighted according to an increase/decrease in population size. The use of geographical (as opposed to non-geographical) weights in variogram estimation can be found in Johannesson and Cressie (2004), where a GWV estimator was used to explore local spatial dependence in residual data from a detrending of a very large ozone dataset sampled across the globe. This study now develops the use of this GWV estimator with kriging. A GWV can be considered a part of a wider set of geographically weighted (GW) models that includes GW summary statistics, GW principle component analysis and GW regression (Fotheringham et al. 2002).

A GWV estimator can be constructed in various ways, but after some experimentation a form that relates closely to that described in Johannesson and Cressie (2004, pp. 27–30) is adopted. This GWV estimator can be defined by re-writing expression 1 in terms of a box–car weighting function:

$$2\hat{\gamma}(h;\mathbf{u}) = \frac{1}{W(h;\mathbf{u})} \sum_{i,j} W(\mathbf{x}_i;\mathbf{x}_j;h;\mathbf{u})\,[z(\mathbf{x}_i) - z(\mathbf{x}_j)]^2; \quad h[0 \tag{4}$$

where $W(\mathbf{x}_i, \mathbf{x}_j; h, \mathbf{u})$ is a weight taking the value of 1 if $(i, j) \in N(h; \mathbf{u})$, and 0 otherwise, and $W(h; \mathbf{u}) : \sum_{i,j} W(\mathbf{x}_i, \mathbf{x}_j; h, \mathbf{u})$. Next write $W(\mathbf{x}_i, \mathbf{x}_j; h, \mathbf{u})$ in a more general form, for weights other than and including 0–1:

$$W(\mathbf{x}_i;\mathbf{x}_j;h;\mathbf{u}) = w_l(\|\mathbf{x}_i - \mathbf{x}_j\| - h)\, w_d(\|\mathbf{x}_i - \mathbf{u}\|) w_d(\|\mathbf{x}_j - \mathbf{u}\|) \tag{5}$$

where $w_l(\cdot)$ is a box–car weighting function that defines the lag intervals and $w_d(\cdot)$ is a (geographical) weighting function that defines what is meant to be in the neighbourhood of $\mathbf{u}$. Expressions 4 and 5 require the calculation of distances d between location $\mathbf{u}$ and each pair of points defining each (semi)variance value. Each (semi)variance value is then geographically weighted according to the weight of the first (say shortest) distance measure $w_d(\|\mathbf{x}_i - \mathbf{u}\|)$ multiplied by the weight of the second (say longest) distance measure $w_d(\|\mathbf{x}_j - \mathbf{u}\|)$. Again for simplicity, only omni-directional (isotropic) GWV estimators are considered.

For this study, inverse distance weighting (IDW) is used to find the geographical weights, i.e. $w_d(\cdot) = 1/d^q$. Other kernel functions can be used, but an IDW function defined by distance (i.e. a fixed kernel-type) is chosen where small values of the controlling parameter q give a weighting function that decays gradually with distance, whereas large values result in a weighting function that decays steeply. All of the semivariance data are always used with this weighting function and if q is pre-set at 0, the global variogram would be found at every location $\mathbf{u}$. This type of GWV specification also enables the kriging neighbourhood to be taken as global. Akin to finding an optimal window size for MWK–CV in Sect. 3.5, an optimal value of q is found by first calibrating MWK–GWV models for six pre-set values of q (taken as: 0.5, 1, 1.5, 2, 2.5 and 3) and then ranking models with respect to their average ranking performance for each of the same three model bias statistics. The MWK–GWV model with the best ranking is taken to have the optimal value of q.

3.7 MWK–GWV: GWV/WLS variogram fits

Once GWVs have been estimated at every location $\mathbf{u}$, the resulting MWK–GWV algorithm is identical to the MWK–CV algorithm. That is the same four variogram models are fitted by WLS and the model that provides the smallest WSS is used to parameterise a SK system. Furthermore, the same set of heuristics is used to define the GWV lag intervals; the minimum number of pairs allowed at the first GWV lag; truncation at long GWV lags; and starting parameters for the WLS fit.

Truncation of the GWV estimator at long lags can be crucial as the GWV tends to behave in a nonstationary manner at shorter lags (i.e. akin to a local CV) and a stationary manner at longer lags (i.e. akin to the global CV). This means that a GWV can sometimes resemble a CV computed from data where the underlying process has some form of trend, which in turn can result in problems when

estimating the (total) sill and range parameters by WLS. Such problems may not affect prediction accuracy adversely, but can have a detrimental effect on the accuracy of prediction uncertainty. This behaviour in the GWV is primarily a consequence of using the same weighting scheme for each lag interval, where the distribution of weights at any short lag will often show much variation, whereas the distribution of weights at any long lag will not (as all pairs separated by a long lag will tend to lie at a similar distance from any location **u**). As is intuitively expected, this behaviour in the GWV is weakest when the geographical weights decay steeply with distance (i.e. for a large value of q). All such issues stem from distance-weighting a statistic that is already a function of distance and these issues are complicated further when a GWV also reflects any true trend in the process being modelled. Aside from (or in addition to) truncation, other possible approaches to guard against model fitting problems at long GWV lags include:

1. Using different weighting schemes for different lag distances in the GWV estimation. For example, use an IDW scheme but let the controlling parameter q increase as the lag distance increases.
2. Fixing the total sill for each GWV/WLS variogram fit as the GW variance (GWVAR) assigned to the same location **u**. Here a GWVAR is found using $s^2(\mathbf{u}) = \sum_{i=1}^{n} w_i(z(\mathbf{x}_i) - m(u))^2 \cdot \sum_{i=1}^{n} w_i$ where $m(\mathbf{u})$ is the local mean at **u** and the weights w_i accord to the same IDW scheme as that used for the GWV. Effectively, this approach entails that a GWV is used only to improve the nugget variance and range estimate over that found with the usual local CV.
3. Tailoring the (non-geographical) weighting function of the WLS variogram fit so that less importance is attached to long variogram lags. Here the chosen WLS fit uses N_h/h^2. weights with N_h the number of point pairs, which already reduces the influence of long lags on the variogram fit. Thus, weights of say, N_h/h^3, N_h/h^4 can be specified to reduce the influence of long lags on the variogram fit further.

For this study, the two latter approaches were experimented with, but as neither could offer a noticeable or consistent improvement over the GWV/WLS variogram fits specified, they were not pursued any further. Further work on this issue is considered worthwhile.

3.8 MWK–GWV: example GWV/WLS variogram fits with dataset four

For clarity, a spatially-representative selection of GWVs at four target locations are given in Fig. 2 using the small critical load data subset (dataset four). Here q is set at 0.5, 1 and 3 for GWV estimation in Fig. 2a–c, respectively. All variogram plots are shown with the *complete* GWV, but the WLS fits are to truncated GWVs (in this case set at a distance that is half the sampled area). For this particular example, the GWVs at locations two and three tend towards the global variogram (Fig. 2e) at their longest lags and this behaviour weakens as q is increased. For GWVs at locations one and four, such an effect is only really apparent at their *very* longest lags, reflecting the extreme locations of these example GWVs in the sample area (i.e. their location promotes variation in the geographical weights at the longer lags). For all twelve GWVs shown, the WLS variogram fits appear reasonable, providing some justification for GWV/WLS variogram fitting approach adopted.

As q is increased from 0.5 to 1 the GWVs become rather erratic, but are little different when q is increased from 1 to 3; and despite this change in behaviour of the estimated GWVs, the corresponding WLS fits are fairly stable for all three values of q. This relative stability is reassuring because if large discrepancies exist between corresponding WLS fits (at a given location), then it is possible that spurious local variogram structures have been induced by the geographic weighting schemes specified. In this respect, it is recommended that MWK–GWV models should always be complemented with some standard exploratory variography. For reference, Rivoirard (2002), Goovaerts et al. (2005) discuss the likelihood of such unwanted side-effects when weighting stationary variograms.

4 Model assessment

Two measures are used to assess the accuracy of the four kriging models examined. These are the prediction accuracy as measured by a relative root mean square error (RMSE) term and the prediction uncertainty accuracy measured by: (a) two prediction confidence interval (PCI) statistics; and (b) the correlation between the actual and estimated prediction errors. These are now defined briefly.

4.1 Prediction accuracy

A relative RMSE diagnostic is defined as relRMSE $= \mathrm{RMSE}_{\mathrm{Model}}/\mathrm{RMSE}_{\mathrm{Mean}}$. Here $\mathrm{RMSE} = \sqrt{(1/(n_v)) \sum_{k=1}^{n_v} \{z(\mathbf{u}_k) - \hat{z}(\mathbf{u}_k)\}^2}$ where $\hat{z}(\mathbf{u}_k)$ is the predicted and $z(\mathbf{u}_k)$ is the actual value at validation site $\mathbf{u}_k$; and n_v is the size of the validation data set. Here we would expect relRMSE $\lesssim 1$ for each kriging model examined, whereas if relRMSE ≥ 1 then the calibration data mean predicts as well or better.

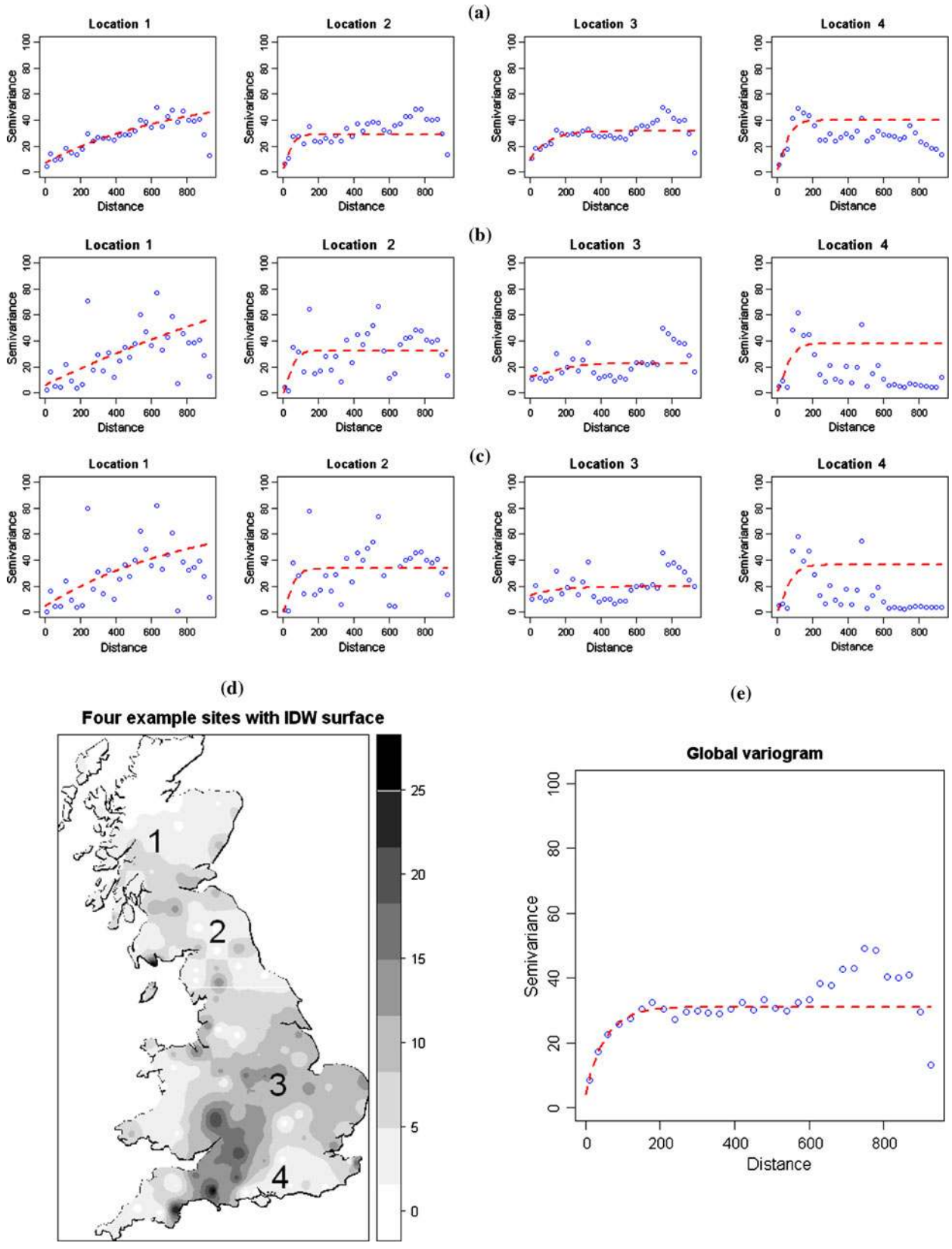

Fig. 2 Example GWV/WLS variogram fits at four locations using the small critical load data subset: **a** GWV/WLS variogram fits with q = 0.5, **b** GWV/WLS variogram fits with q = 1, **c** GWV/WLS variogram fits with q = 3, **d** four locations with IDW mean surface (to small data subset), and **e** the global variogram with WLS fit. All variograms are shown with all lag distances but where the WLS fit is to a truncated variogram

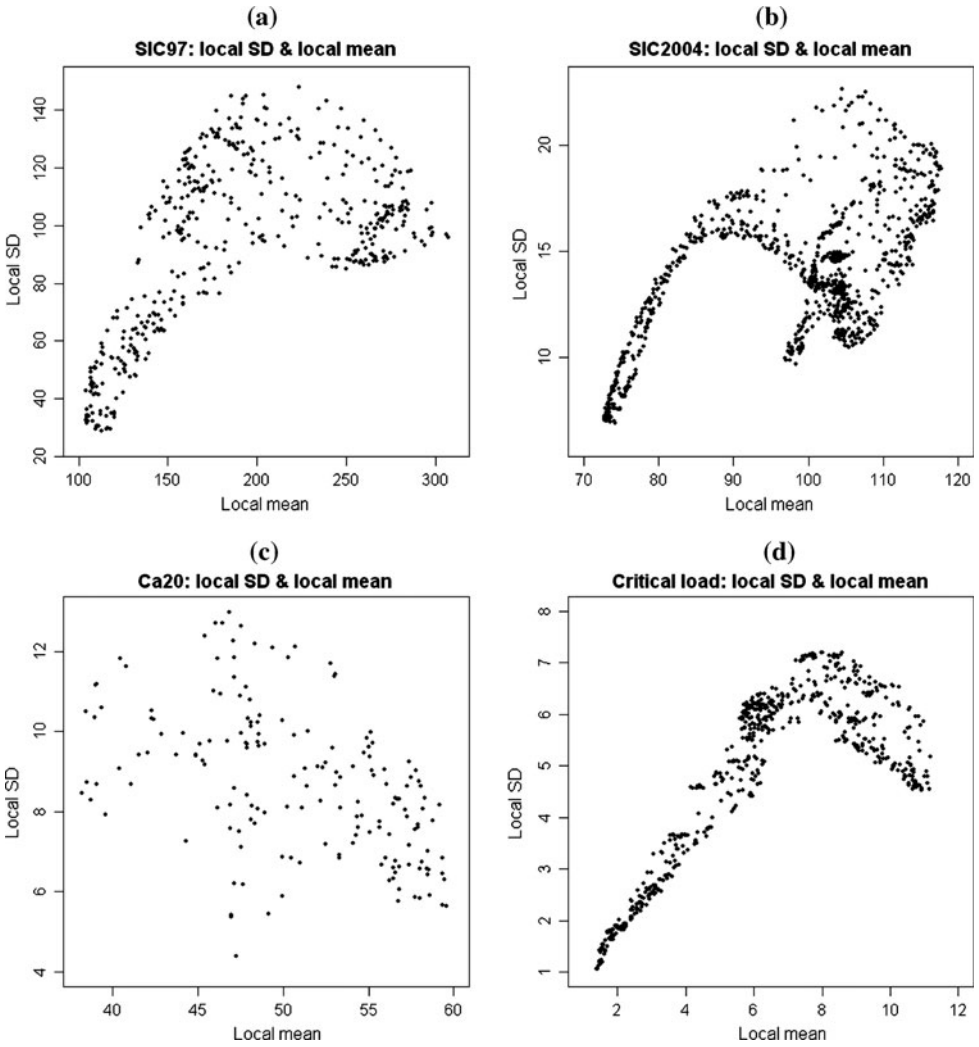

Fig. 3 Relationships between local means and local SDs: **a** dataset one, SIC97; **b** dataset two, SIC2004; **c** dataset three, Ca20; **d** dataset four, critical load. Local statistics are calculated using a simple moving window algorithm to the full datasets (window size is taken as 15% of the data)

4.2 Prediction uncertainty accuracy

For all models, the usual assumption of multivariate normality is adopted at the scale of the kriging neighbourhood. From a classical geostatistics viewpoint, this entails that $\hat{z}(\mathbf{u})$ and its corresponding prediction standard error r (**u**) can be taken as the two defining moments of a normal distribution at location **u**. This in turn enables the calculation of a PCI whose accuracy can be assessed using coverage probabilities. For example, if symmetric 95% PCIs were calculated at each validation site (i.e. using $\hat{z}(\mathbf{u}_k) \pm 1{:}96\mathrm{r}\,(\mathbf{u}_k)$), then a correct modelling of local uncertainty would entail that there is a 0.95 (expected coverage) probability that the actual value $z(\mathbf{u}_k)$ falls within the interval. If the actual (not expected) coverage probabilities are 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 (expected coverage) probability interval p, then an accuracy plot is obtained for a given model (e.g. see Fig. 7c).

Accuracy plots are not presented for all models, but a complementary 'goodness' statistic is reported that provides value to a model's accuracy plot. This G-statistic can be defined as $G = 1 - \int_0^1 [3a(p) - 2]\,|\bar{n}(p) - p|\,dp$, where $\bar{n}$ 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

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

which implies that twice the importance is given to deviations when $\bar{n}(p) < p$. For cases where two models result in similar accuracy plots and G-statistics, one model can be preferred if its PCI widths that contain the actual value are smaller (i.e. a model that consistently provides *narrow* and accurate PCIs should be preferred to a model that

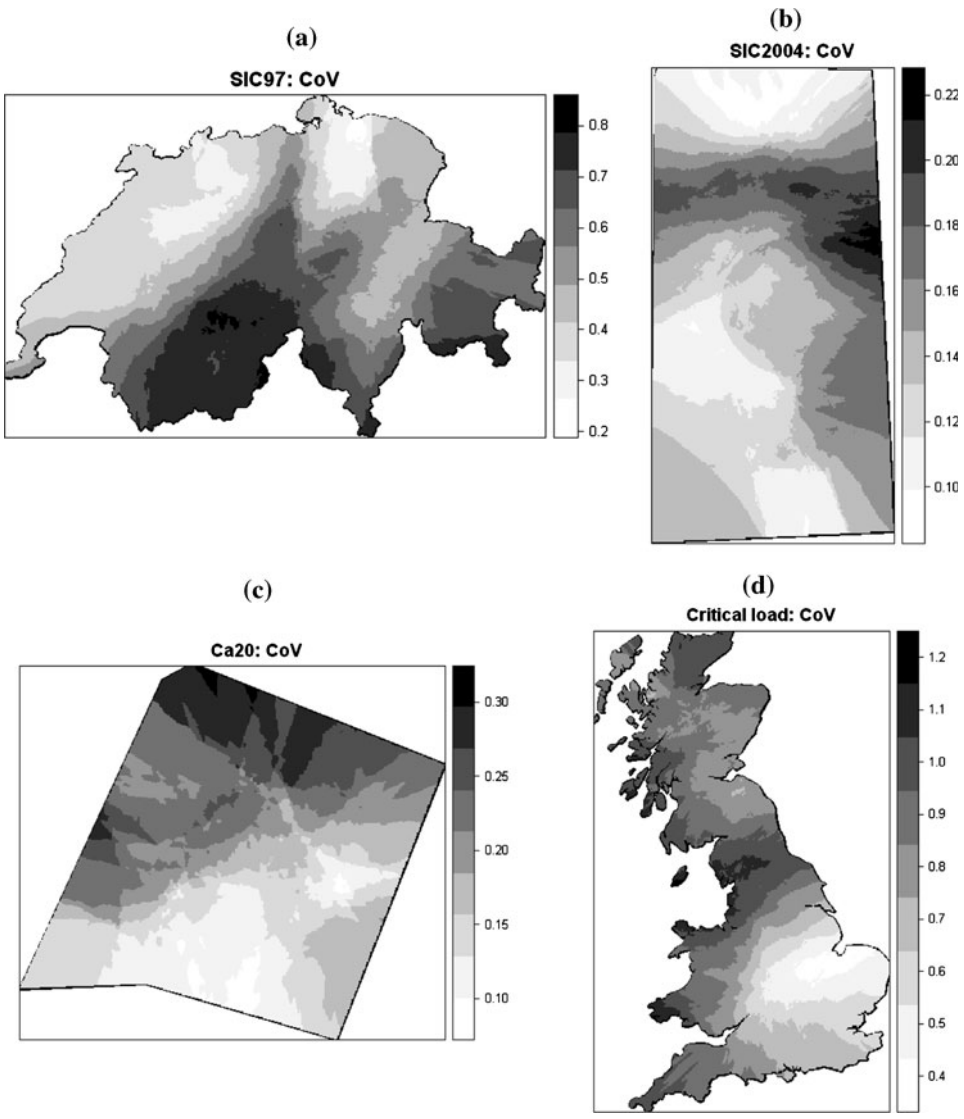

Fig. 4 CoV surfaces for **a** full dataset one, SIC97; **b** full dataset two, SIC2004; **c** full dataset three, Ca20; **d** full dataset four, critical load. Surfaces are calculated using a moving window algorithm, where window size is taken as 15% of the data

consistently provides *wide* and accurate PCIs). Here it is usual to construct the corresponding PCI width plots to the given accuracy plots and compare (e.g. see Fig. 7c, d). However, in this study, only an average PCI width for all p is reported (M-PCI-W) which acts as a rough guide to this important aspect of a model's PCIs. Details of accuracy plots and associated diagnostics are given in Goovaerts (2001).

A fourth model bias diagnostic is taken as the linear correlation coefficient between the absolute (actual) prediction errors and the estimated prediction (kriging) standard errors (Error-Corr). This correlation should be strong and positive if the estimated prediction standard errors actually reflect the local variability that is present in the sample data. The SK models are expected to perform consistently badly in this respect, whereas the BCK and MWK models are expected to perform well.

5 Results

5.1 Exploratory evidence of a proportional effect (for BCK)

To assess if a regionalised variable follows a lognormal distribution, local means can be plotted against the local standard deviations (SDs). If a linear relationship is evident then lognormality can be assumed (Chilès and Delfiner 1999, p. 56 and pp. 107–108). Similar relationships can be used to guide the use of the (more general) Box–Cox transform and in this respect local means are plotted against local SDs for each study dataset in Fig. 3. Here datasets one, two and four show some evidence of an increasing mean coinciding with an increasing SD and thus BCK (with k ? 0) may perform relatively well with these datasets. However, in all three cases, there is a tail-off in

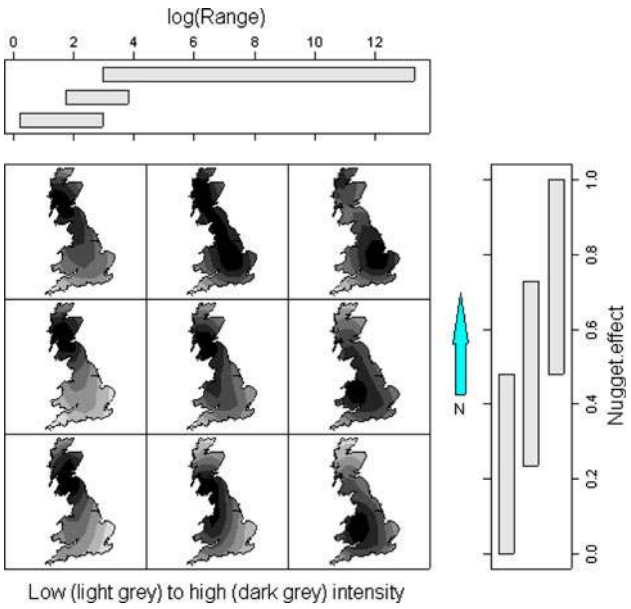

Fig. 5 Comap (specified using KDE) for the local nugget effect versus the local correlation range (in logs) for full dataset four (critical load). The logarithm of the correlation range enables a better visualisation. The underlying MWK–CV algorithm was specified with a 5% window size (the minimum allowed)

linearity at high mean values; and there is no discernable relationship for dataset four.

The spatial nature of a proportional effect can be explored with a local coefficient of variation (CoV) surface which should tend to some stationary constant if the mean scales linearly with the SD for the whole of the sampled region (i.e. an endorsement of lognormality). If, however, the CoV is clearly nonstationary in behaviour (i.e. reflecting a nonlinear relationship between means and SDs such as that observed in all four datasets), then the adoption of a technique such as MWK (in preference to BCK) may be of value. The CoV surfaces for all datasets provide evidence of CoV nonstationarity (see Fig. 4). More in depth investigations can be useful here. For example with dataset four (critical load), the tail-off in linearity at high mean values observed in Fig. 3d relates directly to an unusually low CoV in SE England (an area of high critical loads that varies little spatially, see Fig. 1d).

5.2 Exploratory evidence of variogram nonstationarity with comaps (for MWK)

To gauge the likely worth of a MWK application, variogram nonstationarity can be tested (Haas 1998; Fuentes 2005) or more simply, it can be verified informally using exploratory data analysis (EDA). Here a useful spatial representation of local variogram parameters obtained with MWK is possible with the comap (Brunsdon 2001). The comap is a spatial variant of the co-plot for assessing conditional distributions, where the relationship between a primary and secondary variable is plotted conditional to a third (secondary) variable (or a third and fourth secondary variable together). In the comap, geographical location is the first pair of variables conditioned on the primary variable of interest and a chosen secondary variable. The comap can be specified showing, (a) the geographic locations of the conditional data or (b) using kernel density estimation (KDE) to visualise the clustering of the same geographic locations.

For MWK, the nugget effect parameter (defined as the ratio $c_0/(c_0 + c)$) can be related to the range using a comap. This not only provides an assessment of variogram nonstationarity but also identifies regions most suited to kriging (kriging performs well in a relative sense when the nugget effect is small and the range is large). Such a comap is given in Fig. 5 for dataset four (critical load) using parameters obtained from an application of MWK–CV. There is clear evidence of variogram nonstationarity, as both the nugget effect and the range vary across space. Furthermore, critical load prediction accuracy should be relatively strong in areas that traverse the borders of Wales and England (bottom-right map), but relatively weak in areas of W Scotland and NW England (top-left map). In the latter case, weak spatial dependence is in part, a consequence of outlying (high-valued) critical load observations. Comaps (not given) using MWK–CV parameters for the other three datasets also portrayed evidence of variogram nonstationarity.

5.3 Example GWV/WLS and local CV/WLS variogram fits with dataset four

Complementing the example and exploratory variography conducted in Sect. 3.8 with the critical load data, it is prudent to continue this investigation where GWV/WLS variogram fits are compared with local CV/WLS variogram fits at the same four locations (similar investigations should be conducted using the other datasets also). In this respect, GWV/WLS and local CV/WLS variogram fits are given in Fig. 6a and b, respectively; and Fig. 6c shows the global variogram that will be used in SK. Clearly for both local variogram techniques, there is visual evidence of variogram nonstationarity and as a consequence, a MWK application may be of value. Again, reassuringly, corresponding GWV/WLS and CV/WLS variogram fits depict a broadly similar structure to each other, providing little evidence that spurious variographic structures have resulted from GWV estimation.

For the GWV/WLS and local CV/WLS variogram fits shown, the GWVs are specified with q = 0.5 and the local CVs are specified with a 30% window size; values which are taken as optimal on application of the corresponding

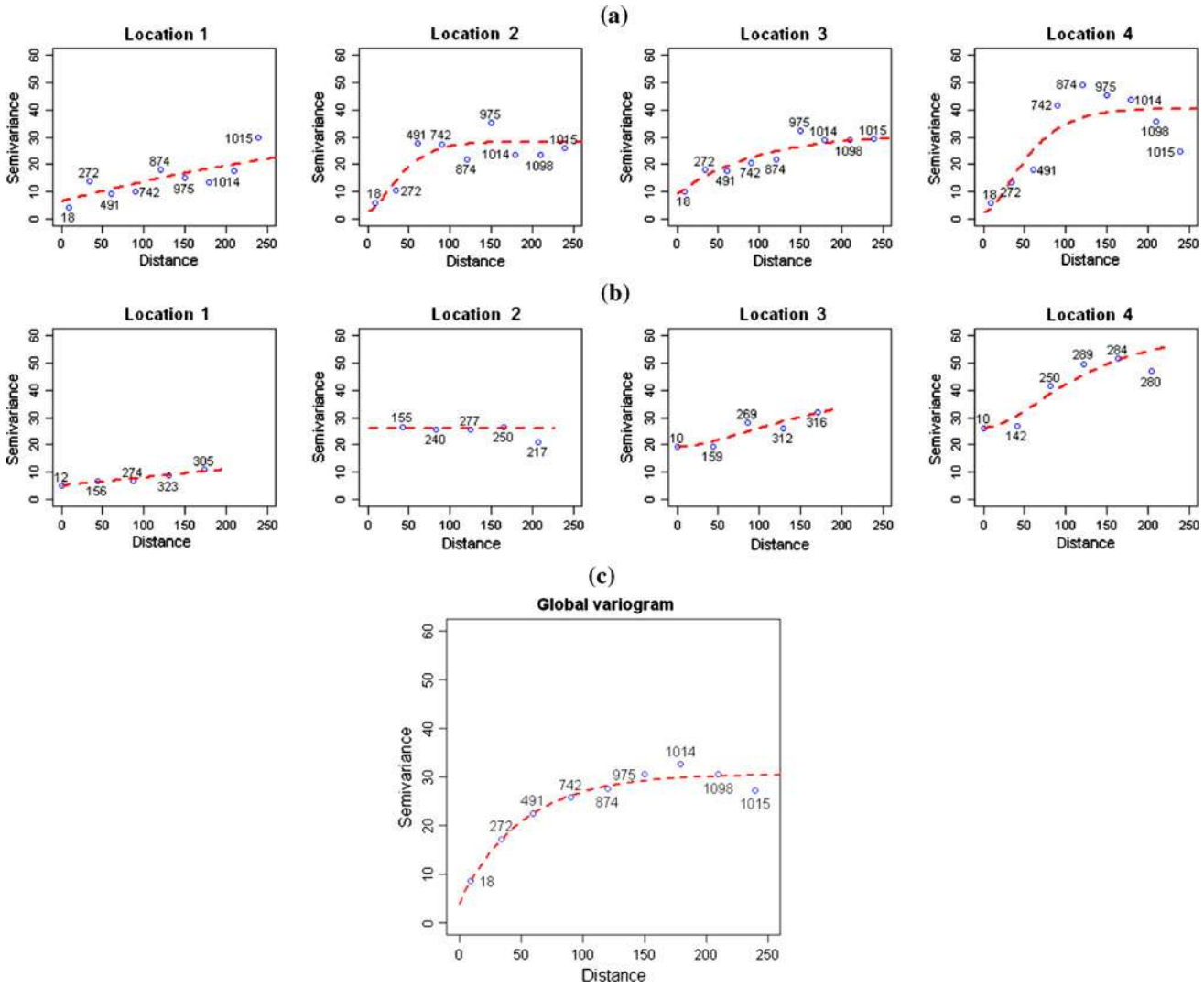

Fig. 6 Example local variograms at four locations using the small critical load data subset (see Fig. 2d): **a** GWV/WLS variogram fits, **b** CV/WLS variogram fits, and **c** global variogram with WLS fit. GWVs are specified with q = 0.5 and local CVs are specified with a 30% window size (i.e. optimal values, see Table 1 for dataset 4S). All variograms are shown with the number of data pairs at each lag

MWK model (see Table 1, dataset 4S). Such controlling parameters indicate a mildly nonstationary variogram process (as the GWV weighting function decays gradually with distance and the local CV window size is fairly large). Here it is also observed that as the IDW scheme used for GWV estimation distance-decays more steeply (see Sect. 3.8) or as the window size for local CV estimation becomes smaller, the corresponding WLS fits and MWK results tend to become poorer. Thus, the controlling parameters in MWK–GWV and MWK–CV have an analogous influence on their respective MWK models.

At this stage, it is useful to highlight the differences in information that is available to a GWV and a corresponding local CV. In Fig. 6, all variograms are shown with the number of data pairs available at each lag distance. Clearly, a GWV benefits from the same information as the global variogram (and continues at all lag distances, i.e. those above 250 km too). In contrast, the local CVs can suffer from reduced information at each lag distance and from no information at all at longer lag distances. For example, the local CV at location one has information at only five lags up to distances of around 200 km. Such local CV information is of course dependent on (a) the heuristics used to estimate the local CV and (b) window size. However, (a) altering the heuristics to provide more information is counter-productive as poor WLS fits (to local CVs with much scatter) will invariably result and (b) increasing window size simply results in local CVs that tend to the global CV.

Whilst on this subject, observe that the local CV at location two is modelled as a pure nugget variogram. This is in part a consequence of the heuristic used to define the

Table 1 Model performance results

Dataset, calibration size and model	Box–Cox trans, window size or IDW exponent	relRMSE	*G*-statistic	M-PCI-W	G-MPCIW-RANK	Error-Corr[c]
Data 1S-100						
SK	–	0.532	0.946	89.38	2	0.10
BCK	0.35	0.530	**0.977**	92.90	**1**	**0.21**
MWK–CV	80%	0.526	0.948	89.86	3	0.12
MWK–GWV	1	**0.516**	0.947	108.66	4	0.20
Data 1L-367						
SK	–	0.491	0.968	74.23	**1**	0.34
BCK	0.58	0.474	0.966	82.13	3	0.49
MWK–CV	20%	0.481	**0.970**	75.68	2	0.37
MWK–GWV	2	**0.473**	0.722	73.08	4	**0.52**
Data 2S-200						
SK	–	0.629	0.983	17.90	2	0.09
BCK	0.80	**0.628**	0.985	17.92	3	**0.30**
MWK–CV	90%	0.629	0.975	18.54	4	0.20
MWK–GWV	0.5	0.637	**0.986**	17.82	**1**	0.16
Data 2L-808						
SK	–	0.593	0.910	13.09	4	0.03
BCK	0.21	0.588	0.910	12.85	3	**0.24**
MWK–CV	20%	0.576	**0.972**	16.49	**1**	0.15
MWK–GWV	0.5	**0.570**	0.959	16.65	2	0.19
Data 3S-78						
SK[a]	–	0.822	**0.930**	12.33	2	0.09
MWK–CV	45%[b]	**0.812**	0.894	11.08	3	0.16
MWK–GWV	1	**0.812**	0.929	11.65	**1**	**0.31**
Data 3L-100						
SK[a]	–	0.694	**0.972**	12.92	2	0.21
MWK–CV	50%	0.717	0.965	12.13	**1**	**0.41**
MWK–GWV	0.5	**0.687**	0.935	12.56	3	0.35
Data 4S-189						
SK	–	**0.829**	0.971	7.00	2	0.07
BCK	0.19	0.840	0.873	6.32	4	0.28
MWK–CV	30%	0.847	**0.981**	6.20	**1**	**0.37**
MWK–GWV	0.5	0.874	0.936	5.52	3	**0.37**
Data 4L-497						
SK	–	**0.781**	0.917	6.83	4	0.11
BCK	0.15	0.816	0.946	6.31	3	0.34
MWK–CV	10%	0.789	**0.962**	6.54	2	0.39
MWK–GWV	1	0.827	0.960	5.78	**1**	**0.40**

Best results are in bold

[a] No data transformation necessary

[b] The minimum window size allowed

[c] This correlation is always checked against its corresponding scatterplot to guard against any severe scaling issues

minimum number of pairs allowed at the first lag, where in this case, the actual number pairs was below the set limit of ten. This results in the local CV at location two having no semivariance value at its first lag and in turn, a nugget variogram is fitted. Conversely, the GWV does not suffer from such problems, and as a result is fitted with a structured variogram. Of course, the actual variogram at location two is never known, but in this case, one option for it is much more likely to be representative of the local spatial process than the other.

Table 2 Summary of model performance results: *best* (all models)

Dataset and size	Lowest relRMSE	Lowest G-MPCIW-RANK	Strongest Error-Corr	*Best* overall
1S-100	**MWK–GWV**	BCK	BCK	BCK
1L-367	**MWK–GWV**	SK	**MWK–GWV**	**MWK–GWV**
2S-200	BCK	**MWK–GWV**	BCK	BCK
2L-808	**MWK–GWV**	MWK–CV	BCK	**MWK–GWV**
3S-78	MWK–CV/**MWK–GWV**	**MWK–GWV**	**MWK–GWV**	**MWK–GWV**
3L-100	**MWK–GWV**	MWK–CV	MWK–CV	MWK–CV
4S-189	SK	MWK–CV	MWK–CV/**MWK–GWV**	MWK–CV
4L-497	SK	**MWK–GWV**	**MWK–GWV**	MWK–CV/**MWK–GWV**

MWK–GWV models are in bold

Table 3 Summary of model performance results (MWK models only)

Dataset and size	Lowest relRMSE	Lowest G-MPCIW-RANK	Strongest Error-Corr	*Best* overall
1S-100	**MWK–GWV**	MWK–CV	**MWK–GWV**	**MWK–GWV**
1L-367	**MWK–GWV**	MWK–CV	**MWK–GWV**	**MWK–GWV**
2S-200	MWK–CV	**MWK–GWV**	MWK–CV	EQUAL
2L-808	**MWK–GWV**	MWK–CV	**MWK–GWV**	**MWK–GWV**
3S-78	EQUAL	**MWK–GWV**	**MWK–GWV**	**MWK–GWV**
3L-100	**MWK–GWV**	MWK–CV	MWK–CV	MWK–CV
4S-189	MWK–CV	MWK–CV	EQUAL	MWK–CV
4L-497	MWK–CV	**MWK–GWV**	**MWK–GWV**	EQUAL

MWK–GWV models are in bold

5.4 SK, BCK, MWK–CV and MWK–GWV model performance

Details of the performance of the four study models in the eight different data situations are given in Table 1 using the four goodness-of-fit measures described in Sect. 4 (i.e. relRMSE, *G*-statistic, M-PCI-W and Error-Corr). Here an additional goodness-of-fit measure is reported to discriminate better between models with respect to their performance in estimating PCIs. As indicated in Sect. 4, for cases where models provide similar *G*-statistics, one model can be preferred if its PCI widths that contain the actual value are smaller. Thus, to account for this aspect of a model's PCIs, a ranked performance diagnostic is reported. Here for each data situation, the four kriging models are first ranked according to their *G*-statistic value. This ranking is then adjusted for cases when the *G*-statistics are assumed similar (in this case, lying within 0.01 units of each of other). In such cases, the model ranking depends on the M-PCI-W value instead, and a lower ranking is given to models with the smaller (tighter) M-PCI-W values. This ranked diagnostic is named G-MPCIW-RANK (where a rank of 1 is best).

The results of Table 1 are summarised in Table 2 (all models compared) and Table 3 (only MWK models compared). Here the *best* performing models are listed with respect to the lowest relRMSE, the lowest G-MPCIW-RANK and the strongest Error-Corr. The *best* model overall is also given for each data situation, using a similar ranking system to that used for finding the optimal controlling parameter of a MWK model (see Sect. 3). However, in this ranking system, the *G*-statistic ranking is replaced by G-MPCIW-RANK. Ideally, this improved ranking system should have been used to discriminate between different MWK–CV and different MWK–GWV models, but G-MPCIW-RANK is awkward to calculate when there are numerous models to choose from and as such, the initial, more pragmatic and objective ranking system was followed.

From Tables 1 and 2, it is clear that datasets one and two appear to benefit the most from a geostatistical approach. Interestingly, the choice of calibration data subset can have marked effect on model parameterisation (e.g. see parameters found for dataset one). This can be attributable to changes in sample variation, size and configuration. This observation alone goes some way to justify the use of *eight* different data situations to establish empirically the utility of our MWK–GWV approach. From Table 2, it is evident that most datasets tend to suit a MWK model. This confirms the findings of the exploratory investigations of Sects. 5.1 and 5.2, demonstrating the value of a judged EDA for nonstationary effects.

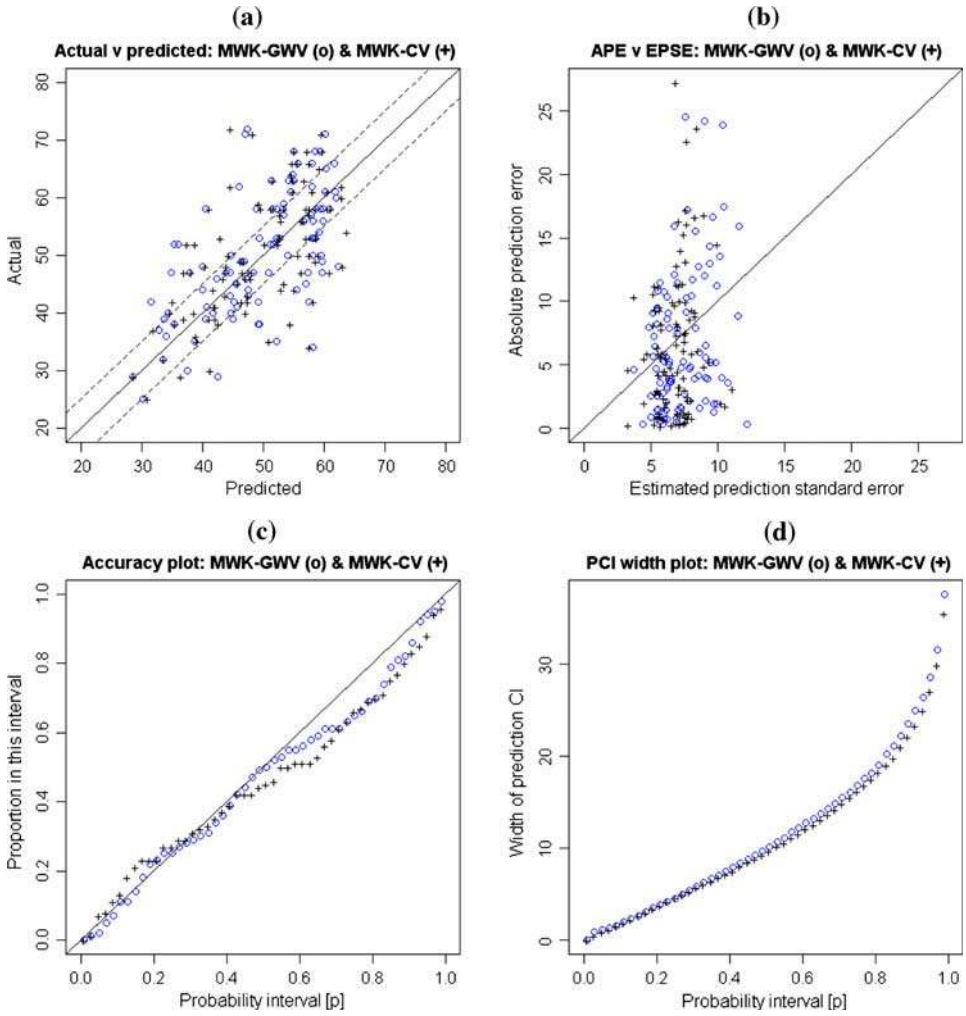

Fig. 7 MWK–CV versus MWK–GWV for dataset 3S (Ca20): **a** actual versus predicted plots, **b** absolute prediction error (APE) versus estimated prediction standard error (EPSE) plots, **c** accuracy plots, and **d** PCI width plots

From Table 2, a MWK–GWV, SK, BCK or MWK–CV model is the most accurate predictor (lowest relRMSE) on five, two, one and one occasion, respectively. Similarly, a MWK–CV, MWK–GWV, SK or BCK model provides the most accurate estimates of prediction uncertainty (with respect to G-MPCIW-RANK) on three, three, one and one occasion, respectively. Alternatively, a MWK–GWV, BCK or MWK–CV model provides the most accurate estimates of prediction uncertainty (with respect to Error-Corr) on four, three and two occasions, respectively. Thus, a MWK–GWV model performs well, both in terms of accuracy of prediction *and* of prediction uncertainty. From Table 2, a MWK–GWV, MWK–CV or BCK model is the *best* overall model on four, three and two occasions, respectively. From Table 3, a MWK–GWV is the *best* overall on four occasions, MWK–CV is the *best* overall on two occasions and on two occasions the MWK models perform equally.

It is also useful to assess the effect (if any) of calibration sample size on MWK performance. Intuitively, MWK–GWV would be expected to out-perform MWK–CV when sample size is small. For the smallest calibration data subset (dataset 3S, where sample size is 78), this intuition is realised and MWK–GWV performs better than MWK–CV. However, for the next two smallest calibration data subsets (datasets 1S and 3L, where sample size is 100), MWK–GWV does not perform better than MWK–CV on both occasions. Closer inspection of the relevant results in Table 1, does not reveal any reportable pattern and as such, this perceived benefit of a MWK–GWV approach is left unanswered. Example model performance plots are given for dataset 3S in Fig. 7 for the MWK models only. Here MWK–GWV provides a slightly stronger relationship between its absolute prediction errors and its estimated prediction standard errors (Fig. 7b). This is in part, a consequence of MWK–GWV providing a higher variation in its estimated prediction standard errors. More accurate prediction uncertainty estimates are similarly reflected in the accuracy plot for MWK–GWV which tends more strongly to the 45° line than the accuracy plot for MWK–CV does (Fig. 7c). The PCI width plots are little different between models (Fig. 7d).

6 Discussion and conclusions

The results provide a useful guide to the likely value of MWK–GWV. For four different datasets and eight different data situations, MWK–GWV performed the best of all four models. Thus, the MWK–GWV model has performed with merit and as such, justifies further development. The additional complexity of MWK–GWV is not considered a problem and it should be viewed as a general form of MWK, where MWK–CV is MWK–GWV specified with an adaptive box–car weighting function for local variogram estimation. Computationally, MWK–GWV is slightly more demanding than MWK–CV; and both models are affected by large datasets (a problem routinely encountered with numerous spatial methods, see Banerjee et al. 2008, for recent advances).

Additional work could consider the effects of sample configuration on MWK performance, since if variogram inference is difficult globally due to nuances in the sample layout, then locally it is likely to be worse. The model comparison could also be expanded to include not only the alternatives listed in Sect. 1 which mainly operate within the classical geostatistics framework, but also more modern nonstationary predictors such as the kernel-based (KS–StK) models of Fuentes (2001), Fuentes et al. (2003), the (different) kernel-based models of Higdon et al. (1999), Paciorek and Schervish (2006), and (possibly) the (single realisation) deformation models of Anderes and Stein (2005); most of which benefit from a Bayesian viewpoint to uncertainty.

Only basic MWK–CV and MWK–GWV models have been calibrated and assessed in this study. However, for both MWK forms, more elaborate models could be specified. For example, models could use: different underlying variogram estimators (e.g. a robust variogram for outlying data or a weighted variogram for clustered data); additional variogram model-types (e.g. a spherical model); different (least squares) variogram model fitting techniques; different statistics to choose the best fitting variogram model (e.g. AIC); different underlying kriging algorithms (e.g. OK); techniques to incorporate any anisotropic effects; etc. Furthermore, just as variogram model-type can be locally-defined, so can the variogram estimator (e.g. only use a robust variogram in areas of outlying data) and other MWK specifications (see below). It is also possible to extend both MWK forms to a KS–MWK version.

It is useful to observe how the choice of weighting function in GWV estimation tends to dictate other specifications of the MWK–GWV model. For example, with the MWK–GWV model defined in this study, it was natural to specify a global kriging neighbourhood and it would be similarly natural to globally-specify any data detrending or data transform if required (i.e. specifications relate to the same scale as the variography). Conversely, for a MWK–CV model or a MWK–GWV model where only the closest data pairs are weighted in the GWV estimation (i.e. use a discontinuous distance-decay function, such as a bi-square), it is natural to specify a local kriging neighbourhood and (if required) any data detrending or data transform should also be done at this same local scale (as the variography).

For this study, specifying a global kriging neighbourhood with MWK–GWV may have been detrimental to its performance relative to the local neighbourhoods of MWK–CV. However, both MWK models could have been specified with neighbourhoods smaller than that chosen (i.e. assume quasi-stationarity). In this respect, the MWK–GWV model, as defined in this study, has a much wider choice of kriging neighbourhood than any MWK–CV model, which should be considered an advantage.

Issues of data detrending were deliberately avoided in this study, so as not to over-complicate the model comparison. However, future work should consider adapting the MWK–GWV model so that both local and global trends in data are catered for. For some processes, such an adaptation may improve the WLS variogram model fits to (better-behaved, residual) GWVs. Adaptations are also vital to data that are accompanied by explanatory variables which can inform trend prediction. A recent example of MWK modelling that caters for both local and global trends in (precipitation) data can be found in Lloyd (2010).

Finally, one area that is vital to MWK–CV and MWK–GWV model performance analysis are the heuristics used to define: (a) the lag intervals, the minimum number of pairs allowed at the first lag and the truncation of the variogram estimator; and (b) the starting parameters for the WLS fit. Further work could investigate this particular aspect of MWK model specification more deeply. Ways to counter any problems here are to specify MWK models using ML fits (see the KS–MWK model of Pardo-Igúzquiza et al. 2005) or (possibly) use some nonparametric variogram model selection method (see Ploner and Dutter 2000) locally.

Acknowledgements 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 at Newcastle University and to Professor Chris Brunsdon for kindly providing the R code for the comap. We also thank the anonymous referees whose comments and insights helped improve the paper.

References

Anderes EB, Stein ML (2005) Estimating deformations of isotropic Gaussian random fields on the plane. Technical report no. 27, Centre for Integrating Statistics and Environmental Sciences, University of Chicago

Banerjee S, Gelfand AE, Finley AO, Sang H (2008) Gaussian predictive processes for large spatial datasets. J R Stat Soc B Stat Methodol 70:825–848

Brunsdon C (2001) The Comap: exploring spatial pattern via conditional distributions. Comput Environ Urban Syst 25:53–68

Cattle JA, McBratney AB, Minasny B (2002) Kriging methods evaluation for assessing the spatial distribution of urban soil lead contamination. J Environ Qual 31:1576–1588

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

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

Cressie N (1993) Statistics for spatial data. Wiley, New Jersey

Dubois G (2008) Editorial. Advances in automatic interpolation for real-time mapping. Stoch Environ Res Risk Assess 22:597–599

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

Fuentes M (2001) A new high frequency kriging approach for nonstationary environmental processes. Environmetrics 12:469–483

Fuentes M (2005) A formal test for nonstationarity of spatial stochastic processes. J Multivar Anal 96:30–54

Fuentes M, Guttorp P, Challenor P (2003) Statistical assessment of numerical models. Int Stat Rev 71:201–221

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

Goovaerts P, Jacquez GM, Greiling D (2005) Exploring scale-dependent correlations between cancer mortality rates using factorial kriging and population-weighted semivariograms. Geogr Anal 37:152–182

Haas TC (1990) Lognormal and moving window methods of estimating acid deposition. J Am Stat Assoc 85:950–963

Haas TC (1996) Multivariate spatial prediction in the presence of non-linear trend and covariance non-stationarity. Environmetrics 7:145–165

Haas TC (1998) Statistical assessment of spatio-temporal pollutant trends and meteorological transport models. Atmos Environ 32:1865–1879

Haas TC (2002) New systems for modelling, estimating, and predicting a multivariate spatio-temporal process. Environmetrics 13:311–332

Harris P, Brunsdon C (2010) Exploring spatial variation and spatial relationships in a freshwater acidification critical load data set for Great Britain using geographically weighted summary statistics. Comput Geosci 36:54–70

Higdon D, Swall J, Kern J (1999) Nonstationary spatial modelling. In: Berardo JM, Berger JO, Dawid AP, Smith AFM (eds) Bayesian statistics 6. Oxford University Press, Oxford, pp 761–768

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

Johannesson G, Cressie N (2004) Finding large-scale spatial trends in massive, global, environmental datasets. Environmetrics 15:1–44

Journel AG (1983) Nonparametric estimation of spatial distributions. Math Geol 15:445–468

Journel AG (1989) Fundamentals of geostatistics in five lessons. Short course in geology. American Geophysical Union Press, Washington, DC

Journel AG, Huijbregts CJ (1978) Mining geostatistics. Academic Press, London

Kitanidis PK, Shen K-F (1996) Geostatistical interpolation of chemical concentration. Adv Water Resour 19:369–378

Lahiri SN, Lee Y, Cressie N (2002) On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters. J Stat Plan Inference 103:65–85

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

Lloyd CD, Atkinson PM (2002) Nonstationary approaches for mapping terrain and assessing prediction uncertainty. Trans GIS 6:17–30

Matheron G (1976) A simple substitute for conditional expectation: the disjunctive kriging. In: Guarascio M, David M, Huijbrechts C (eds) Advanced geostatistics for the mining industry. Reidel, Dordrecht, pp 221–236

Müller WG (1999) Least-squares fitting from the variogram cloud. Stat Probab Lett 43:93–98

Omre H (1984) The variogram and its estimation. In Verly G, David M, Journel AG, Marachal A (eds) Geostatistics for natural resources characterization. ASI. Reidel, Hingham, pp 107–125

Paciorek CJ, Schervish MJ (2006) Spatial modelling using a new class of nonstationary covariance functions. Environmetrics 17:483–506

Pardo-Igúzquiza E, Dowd P, Grimes D (2005) An automatic moving window approach for mapping meteorological data. Int J Climatol 26:665–678

Parker HM (1991) Statistical treatment of outlier data in epithermal gold deposit reserve estimation. Math Geol 23:175–199

Pebesma EJ (2004) Multivariate geostatistics in S: the gstat package. Comput Geosci 30:683–691

Ploner A, Dutter R (2000) New directions in geostatistics. J Stat Plan Inference 91:499–509

Ribeiro PJ, Diggle PJ (2001) geoR: a package for geostatistical analysis. R News 1:15–18

Richmond A (2002) Two-point declustering for weighting data pairs in experimental variogram calculations. Comput Geosci 28:231–241

Rivoirard J (2002) Weighted variograms. In Kleingold WJ, Krige DG (eds) Geostatistics 2000 Cape Town. Geostatistical Association of Southern Africa, South Africa

Sampson PD, Damian D, Guttorp P (2001) Advances in modeling and inference for environmental processes with nonstationary spatial covariance. NRCSE-TRS No 61, National Research Centre for Statistics and the Environment Technical Report Series, 2001

Stein A, Hoogerwerf M, Bouma J (1988) Use of soil map delineations to improve (co)kriging of point data on moisture deficits. Geoderma 43:163–177

Van Tooren CF, Haas TC (1993) A site investigation strategy using moving window kriging and automated semivariogram modelling. In: Contaminated soil '93. Kluwer Academic Press, Dordrecht, pp 609–622

Walter C, McBratney AB, Douaoui A, Minasny B (2001) Spatial prediction of topsoil salinity in the Chelif Valley, Algeria, using local ordinary kriging with local variograms versus whole-area variogram. Aust J Soil Res 39:259–272

Yamamoto JK (2000) An alternative measure of the reliability of ordinary kriging estimates. Math Geol 32:489–509

Zhang X, Eijkeren JC, Heemink AW (1995) On the weighted least-squares method for fitting a semivariogram model. Comput Geosci 21:605–608